LECTURENOTES

Computer Integrated Manufacturing and FMS

B.Tech, 6th Semester, ME

Preparedby:

Dr. Chinmay Deheri

Associate Professor

Department of Mechanical Engineering

VikashInstitute ofTechnology, Bargarh

(ApprovedbyAICTE,NewDelhi&AffiliatedtoBPUT,Odisha)

BarahagudaCanalChowk,Bargarh,Odisha-768040

www.vitbargarh.ac.in

DISCLAIMER

- This document does not claim any originality and cannot be used as a substitute for prescribed textbooks.
- The information presented here is merely a collection by Dr. Chinmay Deheri with the inputs of students for their respective teaching assignments as an additional tool for the teaching-learning process.
- Various sources as mentioned at the reference of the document as well as freely available materials from internet were consulted for preparing this document.
- Further, this document is not intended to be used for commercial purpose and the authors are not accountable for any issues, legal or otherwise, arising out of use of this document.
- The author makes no representations or warranties with respect to the accuracy or completeness of the contents of this document and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose.

****

COURSE CONTENT

Computer Integrated Manufacturing and FMS

B.Tech, 6th Semester, ME

➢ MODULE – I

Fundamentals of Manufacturing and Automation: Production systems, automation principles and its strategies; Manufacturing industries; Types of production function in manufacturing;Automation principles and strategies, elements of automated system, automation functions and level of automation; product/production relationship, Production concept and mathematicalmodels for production rate, capacity, utilization and availability; Cost-benefit analysis.Computer Integrated Manufacturing: Basics of product design, CAD/CAM, Concurrent engineering, CAPP and CIM.

➢ MODULE – II

Industrial Robotics: Robot anatomy, control systems, end effectors, sensors and actuators; fundamentals of NC technology, CNC, DNC, NC part programming; Robotic programming, Robotic languages, work cell control, Robot cleft design, types of robot application, Processing operations, Programmable Logic controllers: Parts of PLC, Operation and application of PLC, Fundamentals of Net workings; Material Handling and automated storage and retrieval systems, automatic data capture, identification methods, bar code and other technologies.

➢ MODULE – III

Introduction to manufacturing systems: Group Technology and cellular manufacturing, Partfamilies, Part classification and coding, Production flow analysis, Machine cell design, Applications and Benefits of Group Technology. Flexible Manufacturing system: Basics of FMS, components of FMS, FMS planning and implementation, flexibility, quantitative analysis of flexibility, application and benefits of FMS. Computer Aided Quality Control: objectives of CAQC, QC and CIM, CMM and Flexible Inspection systems.

****

REFERENCES

Computer Integrated Manufacturing and FMS

B.Tech, 6th Semester, ME

Books:

1. Automation, Production Systems and Computer Integrated Manufacturing: M.P. Groover, Pearson Publication.
2. Automation, Production systems & Computer Integrated Manufacturing, M.P Groover, PHI.
3. CAD/CAM/CIM, P.Radhakrishnan, S.Subramanyam and V.Raju, New Age International
4. Flexible Manufacturing Systems in Practice, J Talavage and R.G. Hannam, Marcell Decker
5. CAD/CAM Theory and Practice, Zeid and Subramanian, TMH Publication
6. CAD/CAM Theory and Concepts, K. Sareen and C. Grewal, S Chand publication
7. Computer Aided Design and Manufacturing, L. Narayan, M. Rao and S. Sarkar, PHI.
8. Principles of Computer Integrated Manufacturing, S.K.Vajpayee, PHI
9. Computer Integrated Manufacturing, J.A.Rehg and H.W.Kraebber, Prentice Hall

Digital Learning Resources:

Course Name:	Computer Integrated Manufacturing
Course Link:	https://onlinecourses.nptel.ac.in/noc21_me65/preview
Course Instructor:	Prof. J Ramkumar, Prof Amandeep Singh, IIT Kanpur

****

became more complex, and so did the processes to make them. Workers had to specialize in their tasks. Rather than overseeing the fabrication of the entire product, they were responsible for only a small part of the total work. More up-front planning was required, and more coordination of the operations was needed to keep track of the work flow in the factories. Slowly but surely, the systems of production were being developed.

The systems of production are essential in modern manufacturing. This book is all about these production systems and how they are sometimes automated and computerized.

1.1 PRODUCTION SYSTEMS

A production system is a collection of people, equipment, and procedures organized to perform the manufacturing operations of a company. It consists of two major components as indicated in Figure 1.1:

1. *Facilities*. The physical facilities of the production system include the equipment, the way the equipment is laid out, and the factory in which the equipment is located.
2. *Manufacturing support systems*. These are the procedures used by the company to manage production and to solve the technical and logistics problems encountered in ordering materials, moving the work through the factory, and ensuring that products meet quality standards. Product design and certain business functions are included in the manufacturing support systems.

In modern manufacturing operations, portions of the production system are automated and/or computerized. In addition, production systems include people. People make these systems work. In general, direct labor people (blue-collar workers)

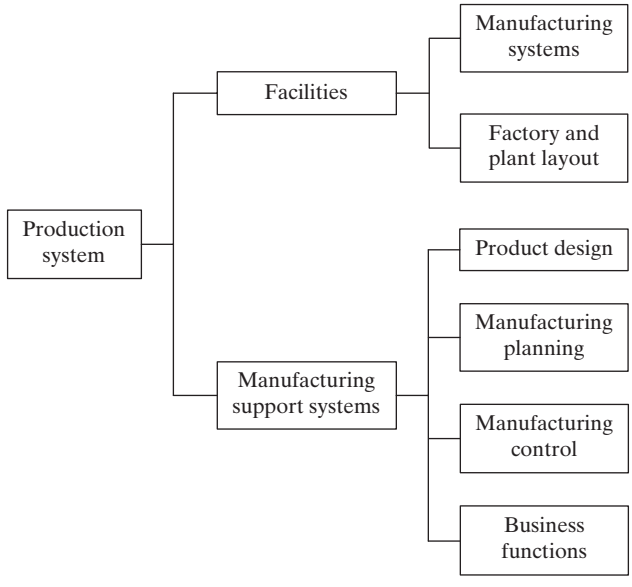


Figure 1.1 The production system consists of facilities and manufacturing support systems.

are responsible for operating the facilities, and professional staff people (white-collar workers) are responsible for the manufacturing support systems.

1.1.1 Facilities

The facilities in the production system consist of the factory, production machines and tooling, material handling equipment, inspection equipment, and computer systems that control the manufacturing operations. Facilities also include the ***plant layout***, which is the way the equipment is physically arranged in the factory. The equipment is usually organized into ***manufacturing systems***, which are the logical groupings of equipment and workers that accomplish the processing and assembly operations on parts and products made by the factory. Manufacturing systems can be individual work cells consisting of a single production machine and a worker assigned to that machine. More complex manufacturing systems consist of collections of machines and workers, for example, a production line. The manufacturing systems come in direct physical contact with the parts and/or assemblies being made. They "touch" the product.

In terms of human participation in the processes performed by the manufacturing systems, three basic categories can be distinguished, as portrayed in Figure 1.2: (a) manual work systems, (b) worker-machine systems, and (c) automated systems.

Manual Work Systems. A manual work system consists of one or more workers performing one or more tasks without the aid of powered tools. Manual material handling tasks are common activities in manual work systems. Production tasks commonly require the use of hand tools, such as screwdrivers and hammers. When using hand tools, a workholder is often employed to grasp the work part and position it securely for processing. Examples of production-related manual tasks involving the use of hand tools include

- A machinist using a file to round the edges of a rectangular part that has just been milled
- A quality control inspector using a micrometer to measure the diameter of a shaft
- A material handling worker using a dolly to move cartons in a warehouse
- A team of assembly workers putting together a piece of machinery using hand tools.

Worker-Machine Systems. In a worker-machine system, a human worker operates powered equipment, such as a machine tool or other production machine. This is one of the most widely used manufacturing systems. Worker-machine systems include

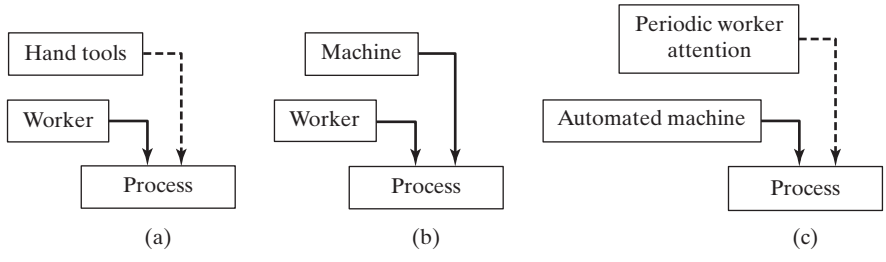


Figure 1.2 Three categories of manufacturing systems: (a) manual work system, (b) worker-machine system, and (c) fully automated system.

combinations of one or more workers and one or more pieces of equipment. The workers and machines are combined to take advantage of their relative strengths and attributes, which are listed in Table 1.1. Examples of worker-machine systems include the following:

- A machinist operating an engine lathe to fabricate a part for a product
- A fitter and an industrial robot working together in an arc–welding work cell
- A crew of workers operating a rolling mill that converts hot steel slabs into flat plates
- A production line in which the products are moved by mechanized conveyor and the workers at some of the stations use power tools to accomplish their processing or assembly tasks.

Automated Systems. An automated system is one in which a process is performed by a machine without the direct participation of a human worker. Automation is implemented using a program of instructions combined with a control system that executes the instructions. Power is required to drive the process and to operate the program and control system (these terms are defined more completely in Chapter 4).

There is not always a clear distinction between worker-machine systems and automated systems, because many worker-machine systems operate with some degree of automation. Two levels of automation can be identified: semiautomated and fully automated. A ***semiautomated machine*** performs a portion of the work cycle under some form of program control, and a human worker tends to the machine for the remainder of the cycle, by loading and unloading it, or by performing some other task each cycle. A ***fully automated machine*** is distinguished from its semiautomated counterpart by its capacity to operate for an extended period of time with no human attention. Extended period of time means longer than one work cycle; a worker is not required to be present during each cycle. Instead, the worker may need to tend the machine every tenth cycle, or every hundredth cycle. An example of this type of operation is found in many injection molding plants, where the molding machines run on automatic cycles, but periodically the molded parts at the machine must be collected by a worker. Figure 1.2(c) depicts a fully automated system. The semiautomated system is best portrayed by Figure 1.2(b).

In certain fully automated processes, one or more workers are required to be present to continuously monitor the operation, and make sure that it performs according to the intended specifications. Examples of these kinds of automated processes include complex

TABLE 1.1 Relative Strengths and Attributes of Humans and Machines

Humans	Machines
Sense unexpected stimuli	Perform repetitive tasks consistently
Develop new solutions to problems	Store large amounts of data
Cope with abstract problems	Retrieve data from memory reliably
Adapt to change	Perform multiple tasks simultaneously
Generalize from observations	Apply high forces and power
Learn from experience	Perform simple computations quickly
Make decisions based on incomplete data	Make routine decisions quickly

chemical processes, oil refineries, and nuclear power plants. The workers do not actively participate in the process except to make occasional adjustments in the equipment settings, perform periodic maintenance, and spring into action if something goes wrong.

1.1.2 Manufacturing Support Systems

To operate the production facilities efficiently, a company must organize itself to design the processes and equipment, plan and control the production orders, and satisfy product quality requirements. These functions are accomplished by manufacturing support systems—people and procedures by which a company manages its production operations. Most of these support systems do not directly contact the product, but they plan and control its progress through the factory.

Manufacturing support involves a sequence of activities, as depicted in Figure 1.3. The activities consist of four functions that include much information flow and data processing: (1) business functions, (2) product design, (3) manufacturing planning, and (4) manufacturing control.

Business Functions. The business functions are the principal means by which the company communicates with the customer. They are, therefore, the beginning and the end of the information-processing sequence. Included in this category are sales and marketing, sales forecasting, order entry, and customer billing.

The order to produce a product typically originates from the customer and proceeds into the company through the sales department of the firm. The production order will be in one of the following forms: (1) an order to manufacture an item to the customer's specifications, (2) a customer order to buy one or more of the manufacturer's proprietary products, or (3) an internal company order based on a forecast of future demand for a proprietary product.

Product Design. If the product is manufactured to customer design, the design has been provided by the customer, and the manufacturer's product design department is not involved. If the product is to be produced to customer specifications, the manufacturer's product design department may be contracted to do the design work for the product as well as to manufacture it.

If the product is proprietary, the manufacturing firm is responsible for its development and design. The sequence of events that initiates a new product design often originates in the sales department; the direction of information flow is indicated in Figure 1.3. The departments of the firm that are organized to accomplish product design might include research and development, design engineering, and perhaps a prototype shop.

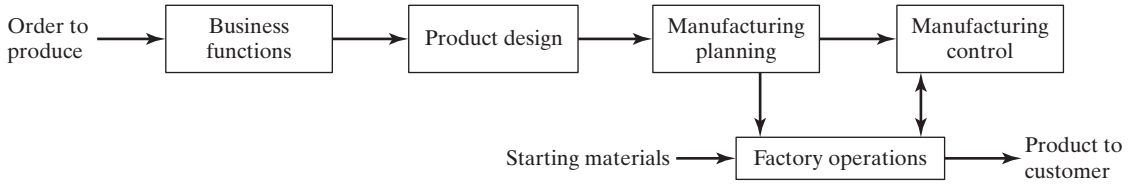


Figure 1.3 Sequence of information-processing activities in a typical manufacturing firm.

Manufacturing Planning. The information and documentation that constitute the product design flows into the manufacturing planning function. The information-processing activities in manufacturing planning include process planning, master scheduling, material requirements planning, and capacity planning.

Process planning consists of determining the sequence of individual processing and assembly operations needed to produce the part. The manufacturing engineering department is responsible for planning the processes and related technical details such as tooling. Manufacturing planning includes logistics issues, commonly known as production planning. The authorization to produce the product must be translated into the ***master production schedule***, which is a listing of the products to be made, the dates on which they are to be delivered, and the quantities of each. Based on this master schedule, the individual components and subassemblies that make up each product must be scheduled. Raw materials must be purchased or requisitioned from storage, parts must be ordered from suppliers, and all of these items must be planned so they are available when needed. The computations for this planning are made by ***material requirements planning***. In addition, the master schedule must not list more quantities of products than the factory is capable of producing each month with its given number of machines and manpower. ***Capacity planning*** is concerned with determining the human and equipment resources of the firm and checking to make sure that the production plan is feasible.

Manufacturing Control. Manufacturing control is concerned with managing and controlling the physical operations in the factory to implement the manufacturing plans. The flow of information is from planning to control as indicated in Figure 1.3. Information also flows back and forth between manufacturing control and the factory operations. Included in this function are shop floor control, inventory control, and quality control.

Shop floor control deals with the problem of monitoring the progress of the product as it is being processed, assembled, moved, and inspected in the factory. Shop floor control is concerned with inventory in the sense that the materials being processed in the factory are work-in-process inventory. Thus, shop floor control and inventory control overlap to some extent. ***Inventory control*** attempts to strike a proper balance between the risk of too little inventory (with possible stock-outs of materials) and the carrying cost of too much inventory. It deals with such issues as deciding the right quantities of materials to order and when to reorder a given item when stock is low. The function of ***quality control*** is to ensure that the quality of the product and its components meet the standards specified by the product designer. To accomplish its mission, quality control depends on inspection activities performed in the factory at various times during the manufacture of the product. Also, raw materials and component parts from outside sources are sometimes inspected when they are received, and final inspection and testing of the finished product is performed to ensure functional quality and appearance. Quality control also includes data collection and problem-solving approaches to address process problems related to quality, such as statistical process control (SPC) and Six Sigma.

1.2 AUTOMATION IN PRODUCTION SYSTEMS

Some components of the firm's production system are likely to be automated, whereas others will be operated manually or clerically. The automated elements of the production system can be separated into two categories: (1) automation of the manufacturing

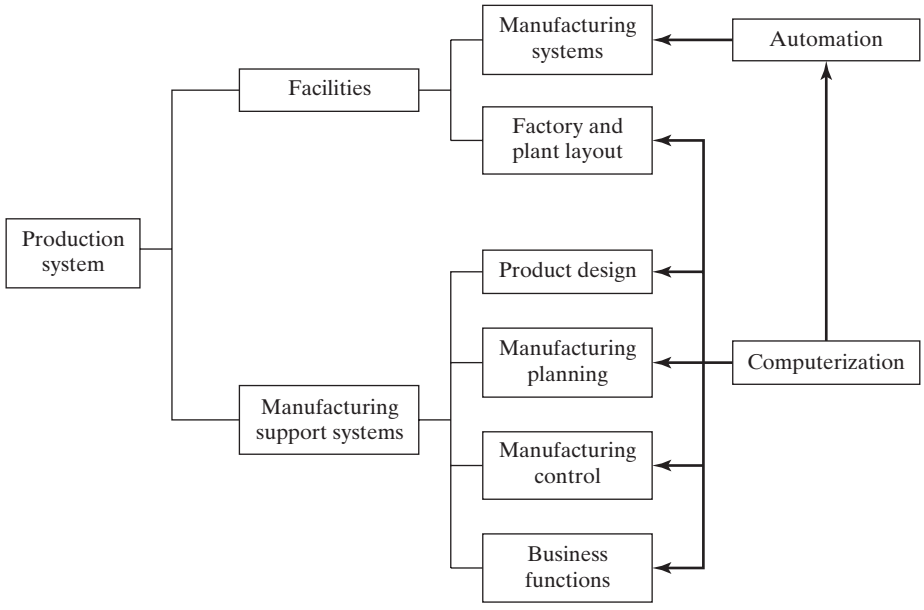


Figure 1.4 Opportunities for automation and computerization in a production system.

systems in the factory, and (2) computerization of the manufacturing support systems. In modern production systems, the two categories are closely related, because the automated manufacturing systems on the factory floor are themselves usually implemented by computer systems that are integrated with the manufacturing support systems and management information system operating at the plant and enterprise levels. The two categories of automation are shown in Figure 1.4 as an overlay on Figure 1.1.

1.2.1 Automated Manufacturing Systems

Automated manufacturing systems operate in the factory on the physical product. They perform operations such as processing, assembly, inspection, and material handling, in many 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:

- Automated machine tools that process parts
- Transfer lines that perform a series of machining operations
- Automated assembly systems
- Manufacturing systems that use industrial robots to perform processing or assembly operations
- Automatic material handling and storage systems to integrate manufacturing operations
- Automatic inspection systems for quality control.

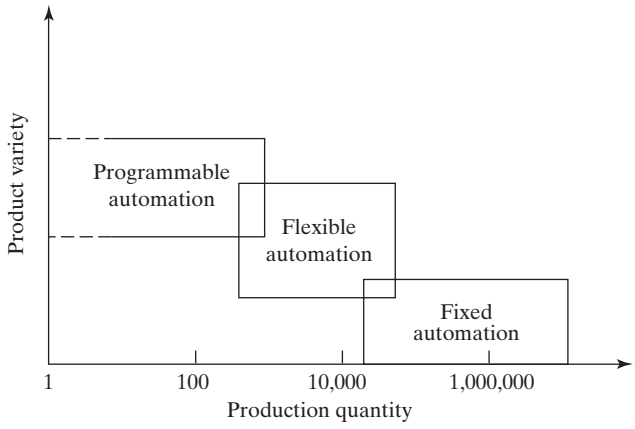


Figure 1.5 Three types of automation relative to production quantity and product variety.

Automated manufacturing systems can be classified into three basic types: (1) fixed automation, (2) programmable automation, and (3) flexible automation. They generally operate as fully automated systems although semiautomated systems are common in programmable automation. The relative positions of the three types of automation for different production volumes and product varieties are depicted in Figure 1.5.

Fixed Automation. Fixed automation is a system in which the sequence of processing (or assembly) operations is fixed by the equipment configuration. Each operation in the sequence is usually simple, involving perhaps a plain linear or rotational motion or an uncomplicated combination of the two, such as feeding a rotating spindle. It is the integration and coordination of many such operations in one piece of equipment that makes the system complex. Typical features of fixed automation are (1) high initial investment for custom-engineered equipment, (2) high production rates, and (3) inflexibility of the equipment to accommodate product variety.

The economic justification for fixed automation is found in products that are made 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 minimizing the unit cost relative to alternative methods of production. Examples of fixed automation include machining transfer lines and automated assembly machines.

Programmable Automation. In programmable automation, the production equipment is designed with the capability to change the sequence of operations to accommodate different product configurations. 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 (1) high investment in general-purpose equipment, (2) lower production rates than fixed automation, (3) flexibility to deal with variations and changes in product configuration, and (4) high suitability for batch production.

Programmable automated 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 item, the system must be reprogrammed with the set of machine instructions that correspond to the new item. The physical setup of the machine must also be changed: Tools must be loaded, fixtures must be attached to the machine table, and any required machine settings must be entered. This changeover takes time. Consequently, the typical cycle for a given batch includes a period during which the setup and reprogramming take place, followed by a period in which the parts are produced. Examples of programmable automation include numerically controlled (NC) machine tools, industrial robots, and programmable logic controllers.

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 design to the next. There is no lost production time while reprogramming the system and altering the physical setup (tooling, fixtures, machine settings). Accordingly, the system can produce various mixes 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, so the amount of changeover between designs is minimal. Features of flexible automation include (1) high investment for a custom-engineered system, (2) continuous production of variable mixtures of parts or products, (3) medium production rates, and (4) flexibility to deal with product design variations. Examples of flexible automation are flexible manufacturing systems that perform machining processes.

1.2.2 Computerized Manufacturing Support Systems

Automation of the manufacturing support systems is aimed at reducing the amount of manual and clerical effort in product design, manufacturing planning and control, and the business functions of the firm. Nearly all modern manufacturing support systems are implemented using computers. Indeed, computer technology is used to implement automation of the manufacturing systems in the factory as well. ***Computer-integrated manufacturing*** (CIM) denotes the pervasive use of computer systems to design the products, plan the production, control the operations, and perform the various information-processing functions needed in a manufacturing firm. True CIM involves integrating all of these functions in one system that operates throughout the enterprise. Other terms are used to identify specific elements of the CIM system; for example, *computer-aided design* (CAD) supports the product design function. *Computer-aided manufacturing* (CAM) is used for functions related to manufacturing engineering, such as process planning and numerical control part programming. Some computer systems perform both CAD and CAM, and so the term *CAD/CAM* is used to indicate the integration of the two into one system.

Computer-integrated manufacturing involves the information-processing activities that provide the data and knowledge required to successfully produce the product. These activities are accomplished to implement the four basic manufacturing support functions identified earlier: (1) business functions, (2) product design, (3) manufacturing planning, and (4) manufacturing control.

1.2.3 Reasons for Automating

Companies undertake projects in automation and computer-integrated manufacturing for good reasons, some of which are the following:

1. *Increase labor productivity*. Automating a manufacturing operation invariably increases production rate and labor productivity. This means greater output per hour of labor input.
2. *Reduce labor cost*. Increasing labor 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 labor to reduce unit product cost.
3. *Mitigate 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 for labor.
4. *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 improves the general level of working conditions.
5. *Improve worker safety*. Automating a given operation and transferring the worker from active participation in the process to a monitoring role, or removing the worker from the operation altogether, makes the work 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. *Improve product quality*. Automation not only results in higher production rates than manual operation, it also performs the manufacturing process with greater consistency and conformity to quality specifications.
7. *Reduce manufacturing lead time*. Automation helps reduce the elapsed time between customer order and product delivery, providing a competitive advantage to the manufacturer for future orders. By reducing manufacturing lead time, the manufacturer also reduces work-in-process inventory.
8. *Accomplish processes that cannot be done manually*. Certain operations cannot be accomplished without the aid of a machine. These processes require 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. *Avoid the high cost of not automating*. There is a significant competitive advantage gained in automating a manufacturing plant. The advantage cannot always 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.

humans are much better suited than machines, according to Table 1.1. Even if all of the manufacturing systems in the factory are automated, there is still a need for the following kinds of work to be performed by humans:

- *Equipment maintenance*. Skilled technicians are required to maintain and repair the automated systems in the factory when these systems break down. To improve the reliability of the automated systems, preventive maintenance programs are implemented.
- *Programming and computer operation*. There will be a continual demand to upgrade software, install new versions of software packages, and execute the programs. It is anticipated that much of the routine process planning, numerical control part programming, and robot programming may be highly automated using artificial intelligence (AI) in the future. But the AI programs must be developed and operated by people.
- *Engineering project work*. The computer-automated and integrated factory is likely never to be finished. There will be a continual need to upgrade production machines, design tooling, solve technical problems, and undertake continuous improvement projects. These activities require the skills of engineers working in the factory.
- *Plant management*. Someone must be responsible for running the factory. There will be a staff of professional managers and engineers who are responsible for plant operations. There is likely to be an increased emphasis on managers' technical skills compared with traditional factory management positions, where the emphasis is on personnel skills.

1.4 AUTOMATION PRINCIPLES AND STRATEGIES

The preceding section leads one to conclude that automation is not always the right answer for a given production situation. A certain caution and respect must be observed in applying automation technologies. This section offers three approaches for dealing with automation projects:[1] (1) the USA Principle, (2) Ten Strategies for Automation and Process Improvement, and (3) an Automation Migration Strategy.

1.4.1 The USA Principle

The USA Principle is a commonsense approach to automation and process improvement projects. Similar procedures have been suggested in the manufacturing and automation trade literature, but none has a more captivating title than this one. USA stands for (1) understand the existing process, (2) simplify the process, and (3) automate the process. A statement of the USA Principle appeared in an article published by the American Production and Inventory Control Society [5]. The article is concerned with implementing enterprise resource planning (ERP, Section 25.7), but the USA approach is so general that it is applicable to nearly any automation project. Going through each step of the procedure for an automation project may in fact reveal that simplifying the process is sufficient and automation is not necessary.

[1]There are additional approaches not discussed here, but in which the reader may be interested—for example, the ten steps to integrated manufacturing production systems discussed in J. Black's book *The Design of the Factory with a Future* [1]. Much of Black's book deals with lean production and the Toyota Production System, which is covered in Chapter 26 of the present book.

Understand the Existing Process. The first step in the USA approach is to comprehend the current process in all of its details. What are the inputs? What are the outputs? What exactly happens to the work unit[2] between input and output? What is the function of the process? How does it add value to the product? What are the upstream and downstream operations in the production sequence, and can they be combined with the process under consideration?

Some of the traditional industrial engineering charting tools used in methods analysis are useful in this regard, such as the operation chart and the flow process chart [3]. Application of these tools to the existing process provides a model of the process that can be analyzed and searched for weaknesses (and strengths). The number of steps in the process, the number and placement of inspections, the number of moves and delays experienced by the work unit, and the time spent in storage can be ascertained by these charting techniques.

Mathematical models of the process may also be useful to indicate relationships between input parameters and output variables. What are the important output variables? How are these output variables affected by inputs to the process, such as raw material properties, process settings, operating parameters, and environmental conditions? This information may be valuable in identifying what output variables need to be measured for feedback purposes and in formulating algorithms for automatic process control.

Simplify the Process. Once the existing process is understood, then the search begins for ways to simplify. This often involves a checklist of questions about the existing process. What is the purpose of this step or this transport? Is the step necessary? Can it be eliminated? Does it use the most appropriate technology? How can it be simplified? Are there unnecessary steps in the process that might be eliminated without detracting from function?

Some of the ten strategies for automation and process improvement (Section 1.4.2) can help simplify the process. Can steps be combined? Can steps be performed simultaneously? Can steps be integrated into a manually operated production line?

Automate the Process. Once the process has been reduced to its simplest form, then automation can be considered. The possible forms of automation include those listed in the ten strategies discussed in the following section. An automation migration strategy (such as the one in Section 1.4.3) might be implemented for a new product that has not yet proven itself.

1.4.2 Ten Strategies for Automation and Process Improvement

Applying the USA Principle is a good approach in any automation project. As suggested previously, it may turn out that automation of the process is unnecessary or cannot be cost justified after the process has been simplified.

If automation seems a feasible solution to improving productivity, quality, or other measure of performance, then the following ten strategies provide a road map to search for these improvements. These ten strategies were originally published in the author's first book.[3] They seem as relevant and appropriate today as they did in 1980. They

[2]The ***work unit*** is the part or product being processed or assembled.

[3]M. P. Groover, *Automation, Production Systems, and Computer-Aided Manufacturing,* Prentice Hall, Englewood Cliffs, NJ, 1980.

are referred to as strategies for automation and process improvement because some of them are applicable whether the process is a candidate for automation or just for simplification.

1. *Specialization of operations*. The first strategy involves the use of special-purpose equipment designed to perform one operation with the greatest possible efficiency. This is analogous to the specialization of labor, which is employed to improve labor productivity.
2. *Combined operations*. Production occurs as a sequence of operations. Complex parts may require dozens or even hundreds of processing steps. The strategy of combined operations involves reducing the number of distinct production machines or workstations through which the part must be routed. This is accomplished by performing more than one operation at a given machine, thereby reducing the number of separate machines needed. Since each machine typically involves a setup, setup time can usually be saved by this strategy. Material handling effort, nonoperation time, waiting time, and manufacturing lead time are all reduced.
3. *Simultaneous operations*. A logical extension of the combined operations strategy is to simultaneously perform the operations that are combined at one workstation. In effect, two or more processing (or assembly) operations are being performed simultaneously on the same work part, thus reducing total processing time.
4. *Integration of operations*. This strategy involves linking several workstations together into a single integrated mechanism, using automated work handling devices to transfer parts between stations. In effect, this reduces the number of separate work centers through which the product must be scheduled. With more than one workstation, several parts can be processed simultaneously, thereby increasing the overall output of the system.
5. *Increased flexibility*. This strategy attempts to achieve maximum utilization of equipment for job shop and medium-volume situations by using the same equipment for a variety of parts or products. It involves the use of programmable or flexible automation (Section 1.2.1). Prime objectives are to reduce setup time and programming time for the production machine. This normally translates into lower manufacturing lead time and less work-in-process.
6. *Improved material handling and storage*. A great opportunity for reducing nonproductive time exists in the use of automated material handling and storage systems. Typical benefits include reduced work-in-process, shorter manufacturing lead times, and lower labor costs.
7. *On-line inspection*. Inspection for quality of work is traditionally performed after the process is completed. This means that any poor-quality product has already been produced by the time it is inspected. Incorporating inspection into the manufacturing process permits corrections to the process as the product is being made. This reduces scrap and brings the overall quality of the product closer to the nominal specifications intended by the designer.
8. *Process control and optimization*. This includes a wide range of control schemes intended to operate the individual processes and associated equipment more efficiently. By this strategy, the individual process times can be reduced and product quality can be improved.

9. *Plant operations control*. Whereas the previous strategy is concerned with the control of individual manufacturing processes, this strategy is concerned with control at the plant level. It attempts to manage and coordinate the aggregate operations in the plant more efficiently. Its implementation involves a high level of computer networking within the factory.
10. *Computer-integrated manufacturing* (CIM). Taking the previous strategy one level higher, CIM involves extensive use of computer systems, databases, and networks throughout the enterprise to integrate the factory operations and business functions.

The ten strategies constitute a checklist of possibilities for improving the production system through automation or simplification. They should not be considered mutually exclusive. For most situations, multiple strategies can be implemented in one improvement project. The reader will see these strategies implemented in the many systems discussed throughout the book.

1.4.3 Automation Migration Strategy

Owing to competitive pressures in the marketplace, a company often needs to introduce a new product in the shortest possible time. As mentioned previously, the easiest and least expensive way to accomplish this objective is to design a manual production method, using a sequence of workstations operating independently. The tooling for a manual method can be fabricated quickly and at low cost. If more than a single set of workstations is required to make the product in sufficient quantities, as is often the case, then the manual cell is replicated as many times as needed to meet demand. If the product turns out to be successful, and high future demand is anticipated, then it makes sense for the company to automate production. The improvements are often carried out in phases. Many companies have an automation migration strategy, that is, a formalized plan for evolving the manufacturing systems used to produce new products as demand grows. A typical automation migration strategy is the following:

Phase 1: *Manual production* using single-station manned cells operating independently. This is used for introduction of the new product for reasons already mentioned: quick and low-cost tooling to get started.

Phase 2: *Automated production* using single-station automated cells operating independently. As demand for the product grows, and it becomes clear that automation can be justified, then the single stations are automated to reduce labor and increase production rate. Work units are still moved between workstations manually.

Phase 3: *Automated integrated production* using a multi-station automated system with serial operations and automated transfer of work units between stations. When the company is certain that the product will be produced in mass quantities and for several years, then integration of the single-station automated cells is warranted to further reduce labor and increase production rate.

This strategy is illustrated in Figure 1.6. Details of the automation migration strategy vary from company to company, depending on the types of products they make and the manufacturing processes they perform. But well-managed manufacturing companies

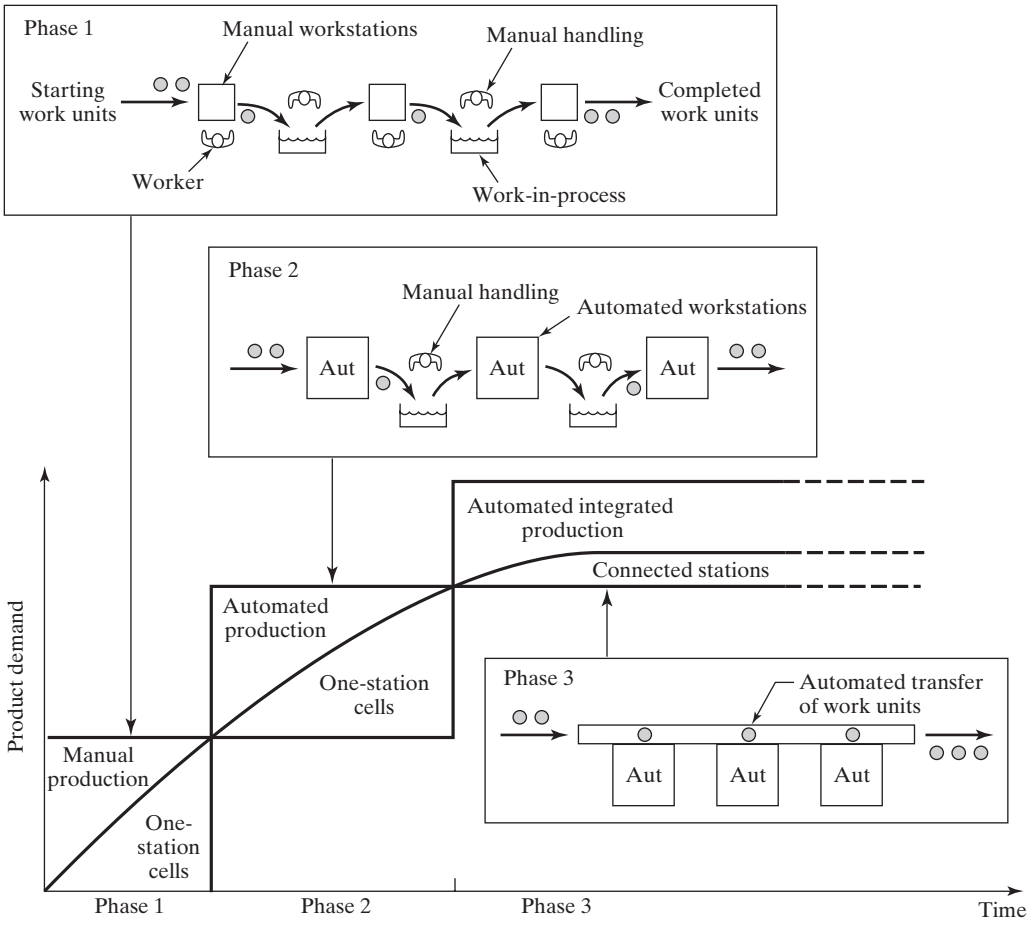


Figure 1.6 A typical automation migration strategy. Phase 1: manual production with single independent workstations. Phase 2: automated production stations with manual handling between stations. Phase 3: automated integrated production with automated handling between stations. Key: Aut = automated workstation.

have policies like the automation migration strategy. There are several advantages of such a strategy:

- It allows introduction of the new product in the shortest possible time, since production cells based on manual workstations are the easiest to design and implement.
- It allows automation to be introduced gradually (in planned phases), as demand for the product grows, engineering changes in the product are made, and time is provided to do a thorough design job on the automated manufacturing system.
- It avoids the commitment to a high level of automation from the start, because there is always a risk that demand for the product will not justify it.

2.1 MANUFACTURING INDUSTRIES AND PRODUCTS

Manufacturing is an important commercial activity, carried out by companies that sell products to customers. The type of manufacturing performed by a company depends on the kinds of products it makes.

Manufacturing Industries. Industry consists of enterprises and organizations that produce and/or supply goods and/or services. Industries can be classified as primary, secondary, and tertiary. ***Primary industries*** are those that cultivate and exploit natural resources, such as agriculture and mining. ***Secondary industries*** convert the outputs of the primary industries into products. Manufacturing is the principal activity in this category, but the secondary industries also include construction and power utilities. ***Tertiary industries*** constitute the service sector of the economy. A list of specific industries in these categories is presented in Table 2.1.

This book is concerned with the secondary industries (middle column in Table 2.1), which are composed of the companies engaged in manufacturing. It is useful to

TABLE 2.1 Specific Industries in the Primary, Secondary, and Tertiary Categories, Based Roughly on the International Standard Industrial Classification (ISIC) Used by the United Nations

Primary	Secondary	Tertiary (Service)
Agriculture	Aerospace	Banking
Forestry	Apparel	Communications
Fishing	Automotive	Education
Livestock	Basic metals	Entertainment
Quarrying	Beverages	Financial services
Mining	Building materials	Government
Petroleum	Chemicals	Health and medical services
	Computers	Hotels
	Construction	Information
	Consumer appliances	Insurance
	Electronics	Legal services
	Equipment	Real estate
	Fabricated metals	Repair and maintenance
	Food processing	Restaurants
	Glass, ceramics	Retail trade
	Heavy machinery	Tourism
	Paper	Transportation
	Petroleum refining	Wholesale trade
	Pharmaceuticals	
	Plastics (shaping)	
	Power utilities	
	Publishing	
	Textiles	
	Tire and rubber	
	Wood and furniture	

TABLE 2.2 International Standard Industrial Classification Codes for Various Industries in the Manufacturing Sector

Basic Code	Products Manufactured
31	Food, beverages (alcoholic and nonalcoholic), tobacco
32	Textiles, clothing, leather goods, fur products
33	Wood and wood products (e.g., furniture), cork products
34	Paper, paper products, printing, publishing, bookbinding
35	Chemicals, coal, petroleum, plastic, rubber, products made from these materials, pharmaceuticals
36	Ceramics (including glass), nonmetallic mineral products (e.g., cement)
37	Basic metals (steel, aluminum, etc.)
38	Fabricated metal products, machinery, equipment (e.g., aircraft, cameras, computers and other office equipment, machinery, motor vehicles, tools, televisions)
39	Other manufactured goods (e.g., jewelry, musical instruments, sporting goods, toys)

distinguish the process industries from the industries that make discrete parts and products. The process industries include chemicals, pharmaceuticals, petroleum, basic metals, food, beverages, and electric power generation. The discrete product industries include automobiles, aircraft, appliances, computers, machinery, and the component parts from which these products are assembled. The International Standard Industrial Classification (ISIC) of industries according to types of products manufactured is listed in Table 2.2. In general, the process industries are included within ISIC codes 31–37, and the discrete product manufacturing industries are included in ISIC codes 38 and 39. However, it must be acknowledged that many of the products made by the process industries are finally sold to the consumer in discrete units. For example, beverages are sold in bottles and cans. Pharmaceuticals are often purchased as pills and capsules.

Production operations in the process industries and the discrete product industries can be divided into continuous production and batch production. The differences are shown in Figure 2.2.

Continuous production occurs when the production equipment is used exclusively for the given product, and the output of the product is uninterrupted. In the process industries, continuous production means that the process is carried out on a continuous stream of material, with no interruptions in the output flow, as suggested by Figure 2.2(a). The material being processed is likely to be in the form of a liquid, gas, powder, or similar physical state. In the discrete manufacturing industries, continuous production means 100% dedication of the production equipment to the part or product, with no breaks for product changeovers. The individual units of production are identifiable, as in Figure 2.2(b).

Batch production occurs when the materials are processed in finite amounts or quantities. The finite amount or quantity of material is called a *batch* in both the process and discrete manufacturing industries. Batch production is discontinuous because there are interruptions in production between batches. Reasons for using batch production include (1) differences in work units between batches necessitate changes in methods, tooling, and equipment to accommodate the part differences; (2) the capacity of the

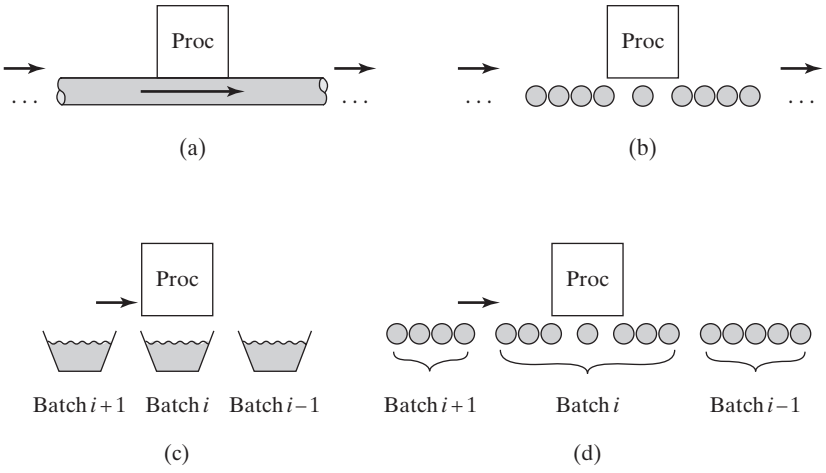


Figure 2.2 Continuous and batch production in the process and discrete manufacturing industries, including (a) continuous production in the process industries, (b) continuous production in the discrete manufacturing industries, (c) batch production in the process industries, and (d) batch production in the discrete manufacturing industries. Key: Proc = process.

equipment limits the amount or quantity of material that can be processed at one time; and (3) the production rate of the equipment is greater than the demand rate for the parts or products, and it therefore makes sense to share the equipment among multiple parts or products. The differences in batch production between the process and discrete manufacturing industries are portrayed in Figure 2.2(c) and (d). Batch production in the process industries generally means that the starting materials are in liquid or bulk form, and they are processed altogether as a unit. By contrast, in the discrete manufacturing industries, a batch is a certain quantity of work units, and the work units are usually processed one at a time rather than all together at once. The number of parts in a batch can range from as few as one to as many as thousands of units.

Manufactured Products. As indicated in Table 2.2, the secondary industries include food, beverages, textiles, wood, paper, publishing, chemicals, and basic metals (ISIC codes 31–37). The scope of this book is primarily directed at the industries that produce discrete products. Table 2.3 lists the manufacturing industries and corresponding products for which the production systems in this book are most applicable.

Final products made by the industries listed in Table 2.3 can be divided into two major classes: consumer goods and capital goods. *Consumer goods* are products purchased directly by consumers, such as cars, personal computers, TVs, tires, toys, and tennis rackets. *Capital goods* are products purchased by other companies to produce goods and supply services. Examples of capital goods include commercial aircraft, process control computers, machine tools, railroad equipment, and construction machinery.

In addition to final products, which are usually assembled, there are companies in industry whose business is primarily to produce materials, components, and supplies for the companies that make the final products. Examples of these items include sheet steel, bar

TABLE 2.3 Manufacturing Industries Whose Products Are Likely to Be Produced by the Production Systems Discussed in This Book

Industry	Typical Products
Aerospace	Commercial and military aircraft
Automotive	Cars, trucks, buses, motorcycles
Computers	Mainframe and personal computers
Consumer appliances	Large and small household appliances
Electronics	TVs, DVD players, audio equipment, video game consoles
Equipment	Industrial machinery, railroad equipment
Fabricated metals	Machined parts, metal stampings, tools
Glass, ceramics	Glass products, ceramic tools, pottery
Heavy machinery	Machine tools, construction equipment
Plastics (shaping)	Plastic moldings, extrusions
Tire and rubber	Tires, shoe soles, tennis balls

stock, metal stampings, machined parts, plastic moldings, cutting tools, dies, molds, and lubricants. Thus, the manufacturing industries consist of a complex infrastructure with various categories and layers of intermediate suppliers with whom the final consumer never deals.

2.2 MANUFACTURING OPERATIONS

There are certain basic activities that must be carried out in a factory to convert raw materials into finished products. For a plant engaged in making discrete products, the factory activities are (1) processing and assembly operations, (2) material handling, (3) inspection and test, and (4) coordination and control.

The first three activities are the physical activities that "touch" the product as it is being made. Processing and assembly operations alter the geometry, properties, and/or appearance of the work unit. They add value to the product. The product must be moved from one operation to the next in the manufacturing sequence, and it must be inspected and/or tested to ensure high quality. It is sometimes argued that material handling and inspection activities do not add value to the product. However, material handling and inspection may be required to accomplish the necessary processing and assembly operations, for example, loading parts into a production machine and assuring that a starting work unit is of acceptable quality before processing begins.

2.2.1 Processing and Assembly Operations

Manufacturing processes can be divided into two basic types: (1) processing operations and (2) assembly operations. A ***processing operation*** transforms a work material from one state of completion to a more advanced state that is closer to the final desired part or product. It adds value by changing the geometry, properties, or appearance of the starting material. In general, processing operations are performed on discrete work parts, but some processing operations are also applicable to assembled items, for example, painting a welded sheet metal car body. An ***assembly operation*** joins two or more components to create a new entity, which is called an assembly, subassembly, or some other term that

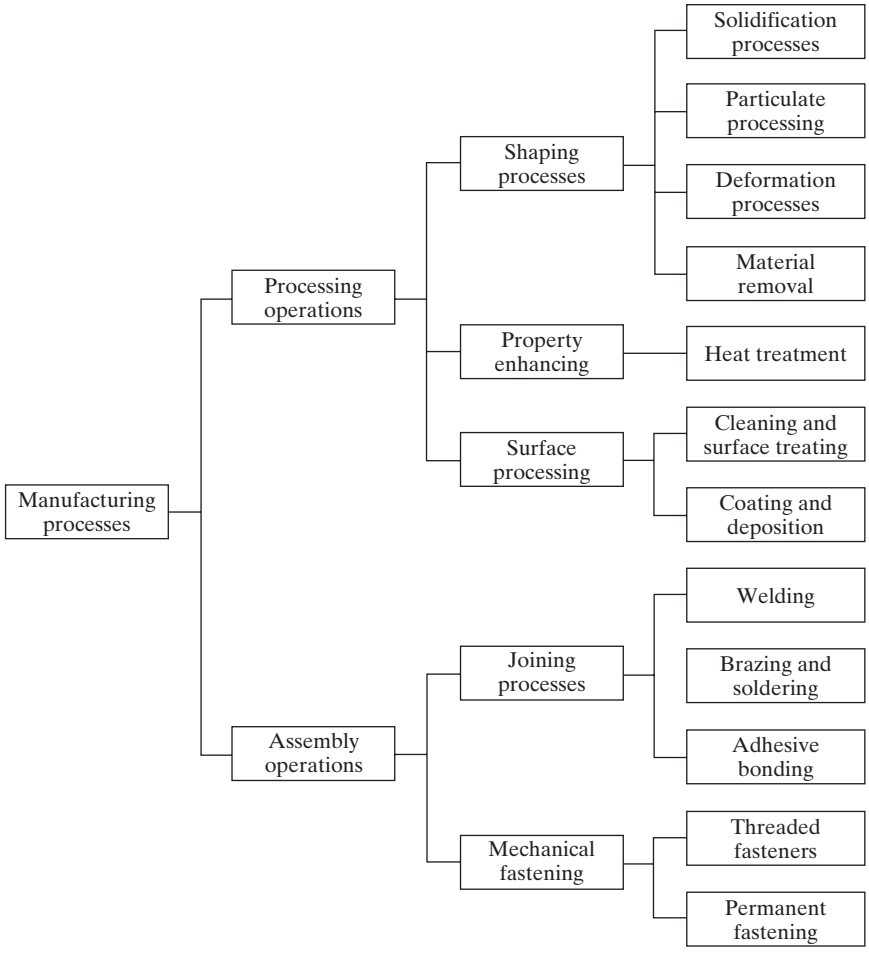


Figure 2.3 Classification of manufacturing processes.

refers to the specific joining process. Figure 2.3 shows a classification of manufacturing processes and how they divide into various categories.

Processing Operations. A processing operation uses energy to alter a work part's shape, physical properties, or appearance to add value to the material. The energy is applied in a controlled way by means of machinery and tooling. Human energy may also be required, but human workers are generally employed to control the machines, to oversee the operations, and to load and unload parts before and after each cycle of operation. A general model of a processing operation is illustrated in Figure 2.1(a). Material is fed into the process, energy is applied by the machinery and tooling to transform the material, and the completed work part exits the process. As shown in the model, most production operations produce waste or scrap, either as a natural by-product of the process (e.g., removing material as in machining) or in the form of occasional defective pieces. A desirable objective in manufacturing is to reduce waste in either of these forms.

More than one processing operation is usually required to transform the starting material into final form. The operations are performed in the particular sequence to achieve the geometry and/or condition defined by the design specification.

Three categories of processing operations are distinguished: (1) shaping operations, (2) property-enhancing operations, and (3) surface processing operations. ***Part-shaping operations*** apply mechanical force and/or heat or other forms and combinations of energy to change the geometry of the work material. There are various ways to classify these processes. The classification used here is based on the state of the starting material. There are four categories:

1. *Solidification processes.* The important processes in this category are casting (for metals) and molding (for plastics and glasses), in which the starting material is a heated liquid or semifluid, and it can be poured or otherwise forced to flow into a mold cavity where it cools and solidifies, taking a solid shape that is the same as the cavity.
2. *Particulate processing*. The starting material is a powder. The common technique involves pressing the powders in a die cavity under high pressure to cause the powders to take the shape of the cavity. However, the compacted work part lacks sufficient strength for any useful application. To increase strength, the part is then sintered—heated to a temperature below the melting point, which causes the individual particles to bond together. Both metals (powder metallurgy) and ceramics (e.g., clay products) can be formed by particulate processing.
3. *Deformation processes*. In most cases, the starting material is a ductile metal that is shaped by applying stresses that exceed the metal's yield strength. To increase ductility, the metal is often heated prior to forming. Deformation processes include forging, extrusion, and rolling. Also included in this category are sheet metal processes such as drawing, forming, and bending.
4. *Material removal processes*. The starting material is solid (commonly a metal, ductile or brittle), from which excess material is removed from the starting workpiece so that the resulting part has the desired geometry. Most important in this category are *machining* operations such as turning, drilling, and milling, accomplished using sharp cutting tools that are harder and stronger than the work metal. Grinding is another common process in this category, in which an abrasive grinding wheel is used to remove material. Other material removal processes are known as nontraditional processes because they do not use traditional cutting and grinding tools. Instead, they are based on lasers, electron beams, chemical erosion, electric discharge, or electrochemical energy.

In addition to these four categories based on starting material, there is also a family of part fabrication technologies called *additive manufacturing*. Also known as *rapid prototyping* (Section 23.1.2), these technologies operate on a variety of material types by building the part as a sequence of thin layers each on top of the previous until the entire solid geometry has been completed.

Property-enhancing operations are designed to improve mechanical or physical properties of the work material. The most important property-enhancing operations involve heat treatments, which include various temperature-induced strengthening and/or toughening processes for metals and glasses. Sintering of powdered metals and ceramics, mentioned previously, is also a heat treatment, which strengthens a pressed powder work part. Property-enhancing operations do not alter part shape, except unintentionally in

some cases, for example, warping of a metal part during heat treatment or shrinkage of a ceramic part during sintering.

Surface processing operations include (1) cleaning, (2) surface treatments, and (3) coating and thin film deposition processes. Cleaning includes both chemical and mechanical processes to remove dirt, oil, and other contaminants from the surface. Surface treatments include mechanical working, such as shot peening and sand blasting, and physical processes like diffusion and ion implantation. Coating and thin film deposition processes apply a coating of material to the exterior surface of the work part. Common coating processes include electroplating, anodizing of aluminum, and organic coating (call it painting). Thin film deposition processes include physical vapor deposition and chemical vapor deposition to form extremely thin coatings of various substances. Several surface processing operations have been adapted to fabricate semiconductor materials (most commonly silicon) into integrated circuits for microelectronics. These processes include chemical vapor deposition, physical vapor deposition, and oxidation. They are applied to very localized regions on the surface of a thin wafer of silicon (or other semiconductor material) to create the microscopic circuit.

Assembly Operations. The second basic type of manufacturing operation is assembly, in which two or more separate parts are joined to form a new entity. Components of the new entity are connected together either permanently or semipermanently. Permanent joining processes include welding, brazing, soldering, and adhesive bonding. They combine parts by forming a joint that cannot be easily disconnected. Mechanical assembly methods are available to fasten two or more parts together in a joint that can be conveniently disassembled. The use of threaded fasteners (e.g., screws, bolts, nuts) are important traditional methods in this category. Other mechanical assembly techniques that form a permanent connection include rivets, press fitting, and expansion fits. Special assembly methods are used in electronics. Some of the methods are identical to or adaptations of the above techniques. For example, soldering is widely used in electronics assembly. Electronics assembly is concerned primarily with the assembly of components (e.g., integrated circuit packages) to printed circuit boards to produce the complex circuits used in so many of today's products.

2.2.2 Other Factory Operations

Other activities that must be performed in the factory include material handling and storage, inspection and testing, and coordination and control.

Material Handling and Storage. Moving and storing materials between processing and/or assembly operations are usually required. In most manufacturing plants, materials spend more time being moved and stored than being processed. In some cases, the majority of the labor cost in the factory is consumed in handling, moving, and storing materials. It is important that this function be carried out as efficiently as possible. Part III of this book considers the material handling and storage technologies that are used in factory operations.

Eugene Merchant, an advocate and spokesman for the machine tool industry for many years, observed that materials in a typical metal machining batch factory or job shop spend more time waiting or being moved than being processed [4]. His observation is illustrated in Figure 2.4. About 95% of a part's time is spent either moving or waiting (temporary storage). Only 5% of its time is spent on the machine tool. Of this 5%, less

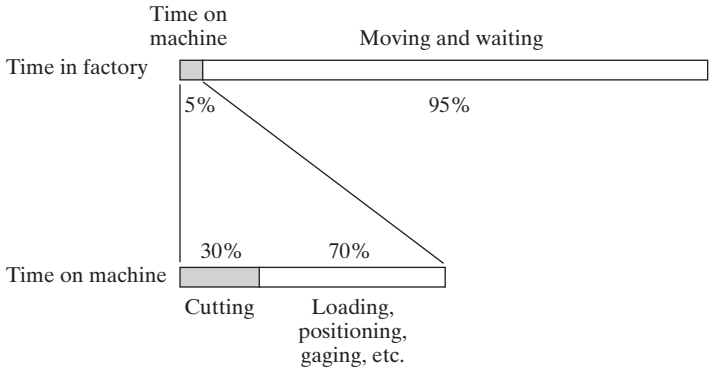


Figure 2.4 How time is spent by a typical part in a batch production machine shop [4].

than 30% of the time on the machine (1.5% of the total time of the part) is time during which actual cutting is taking place. The remaining 70% (3.5% of the total) is required for loading and unloading, part handling and positioning, tool positioning, gaging, and other elements of nonprocessing time. These time proportions indicate the significance of material handling and storage in a typical factory.

Inspection and Testing. Inspection and testing are quality control activities. The purpose of inspection is to determine whether the manufactured product meets the established design standards and specifications. For example, inspection examines whether the actual dimensions of a mechanical part are within the tolerances indicated on the engineering drawing for the part. Testing is generally concerned with the functional specifications of the final product rather than with the individual parts that go into the product. For example, final testing of the product ensures that it functions and operates in the manner specified by the product designer. Part V of the text examines inspection and testing.

Coordination and Control. Coordination and control in manufacturing include both the regulation of individual processing and assembly operations and the management of plant-level activities. Control at the process level involves the achievement of certain performance objectives by properly manipulating the inputs and other parameters of the process. Control at the process level is discussed in Part II of the book.

Control at the plant level includes effective use of labor, maintenance of the equipment, moving materials in the factory, controlling inventory, shipping products of good quality on schedule, and keeping plant operating costs to a minimum. The manufacturing control function at the plant level represents the major point of intersection between the physical operations in the factory and the information-processing activities that occur in production. Many of these plant- and enterprise-level control functions are discussed in Parts V and VI.

2.3 PRODUCTION FACILITIES

A manufacturing company attempts to organize its facilities in the most efficient way to serve the particular mission of each plant. Over the years, certain types of production facilities have come to be recognized as the most appropriate way to organize for a given

type of manufacturing. Of course, one of the most important factors that determine the type of manufacturing is the type of products that are made. As mentioned previously, this book is concerned primarily with the production of discrete parts and products. The quantity of parts and/or products made by a factory has a very significant influence on its facilities and the way manufacturing is organized. ***Production quantity*** refers to the number of units of a given part or product produced annually by the plant. The annual part or product quantities produced in a given factory can be classified into three ranges:

1. *Low production*: Quantities in the range of 1 to 100 units
2. *Medium production*: Quantities in the range of 100 to 10,000 units
3. *High production*: Production quantities are 10,000 to millions of units.

The boundaries between the three ranges are somewhat arbitrary (author's judgment). Depending on the types of products, these boundaries may shift by an order of magnitude or so.

Some plants produce a variety of different product types, each type being made in low or medium quantities. Other plants specialize in high production of only one product type. It is instructive to identify product variety as a parameter distinct from production quantity. ***Product variety*** refers to the different product designs or types that are produced in a plant. Different products have different shapes and sizes and styles, they perform different functions, they are sometimes intended for different markets, some have more components than others, and so forth. The number of different product types made each year can be counted. When the number of product types made in a factory is high, this indicates high product variety.

There is an inverse correlation between product variety and production quantity in terms of factory operations. When product variety is high, production quantity tends to be low, and vice versa. This relationship is depicted in Figure 2.5. Manufacturing plants tend to specialize in a combination of production quantity and product variety that lies somewhere inside the diagonal band in Figure 2.5. In general, a given factory tends to be limited to the product variety value that is correlated with that production quantity.

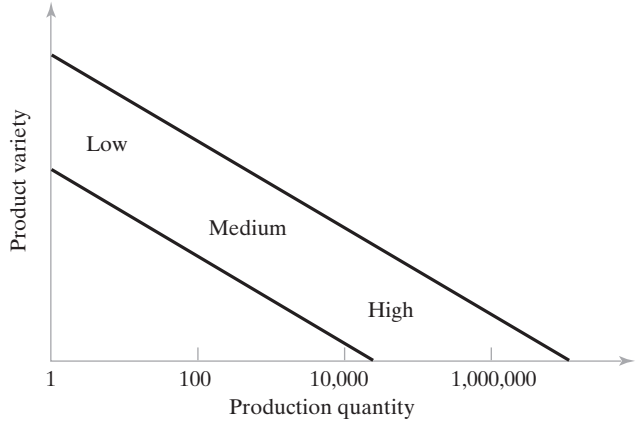


Figure 2.5 Relationship between product variety and production quantity in discrete product manufacturing.

Although product variety has been identified as a quantitative parameter (the number of different product types made by the plant or company), this parameter is much less exact than production quantity, because details on how much the designs differ are not captured simply by the number of different designs. The differences between an automobile and an air conditioner are far greater than between an air conditioner and a heat pump. Products can be different, but the extent of the differences may be small or great. The automotive industry provides some examples to illustrate this point. Each of the U.S. automotive companies produces cars with two or three different nameplates in the same assembly plant, although the body styles and other design features are nearly the same. In different plants, the same company builds trucks. Let the terms "hard" and "soft" be used to describe these differences in product variety. ***Hard product variety*** is when the products differ substantially. In an assembled product, hard variety is characterized by a low proportion of common parts among the products; in many cases, there are no common parts. The difference between a car and a truck is hard. ***Soft product variety*** is when there are only small differences between products, such as the differences between car models made on the same production line. There is a high proportion of common parts among assembled products whose variety is soft. The variety between different product categories tends to be hard; the variety between different models within the same product category tends to be soft.

The three production quantity ranges can be used to identify three basic categories of production plants. Although there are variations in the work organization within each category, usually depending on the amount of product variety, this is nevertheless a reasonable way to classify factories for the purpose of this discussion.

2.3.1 Low Production

The type of production facility usually associated with the quantity range of 1–100 units/year is the ***job shop***, which makes low quantities of specialized and customized products. The products are typically complex, such as experimental aircraft and special machinery. Job shop production can also include fabricating the component parts for the products. Customer orders for these kinds of items are often special, and repeat orders may never occur. Equipment in a job shop is general purpose and the labor force is highly skilled.

A job shop must be designed for maximum flexibility to deal with the wide part and product variations encountered (hard product variety). If the product is large and heavy, and therefore difficult to move in the factory, it typically remains in a single location, at least during its final assembly. Workers and processing equipment are brought to the product, rather than moving the product to the equipment. This type of layout is a ***fixed-position layout***, shown in Figure 2.6(a), in which the product remains in a single location during its entire fabrication. Examples of such products include ships, aircraft, railway locomotives, and heavy machinery. In actual practice, these items are usually built in large modules at single locations, and then the completed modules are brought together for final assembly using large-capacity cranes.

The individual parts that comprise these large products are often made in factories that have a ***process layout***, in which the equipment is arranged according to function or type. The lathes are in one department, the milling machines are in another department, and so on, as in Figure 2.6(b). Different parts, each requiring a different operation sequence, are routed through the departments in the particular order needed for their processing, usually in batches. The process layout is noted for its flexibility; it can accommodate a great variety of alternative operation sequences for different part configurations. Its disadvantage is that the machinery and methods to produce a part are not

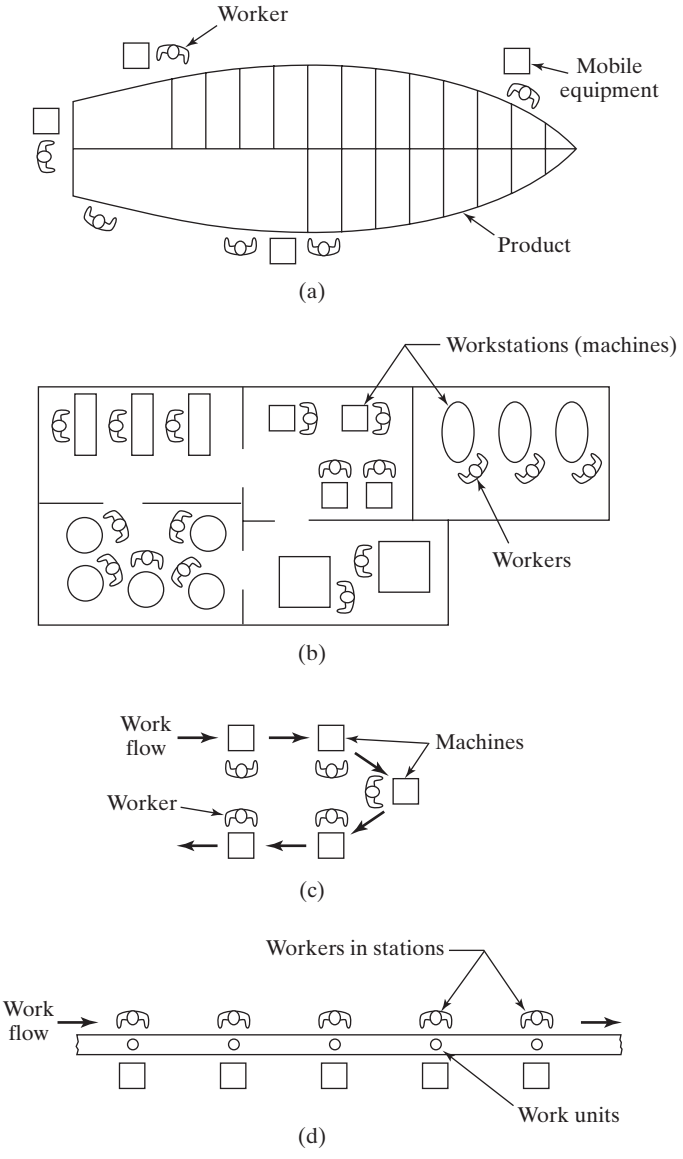


Figure 2.6 Various types of plant layout: (a) fixed-position layout, (b) process layout, (c) cellular layout, and (d) product layout.

designed for high efficiency. Much material handling is required to move parts between departments, so in-process inventory tends to be high.

2.3.2 Medium Production

In the medium quantity range (100–10,000 units annually), two different types of facility can be distinguished, depending on product variety. When product variety is hard, the traditional approach is ***batch production***, in which a batch of one product is made,

after which the facility is changed over to produce a batch of the next product, and so on. Orders for each product are frequently repeated. The production rate of the equipment is greater than the demand rate for any single product type, and so the same equipment can be shared among multiple products. The changeover between production runs takes time. Called the *setup time* or *changeover time*, it is the time to change tooling and to set up and reprogram the machinery. This is lost production time, which is a disadvantage of batch manufacturing. Batch production is commonly used in make-to-stock situations, in which items are manufactured to replenish inventory that has been gradually depleted by demand. The equipment for batch production is usually arranged in a process layout Figure 2.6(b).

An alternative approach to medium range production is possible if product variety is soft. In this case, extensive changeovers between one product style and the next may not be required. It is often possible to configure the equipment so that groups of similar parts or products can be made on the same equipment without significant lost time for changeovers. The processing or assembly of different parts or products is accomplished in cells consisting of several workstations or machines. The term *cellular manufacturing* is often associated with this type of production. Each cell is designed to produce a limited variety of part configurations; that is, the cell specializes in the production of a given set of similar parts or products, according to the principles of group technology (Chapter 18). The layout is called a *cellular layout*, depicted in Figure 2.6(c).

2.3.3 High Production

The high quantity range (10,000 to millions of units per year) is often referred to as *mass production*. The situation is characterized by a high demand rate for the product, and the production facility is dedicated to the manufacture of that product. Two categories of mass production can be distinguished: (1) quantity production and (2) flow-line production. ***Quantity production*** involves the mass production of single parts on single pieces of equipment. The method of production typically involves standard machines (such as stamping presses) equipped with special tooling (e.g., dies and material handling devices), in effect dedicating the equipment to the production of one part type. The typical layout used in quantity production is the process layout [Figure 2.6(b)].

Flow-line production involves multiple workstations arranged in sequence, and the parts or assemblies are physically moved through the sequence to complete the product. The workstations consist of production machines and/or workers equipped with specialized tools. The collection of stations is designed specifically for the product to maximize efficiency. This is a ***product layout***, in which the workstations are arranged into one long line, as depicted in Figure 2.6(d), or into a series of connected line segments. The work is usually moved between stations by powered conveyor. At each station, a small amount of the total work is completed on each unit of product.

The most familiar example of flow-line production is the assembly line, associated with products such as cars and household appliances. The pure case of flow-line production is where there is no variation in the products made on the line. Every product is identical, and the line is referred to as a *single-model production line*. However, to successfully market a given product, it is often necessary to introduce model variations so that individual customers can choose the exact style and options that appeal to them. From a production viewpoint, the model differences represent a case of soft product variety. The term *mixed-model production line* applies to those situations where there is soft

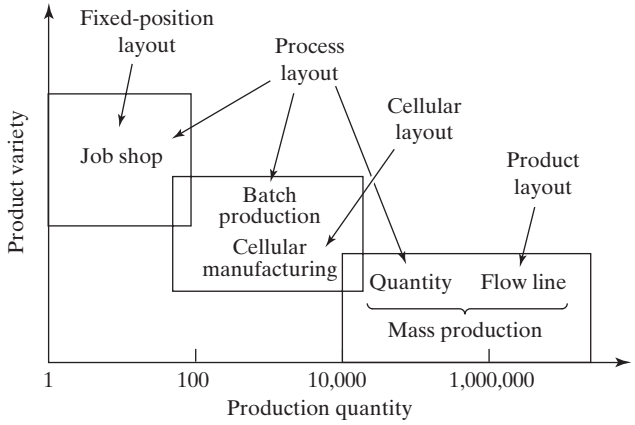


Figure 2.7 Types of facilities and layouts used for different levels of production quantity and product variety.

variety in the products made on the line. Modern automobile assembly is an example. Cars coming off the assembly line have variations in options and trim representing different models (and, in many cases, different nameplates) of the same basic car design. Other examples include small and major appliances. The Boeing Commercial Airplane Company uses production line techniques to assemble its 737 model.

Much of the discussion of the types of production facilities is summarized in Figure 2.7, which adds detail to Figure 2.5 by identifying the types of production facilities and plant layouts used. As Figure 2.7 shows, some overlap exists among the different facility types. Also note the comparison with earlier Figure 1.5, which indicates the type of automation that would be used in each facility type if the facility were automated.

2.4 PRODUCT/PRODUCTION RELATIONSHIPS

As noted in the preceding section, companies organize their production facilities and manufacturing systems in the most efficient manner for the particular products they make. It is instructive to recognize that there are certain product parameters that are influential in determining how the products are manufactured. Consider the following parameters: (1) production quantity, (2) product variety, (3) product complexity (of assembled products), and (4) part complexity.

2.4.1 Production Quantity and Product Variety

Production quantity and product variety were previously discussed in Section 2.3. The symbols Q and P can be used to represent these important parameters, respectively. Q refers to the number of units of a given part or product that are produced annually by a plant, both the quantities of each individual part or product style and the total quantity of all styles. Let each part or product style be identified using the subscript j, so that

Q_j = annual quantity of style j. Then let Q_f = total quantity of all parts or products made in the factory (the subscript f refers to factory). Q_j and Q_f are related as

$$Q_f = \sum_{j=1}^{P} Q_j \tag{2.1}$$

where P = total number of different part or product styles, and j is a subscript to identify products, $j = 1, 2, \ldots, P$.

P refers to the different product designs or types that are produced in a plant. It is a parameter that can be counted, and yet it must be recognized that the difference between products can be great or small, for example, the difference between hard product variety and soft product variety discussed in Section 2.3. Hard product variety is when the products differ substantially. Soft product variety is when there are only small differences between products. The parameter P can be divided into two levels, as in a tree structure. Call them P_1 and P_2. P_1 refers to the number of distinct product lines produced by the factory, and P_2 refers to the number of models in a product line. P_1 represents hard product variety and P_2 soft variety. The total number of product models is given by

$$P = \sum_{j=1}^{P_1} P_{2j} \tag{2.2}$$

where the subscript j identifies the product line: $j = 1, 2, \ldots, P_1$.

EXAMPLE 2.1 Product Lines and Product Models

A company specializes in home entertainment products. It produces only TVs and audio systems. Thus $P_1 = 2$. In its TV line it offers 15 different models, and in its audio line it offers 5 models. Thus for TVs, $P_2 = 15$, and for audio systems, $P_2 = 5$. The totality of product models offered is given by Equation (2.2):

$$P = \sum_{j=1}^{2} P_{2j} = 15 + 5 = 20$$

2.4.2 Product and Part Complexity

How complex is each product made in the plant? Product complexity is a complicated issue. It has both qualitative and quantitative aspects. For an assembled product, one possible quantitative indicator of product complexity is its number of components—the more parts, the more complex the product is. This is easily demonstrated by comparing the numbers of components in various assembled products, as in Table 2.4. The list demonstrates that the more components a product has, the more complex it tends to be.

For a manufactured component, a possible measure of part complexity is the number of processing steps required to produce it. An integrated circuit, which is technically a monolithic silicon chip with localized alterations in its surface chemistry, requires hundreds of processing steps in its fabrication. Although it may measure only 12 mm (0.5 in) on a side and 0.5-mm (0.020 in) thick, its complexity is orders of magnitude greater than a round

TABLE 2.4 Typical Number of Separate Components in Various Assembled Products (Compiled from [1], [3], and Other Sources)

Product (Approx. Date or Circa)	Approx. Number of Components
Mechanical pencil (modern)	10
Ball bearing (modern)	20
Rifle (1800)	50
Sewing machine (1875)	150
Bicycle chain	300
Bicycle (modern)	750
Early automobile (1910)	2,000
Automobile (modern)	10,000
Commercial airplane (1930)	100,000
Commercial airplane (modern)	4,000,000

TABLE 2.5 Typical Number of Processing Operations Required to Fabricate Various Parts

Part	Approx. Number of Processing Operations	Typical Processing Operations Used
Plastic molded part	1	Injection molding
Washer (stainless steel)	1	Stamping
Washer (plated steel)	2	Stamping, electroplating
Forged part	3	Heating, forging, trimming
Pump shaft	10	Machining (from bar stock)
Coated carbide cutting tool	15	Pressing, sintering, coating, grinding
Pump housing, machined	20	Casting, machining
V-6 engine block	50	Casting, machining
Integrated circuit chip	Hundreds	Photolithography, various thermal and chemical processes

washer of 12-mm (1/2-in) outside diameter, stamped out of 0.8-mm (1/32-in) thick stainless steel in one step. Table 2.5 is a list of manufactured parts with the typical number of processing operations required for each.

So, complexity of an assembled product can be defined as the number of distinct components; let n_p = the number of parts per product. And processing complexity of each part can be defined as the number of operations required to make it; let n_o = the number of operations or processing steps to make a part. As defined in Figure 2.8, three different types of production plant can be identified on the basis of n_p and n_o: parts producers, pure assembly plants, and vertically integrated plants.

Several relationships can be developed among the parameters P, Q, n_p, and n_o that indicate the level of activity in a manufacturing plant. Ignore the differences between P_1 and P_2 here, although Equation (2.2) could be used to convert these parameters into the corresponding P value. The total number of products made annually in a plant is the sum of the quantities of the individual product designs, as expressed in Equation (2.1). Assuming that the products are all assembled and that all component parts used in these

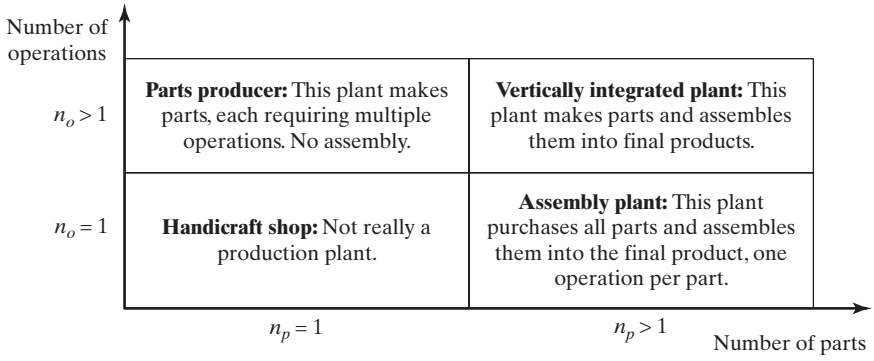


Figure 2.8 Production plants distinguished by number of parts n_p (for assembled products) and number of operations n_o (for manufactured parts).

products are made in the plant (no purchased components), the total number of parts manufactured by the plant per year is given by

$$n_{pf} = \sum_{j=1}^{P} Q_j n_{pj} \tag{2.3}$$

where n_{pf} = total number of parts made in the factory, pc/yr; Q_j = annual quantity of product style j, products/yr; and n_{pj} = number of parts in product j, pc/product.

Finally, if all parts are manufactured in the plant, then the total number of processing operations performed by the plant is given by

$$n_{of} = \sum_{j=1}^{P} Q_j \sum_{k=1}^{n_{pj}} n_{ojk} \tag{2.4}$$

where n_{of} = total number of operation cycles performed in the factory, ops/yr; and n_{ojk} = number of processing operations for each part k, summed over the number of parts in product j, n_{pj}. Parameter n_{of} provides a numerical value for the total level of part processing activity in the factory.

Average values of the four parameters P, Q, n_p, and n_o might be used to simplify and better conceptualize the factory model represented by Equations (2.1), (2.3), and (2.4). In this case, the total number of product units produced by the factory is given by

$$Q_f = PQ \tag{2.5}$$

where P = total number of product styles, Q_f = total quantity of products made in the factory and the average Q value is given by the following:

$$Q = \frac{\sum_{j=1}^{P} Q_j}{P} \tag{2.6}$$

The total number of parts produced by the factory is given by

$$n_{pf} = PQn_p \tag{2.7}$$

where the average n_p value is given by the following:

$$n_p = \frac{\sum_{j=1}^{P} Q_j n_{pj}}{PQ} \tag{2.8}$$

The total number of manufacturing operations performed by the factory is given by

$$n_{of} = PQn_p n_o \tag{2.9}$$

where the average n_o value is given by the following:

$$n_o = \frac{\sum_{j=1}^{P} Q_j \sum_{k=1}^{n_{pj}} n_{ojk}}{PQn_{pf}} \tag{2.10}$$

Using the simplified equations based on average values of the parameters, consider the following example.

EXAMPLE 2.2 A Production System Problem

Suppose a company has designed a new product line and is planning to build a new plant to manufacture this product line. The new line consists of 100 different product types, and for each product type the company wants to produce 10,000 units annually. The products average 1,000 components each, and the average number of processing steps required for each component is 10. All parts will be made in the factory. Each processing step takes an average of 1 min. Determine (a) how many products, (b) how many parts, and (c) how many production operations will be required each year, and (d) how many workers will be needed in the plant, if each worker works 8 hr per shift for 250 days/yr (2,000 hr/yr)?

Solution: (a) The total number of units to be produced by the factory annually is given by

$$Q = PQ = 100 \times 10{,}000 = \textbf{1,000,000 products}$$

(b) The total number of parts produced annually is

$$n_{pf} = PQn_p = 1{,}000{,}000 \times 1{,}000 = \textbf{1,000,000,000 parts}$$

(c) The number of distinct production operations is

$$n_{of} = PQn_p n_o = 1{,}000{,}000{,}000 \times 10 = \textbf{10,000,000,000 operations}$$

(d) First consider the total time TT to perform these operations. If each operation takes 1 min (1/60 hr),

$$TT = 10{,}000{,}000{,}000 \times 1/60 = \textbf{166,666,667 hr}$$

If each worker works 2,000 hr/yr, then the total number of workers required is

$$w = \frac{166{,}666{,}667}{2000} = \textbf{83,333 workers}$$

The factory in this example is a parts producer. If product assembly were accomplished in addition to parts production, then it would be a vertically integrated plant. In either case, it would be a big factory. The calculated number of workers only includes direct labor for parts production. Add indirect labor, staff, and management, and the number increases to well over 100,000 employees. Imagine the parking lot. And inside the factory, the logistics problems of dealing with all of the products, parts, and operations would be overwhelming. No organization in its right mind would consider building or operating such a plant today—not even the federal government.

2.4.3 Limitations and Capabilities of a Manufacturing Plant

Companies do not attempt the kind of factory in Example 2.2. Instead, today's factories are designed with much more specific missions. Referred to as *focused factories*, they are plants that concentrate "on a limited, concise, manageable set of products, technologies, volumes, and markets" [5]. It is a recognition that a manufacturing plant cannot do everything. It must limit its mission to a certain scope of products and activities in which it can best compete. Its size is typically about 500 workers or fewer, although the number may vary for different types of products and manufacturing operations.

Consider how a plant, or its parent company, limits the scope of its manufacturing operations and production systems. In limiting its scope, the plant in effect makes a set of deliberate decisions about what it will not try to do. Certainly one way to limit a plant's scope is to avoid being a fully integrated factory. Instead, the plant specializes in being either a parts producer or an assembly plant. Just as it decides what it will not do, the plant must also decide on the specific technologies, products, and volumes in which it will specialize. These decisions determine the plant's intended ***manufacturing capability***, which refers to the technical and physical limitations of a manufacturing firm and each of its plants. Several dimensions of this capability can be identified: (1) technological processing capability, (2) physical size and weight of product, and (3) production capacity.

Technological Processing Capability. The technological processing capability of a plant (or company) is its available set of manufacturing processes. Certain plants perform machining operations, others roll steel billets into sheet stock, and others build automobiles. A machine shop cannot roll steel, and a rolling mill cannot build cars. The underlying feature that distinguishes these plants is the set of processes they can perform. Technological processing capability is closely related to the material being processed. Certain manufacturing processes are suited to certain materials, while other processes are suited to other materials. By specializing in a certain process or group of processes, the plant is simultaneously specializing in a certain material type or range of materials.

Technological processing capability includes not only the physical processes, but also the expertise possessed by plant personnel in these processing technologies. Companies are limited by their available processes. They must focus on designing and manufacturing products for which their technological processing capability provides a competitive advantage.

Physical Product Limitations. A second aspect of manufacturing capability is imposed by the physical product. Given a plant with a certain set of processes, there are size and weight limitations on the products that can be accommodated in the plant. Big, heavy products are difficult to move. To move such products, the plant must be equipped with cranes of large load capacity. Smaller parts and products made in large quantities can be

moved by conveyor or fork lift truck. The limitation on product size and weight extends to the physical capacity of the manufacturing equipment as well. Production machines come in different sizes. Larger machines can be used to process larger parts. Smaller machines limit the size of the work that can be processed. The set of production equipment, material handling, storage capability, and plant size must be planned for products that lie within a certain size and weight range.

Production Capacity. A third limitation on a plant's manufacturing capability is the production quantity that can be produced in a given time period (e.g., month or year). Production capacity is defined as the maximum rate of production per period that a plant can achieve under assumed operating conditions. The operating conditions refer to the number of shifts per week, hours per shift, direct labor manning levels in the plant, and similar conditions under which the plant has been designed to operate. These factors represent inputs to the manufacturing plant. Given these inputs, how much output can the factory produce?

Plant capacity is often measured in terms of output units, such as annual tons of steel produced by a steel mill, or number of cars produced by a final assembly plant. In these cases, the outputs are homogeneous, more or less. In cases where the output units are not homogeneous, other factors may be more appropriate measures, such as available labor hours of productive capacity in a machine shop that produces a variety of parts.

REFERENCES

[1] Black, J. T., *The Design of the Factory with a Future*, McGraw-Hill, Inc., New York, NY, 1991.

[2] Groover, M. P., *Fundamentals of Modern Manufacturing: Materials, Processes, and Systems*, 5th ed., John Wiley & Sons, Inc., Hoboken, NJ, 2013.

[3] Hounshell, D. A., *From the American System to Mass Production, 1800–1932*, The Johns Hopkins University Press, Baltimore, MD, 1984.

[4] Merchant, M. E., "The Inexorable Push for Automated Production," *Production Engineering*, January 1977, pp. 45–46.

[5] Skinner, W., "The Focused Factory," *Harvard Business Review*, May–June 1974, pp. 113–121.

REVIEW QUESTIONS

2.1 What is manufacturing?

2.2 What are the three basic industry categories?

2.3 What is the difference between consumer goods and capital goods?

2.4 What is the difference between a processing operation and an assembly operation?

2.5 Name the four categories of part-shaping operations, based on the state of the starting work material.

2.6 Assembly operations can be classified as permanent joining methods and mechanical assembly. What are the four types of permanent joining methods?

2.7 What is the difference between hard product variety and soft product variety?

2.8 What type of production does a job shop perform?

4.1 BASIC ELEMENTS OF AN AUTOMATED SYSTEM

An automated system consists of three basic elements: (1) power to accomplish the process and operate the system, (2) a program of instructions to direct the process, and (3) a control system to actuate the instructions. The relationship among these elements is illustrated in Figure 4.2. All systems that qualify as being automated include these three basic elements in one form or another. They are present in the three basic types of automated manufacturing systems: fixed automation, programmable automation, and flexible automation (Section 1.2.1).

4.1.1 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 nonautomated processes:

- Electric power is widely available at moderate cost. It is an important part of the industrial infrastructure.
- Electric power can be readily converted to alternative energy forms: mechanical, thermal, light, acoustic, hydraulic, and pneumatic.
- Electric power at low levels can be used to accomplish functions such as signal transmission, information processing, and data storage and communication.
- Electric 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, atomic, solar, 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 4.1, a list of common

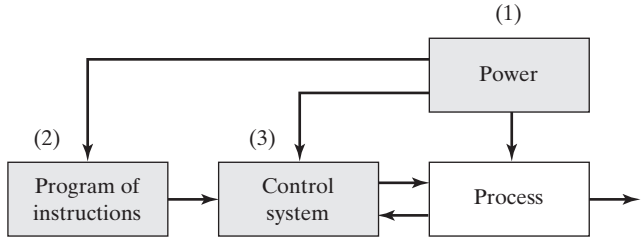


Figure 4.2 Elements of an automated system: (1) power, (2) program of instructions, and (3) control systems.

TABLE 4.1 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	Electrical	Metal removal is accomplished by a series of discrete electrical discharges between electrode (tool) and workpiece. The electric discharges cause very high localized temperatures that melt the metal.
Forging	Mechanical	Metal work part is deformed by opposing dies. Work parts are often heated in advance of deformation, thus thermal power is also required.
Heat-treating	Thermal	Metallic work unit is heated to temperature below melting point to effect microstructural 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 workpiece.
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.

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 earlier, the power source for each operation is often 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 4.1 are accomplished on discrete parts. These parts must be moved into the proper position and orientation for the process to be performed, and power is required for this transport and placement function. At the conclusion of the process, the work unit must be removed. If the process is completely automated, then some form of mechanized power is used. If the process is manually operated or semiautomated, 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. The material handling technologies associated with this transport function are covered in Chapter 10.

Power for Automation. Above and beyond the basic power requirements for the manufacturing operation, additional power is required for automation. The additional 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, perform the control calculations, and execute the instructions by transmitting the proper commands to 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* (Section 6.2). 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 be 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, for some processes, it is a legal requirement that records be kept of process performance and/or product quality. These data acquisition and record-keeping functions require power, although in modest amounts.

4.1.2 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, each part or product requires one or more processing steps that are unique to that part or product. These processing steps are performed during a work cycle. A new part is completed at the end of each work cycle (in some manufacturing operations, more than one part is produced during the work cycle: for example, 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, called *part programs* in numerical control (Chapter 7). 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. (It is assumed 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, so technically this is not a fully automated process.) In this case, programming simply involves setting the temperature dial on the furnace. This type of program is ***set-point control***, in which the set point is the value of the process parameter or desired value of the controlled variable in the process (furnace temperature in this example). A ***process parameter*** is an input to the process, such as the temperature dial setting, whereas a ***process variable*** is the corresponding output of the process, which is the actual temperature of the furnace.[1]

[1]Other examples of process parameters include desired coordinate axis value in a positioning system, valve open or closed in a fluid flow system, and motor on or off. Examples of corresponding process variables include the actual position of the coordinate axis, flow rate of fluid in the pipe, and rotational speed of the motor.

To change the program, the operator simply changes the dial setting. In an extension of this simple case, the one-step process is defined by more than one process parameter, for example, a furnace in which both temperature and atmosphere are controlled. Because of dynamics in the way the process operates, the process variable is not always equal to the process parameter. For example, if the temperature setting suddenly were to be increased or decreased, it would take time for the furnace temperature to reach the new set-point value. (This is getting into control system issues, which is the topic of Section 4.1.3.)

Work cycle programs are usually much more complicated than in the furnace example described. Following are five categories of work cycle programs, arranged in approximate order of increasing complexity and allowing for more than one process parameter in the program:

- *Set-point control,* in which the process parameter value is constant during the work cycle (as in the furnace example).
- *Logic control*, in which the process parameter value depends on the values of other variables in the process. Logic control is described in Section 9.1.1.
- *Sequence control*, in which the value of the process parameter changes as a function of time. The process parameter values can be either discrete (a sequence of step values) or continuously variable. Sequence control, also called *sequencing*, is discussed in Section 9.1.2.
- *Interactive program*, in which interaction occurs between a human operator and the control system during the work cycle.
- *Intelligent program*, in which the control system exhibits aspects of human intelligence (e.g., logic, decision making, cognition, learning) as a result of the work cycle program. Some capabilities of intelligent programs are discussed in Section 4.2.

Most processes involve 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 the following: (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.

EXAMPLE 4.1 An Automated Turning Operation

Consider an automated turning operation that generates a cone-shaped product. The system is automated and a robot loads and unloads the work units. The work cycle consists of the following steps: (1) load starting workpiece, (2) position cutting tool prior to turning, (3) turn, (4) reposition tool to a safe location at end of turning, and (5) unload finished workpiece. Identify the activities and process parameters for 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 retreating to a safe position to await 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 is the movement of the cutting tool to a "ready" position. The process parameters associated with this activity are the x- 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 workpiece (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 workpiece. For a consistent finish on the surface, the rotational speed must be continuously adjusted to maintain a constant surface speed (m/min); and for 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) are the reverse of steps (2) and (1), respectively, and the process parameters are the same.

Many production operations consist of multiple steps, sometimes more complicated than in the turning example. Examples of these operations include automatic screw machine cycles, sheet metal stamping, plastic injection molding, and die casting. Each of these manufacturing processes has been used for many decades. In earlier versions of these operations, work cycles were controlled by hardware components, such as limit switches, timers, cams, and electromechanical relays. In effect, the assemblage of hardware components served as the program of instructions that directed the sequence of steps in the processing cycle. Although these devices were quite adequate in performing their logic and sequencing functions, they suffered from the following disadvantages: (1) They often required considerable time to design and fabricate, 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 was 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 components, the programs for computer-controlled equipment are contained in compact disks (CD-ROMs), computer memory, and other modern storage technologies. Virtually all modern production equipment is designed with some form of computer controller to execute its respective processing cycles. The use of digital computers 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 hardware components mentioned earlier.

A work cycle may include manual steps, in which the operator performs certain activities during the work cycle, and the automated system performs the rest. These are referred to as *semiautomated* work cycles. A common example is the loading and unloading of parts by an operator into and from a numerical control machine between machining cycles, while the machine performs the cutting operation under part program control. Initiation of the cutting operation in 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 Example 4.1, the only two features of the work cycle were (1) the number and sequence of processing steps and (2) the process parameter changes in each step. Each work cycle consisted of the same steps and associated process parameter changes with no variation from one cycle to the next. The program of instructions is repeated 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). After the characters are entered, the system accomplishes the engraving automatically. (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 the teller machine must accomplish.)
- *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 some 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. These measures are discussed later in the chapter in the context of advanced automation functions (Section 4.2).

Various production situations and work cycle programs have been discussed here. The following summarizes the features of work cycle programs (part programs) used to direct the operations of an automated system:

- *Process parameters.* How many process parameters must be controlled during each step? Are the process parameters continuous or discrete? Do they change during the step, for example, a positioning system whose axis values change during the processing step?
- *Number of steps in work cycle.* How many distinct steps or work elements are included in the work cycle? A general sequence in discrete production operations is (1) load, (2), process, (3) unload, but the process may include multiple steps.
- *Manual participation in the work cycle.* Is a human worker required to perform certain steps in the work cycle, such as loading and unloading a production machine, or is the work cycle fully automated?
- *Operator interaction.* For example, is the operator required to enter processing data for each work cycle?
- *Variations in part or product styles.* Are the work units identical each cycle, as in mass production (fixed automation) or batch production (programmable automation), or are different part or product styles processed each cycle (flexible automation)?
- *Variations in starting work units.* Variations can occur in starting dimensions or materials. If the variations are significant, some adjustments may be required during the work cycle.

4.1.3 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 is to perform some manufacturing operation. A brief introduction to control systems is provided here. The following chapter describes this 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 4.3, a closed-loop control system consists of six basic elements: (1) input parameter, (2) process, (3) output

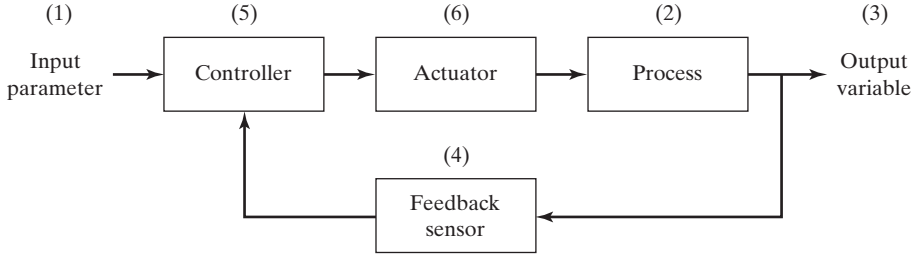


Figure 4.3 A feedback control system.

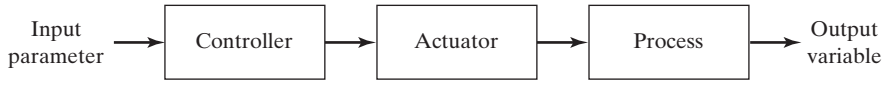


Figure 4.4 An open-loop control system.

variable, (4) feedback sensor, (5) controller, and (6) actuator. The input parameter (i.e., 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 electric motors or flow valves. It should be mentioned that Figure 4.3 shows only one loop. Most industrial processes require multiple loops, one for each process variable that must be controlled.

In contrast to a closed-loop control system, an ***open-loop control system*** operates without the feedback loop, as in Figure 4.4. In this case, the controls operate without measuring the output variable, so no comparison is made between the actual value of the output and the desired input parameter. The controller relies on an accurate model of the effect of its actuator on 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 actuator 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. Figure 4.5 illustrates the case of a closed-loop positioning system. 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

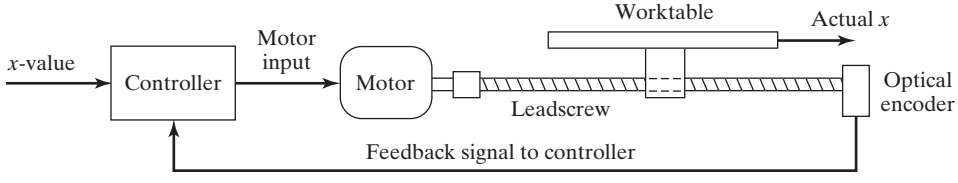


Figure 4.5 A (one-axis) positioning system consisting of a leadscrew driven by a dc servomotor.

control system for each axis, but the diagram only illustrates one of these axes. A dc servomotor connected to a leadscrew 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 leadscrew, whose rotation is converted into linear motion of the positioning table. The actual x-position is measured by a feedback sensor (e.g., an optical encoder). As the table moves closer to the desired x-coordinate value, the difference between the actual x-position and the input x-value decreases. The controller continues to drive the motor until the actual table position corresponds to the input position value.

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 would be 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 leadscrew, and the leadscrew 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 characteristics match the preceding list of operating conditions, an open-loop positioning system works with high reliability.

The engineering analysis of closed-loop and open-loop positioning systems is discussed in the context of numerical control in Section 7.4.

4.2 ADVANCED AUTOMATION FUNCTIONS

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 safety and performance 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 instructions. In some cases, the functions provide information only and do not involve any physical actions by the control system, for example, 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 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 equipment.

4.2.1 Safety Monitoring

One of the significant reasons for automating a manufacturing operation is to remove workers 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 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 comprising 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 of 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 include one or more of the following: (1) completely stopping the automated system, (2) sounding an alarm, (3) reducing the operating speed of the process, and (4) 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 (Section 4.2.3).

Sensors for safety monitoring range from very simple devices to highly sophisticated systems. Sensors are discussed in Section 6.1. 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 workholding 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 in 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 might 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 in the work cell.
- Machine vision systems to perform surveillance of 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.

4.2.2 Maintenance and Repair Diagnostics

Modern automated production systems are becoming increasingly complex and sophisticated, complicating the problem of maintaining and repairing them. Maintenance and repair diagnostics refers to the capabilities of an automated system to assist in identifying 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 monitors 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 its cause can be identified.
3. *Recommendation of repair procedure.* In the third mode of operation, the subsystem recommends to the repair crew 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 another time when the process is not operating, so the system experiences no loss of regular operation.

4.2.3 Error Detection and Recovery

In the operation of any automated system, there are hardware malfunctions and unexpected events. 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 a 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. The error detection step uses the automated system's available sensors to determine when a deviation or malfunction has occurred, interpret the sensor signal(s), and classify the error. Design of the error detection subsystem must begin with a systematic enumeration of all 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 (Section 20.4). 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 or drift in an equipment setting. These errors usually cause the product to deviate from specifications so as to be of unacceptable quality. 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, burst 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) anticipating all of the possible errors that can occur in a given process, and (2) specifying 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 detect and identify it.

EXAMPLE 4.2 Error Detection in an Automated Machining Cell

Consider an automated cell consisting of a CNC (computer numerical control) machine tool, a parts 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) workholding 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: Table 4.2 provides a list of the possible errors in the machining cell for each of the five categories.

TABLE 4.2 Possible Errors in the Automated Machining Cell

Category	Possible Errors
Machine and process	Loss of power, power overload, thermal deflection, cutting temperature too high, vibration, no coolant, chip fouling, wrong part program, defective part
Cutting tools	Tool breakage, tool wear-out, vibration, tool not present, wrong tool
Workholding fixture	Part not in fixture, clamps not actuated, part dislodged during machining, part deflection during machining, part breakage, chips causing location problems
Part storage unit	Work part not present, wrong work part, oversized or undersized work part
Load/unload robot	Improper grasping of work part, dropping of 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 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 detected error, 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 it is detected. However, the designated corrective action must be possible to accomplish while the work cycle is still being executed. If that is not possible, then the process must be stopped.
3. *Stop the process to invoke corrective action.* In this case, the deviation or malfunction requires that 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 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 (Section 5.3.2). 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 4.3 Error Recovery in an Automated Machining Cell

For the automated cell of Example 4.2, develop a list of possible corrective actions that might be taken by the system to address some of the errors.

Solution: A list of possible corrective actions is presented in Table 4.3.

TABLE 4.3 Error Recovery in an Automated Machining Cell: Possible Corrective Actions That Might Be Taken in Response to Errors Detected During the Operation

Error Detected	Possible Corrective Action 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)
Starting work part is 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)
Cutting tool failed	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)

4.3 LEVELS OF AUTOMATION

Automated systems can be applied to various levels of factory operations. One normally associates automation with the individual production machines. However, the production machine itself is made up of subsystems that may themselves be automated. For example, one of the important automation technologies discussed in this part of the book is computer numerical control (CNC, Chapter 7). A modern CNC machine tool is a highly automated system that is composed of multiple control systems. Any CNC machine has at least two axes of motion, and some machines have more than five axes. Each of these axes operates as a positioning system, as described in Section 4.1.3., and is, in effect, an automated system. Similarly, a CNC machine is often part of a larger manufacturing system, and the larger system may be automated. For example, two or three machine tools may be connected by an automated part handling system operating under computer control. The machine tools also receive instructions (e.g., part programs) from the computer. Thus three levels of automation and control are included here (the positioning system level, the machine tool level, and the manufacturing system level). For the purposes of this text, five levels of automation can be identified, and their hierarchy is depicted in Figure 4.6:

1. *Device level.* This is the lowest level in the automation hierarchy. It includes the actuators, sensors, and other hardware components that comprise the machine level. The devices are combined into the individual control loops of the machine, for example, the feedback control loop for one axis of a CNC machine or one joint of an industrial robot.

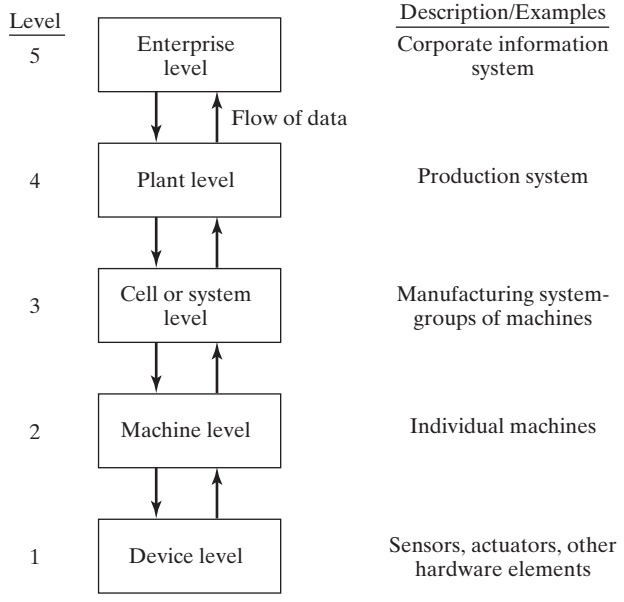


Figure 4.6 Five levels of automation and control in manufacturing.

2. *Machine level.* Hardware at the device level is assembled into individual machines. Examples include CNC machine tools and similar production equipment, industrial robots, powered conveyors, and automated guided vehicles. Control functions at this level include performing the sequence of steps in the program of instructions in the correct order and making sure that each step is properly executed.
3. *Cell or system level.* This is the manufacturing cell or system level, which operates under instructions from the plant level. A manufacturing cell or system is a group of machines or workstations connected and supported by a material handling system, computer, and other equipment appropriate to the manufacturing process. Production lines are included in this level. Functions include part dispatching and machine loading, coordination among machines and material handling system, and collecting and evaluating inspection data.
4. *Plant level.* This is the factory or production systems level. It receives instructions from the corporate information system and translates them into operational plans for production. Likely functions include order processing, process planning, inventory control, purchasing, material requirements planning, shop floor control, and quality control.
5. *Enterprise level.* This is the highest level, consisting of the corporate information system. It is concerned with all of the functions necessary to manage the company: marketing and sales, accounting, design, research, aggregate planning, and master production scheduling. The corporate information system is usually managed using Enterprise Resource Planning (Section 25.7).

Most of the technologies discussed in this part of the book are at levels 2 and 3 (machine level and cell level), although level 1 automation technologies (the devices that make up a control system) are discussed in Chapter 6. Level 2 technologies include the individual controllers (e.g., programmable logic controllers and digital computer controllers), numerical control machines, and industrial robots. The material handling equipment discussed in Part III also represent technologies at level 2, although some pieces of handling equipment are themselves sophisticated automated systems. The automation and control issues at level 2 are concerned with the basic operation of the equipment and the physical processes they perform.

Controllers, machines, and material handling equipment are combined into manufacturing cells, production lines, or similar systems, which make up level 3, considered in Part IV. A ***manufacturing system*** is defined in this book as a collection of integrated equipment designed for some special mission, such as machining a defined part family or assembling a certain product. Manufacturing systems include people. Certain highly automated manufacturing systems can operate for extended periods of time without humans present to attend to their needs. But most manufacturing systems include workers as important participants in the system, for example, assembly workers on a conveyorized production line or part loaders/unloaders in a machining cell. Thus, manufacturing systems are designed with varying degrees of automation; some are highly automated, others are completely manual, and there is a wide range between the two.

The manufacturing systems in a factory are components of a larger ***production system***, which is defined as the people, equipment, and procedures that are organized for the combination of materials and processes that comprise a company's manufacturing operations. Production systems are at level 4, the plant level, while manufacturing systems are at level 3 in the automation hierarchy. Production systems include not only the groups of machines and workstations in the factory but also the support procedures that make them work. These procedures include process planning, production control, inventory control, material requirements planning, shop floor control, and quality control, all of which are discussed in Parts V and VI. They are implemented not only at the plant level but also at the corporate level (level 5).

REFERENCES

[1] Boucher, T. O., *Computer Automation in Manufacturing*, Chapman & Hall, London, UK, 1996.

[2] Groover, M. P., "Automation," *Encyclopaedia Britannica, Macropaedia*, 15th ed., Chicago, IL, 1992. Vol. 14, pp. 548–557.

[3] Groover, M. P., "Automation," *Handbook of Design, Manufacturing, and Automation*, R. C. Dorf and A. Kusiak (eds.), John Wiley & Sons, Inc., New York, 1994, pp. 3–21.

[4] Groover, M. P., "Industrial Control Systems," *Maynard's Industrial Engineering Handbook*, 5th ed., K. Zandin (ed.), McGraw-Hill Book Company, New York, 2001.

[5] Platt, R., *Smithsonian Visual Timeline of Inventions,* Dorling Kindersley Ltd., London, UK, 1994.

[6] "The Power of Invention," *Newsweek Special Issue*, Winter 1997–98, pp. 6–79.

[7] www.wikipedia.org/wiki/Automation

equipment availability (a reliability measure), and manufacturing lead time. Manufacturing costs that are important to a company include labor and material costs, overhead costs, the cost of operating a given piece of equipment, and unit part and product costs.

3.1 PRODUCTION PERFORMANCE METRICS

In this section, various metrics of production performance are defined. The logical starting point is the cycle time for a unit operation, from which the production rate for the operation is derived. These unit operation metrics can be used to develop measures of performance at the factory level: production capacity, utilization, manufacturing lead time, and work-in-process.

3.1.1 Cycle Time and Production Rate

As described in the introduction to Chapter 2, manufacturing is almost always carried out as a sequence of unit operations, each of which transforms the part or product closer to its final form as defined by the engineering specifications. Unit operations are usually performed by production machines that are tended by workers, either full time or periodically in the case of automated equipment. In flow-line production (e.g., production lines), unit operations are performed at the workstations that comprise the line.

Cycle Time Analysis. For a unit operation, the cycle time T_c is the time that one work unit[1] spends being processed or assembled. It is the time interval between when one work unit begins processing (or assembly) and when the next unit begins. T_c is the time an individual part spends at the machine, but not all of this is processing time. In a typical processing operation, such as machining, T_c consists of (1) actual processing time, (2) work part handling time, and (3) tool handling time per workpiece. As an equation, this can be expressed as:

$$T_c = T_o + T_h + T_t \tag{3.1}$$

where T_c = cycle time, min/pc; T_o = time of the actual processing or assembly operation, min/pc; T_h = handling time, min/pc; and T_t = average tool handling time, min/pc, if such an activity is applicable. In a machining operation, tool handling time consists of time spent changing tools when they wear out, time changing from one tool to the next, tool indexing time for indexable inserts or for tools on a turret lathe or turret drill, tool repositioning for a next pass, and so on. Some of these tool handling activities do not occur every cycle; therefore, they must be apportioned over the number of parts between their occurrences to obtain an average time per workpiece.

Each of the terms, T_o, T_h, and T_t, has its counterpart in other types of discrete-item production. There is a portion of the cycle when the part is actually being processed (T_o); there is a portion of the cycle when the part is being handled (T_h); and there is, on average, a portion when the tooling is being adjusted or changed (T_t). Accordingly, Equation (3.1) can be generalized to cover most processing operations in manufacturing.

Production Rate. The production rate for a unit production operation is usually expressed as an hourly rate, that is, work units completed per hour (pc/hr). Consider

[1]As defined in Chapter 1, the ***work unit*** is the part or product being processed or assembled.

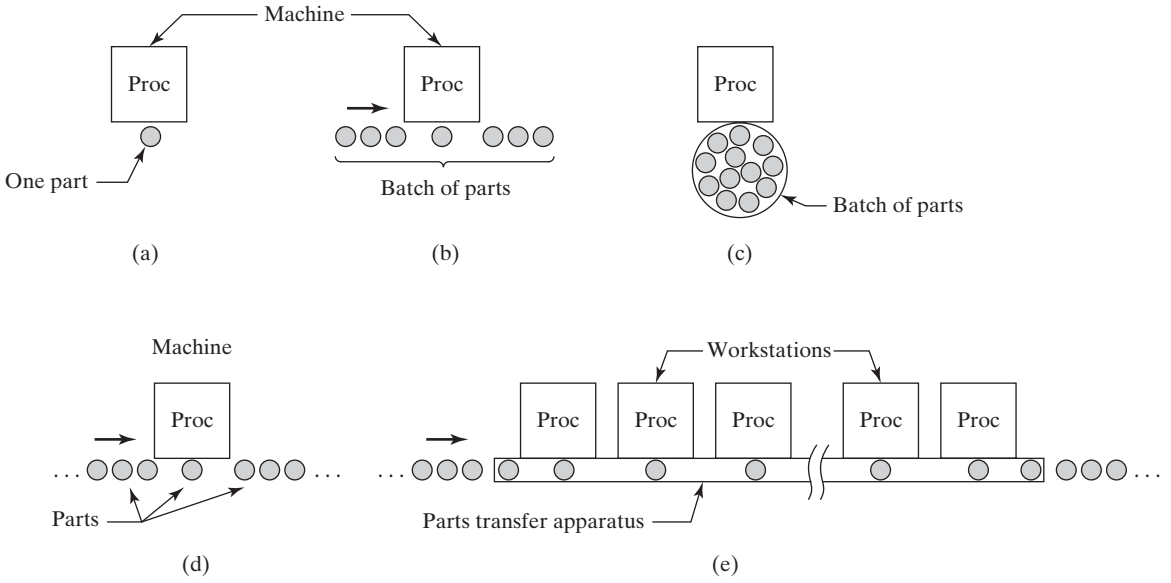


Figure 3.1 Types of production operations: (a) job shop with production quantity $Q = 1$, (b) sequential batch production, (c) simultaneous batch production, (d) quantity mass production, and (e) flow-line mass production. Key: Proc = process.

how the production rate is determined based on the operation cycle time for the three types of production: job shop production, batch production, and mass production. The various categories of production operations are depicted in Figure 3.1.

In job shop production, quantities are low ($1 \leq Q \leq 100$). At the extreme low end of the range, when quantity $Q = 1$, the production time per work unit is the sum of setup and cycle times:

$$T_p = T_{su} + T_c \tag{3.2}$$

where T_p = average production time, min/pc; T_{su} = setup time to prepare the machine to produce the part, min/pc; and T_c = cycle time from Equation (3.1). The production rate for the unit operation is simply the reciprocal of production time, usually expressed as an hourly rate:

$$R_p = \frac{60}{T_p} \tag{3.3}$$

where R_p = hourly production rate, pc/hr; T_p = production time from Equation (3.2), and the constant 60 converts minutes to hours. When the production quantity is greater than one, the analysis is the same as in batch production.

As noted in Section 2.1, batch production usually involves work units that are processed one at a time, referred to as sequential batch processing. Examples include machining, sheet metal stamping, and plastic injection molding. However, some batch production involves all work units in the batch being processed together, called simultaneous batch processing. Examples include most heat-treating and electroplating operations, in which all of the parts in the batch are processed at once.

In sequential batch processing, the time to process one batch consisting of Q work units is the sum of the setup time and processing time, where the processing time is the batch quantity multiplied by the cycle time; that is,

$$T_b = T_{su} + QT_c \tag{3.4a}$$

where T_b = batch processing time, min/batch; T_{su} = setup time to prepare the machine for the batch, min/batch; Q = batch quantity, pc/batch; and T_c = cycle time per work unit, min/cycle. If one work unit is completed each cycle, then T_c has units of min/pc. If more than one part is produced each cycle, then Equation (3.4) must be adjusted accordingly. An example of this situation is when the mold in a plastic injection molding operation contains two cavities, so that two moldings are produced each cycle.

In simultaneous batch processing, the time to process a batch consisting of Q work units is the sum of the setup time and processing time, where the processing time is the time to simultaneously process all of the parts in the batch; that is,

$$T_b = T_{su} + T_c \tag{3.4b}$$

where T_b = batch processing time, min/batch; T_{su} = setup time, min/batch; and T_c = cycle time per batch, min/cycle.

To obtain the average production time per work unit T_p for the unit operation, the batch time in Equation (3.4a) or (3.4b) is divided by the batch quantity:

$$T_p = \frac{T_b}{Q} \tag{3.5}$$

and production rate is calculated using Equation (3.3).

For quantity-type mass production, the production rate equals the cycle rate of the machine (reciprocal of operation cycle time) after production is underway and the effects of setup time become insignificant. That is, as Q becomes very large, $(T_{su}/Q) \rightarrow 0$ and

$$R_p \rightarrow R_c = \frac{60}{T_c} \tag{3.6}$$

where R_c = operation cycle rate of the machine, pc/hr, and T_c = operation cycle time, min/pc.

For flow-line mass production, the production rate approximates the cycle rate of the production line, again neglecting setup time. However, the operation of production lines is complicated by the interdependence of the workstations on the line. One complication is that it is usually impossible to divide the total work equally among all of the workstations on the line; therefore, one station ends up with the longest operation time, and this station sets the pace for the entire line. The term *bottleneck station* is sometimes used to refer to this station. Also included in the cycle time is the time to move parts from one station to the next at the end of each operation. In many production lines, all work units on the line are moved synchronously, each to its respective next station. Taking these factors into account, the cycle time of a production line is the longest processing (or assembly) time plus the time to transfer work units between stations. This can be expressed as

$$T_c = \text{Max } T_o + T_r \tag{3.7}$$

where T_c = cycle time of the production line, min/cycle; Max T_o = the operation time at the bottleneck station (the maximum of the operation times for all stations on the

line, min/cycle); and T_r = time to transfer work units between stations each cycle, min/cycle. T_r is analogous to T_h in Equation (3.1). The tool handling time T_t is usually accomplished as a maintenance function and is not included in the calculation of cycle time. Theoretically, the production rate can be determined by taking the reciprocal of T_c as

$$R_c = \frac{60}{T_c} \tag{3.8}$$

where R_c = theoretical or ideal production rate, but call it the cycle rate to be more precise, cycles/hr, and T_c = cycle time from Equation (3.7).

The preceding equations for cycle time and production rate ignore the issue of defective parts and products made in the operation. Although perfect quality is an ideal goal in manufacturing, the reality is that some processes produce defects. The issue of scrap rates and their effects on production quantities and costs in both unit operations and sequences of unit operations is considered in Chapter 21 on inspection principles and practices.

Equipment Reliability. Lost production time due to equipment reliability problems reduces the production rates determined by the previous equations. The most useful measure of reliability is ***availability***, defined as the uptime proportion of the equipment; that is, the proportion of time that the equipment is capable of operating (not broken down) relative to the scheduled hours of production. The measure is especially appropriate for automated production equipment.

Availability can also be defined using two other reliability terms, *mean time between failures* (*MTBF*) and *mean time to repair* (*MTTR*). As depicted in Figure 3.2, *MTBF* is the average length of time the piece of equipment runs between breakdowns, and *MTTR* is the average time required to service the equipment and put it back into operation when a breakdown occurs. In equation form,

$$A = \frac{MTBF - MTTR}{MTBF} \tag{3.9}$$

where A = availability (proportion); $MTBF$ = mean time between failures, hr; and $MTTR$ = mean time to repair, hr. The mean time to repair may include waiting time of the broken-down equipment before repairs begin. Availability is typically expressed as

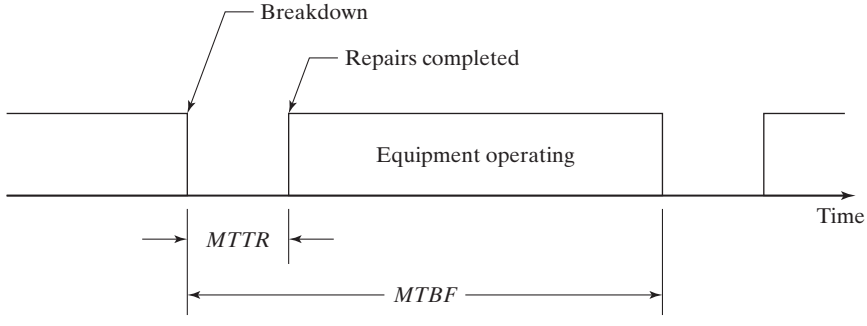


Figure 3.2 Time scale showing *MTBF* and *MTTR* used to define availability *A*.

a percentage. When a piece of equipment is brand new (and being debugged), and later when it begins to age, its availability tends to be lower.

Taking availability into account, the actual average production rate of the equipment is its availability multiplied by R_p from any of the preceding production rate equations (i.e., average production rate $= AR_p$), based on the assumption that setup time is also affected by the availability.

Reliability is particularly bothersome in the operation of automated production lines. This is because of the interdependence of workstations in an automated line, in which the entire line is forced to stop when one station breaks down. The actual average production rate R_p is reduced to a value that is often substantially below the ideal R_c given by Equation (3.8). The effect of reliability on manual and automated production lines and automated assembly systems is examined in Chapters 15 through 17.

3.1.2 Production Capacity and Utilization

Production capacity was discussed in the context of manufacturing capabilities in Section 2.4.3. It is defined as the maximum rate of output that a production facility (or production line, or group of machines) is able to produce under a given set of assumed operating conditions. The production facility usually refers to a plant or factory, and so the term *plant capacity* is often used for this measure. One might say that plant capacity is to the aggregate plant operation as production rate is to the unit operation. As mentioned before, the assumed operating conditions refer to the number of shifts per day (one, two, or three), number of days in the week that the plant operates, employment levels, and so forth.

The number of hours of plant operation per week is a critical issue in defining plant capacity. For continuous chemical production in which the reactions occur at elevated temperatures, the plant is usually operated 24 hours per day, seven days per week (168 hours per week). On the other hand, many discrete product plants operate one shift per day, five days per week. For an automobile final assembly plant, capacity is typically defined as one or two shifts, depending on the demand for the cars made in the plant. In situations when demand is very high, three production shifts may be used. A trend in manufacturing is to define plant capacity for the full 7-day week, 24 hours per day. This is the maximum time available, and if the plant operates fewer hours, then it is operating at less than its full capacity.

Determining Plant Capacity. Quantitative measures of plant capacity can be developed based on the production rate models derived earlier. Let $PC =$ the production capacity of a given facility, where the measure of capacity is the number of units produced per time period (e.g., week, month, year). The simplest case is where there are n production machines in the plant and they all produce the same part or product, which implies quantity-type mass production. Each machine is capable of producing at the same rate of R_p units per hour, as defined by Equation (3.6). Each machine operates for the number of hours in the period. These parameters can be combined to calculate the weekly production capacity of the facility,

$$PC = nH_{pc}R_p \tag{3.10}$$

where $PC =$ production capacity, pc/period; $n =$ number of machines; and $H_{pc} =$ the number of hours in the period being used to measure production capacity (or plant capacity).

TABLE 3.1 Number of Hours of Plant Operation for Various Periods and Operating Conditions.

Operating Conditions	Period		
	Week	Month	Year
One 8-hr shift, 5 days/week, 50 weeks/year	40	167	2000
Two 8-hr shifts, 5 days/week, 50 weeks/year	80	333	4000
Three 8-hr shifts, 5 days/week, 50 weeks/year	120	500	6000
One 8-hr shift, 7 days/week, 50 weeks/year	56	233	2800
Two 8-hr shifts, 7 days/week, 50 weeks/year	112	467	5600
Three 8-hr shifts, 7 days/week, 50 weeks/year	168	700	8400
24 hr/day, 7 days/week, 52 weeks/year (24/7)	168	728	8736

Table 3.1 lists the number of hours of plant operation for various periods and operating conditions. Consistent with the definition of production capacity given earlier, Equation (3.10) assumes that all machines are operating full time during the entire period defined by H_{pc}.

EXAMPLE 3.1 Production Capacity

The automatic lathe department has five machines, all devoted to the production of the same product. The machines operate two 8-hr shifts, 5 days/week, 50 weeks/year. Production rate of each machine is 15 unit/hr. Determine the weekly production capacity of the automatic lathe department.

Solution: From Equation (3.10) and Table 3.1,

$$PC = 5(80)(15) = \textbf{6,000 pc/wk}$$

In cases in which different machines produce different parts at different production rates, the following equation applies for quantity-type mass production:

$$PC = H_{pc} \sum_{i=1}^{n} R_{pi} \tag{3.11}$$

where n = number of machines in the plant, and R_{pi} = hourly production rate of machine i, and all machines are operating full time during the entire period defined by H_{pc}.

In job shop and batch production, each machine may be used to produce more than one batch, where each batch is made up of a different part style j. Let f_{ij} = the fraction of time during the period that machine i is processing part style j. Under normal operating conditions, it follows that for each machine i,

$$0 \leq \sum_{j} f_{ij} \leq 1 \text{ where } 0 \leq f_{ij} \leq 1 \text{ for all } i \tag{3.12}$$

The lower limit in Equation (3.12) indicates that the machine is idle during the entire week. Values between 0 and 1 mean that the machine experiences idle time during the week. The upper limit means that the machine is utilized 100% of the time during

the week. If the upper limit is exceeded ($\Sigma f_{ij} > 1$), then this can be interpreted as the machine being used on an overtime basis beyond the number of hours H_{pc} in the definition of plant capacity.

The production output of the plant must include the effect of operation sequence for part or product j. This is accomplished by dividing the production rate for each machine that participates in the production of part j by the number of operations in the operation sequence for that part, n_{oj}. The resulting average hourly production output for the plant is given by:

$$R_{pph} = \sum_{i=1}^{n} \sum_{j} f_{ij} R_{pij} / n_{oj} \tag{3.13}$$

where R_{pph} = average hourly plant production rate, pc/hr; R_{pij} = production rate of machine i when processing part j, pc/hr; n_{oj} = the number of operations required to produce part j, and f_{ij} is defined earlier. The individual values of R_{pij} are determined based on Equations (3.3) and (3.5), specifically:

$$R_{pij} = \frac{60}{T_{pij}} \text{ where } T_{pij} = \frac{T_{suij} + Q_j T_{cij}}{Q_j}$$

where T_{pij} = average production time for part j on machine i, min/pc; T_{suij} = setup time for part j on machine i, min/batch; and Q_j = batch quantity of part j, pc/batch.

The plant output for a given period of interest (e.g., week, month, year) can be determined based on the average hourly production rate given by Equation (3.13). For example, weekly plant output is given by the following:

$$R_{ppw} = H_{pw} R_{pph} \tag{3.14}$$

where R_{ppw} = weekly plant production rate for the plant, pc/wk; R_{pph} = average hourly production rate for the plant, pc/hr, from Equation (3.13); and H_{pw} = number of hours in the week from Table 3.1. If the period of interest is a month, then $R_{ppm} = H_{pm} R_{pph}$, and if the period is a year, then $R_{ppy} = H_{py} R_{pph}$.

Most manufacturing is accomplished in batches, and most manufactured products require a sequence of processing steps on multiple machines. Just as there is a bottleneck station in flow-line production, it is not unusual for certain machines in a given plant to limit the production output of the plant. They determine the plant capacity. These machines operate at 100% utilization while other machines in the sequence have lower utilizations. The net result is that the average equipment utilization in the plant is less than 100%, but the plant is still operating at its maximum capacity due to the limitations of these bottleneck operations. If this is the situation, then weekly production capacity is given by Equation (3.14), that is, $PC = R_{ppw}$.

EXAMPLE 3.2 Weekly Production Rate

A small machine shop has two machines and works 40 hr/wk. During a week of interest, four batches of parts were processed through these machines. Batch quantities, batch times, and operation sequences for the parts are given in the table below. Determine (a) weekly production output of the shop and (b) whether this represents the weekly plant capacity.

	Machine 1		Machine 2	
Part	R_p	Duration	R_p	Duration
A	25 pc/hr	12 hr	30 pc/hr	10 hr
B	10 pc/hr	20 hr		
C			7.5 pc/hr	24 hr
D			20 pc/hr	6 hr

Solution: (a) To determine the weekly production output, the f_{ij} values are determined as follows, given 40 hr per week: $f_{1A} = 12/40 = 0.30$, $f_{1B} = 20/40 = 0.50$, $f_{2A} = 10/40 = 0.25$, $f_{2C} = 24/40 = 0.60$, and $f_{2D} = 6/40 = 0.15$. The fraction of idle time on machine 1 is $= 8/40 = 0.20$. Noting that part A has 2 operations in its operation sequence and the other parts have 1, the hourly production rate of parts completed in the plant is given by Equation (3.13):

$$R_{pph} = \frac{0.3(25)}{2} + 0.5(10) + \frac{0.25(30)}{2} + 0.6(7.5) + 0.15(20) = 20 \text{ pc/hr}$$

Weekly production output $R_{ppw} = 40(20) =$ **800 pc/wk**

(b) Machine 2 is operating the full 40 hr/wk. Given the part mix in the problem, machine 2 is the bottleneck in the plant, and so the 800 pc/wk represents plant capacity: $\boldsymbol{PC_w}$ = **800 pc/wk**.

Comment: Machine 1 is only operating 32 hr/wk, so it might be inferred from the situation that the production of part B could be increased by 80 units (8 hr $\times$ 10 pc/hr) to achieve a plant capacity of 880 pc/wk. The question is whether there would be a demand for those 80 additional units of part B.

Utilization. Utilization is the proportion of time that a productive resource (e.g., a production machine) is used relative to the time available under the definition of plant capacity. Expressing this as an equation,

$$U_i = \sum_j f_{ij} \tag{3.15}$$

where U_i = utilization of machine i, and f_{ij} = the fraction of time during the available hours that machine i is processing part style j. An overall utilization for the plant is determined by averaging the U_i values over the number of machines:

$$U = \frac{\sum_{i=1}^{n} \sum_j f_{ij}}{n} = \frac{\sum_j U_i}{n} \tag{3.16}$$

The trouble with Equation (3.13) is that weekly production rate is the sum of the outputs of a mixture of part or product styles. The mixture is likely to change from week to week, so that parts with different production rates are produced in different weeks.

During one week, output might be higher than average simply because the production rates of the parts produced that week were high. To deal with this possible inconsistency, plant capacity is sometimes reported as the workload corresponding to the output produced during the period. ***Workload*** is defined as the total hours required to produce a given number of units during a given week or other period of interest. That is,

$$WL = \sum_i \sum_j Q_{ij} T_{pij} \tag{3.17}$$

where WL = workload, hr; Q_{ij} = number of work units produced of part style j on machine i during the period of interest; and T_{pij} = average production time of part style j on machine i. In Example 3.2, the workload is the sum of the duration hours listed for machines 1 and 2, a total of 72 hr. When used as the definition of plant capacity, workload refers to the maximum number of hours of work that the plant is capable of completing in the period of interest, which is 80 hr in Example 3.2.

Adjusting Plant Capacity. The preceding equations and examples indicate the operating parameters that affect plant capacity. Changes that can be made to increase or decrease plant capacity over the short term are listed below:

- Increase or decrease the number of machines n in the plant. It is easier to remove machines from operation than to add machines if adding them means purchasing equipment that may require long lead times to procure. Adding workers in the short term may be easier than adding equipment.
- Increase or decrease the number of shifts per week. For example, Saturday shifts might be authorized to temporarily increase capacity, or the plant might operate two shifts per day instead of one.
- Increase or decrease the number of hours worked per shift. For example, overtime on each regular shift might be authorized to increase capacity.

Over the intermediate and longer terms, the following changes can be made to increase plant capacity:

- Increase the number of machines n in the shop. This might be done by using equipment that was formerly not in use, acquiring new machines, and hiring new workers.
- Increase the production rate R_p by making improvements in methods and/or processing technology.
- Reduce the number of operations n_o in the operation sequence of parts by using combined operations, simultaneous operations, and/or integration of operations (Section 1.4.2, strategies 2, 3, and 4).

Other adjustments that can be considered to affect plant capacity in the short term or long term include the following:

- Identify the bottleneck operations in the plant and somehow increase the output rates of these operations, using the USA Principle and other approaches outlined in Section 1.4. Bottleneck operations in a batch manufacturing plant usually reveal themselves in one or both of the following ways: (1) These machines are always busy; they operate at 100% utilization; and (2) they have large queues of work waiting in front of them.

- Stockpile inventory to maintain level employment during slow periods, trusting (and betting) that the goods can later be sold when demand increases.
- Backlogging orders, which means delaying deliveries to customers during busy periods to avoid temporary and potentially costly increases in production capacity.
- Subcontracting work to outside vendors during busy periods or taking in extra work from other firms during slack periods.

3.1.3 Manufacturing Lead Time and Work-In-Process

In the competitive environment of global commerce, the ability of a manufacturing firm to deliver a product to the customer in the shortest possible time often wins the order. This section examines this performance measure, called manufacturing lead time (MLT). Closely correlated with MLT is the amount of inventory located in the plant as partially completed product, called work-in-process (WIP). When there is too much work-in-process, manufacturing lead time tends to be long.

Manufacturing Lead Time. MLT is defined as the total time required to process a given part or product through the plant, including any time due to delays, parts being moved between operations, time spent in queues, and so on. As noted previously, production usually consists of a sequence of unit processing operations. Between the unit operations are these nonproductive elements, which typically consume large blocks of time (recall the Merchant study, Section 2.2.2). Thus, production activities can be divided into two categories, unit operations and nonoperation times.

The reader may be wondering: Why do these nonoperation times occur? Why not just take the parts straightaway from one operation to the next without these delays? Some of the reasons why nonoperation time occurs between unit operations are the following: (1) time spent transporting batches of parts between operations, (2) buildup of queues of parts waiting before each operation, (3) buildup of queues of parts after each operation waiting to be transported to the next operation, (4) less than optimal scheduling of batches, (5) part inspections before and/or after unit operations, (6) equipment breakdowns resulting in lost production time, and (7) workload imbalances among the machines that perform the operations required for a given part or product style, with some machines being 100% utilized while others spend much of the time waiting for work.

Let T_c = the operation cycle time at a given machine, and T_{no} = the nonoperation time associated with each operation. Further, suppose that the number of separate operations (machines) through which the work unit must be routed = n_o. In batch production, there are Q work units in the batch. A setup is generally required to prepare each machine for the particular product, which requires a time = T_{su}. Given these terms, manufacturing lead time for a given batch is defined as

$$MLT_j = \sum_{i=1}^{n_{oj}} (T_{suij} + Q_j T_{cij} + T_{noij}) \tag{3.18}$$

where MLT_j = manufacturing lead time for a batch of part or product j, min; T_{suij} = setup time for operation i on part or product j, min; Q_j = quantity of part or product j in the batch being processed, pc; T_{cij} = cycle time for operation i on part or product j, min/pc; T_{noij} = nonoperation time associated with operation i, min; and i indicates the operation sequence in the processing, $i = 1, 2, \ldots, n_{oj}$. The MLT equation does not include the

time the raw work part spends in storage before its turn in the production schedule begins. Neither does it take into account availability (reliability) of equipment. The effect of equipment availability is assumed to be factored into the nonoperation time between operations.

The average manufacturing lead time over the number of batches to be averaged is given by the following:

$$MLT = \frac{\sum_{j=1}^{n_b} MLT_j}{n_b} \tag{3.19}$$

where MLT = average manufacturing lead time, min, for the n_b batches (parts or products) over which the averaging procedure is carried out, and MLT_j = lead time for batch j from Equation 3.18. In the extreme case in which all of the parts or products are included in the averaging procedure, $n_b = P$, where P = the number of different part or product styles made by the factory.

To simplify matters and enhance conceptualization of this aspect of factory operations, properly weighted average values of batch quantity, number of operations per batch, setup time, operation cycle time, and nonoperation time can be used for the n_b batches being considered. With these simplifications, Equations (3.18) and (3.19) reduce to the following:

$$MLT = n_o(T_{su} + QT_c + T_{no}) \tag{3.20}$$

where MLT = average manufacturing lead time for all parts or products in the plant, min; and the terms Q, n_o, T_{su}, T_c, and T_{no} are all average values for these parameters. Formulas to determine these average values are presented in Appendix 3A.

EXAMPLE 3.3 Manufacturing Lead Time

A certain part is produced in batch sizes of 100 units. The batches must be routed through five operations to complete the processing of the parts. Average setup time is 3.0 hr/batch, and average operation time is 6.0 min/pc. Average nonoperation time is 7.5 hr for each operation. Determine the manufacturing lead time to complete one batch, assuming the plant runs 8 hr/ day, 5 days/wk.

Solution: Given $T_{su} = 3.0$ hr and $T_{no} = 7.5$ hr, the manufacturing lead time for this batch is computed from Equation (3.20), where the symbol j refers to the fact that only one part style is being considered.

$$MLT_j = 5\big(3.0 + 100\big(6.0/60\big) + 7.5\big) = 5(20.5) = \mathbf{102.5\ hr}$$

At 8 hr/day, this amounts to $102.5/8 = 12.81$ days

Equation (3.20) can be adapted for job shop production and mass production by making adjustments in the parameter values. For a job shop in which the batch size is one ($Q = 1$), Equation (3.20) becomes

$$MLT = n_o(T_{su} + T_c + T_{no}) \tag{3.21}$$

For mass production, the Q term in Equation (3.20) is very large and dominates the other terms. In the case of quantity-type mass production in which a large number of units are made on a single machine $(n_o = 1)$, MLT is the operation cycle time for the machine plus the nonoperation time. In this case, T_{no} consists of the time parts spend in queues before and after processing. The transportation of parts into and out of the machine is likely to be accomplished in batches. This definition assumes steady-state operation after the setup has been completed and production begins.

For flow-line mass production, the entire production line is set up in advance. If the workstations are integrated so that all stations are processing their own respective work units, then the time to accomplish all of the operations is the time it takes each work unit to progress through all of the stations on the line plus the nonoperation time. Again, T_{no} consists of the time parts spend in queues before and after processing on the line. The station with the longest operation time sets the pace for all stations:

$$MLT = n_o(\text{Max } T_o + T_r) + T_{no} = n_o T_c + T_{no} \tag{3.22}$$

where MLT = time between start and completion of a given work unit on the line, min; n_o = number of operations on the line; T_r = transfer time, min; Max T_o = operation time at the bottleneck station, min; and T_c = cycle time of the production line, min/pc, $T_c = \text{Max } T_o + T_r$ from Equation (3.7). Because the number of stations on the line is equal to the number of operations $(n = n_o)$, Equation (3.22) can also be stated as

$$MLT = n(\text{Max } T_o + T_r) + T_{no} = nT_c + T_{no} \tag{3.23}$$

where the symbols have the same meaning as above, and n (number of workstations) has been substituted for number of operations n_o.

Work-in-Process. A plant's work-in-process (*WIP*, also known as *work-in-progress*) is the quantity of parts or products currently located in the factory that either are being processed or are between processing operations. *WIP* is inventory that is in the state of being transformed from raw material to finished part or product. An approximate measure of work-in-process can be obtained from the following formula, based on Little's formula,[2] using terms previously defined:

$$WIP = R_{pph}(MLT) \tag{3.24}$$

where WIP = work-in-process in the plant, pc; R_{pph} = hourly plant production rate, pc/hr, from Equation (3.13); and MLT = average manufacturing lead time, hr. Equation (3.24) states that the level of *WIP* equals the rate at which parts flow through the factory multiplied by the length of time the parts spend in the factory. Effects of part queues, equipment availability, and other delays are accounted for in the nonoperation time, which is a component of MLT.

Work-in-process represents an investment by the firm, but one that cannot be turned into revenue until all processing has been completed. Many manufacturing companies sustain major costs because work remains in-process in the factory too long.

[2]This is an equation in queuing theory developed by John D. C. Little that is usually stated as $L = \lambda W$, where L = the expected number of units in the system, λ = processing rate of units in the system, and W = expected time that a unit spends in the system. In Equation (3.24), L becomes WIP, λ becomes R_{pph}, and W becomes MLT. Little's formula assumes that the system being modeled is operating under steady-state conditions.

EXAMPLE 3.4 Work-In-Process

Assume that the part style in Example 3.3 is representative of other parts produced in the factory. Average batch quantity = 100 units, average setup time = 3.0 hr per batch, number of operations per batch = 5, and average operation time is 6.0 min per piece for the population of parts made in the plant. Nonoperation time = 7.5 hr. The plant has 20 production machines that are 100% utilized (setup and run time), and it operates 40 hr/wk. Determine (a) weekly plant production rate and (b) work-in-process for the plant.

Solution: (a) Production rate for the average part can be determined from Equations (3.4) and (3.5):

$$T_p = \frac{3.0(60) + 100(6.0)}{100} = 7.8 \text{ min}$$

Average hourly production rate $R_p = 60/7.8 = 7.69$ pc/hr for each machine. Weekly production rate for the plant can be determined by using this average value of production rate per machine and adapting Equation (3.13) as follows:

$$R_{pph} = n\left(\frac{R_p}{n_o}\right) = 20\left(\frac{7.69}{5}\right) = 30.77 \text{ pc/hr}$$

$$R_{ppw} = 40(30.77) = \mathbf{1{,}231 \text{ pc/wk}}$$

(b) Given $U = 100\% = 1.0$, $WIP = R_{pph}(MLT) = 30.77(102.5) = \mathbf{3{,}154 \text{ pc}}$

Comment: Three observations cry out for attention in this example and the previous one. (1) In part (a), given that the equipment is 100% utilized, the calculated weekly production rate of 1,230 pc/wk must be the plant capacity. Unless the 40 hr of plant operation is increased, the plant cannot produce any more parts than it is currently producing. (2) In part (b), with 20 machines each processing one part at a time, it means that 3,154 − 20 = 3,134 parts are in a nonoperation mode. At any given moment, 3,134 parts in the plant are waiting or being moved. (3) With five operations required for each part, each operation taking 6 min, the total operation time for each part is 30 min. From the previous example, the average total time each part spends in the plant is 102.5 hr or 6,150 min. Thus, each part spends (6,150 − 30)/6,150 = 0.995 or 99.5% of its time in the plant waiting or being moved.

3.2 MANUFACTURING COSTS

Decisions on automation and production systems are usually based on the relative costs of alternatives. This section examines how these costs and cost factors are determined.

3.2.1 Fixed and Variable Costs

Manufacturing costs can be classified into two major categories: (1) fixed costs and (2) variable costs. A ***fixed cost*** is one that remains constant for any level of production output. Examples include the cost of the factory building and production equipment, insurance, and property taxes. All of the fixed costs can be expressed as annual amounts. Expenses such as insurance and property taxes occur naturally as annual costs. Capital investments such as building and equipment can be converted to their equivalent uniform annual costs using interest rate factors.

A ***variable cost*** is one that varies in proportion to production output. As output increases, variable cost increases. Examples include direct labor, raw materials, and electric power to operate the production equipment. The ideal concept of variable cost is that it is directly proportional to output level. Adding fixed and variable costs results in the following total cost equation:

$$TC = C_f + C_v Q \tag{3.25}$$

where TC = total annual cost, \$/yr; C_f = fixed annual cost, \$/yr; C_v = variable cost, \$/pc; and Q = annual quantity produced, pc/yr.

When comparing automated and manual production methods, it is typical that the fixed cost of the automated method is high relative to the manual method, and the variable cost of automation is low relative to the manual method, as pictured in Figure 3.3. Consequently, the manual method has a cost advantage in the low quantity range, while

EXAMPLE 3.5 Manual versus Automated Production

Two production methods are being compared, one manual and the other automated. The manual method produces 10 pc/hr and requires one worker at \$15.00/hr. Fixed cost of the manual method is \$5,000/yr. The automated method produces 25 pc/hr, has a fixed cost of \$55,000/yr, and a variable cost of \$4.50/hr. Determine the break-even point for the two methods; that is, determine the annual production quantity at which the two methods have the same annual cost. Ignore the costs of materials used in the two methods.

Solution: The variable cost of the manual method is $C_v = (\$15.00/\text{hr})/(10\ \text{pc/hr}) = \$1.50/\text{pc}$

Annual cost of the manual method is $TC_m = 5{,}000 + 1.50Q$

The variable cost of the automated method is $C_v = (\$4.50/\text{hr})/(25\ \text{pc/hr}) = \$0.18/\text{pc}$

Annual cost of the automated method is $TC_a = 55{,}000 + 0.18Q$

At the break-even point $TC_m = TC_a$:

$$5{,}000 + 1.50Q = 55{,}000 + 0.18Q$$

$$1.50Q - 0.18Q = 1.32Q = 55{,}000 - 5{,}000 = 50{,}000$$

$$1.32Q = 50{,}000Q = 50{,}000/1.32 = \mathbf{37{,}879\ pc}$$

Comment: It is of interest to note that the manual method operating one shift (8 hr), 250 days per year would produce 8(250)(10) = 20,000 pc/yr, which is less than the break-even quantity of 37,879 pc. On the other hand, the automated method, operating under the same conditions, would produce 8(250)(25) = 50,000 pc, well above the break-even point.

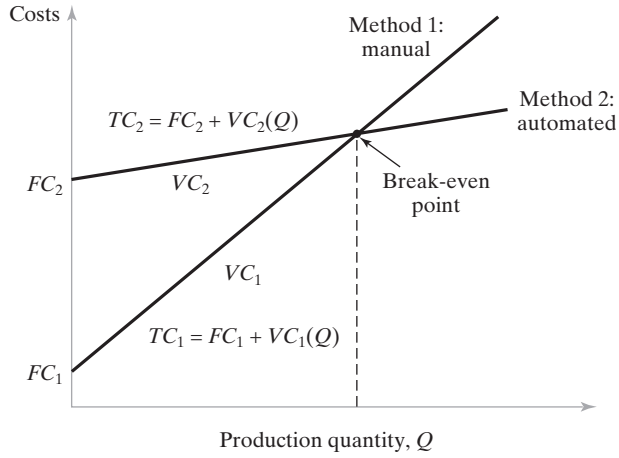


Figure 3.3 Fixed and variable costs as a function of production output for manual and automated production methods.

automation has an advantage for high quantities. This reinforces the arguments presented in Section 1.3.1 on the appropriateness of manual labor for certain production situations.

3.2.2 Direct Labor, Material, and Overhead

Fixed versus variable are not the only possible classifications of costs in manufacturing. An alternative classification separates costs into (1) direct labor, (2) material, and (3) overhead. This is often a more convenient way to analyze costs in production. ***Direct labor cost*** is the sum of the wages and benefits paid to the workers who operate the production equipment and perform the processing and assembly tasks. ***Material cost*** is the cost of all raw materials used to make the product. In the case of a stamping plant, the raw material consists of the sheet stock used to make stampings. For the rolling mill that made the sheet stock, the raw material is the starting slab of metal out of which the sheet is rolled. In the case of an assembled product, materials are the component parts, some of which are produced by supplier firms. Thus, the definition of "raw material" depends on the company and the type of production operations in which it is engaged. The final product of one company can be the raw material for another company. In terms of fixed and variable costs, direct labor and material must be considered as variable costs.

Overhead costs are all of the other expenses associated with running the manufacturing firm. Overhead divides into two categories: (1) factory overhead and (2) corporate overhead. ***Factory overhead*** consists of the costs of operating the factory other than direct labor and materials, such as the factory expenses listed in Table 3.2. Factory overhead is treated as fixed cost, although some of the items in the list could be correlated

TABLE 3.2 Typical Factory Overhead Expenses

Plant supervision	Applicable taxes	Factory depreciation
Line foreman	Insurance	Equipment depreciation
Maintenance crew	Heat and air conditioning	Fringe benefits
Custodial services	Light	Material handling
Security personnel	Power for machinery	Shipping and receiving
Tool crib attendant	Payroll services	Clerical support

TABLE 3.3 Typical Corporate Overhead Expenses

Corporate executives	Engineering	Applicable taxes
Sales and marketing	Research and development	Office space
Accounting department	Other support personnel	Security personnel
Finance department	Insurance	Heat and air conditioning
Legal counsel	Fringe benefits	Lighting

with the output level of the plant. ***Corporate overhead*** is the cost not related to the company's manufacturing activities, such as the corporate expenses in Table 3.3. Many companies operate more than one factory, and this is one of the reasons for dividing overhead into factory and corporate categories. Different factories may have significantly different factory overhead expenses.

J Black [1] provides some typical percentages for the different types of manufacturing and corporate expenses. These are presented in Figure 3.4. Several observations can be made about these data. First, total manufacturing cost represents only about 40% of the product's selling price. Corporate overhead expenses and total manufacturing cost are about equal. Second, materials (including purchased parts) make up the largest percentage of total manufacturing cost, at around 50%. And third, direct labor is a relatively small proportion of total manufacturing cost: 12% of manufacturing cost and only about 5% of final selling price.

Overhead costs can be allocated according to a number of different bases, including direct labor cost, material cost, direct labor hours, and space. Most common in industry is direct labor cost, which will be used here to illustrate how overheads are allocated and subsequently used to compute factors such as selling price of the product.

The allocation procedure (simplified) is as follows. For the most recent year (or several recent years), all costs are compiled and classified into four categories: (1) direct labor, (2) material, (3) factory overhead, and (4) corporate overhead. The objective is to determine an ***overhead rate*** that can be used in the following year to allocate overhead costs to a process or product as a function of the direct labor costs associated with that process or product. Separate overhead rates will be developed for factory and corporate overheads. The ***factory overhead rate*** is calculated as the

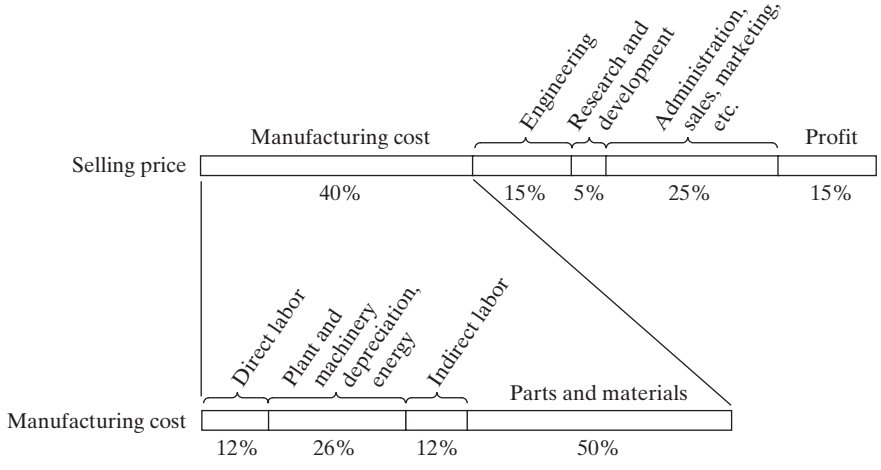


Figure 3.4 Breakdown of costs for a manufactured product [1].

ratio of factory overhead expenses (category 3) to direct labor expenses (category 1); that is,

$$FOHR = \frac{FOHC}{DLC} \tag{3.26}$$

where $FOHR$ = factory overhead rate, $FOHC$ = annual factory overhead costs, \$/yr; and DLC = annual direct labor costs, \$/yr.

The ***corporate overhead rate*** is the ratio of corporate overhead expenses (category 4) to direct labor expenses:

$$COHR = \frac{COHC}{DLC} \tag{3.27}$$

where $COHR$ = corporate overhead rate, $COHC$ = annual corporate overhead costs, \$/yr; and DLC = annual direct labor costs, \$/yr. Both rates are often expressed as percentages. If material cost were used as the allocation basis, then material cost would be used as the denominator in both ratios. The following two examples are presented to illustrate (1) how overhead rates are determined and (2) how they are used to estimate manufacturing cost and establish selling price.

EXAMPLE 3.6 Determining Overhead Rates

Suppose that all costs have been compiled for a certain manufacturing firm for last year. The summary is shown in the table below. The company operates two different manufacturing plants plus a corporate headquarters. Determine (a) the factory overhead rate for each plant, and (b) the corporate overhead rate. These rates will be used by the firm to predict the following year's expenses.

Expense Category	Plant 1 (\$)	Plant 2 (\$)	Headquarters (\$)	Totals (\$)
Direct labor	800,000	400,000		1,200,000
Materials	2,500,000	1,500,000		4,000,000
Factory expense	2,000,000	1,100,000		3,100,000
Corporate expense			7,200,000	7,200,000
Totals	5,300,000	3,000,000	7,200,000	15,500,000

Solution: (a) A separate factory overhead rate must be determined for each plant. For plant 1,

$$FOHR_1 = \frac{\$2{,}000{,}000}{\$800{,}000} = 2.5 = \mathbf{250\%}$$

For plant 2,

$$FOHR_2 = \frac{\$1{,}100{,}000}{\$400{,}000} = 2.75 = \mathbf{275\%}$$

(b) The corporate overhead rate is based on the total labor cost at both plants.

$$COHR = \frac{\$7{,}200{,}000}{\$1{,}200{,}000} = 6.0 = \mathbf{600\%}$$

EXAMPLE 3.7 Estimating Manufacturing Costs and Establishing Selling Price

A customer order of 50 parts is to be processed through plant 1 of the previous example. Raw materials and tooling are supplied by the customer. The total time for processing the parts (including setup and other direct labor) is 100 hr. Direct labor cost is \$15.00/hr. The factory overhead rate is 250% and the corporate overhead rate is 600%. (a) Compute the cost of the job. (b) What price should be quoted to the customer if the company uses a 10% markup?

Solution: (a) The direct labor cost for the job is $(100\text{ hr})(\$15.00/\text{hr}) = \$1{,}500$.
The allocated factory overhead charge, at 250% of direct labor, is $(\$1{,}500)(2.50) = \$3{,}750$. The total factory cost of the job, including allocated factory overhead = \$1,500 + \$3,750 = \$5,250.
The allocated corporate overhead charge, at 600% of direct labor, is $(\$1{,}500)(6.00) = \$9{,}000$. The total cost of the job including corporate overhead = \$5,250 + \$9,000 = **\$14,250**

(b) If the company uses a 10% markup, the price quoted to the customer would be $(1.10)(\$14{,}250) =$ **\$15,675**

3.2.3 Cost of Equipment Usage

The trouble with overhead rates as they have been developed here is that they are based on labor cost alone. A machine operator who runs an old, small engine lathe whose book value is zero will be costed at the same overhead rate as an operator running a new automated lathe just purchased for \$500,000. Obviously, the time on the machining center is more productive and should be valued at a higher rate. If differences in rates of different production machines are not recognized, manufacturing costs will not be accurately measured by the overhead rate structure.

To deal with this difficulty, it is appropriate to divide the cost of a worker running a machine into two components: (1) direct labor cost and (2) machine cost. Associated with each is an applicable overhead rate. These overhead costs apply not to the entire factory operations, but to individual machines.

The direct labor cost consists of the wages and benefits paid to operate the machine. Applicable factory overhead expenses allocated to direct labor cost might include taxes paid by the employer, certain fringe benefits, and line supervision. The machine annual cost is the initial cost of the machine apportioned over the life of the asset at the appropriate rate of return used by the firm. This is done using the capital recovery factor, as

$$UAC = IC(A/P, i, N) \tag{3.28}$$

where UAC = equivalent uniform annual cost, \$/yr; IC = initial cost of the machine, \$; and $(A/P, i, N)$ = capital recovery factor that converts initial cost at year 0 into a series of equivalent uniform annual year-end values, where i = annual interest rate and N = number of years in the service life of the equipment. For given values of i and N, $(A/P, i, N)$ can be computed as follows:

$$(A/P, i, N) = \frac{i(1 + i)^N}{(1 + i)^N - 1} \tag{3.29}$$

Values of $(A/P, i, N)$ can also be found in interest tables that are widely available.

The uniform annual cost can be expressed as an hourly rate by dividing the annual cost by the number of annual hours of equipment use. The machine overhead rate is based on those factory expenses that are directly assignable to the machine. These include power to drive the machine, floor space, maintenance and repair expenses, and so on. In separating the factory overhead items in Table 3.2 between labor and machine, judgment must be used; admittedly, the judgment is sometimes arbitrary. The total cost rate for the machine is the sum of labor and machine costs. This can be summarized for a machine consisting of one worker and one machine as follows:

$$C_o = C_L(1 + FOHR_L) + C_m(1 + FOHR_m) \tag{3.30}$$

where C_o = hourly rate to operate the machine, \$/hr; C_L = direct labor wage rate, \$/hr; $FOHR_L$ = factory overhead rate for labor; C_m = machine hourly rate, \$/hr; and $FOHR_m$ = factory overhead rate applicable to the machine.

It is the author's opinion that corporate overhead expenses should not be included in the analysis when comparing production methods. Including them serves no purpose other than to dramatically inflate the costs of the alternatives. The fact is that these corporate overhead expenses are present whether or not any of the alternatives is selected. On the other hand, when analyzing costs for pricing decisions, corporate overhead must be included because over the long run, these costs must be recovered through revenues generated from selling products.

EXAMPLE 3.8 Hourly Cost of a Machine

The following data are given for a production machine consisting of one worker and one piece of equipment: direct labor rate = \$15.00/hr, applicable factory overhead rate on labor = 60%, capital investment in machine = \$100,000, service life of the machine = 4 yr, rate of return = 10%, salvage value in 4 yr = 0, and applicable factory overhead rate on machine = 50%. The machine will be operated one 8-hr shift, 250 day/yr. Determine the appropriate hourly rate for the machine.

Solution: Labor cost per hour = $C_L(1 + FOHR_L) = \$15.00(1 + 0.60) = \$24.00$/hr. The investment cost of the machine must be annualized, using a 4-yr service life and a rate of return = 10%. First, compute the capital recovery factor:

$$(A/P, 10\%, 4) = \frac{0.10(1 + 0.10)^4}{(1 + 0.10)^4 - 1} = 0.3155$$

Now the uniform annual cost for the \$100,000 initial cost can be determined:

$$UAC = \$100{,}000(A/P, 10\%, 4) = 100{,}000(0.3155) = \$31{,}550/\text{yr}$$

The number of hours per year = (8 hr/day)(250 day/yr) = 2,000 hr/yr. Dividing this into UAC gives 31,550/2,000 = \$15.77/hr. Applying the factory overhead rate,

$$C_m(1 + FOHR_m) = \$15.77(1 + 0.50) = 23.66/\text{hr}$$

Total cost rate for the machine is

$$C_o = 24.00 + 23.66 = \mathbf{\$47.66/hr}$$

3.2.4 Cost of a Manufactured Part

The unit cost of a manufactured part or product is the sum of the production cost, material cost, and tooling cost. As indicated in Example 3.7, overhead costs and profit markup must be added to the unit cost to arrive at a selling price for the product. The unit production cost for each unit operation in the sequence of operations to produce the part or product is given by:

$$C_{oi}T_{pi} + C_{ti}$$

where C_{oi} = cost rate to perform unit operation i, \$/min, defined by Equation (3.30); T_{pi} = production time of operation i, min/pc, as defined by the equations in Section 3.1.1; and C_{ti} = cost of any tooling used in operation i, \$/pc. It should be noted that the cost of tooling is in addition to any tool handling time defined in Equation (3.1), which is included in the value of T_p. Tooling cost is a material cost, whereas tool handling is a time cost at the cost rate of the machine, C_{oi}.

The total unit cost of the part is the sum of the costs of all unit operations plus the cost of raw materials. Summarizing,

$$C_{pc} = C_m + \sum_{i=1}^{n_o} (C_{oi}T_{pi} + C_{ti}) \tag{3.31}$$

where C_{pc} = cost per piece, \$/pc; C_m = cost of starting material, \$/pc; and the summation includes all of the costs of the n_o unit operations in the sequence.

EXAMPLE 3.9 Unit Cost of a Manufactured Part

The machine in Example 3.8 is the first of two machines used to produce a certain part. The starting material cost of the part is \$8.50/pc. As determined in the previous example, the cost rate to operate the first machine is \$47.66/hr, or \$0.794/min. The production time on the first machine is 4.20 min/pc, and there is no tooling cost. The cost rate of the second machine in the process sequence is \$35.80/hr, or \$0.597/min. The production time on the second machine is 2.75 min/pc, and the tooling cost is \$0.20/pc. Determine the unit part cost.

Solution: Using Equation (3.31), the part cost is calculated as follows:

$$C_{pc} = 8.50 + 0.794(4.20) + 0.597(2.75) + 0.20 = \mathbf{\$13.68/pc}$$

REFERENCES

[1] Black, J. T., *The Design of the Factory with a Future*, McGraw-Hill, Inc., New York, NY, 1991.

[2] Blank, L. T., and A. J. Tarquin, *Engineering Economy*, 7th ed., McGraw-Hill, New York, 2011.

[3] Groover, M. P., *Fundamentals of Modern Manufacturing: Materials, Processes, and Systems*, 5th ed., John Wiley & Sons, Inc., Hoboken, NJ, 2013.

[4] Groover, M. P., *Work Systems and the Methods, Measurement, and Management of Work*, Pearson Prentice Hall, Upper Saddle River, NJ, 2007.

23.2.3 Computer-Integrated Manufacturing

Computer-integrated manufacturing includes all of the engineering functions of CAD/CAM, but it also includes the firm's business functions that are related to manufacturing. The ideal CIM system applies computer and communications technology to all the operational functions and information-processing functions in manufacturing from order receipt through design and production to product shipment. The scope of CIM, compared with the more limited scope of CAD/CAM, is depicted in Figure 23.5. Also shown are the components of CAD, CAM, and the business functions.

The CIM concept is that all of the firm's operations related to production are incorporated in an integrated computer system to assist, augment, and automate the operations. The computer system is pervasive throughout the firm, touching all activities that support manufacturing. In this integrated computer system, the output of one activity serves as the input to the next activity, through the chain of events that starts with the sales order and culminates with shipment of the product. Customer orders are initially entered by the company's salesforce or directly by the customer into a computerized order entry system. The orders contain the specifications describing the product. The specifications serve as the input to the product design department. New products are designed on a CAD system. The components that comprise the product are designed, the bill of materials is compiled, and assembly drawings are prepared. The output of the design department serves as the input to manufacturing engineering, where process planning, tool design, and similar activities are accomplished to prepare for production. Process planning is performed using CAPP. Tool and fixture design is done on a CAD system, making use of the product model generated during product design. The output from manufacturing engineering provides the input to production planning and control, where material requirements planning and scheduling are performed using the computer system, and so it goes, through each step in the manufacturing cycle. Full implementation of CIM results

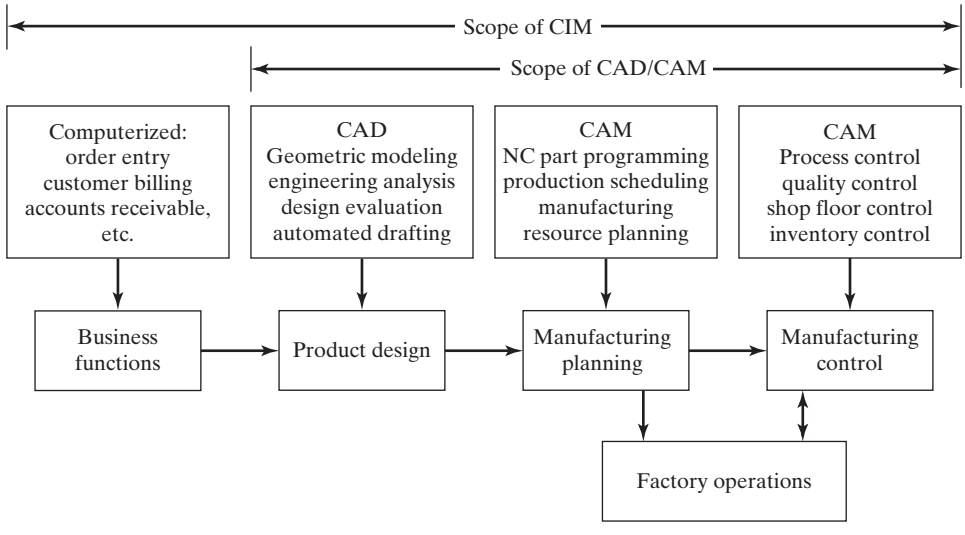


Figure 23.5 The scope of CAD/CAM and CIM, and the computerized elements of a CIM system.

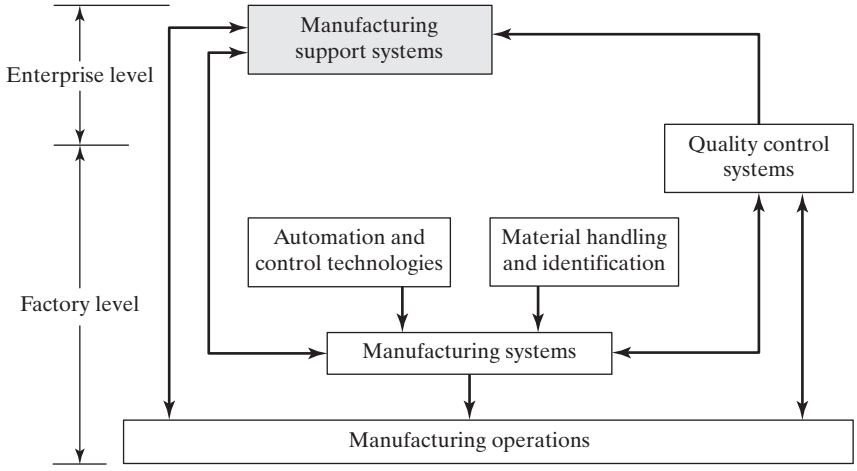


Figure 23.1 The position of the manufacturing support systems in the larger production system.

The present chapter deals with product design and the various technologies that are used to augment and automate the design function. CAD/CAM (computer-aided design and computer-aided manufacturing) is one of those technologies. It uses digital computer systems to accomplish certain functions in product design and production. CAD uses the computer to support the design engineering function, and CAM uses the computer to support manufacturing engineering activities. The combination CAD/CAM is symbolic of efforts to integrate the design and manufacturing functions of a firm into a continuum of activities rather than to treat them as two separate and disparate activities, as they had been considered in the past. CIM (computer-integrated manufacturing) includes all of CAD/CAM but also embraces the business functions of a manufacturing firm. CIM implements computer technology in all of the operational and information-processing activities related to manufacturing. In the final section of the chapter, a systematic method for approaching a product design project, called *quality function deployment*, is described.

Chapters 24 through 26 are concerned with topics in production systems other than product design. Chapter 24 deals with process planning and how it can be automated using computer systems. Included in this discussion are ways in which product design and manufacturing and other functions can be integrated using an approach called *concurrent engineering*. An important issue in concurrent engineering is design for manufacturing; that is, how can a product be designed to make it easier (and cheaper) to produce? Chapter 25 discusses the various methods used to implement production planning and control, through material requirements planning, shop floor control, and enterprise resource planning (ERP). Finally, Chapter 26 is concerned with just-in-time production and lean production, the techniques that were developed and perfected by the Toyota Motor Company in Japan.

23.1 PRODUCT DESIGN AND CAD

Product design is a critical function in the production system. The quality of the product design is probably the single most important factor in determining the commercial success and societal value of a product. If the product design is poor, no matter how well it

is manufactured, the product is very likely doomed to contribute little to the wealth and well-being of the firm that produced it. If the product design is good, there is still the question of whether the product can be produced at sufficiently low cost to contribute to the company's profits and success. One of the facts of life about product design is that a very significant portion of the cost of the product is determined by its design. Design and manufacturing cannot be separated in the production system. They are bound together functionally, technologically, and economically.

23.1.1 The Design Process

The general process of design is characterized as an iterative process consisting of six phases [13]: (1) recognition of need, (2) problem definition, (3) synthesis, (4) analysis and optimization, (5) evaluation, and (6) presentation. These six steps, and the iterative nature of the sequence in which they are performed, are depicted in Figure 23.2(a).

Recognition of need (1) involves the realization by someone that a problem exists which could be solved by a thoughtful design. This recognition might mean identifying some deficiency in a current machine design by an engineer or perceiving some new product opportunity by a salesperson. Problem definition (2) involves a thorough specification of the item to be designed. This specification includes the physical characteristics, function, cost, quality, and operating performance.

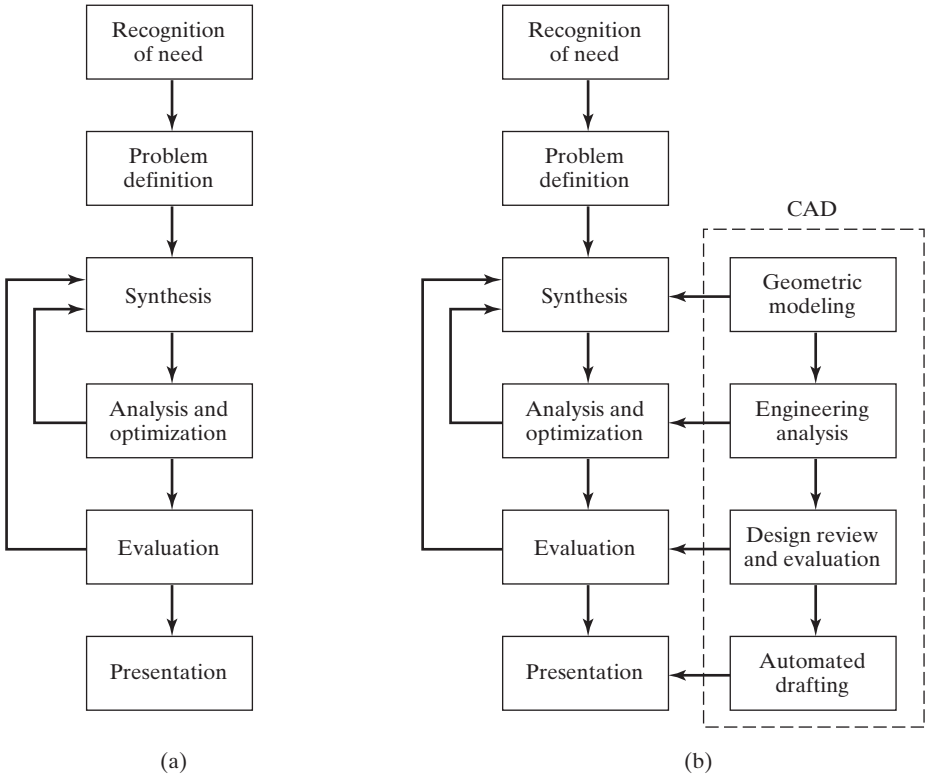


Figure 23.2 (a) Design process as defined by Shigley [13]. (b) The design process using computer-aided design (CAD).

Synthesis (3) and analysis (4) are closely related and highly interactive. Consider the development of a certain product design: Each of the subsystems of the product must be conceptualized by the designer, analyzed, improved through this analysis procedure, redesigned, analyzed again, and so on. The process is repeated until the design has been optimized within the constraints imposed on the designer. The individual components are then synthesized and analyzed into the final product in a similar manner.

Evaluation (5) is concerned with measuring the design against the specifications established in the problem definition phase. This evaluation often requires the fabrication and testing of a prototype model to assess operating performance, quality, reliability, and other criteria. The final phase in the design procedure is the presentation of the design. Presentation (6) is concerned with documenting the design by means of drawings, material specifications, assembly lists, and so on. In essence, documentation means that the design database is created.

23.1.2 Computer-Aided Design

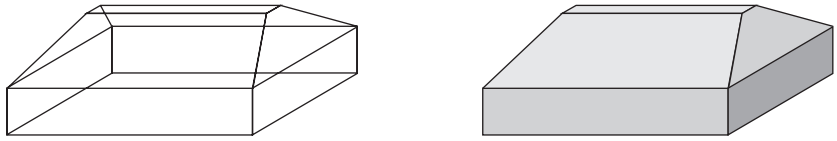
Computer-aided design (CAD) is defined as any design activity that involves the effective use of computer systems to create, modify, analyze, optimize, and document an engineering design. CAD is most commonly associated with the use of an interactive computer graphics system, referred to as a CAD system. The term *CAD/CAM* is also used if the system includes manufacturing applications as well as design applications.

With reference to the six phases of design, a CAD system can facilitate four of the design phases, as illustrated in Figure 23.2(b), as an overlay on the design process.

Geometric Modeling. Geometric modeling involves the use of a CAD system to develop a mathematical description of the geometry of an object. The mathematical description, called a *geometric model*, is contained in computer memory. This permits the user of the CAD system to display an image of the model on a graphics terminal and to perform certain operations on the model. These operations include creating new geometric models from basic building blocks available in the system, moving and reorienting the images on the screen, zooming in on certain features of the image, and so forth. These capabilities permit the designer to construct a model of a new product (or its components) or to modify an existing model.

There are various types of geometric models used in CAD. One classification distinguishes between two-dimensional (2-D) and three-dimensional (3-D) models. Two-dimensional models are best utilized for designing flat objects and building layouts. In the first CAD systems developed in the 1970s, 2-D systems were used principally as automated drafting systems. They were often used for 3-D objects, and it was left to the designers to properly construct the various views as they would have done in manual drafting. Three-dimensional CAD systems are capable of modeling an object in three dimensions according to user instructions. This is helpful in conceptualizing the object since the true 3-D model can be displayed in various views and from different angles.

Geometric models in CAD can also be classified as wire-frame models or solid models. A wire-frame model uses interconnecting lines (straight line segments) to depict the object as illustrated in Figure 23.3(a). Wire-frame models of complicated geometries can become somewhat confusing because all of the lines depicting the shape of the object are usually shown, even the lines representing the other side of the object. These so-called hidden lines can be removed, but even with this improvement, wire-frame representation is still often



(a) (b)

Figure 23.3 Geometric models in CAD: (a) Wire-frame model. (b) Solid model of the same object.

confusing. It is rarely used today. In solid modeling, Figure 23.3(b), an object is modeled in solid three dimensions, providing the user with a vision of the object that is similar to the way it would be seen in real life. More important for engineering purposes, the geometric model is stored in the CAD system as a 3-D solid model, providing a more accurate representation of the object. This is useful for calculating mass properties, in assembly to perform interference checking between mating components, and in other engineering calculations.

Two other features in CAD system models are color and animation. The value of color is largely to enhance the ability of the user to visualize the object on the graphics screen. For example, the various components of an assembly can be displayed in different colors, permitting the parts to be more readily distinguished. And animation capability permits the operation of mechanisms and other moving objects to be displayed on the graphics monitor.

Engineering Analysis. After a particular design alternative has been developed, some form of engineering analysis must often be performed as part of the design process. The analysis may take the form of stress–strain calculations, heat transfer analysis, or dynamic simulation. The computations are often complex and time consuming, and before the advent of the digital computer, these analyses were usually greatly simplified or even omitted in the design procedure. The availability of software for engineering analysis on a CAD system greatly increases the designer's ability and willingness to perform a more thorough analysis of a proposed design. The term *computer-aided engineering* (CAE) applies to engineering analyses performed by computer. Examples of CAE software in common use on CAD systems include:

- *Mass properties analysis.* This involves the computation of such features of a solid object as its volume, surface area, weight, and center of gravity. It is especially applicable in mechanical design. Prior to CAD, determination of these properties often required painstaking and time-consuming calculations by the designer.
- *Interference checking.* This CAD software examines 3-D geometric models consisting of multiple components to identify interferences between components. It is useful in analyzing mechanical assemblies, chemical plant piping systems, and similar multicomponent designs.
- *Tolerance analysis.* Software for analyzing the specified tolerances of a product's components is used (1) to assess how the tolerances may affect the product's function and performance, (2) to determine how tolerances may influence the ease or difficulty of assembling the product, and (3) to assess how variations in component dimensions may affect the overall size of the assembly.
- *Finite element analysis.* Software for finite element analysis (FEA), also known as *finite element modeling* (FEM), is available for use on CAD systems to aid in

stress–strain, heat transfer, fluid flow, and other computations. Finite element analysis is a numerical analysis technique for determining approximate solutions to physical problems described by differential equations that are very difficult or impossible to solve. In FEA, the physical object is modeled by an assemblage of discrete interconnected nodes (finite elements), and the variable of interest (e.g., stress, strain, temperature) in each node can be described by relatively simple mathematical equations. Solving the equations for each node provides the distribution of values of the variable throughout the physical object.

- *Kinematic and dynamic analysis.* Kinematic analysis studies the operation of mechanical linkages and analyzes their motions. A typical kinematic analysis specifies the motion of one or more driving members of the subject linkage, and the resulting motions of the other links are determined by the analysis package. Dynamic analysis extends kinematic analysis by including the effects of the mass of each linkage member and the resulting acceleration forces as well as any externally applied forces.
- *Discrete-event simulation.* This type of simulation is used to model complex operational systems, such as a manufacturing cell or a material handling system, as events occur at discrete moments in time and affect the status and performance of the system. For example, discrete events in the operation of a manufacturing cell include parts arriving for processing and a machine breakdown in the cell. Performance measures include the status of any given machine in the cell (idle or busy), average length of time parts spend in the cell, and overall cell production rate. Current discrete-event simulation software includes animated graphics capability that enhances visualization of the system's operation.

Design Evaluation and Review. Some of the CAD features that are helpful in evaluating and reviewing a proposed design include the following:

- *Automatic dimensioning.* These routines determine precise distance measures between surfaces on the geometric model identified by the user.
- *Error checking.* This term refers to CAD algorithms that are used to review the accuracy and consistency of dimensions and tolerances and to assess whether the proper design documentation format has been followed.
- *Animation of discrete-event simulation solutions.* Discrete-event simulation was described earlier in the context of engineering analysis. Displaying the solution of the discrete-event simulation in animated graphics is a helpful means of presenting and evaluating the solution. Input parameters, probability distributions, and other factors can be changed to assess their effect on the performance of the system being modeled.
- *Plant layout design scores.* A number of software packages are available for facilities design, that is, designing the floor layout and physical arrangement of equipment in a facility. Some of these packages provide one or more numerical scores for each plant layout design, which allow the user to assess the merits of the alternative with respect to material flow, closeness ratings, and similar factors.

The traditional procedure in designing a new product includes fabrication of a prototype before approval and release for production. The prototype serves as the "acid test" of the design, permitting the designer and others to see, feel, operate, and test the product for any last-minute changes or enhancements of the design. The problem with building a prototype is that it is traditionally very time consuming; in some cases, months are required to make and assemble all of the parts. Motivated by the need to reduce this

lead time for building the prototype, engineers have developed several new approaches that rely on the use of the geometric model of the product residing in the CAD data file. Two of these approaches are rapid prototyping and virtual prototyping.

Rapid prototyping (RP) is a family of fabrication technologies that allow engineering prototypes of solid parts to be made in minimum lead time; the common feature of these technologies is that they produce the part directly from the CAD geometric model. This is usually done by dividing the solid object into a series of layers of small thickness and then defining the area shape of each layer. For example, a vertical cone would be divided into a series of circular layers, the circles becoming smaller and smaller toward the vertex of the cone. The RP processes then fabricate the object by starting at the base and building each layer on top of the preceding layer to approximate the solid shape. The fidelity of the approximation depends on the thickness of each layer. As layer thickness decreases, accuracy increases. There are a variety of layer-building processes used in rapid prototyping. One process, called *stereolithography*, uses a photosensitive liquid polymer that cures (solidifies) when subjected to intense light. Curing of the polymer is accomplished using a moving laser beam whose path for each layer is controlled by means of the CAD model. A solid polymer prototype of the part is built up of hardened layers, one on top of another. Another RP process, called *selective laser sintering*, uses a moving laser beam to fuse powders in each layer to form the object layer by layer; work materials include polymers, metals, and ceramics. When used to produce parts rather than prototypes, the term *additive manufacturing* is used for these processing technologies. A comprehensive treatment of rapid prototyping and additive manufacturing is presented in [6]; for a more concise coverage of these technologies, see [8].

Virtual prototyping, based on virtual reality technology, involves the use of the CAD geometric model to construct a digital mock-up of the product, enabling the designer and others to obtain the sensation of the real product without actually building the physical prototype. Virtual prototyping has been used in the automotive industry to evaluate new car style designs. The observer of the virtual prototype is able to assess the appearance of the new design even though no physical model is on display. Other applications of virtual prototyping include checking the feasibility of assembly operations, for example, parts mating, access and clearance of parts during assembly, and assembly sequence.

Automated Drafting. The fourth area where CAD is useful (step 6 in the design process) is presentation and documentation. CAD systems can be used to prepare highly accurate engineering drawings when paper documents are required. It is estimated that a CAD system increases productivity in the drafting function by about fivefold over manual preparation of drawings.

CAD Workstations. The CAD workstation and its available features have an important influence on the convenience, productivity, and quality of the designer's output. The workstation includes a graphics display terminal and one or more user input devices. It is the principal means by which the system communicates with the designer. Two CAD system configurations are depicted in Figure 23.4: (1) engineering workstation and (2) PC-based CAD system.[1] The distinction between the two categories is becoming more and more subtle.

[1]The first CAD systems introduced in the 1970s and 1980s were based on a host-and-terminal configuration, in which the host was a mainframe or minicomputer serving one or more graphics terminals on a time-shared basis. The powerful microprocessors and high-density memory devices so common today were not available at that time. By and large, these host-and-terminal systems have been overtaken by engineering workstations and PC-based CAD systems.

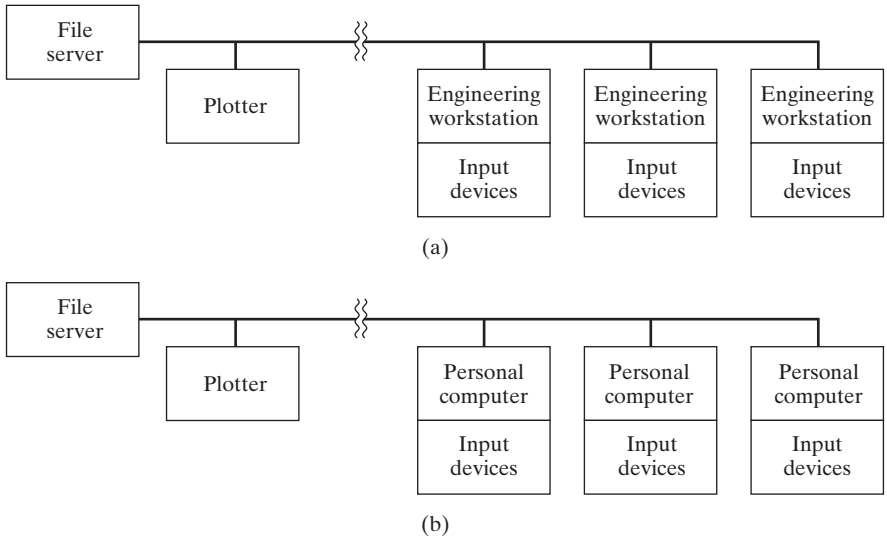


Figure 23.4 Two CAD system configurations: (a) engineering workstation and (b) PC-based CAD system.

An engineering workstation is a stand-alone computer system that is dedicated to one user and capable of executing graphics software and other programs requiring high-speed computational power. The graphics display is a high-resolution monitor with a large screen. As shown in the figure, engineering workstations are often networked to permit exchange of data files and programs between users and to share plotters and data storage devices.

PC-based CAD systems are the most widely used CAD systems today. They consist of a personal computer with a high-performance CPU and high-resolution graphics display screen. The computer is equipped with a large random access memory (RAM), math coprocessor, and large-capacity hard disk for storage of the large applications software packages used for CAD. PC-based CAD systems can be networked to share files, output devices, and for other purposes. CAD software products are based on the graphics environment of Microsoft Windows, and CAD software is also available for Apple's Mac operating system [17], [18]. Although desktop computers are most widely used, some designers prefer laptop PCs to accomplish their creative and analytical tasks.

Managing the Product Design. The output of the creative design process includes huge amounts of data that must be stored and managed. These functions are often accomplished in a modern CAD system using product data management. A ***product data management*** (PDM) system consists of computer software that provides links between users (e.g., designers) and a central database, which stores design data such as geometric models, product structures (e.g., bills of material), and related records. The software also manages the database by tracking the identity of users, facilitating and documenting engineering changes, recording a history of the engineering changes on each part and product, and providing similar documentation functions.

The PDM system is usually considered to be a component of a broader process within a company called ***product lifecycle management*** (PLM), which is concerned

with managing the entire life cycle of a product, starting with the initial concept for it, continuing through its development and design, prototype testing, manufacturing planning, production operations, customer service, and finally its end-of-life disposal. PLM is a business process that begins with product design, but its scope is much broader than product design. Implementing PLM involves the integration of product and production data, business procedures, and people.

Compared with manual design and drafting methods, computer-aided design and management systems provide many advantages, including the following [10], [15]):

- *Increased design productivity.* The use of CAD helps the designer conceptualize the product and its components, which in turn helps reduce the time required by the designer to synthesize, analyze, and document the design. The result is a shorter design cycle and lower product development costs.
- *Increased available geometric forms in the design.* CAD permits the designer to select among a wider range of shapes, such as mathematically defined contours, blended angles, and similar forms that would be difficult to create by manual drafting techniques.
- *Improved quality of the design.* The use of a CAD system permits the designer to do a more complete engineering analysis and to consider a larger number and variety of design alternatives. The quality of the resulting design is thereby improved.
- *Improved design documentation.* The graphical output of a CAD system results in better documentation of the design than what is practical with manual drafting. The engineering drawings are superior, with more standardization among the drawings, fewer drafting errors, and greater legibility. In addition, most CAD packages provide automatic documentation of design changes, which includes who made the changes, as well as when and why the changes were made.
- *Creation of a manufacturing database.* In the process of creating the documentation for the product design (geometric specification of the product, dimensions of the components, materials specifications, bill of materials, etc.), much of the required database to manufacture the product is also created.
- *Design standardization.* Design rules can be included in CAD software to encourage the designer to utilize company-specified models for certain design features—for example, to limit the number of different hole sizes used in the design. This simplifies the hole specification procedure for the designer and reduces the number of drill bit sizes that must be inventoried in manufacturing.

23.2 CAM, CAD/CAM, AND CIM

CAM, CAD/CAM, and CIM were briefly defined in the chapter introduction. CIM is sometimes spoken of interchangeably with CAM and CAD/CAM. Although the terms are closely related, CIM has a broader meaning than CAM or CAD/CAM.

23.2.1 Computer-Aided Manufacturing

Computer-aided manufacturing (CAM) involves the use of computer technology in manufacturing planning and control. CAM is most closely associated with functions in manufacturing engineering, such as process planning and numerical control (NC) part

programming. The applications of CAM can be divided into two broad categories: (1) manufacturing planning and (2) manufacturing control. These two categories are covered in Chapters 24 and 25, but a brief discussion of them here may be helpful to the reader.

Manufacturing Planning. CAM applications for manufacturing planning are those in which the computer is used indirectly to support the production function, but there is no direct connection between the computer and the process. The computer is used to provide information for the effective planning and management of production activities. The following list surveys the important applications of CAM in this category:

- *Computer-aided process planning* (CAPP). Process planning is concerned with the preparation of route sheets that list the sequence of operations and work centers required to produce the product and its components. CAPP systems are available today to prepare these route sheets. CAPP is covered in the following chapter.
- *CAD/CAM NC part programming.* Numerical control part programming was discussed in Chapter 7. For complex part geometries, CAD/CAM part programming represents a much more efficient method of generating the control instructions for the machine tool than manual part programming.
- *Computerized machinability data systems.* One of the problems with operating a metal cutting machine tool is determining the speeds and feeds that should be used for a given operation. Computer programs are available to recommend the appropriate cutting conditions for different materials and operations (e.g., turning, milling, drilling). The recommendations are based on data that have been compiled either in the factory or laboratory that relate tool life to cutting conditions. Machinability data systems are described in [10].
- *Computerized work standards.* The time study department has the responsibility for setting time standards on direct labor jobs performed in the factory. Establishing standards by direct time study can be a tedious and time-consuming task. There are several commercially available computer packages for setting work standards. These computer programs use standard time data that have been developed for basic work elements that comprise any manual task. The program sums the times for the individual elements required to perform a new job in order to calculate the standard time for the job. These packages are discussed in [9].
- *Cost estimating.* The task of estimating the cost of a new product has been simplified in most industries by computerizing several of the key steps required to prepare the estimate. The computer is programmed to apply the appropriate labor and overhead rates to the sequence of planned operations for the components of new products. The program then adds up the individual component costs from the engineering bill of materials to determine the overall product cost.
- *Production and inventory planning.* The production and inventory planning functions include maintenance of inventory records, automatic reordering of stock items when inventory is depleted, production scheduling, maintaining current priorities for the different production orders, material requirements planning, and capacity planning. These functions are described in Chapter 25.
- *Computer-aided line balancing.* Finding the best allocation of work elements among stations on an assembly line is a large and difficult problem if the line is of significant size. Computer programs are available to assist in the solution of the line balancing problem (Section 15.3).

Manufacturing Control. The second category of CAM applications is concerned with computer systems to control and manage the physical operations in the factory. These applications include the following:

- *Process monitoring and control.* Process monitoring and control is concerned with observing and regulating the production equipment and manufacturing processes in the plant. The topic of industrial process control was discussed in Chapter 5. The applications of computer process control are pervasive in modern automated manufacturing systems, which include transfer lines, assembly systems, CNC machine tools, robotics, material handling, and flexible manufacturing systems. All of these topics are covered in earlier chapters.
- *Quality control.* Quality control includes a variety of approaches to ensure the highest possible quality levels in the manufactured product. Quality control systems are covered in Part V.
- *Shop floor control.* Shop floor control refers to production management techniques for collecting data from factory operations and using the data to help control production and inventory in the factory. Shop floor control and factory data collection systems are covered in Chapter 25.
- *Inventory control.* Inventory control is concerned with maintaining the most appropriate levels of inventory in the face of two opposing objectives: minimizing the investment and storage costs of holding inventory, and maximizing service to customers. Inventory control is discussed in Chapter 25.
- *Just-in-time production systems*. Just-in-time (JIT) refers to a production system that is organized to deliver exactly the right number of each component to downstream workstations in the manufacturing sequence just at the time when that component is needed. JIT is one of the pillars of lean production. The term applies not only to production operations but to supplier delivery operations as well. Just-in-time systems and lean production are discussed in Chapter 26.

23.2.2 CAD/CAM

CAD/CAM denotes the integration of design and manufacturing activities by means of computer systems. The method of manufacturing a product is a direct function of its design. With conventional procedures practiced for so many years in industry, engineering drawings were prepared by design draftsmen and later used by manufacturing engineers to develop the process plan. The activities involved in designing the product were separated from the activities associated with process planning. Essentially a two-step procedure was used, which was time-consuming and duplicated the efforts of design and manufacturing personnel. CAD/CAM establishes a direct link between product design and manufacturing engineering. It is the goal of CAD/CAM not only to automate certain phases of design and certain phases of manufacturing, but also to automate the transition from design to manufacturing. In the ideal CAD/CAM system, it is possible to take the design specification of the product as it resides in the CAD database and convert it automatically into a process plan for making the product. Much of the processing might be accomplished on a numerically controlled machine tool. As part of the process plan, the NC part program is generated automatically by the CAD/CAM system, which downloads the program directly to the machine tool. Hence, under this arrangement, product design, NC programming, and physical production are all implemented by computer.

EXAMPLE 24.1 Make or Buy Cost Decision

The quoted price for a certain part is $20.00 per unit for 100 units. The part can be produced in the company's own plant for $28.00. The cost components of making the part are as follows:

$$\begin{aligned} \text{Unit raw material cost} &= \$8.00 \text{ per unit} \\ \text{Direct labor cost} &= \$6.00 \text{ per unit} \\ \text{Labor overhead at 150\%} &= \$9.00 \text{ per unit} \\ \text{Equipment fixed cost} &= \$5.00 \text{ per unit} \\ \text{Total} &= \$28.00 \text{ per unit} \end{aligned}$$

Should the part be bought or made in-house?

Solution: Although the vendor's quote seems to favor a buy decision, consider the possible impact on plant operations if the quote is accepted. Equipment fixed cost of $5.00 is an allocated cost based on an investment that was already made. If the equipment designated for this job is not utilized because of a decision to purchase the part, then the fixed cost continues even if the equipment stands idle. In the same way, the labor overhead cost of $9.00 consists of factory space, utility, and labor costs that remain even if the part is purchased. In addition, there are the costs of purchasing and receiving inspection. By this reasoning, a buy decision is not a good decision because it might cost the company $20.00 + $5.00 + $9.00 = $34.00 per unit (not including purchasing and receiving inspection) if it results in idle time on the machine that would have been used to produce the part. On the other hand, if the equipment in question can be used to produce other parts for which the in-house costs are less than the corresponding outside quotes, then a buy decision is a good decision.

Make or buy decisions are not often as straightforward as in this example. Other factors listed in Table 24.2 also affect the decision. A trend in recent years, especially in the automobile industry, is for companies to stress the importance of building close relationships with parts suppliers.

24.2 COMPUTER-AIDED PROCESS PLANNING

Problems arise when process planning is accomplished manually. Different process planners have different experiences, skills, and knowledge of the available processes in the plant. This means that the process plan for a given part depends on the process planner who developed it. A different planner would likely plan the routing differently. This leads to variations and inconsistencies in the process plans in the plant. Another problem is that the shop-trained people who are familiar with the details of machining and other processes are gradually retiring and will be unavailable in the future to do process planning. As a result of these issues, manufacturing firms are interested in automating the task of process planning using computer-aided process planning (CAPP). The benefits derived from CAPP include the following:

- *Process rationalization and standardization.* Automated process planning leads to more logical and consistent process plans than manual process planning. Standard plans tend to result in lower manufacturing costs and higher product quality.
- *Increased productivity of process planners.* The systematic approach and the availability of standard process plans in the data files permit more work to be accomplished by the process planners.
- *Reduced lead time for process planning.* Process planners working with a CAPP system can provide route sheets in a shorter lead time compared to manual preparation.
- *Improved legibility.* Computer-prepared route sheets are neater and easier to read than manually prepared route sheets.
- *Incorporation of other application programs.* The CAPP program can be interfaced with other application programs, such as cost estimation and work standards.

Computer-aided process planning systems are designed around two approaches: (1) retrieval CAPP systems and (2) generative CAPP systems. Some CAPP systems combine the two approaches in what is known as semi-generative CAPP [10].

24.2.1 Retrieval CAPP Systems

A retrieval CAPP system, also called a *variant CAPP system*, is based on the principles of group technology (GT) and parts classification and coding (Chapter 18). In this form of CAPP, a standard process plan (route sheet) is stored in computer files for each part code number. The standard route sheets are based on current part routings in use in the factory or on an ideal process plan that has been prepared for each family. Developing a database of these process plans requires substantial effort.

A retrieval CAPP system operates as illustrated in Figure 24.3. Before the system can be used for process planning, a significant amount of information must be compiled and entered into the CAPP data files. This is what Chang et al. refer to as the *preparatory phase* [3], [4]. It consists of (1) selecting an appropriate classification and coding scheme for the company, (2) forming part families for the parts produced by the company, and (3) preparing standard process plans for the part families. Steps (2) and (3) are ongoing as new parts are designed and added to the company's design database.

After the preparatory phase has been completed, the system is ready for use. For a new component for which the process plan is to be determined, the first step is to derive the GT code number for the part. With this code number, the user searches the part family file to determine if a standard route sheet exists for the given part code. If the file contains a process plan for the part, it is retrieved (hence, the word "retrieval" for this CAPP system) and displayed for the user. The standard process plan is examined to determine whether any modifications are necessary. It might be that although the new part has the same code number, there are minor differences in the processes required to make it. The user edits the standard plan accordingly. This capacity to alter an existing process plan is what gives the retrieval system its alternative name, "variant" CAPP system.

If the file does not contain a standard process plan for the given code number, the user may search the computer file for a similar or related code number for which a standard route sheet does exist. Either by editing an existing process plan or by starting from scratch, the user prepares the route sheet for the new part. This route sheet becomes the standard process plan for the new part code number.

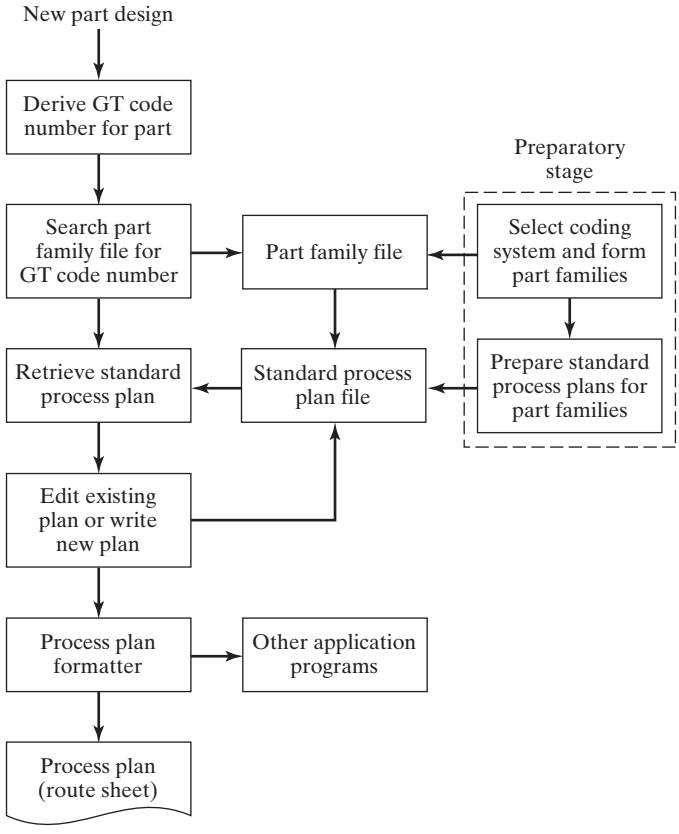


Figure 24.3 General procedure for using one of the retrieval CAPP systems.

The process planning session concludes with the process plan formatter, which prints out the route sheet in the proper format. The formatter may call other application programs into use, for example, to determine machining conditions for the various machine tool operations in the sequence, to calculate standard times for the operations (e.g., for direct labor incentives), or to compute cost estimates for the operations.

24.2.2 Generative CAPP Systems

Generative CAPP systems represent an alternative approach to automated process planning. Instead of retrieving and editing an existing plan contained in a computer database, a generative system creates the process plan based on logical procedures similar to those used by a human planner. In a fully generative CAPP system, the process sequence is planned without human assistance and without a set of predefined standard plans.

Designing a generative CAPP system is usually considered part of the field of expert systems, a branch of artificial intelligence. An ***expert system*** is a computer program that is capable of solving complex problems that normally can only be solved by a human with years of education and experience. Process planning fits within the scope of this definition.

There are several necessary ingredients in a fully generative process planning system. First, the technical knowledge of manufacturing and the logic used by successful process planners must be captured and coded into a computer program. In an expert system applied to process planning, the knowledge and logic of the human process planners is incorporated into a so-called *knowledge base*. The generative CAPP system then uses that knowledge base to solve process planning problems (i.e., create route sheets).

The second ingredient in generative process planning is a computer-compatible description of the part to be produced. This description contains all of the pertinent data and information needed to plan the process sequence. Two possible ways of providing this description are (1) the geometric model of the part that is developed on a CAD system during product design and (2) a GT code number of the part that defines the part features in significant detail.

The third ingredient in a generative CAPP system is the capability to apply the process knowledge and planning logic contained in the knowledge base to a given part description. In other words, the CAPP system uses its knowledge base to solve a specific problem—planning the process for a new part. This problem-solving procedure is referred to as the *inference engine* in the terminology of expert systems. By using its knowledge base and inference engine, the CAPP system synthesizes a new process plan from scratch for each new part it is presented.

24.3 CONCURRENT ENGINEERING AND DESIGN FOR MANUFACTURING

Concurrent engineering is an approach used in product development in which the functions of design engineering, manufacturing engineering, and other departments are integrated to reduce the elapsed time required to bring a new product to market. In the traditional approach to launching a new product, the two functions of design engineering and manufacturing engineering tend to be separated and sequential, as illustrated in Figure 24.4(a). The product design department develops the new design, sometimes without much consideration given to the manufacturing capabilities of the company. There is little opportunity for manufacturing engineers to offer advice on how the design might be altered to make it more manufacturable. It is as if a wall exists between design and manufacturing. When the design engineering department completes the design, it tosses the drawings and specifications over the wall, and only then does process planning begin.

By contrast, in a company that practices concurrent engineering, the manufacturing engineering department becomes involved in the product development cycle early on, providing advice on how the product and its components can be designed to facilitate manufacture and assembly. It also proceeds with the early stages of manufacturing planning for the product. This concurrent engineering approach is pictured in Figure 24.4(b). The product development cycle also involves quality engineering, the manufacturing departments, field service, vendors supplying critical components, and in some cases the customers who will use the product. All of these groups can make contributions during product development to improve not only the new product's function and performance, but also its produceability, inspectability, testability, serviceability, and maintainability. Through early involvement, as opposed to reviewing the final product design after it is too late to conveniently make any changes, the duration of the product development cycle is substantially reduced.

Concurrent engineering includes several elements: (1) design for manufacturing and assembly, (2) design for quality, (3) design for cost, and (4) design for life cycle. In addition, certain enabling technologies such as rapid prototyping, virtual prototyping, and organizational changes are required to facilitate the concurrent engineering approach in a company.

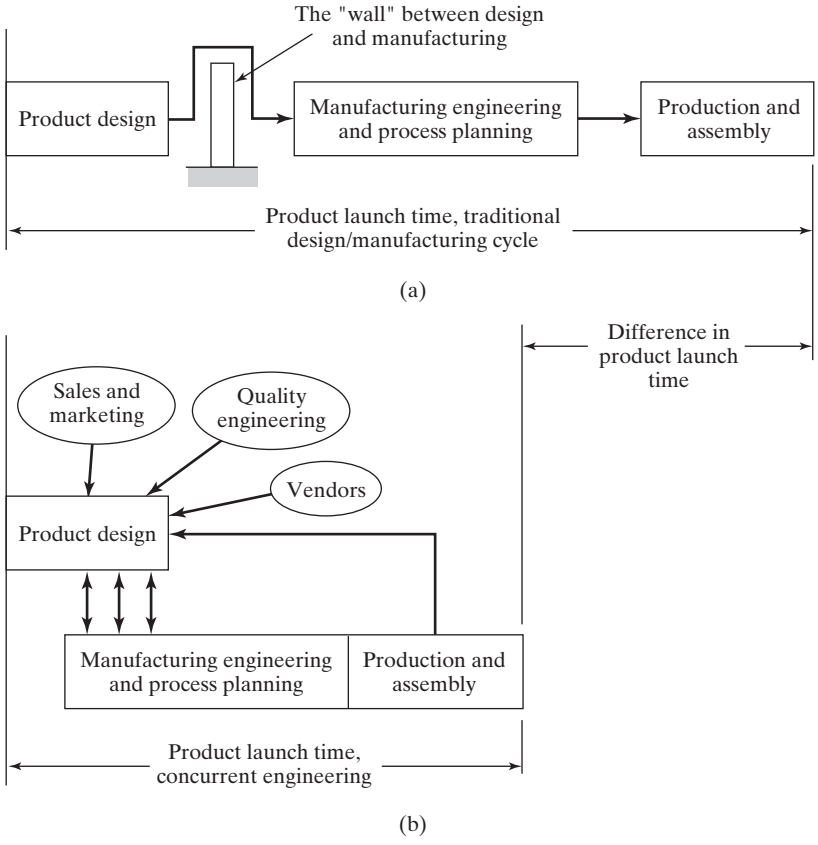


Figure 24.4 (a) Traditional product development cycle and (b) product development using concurrent engineering.

24.3.1 Design for Manufacturing and Assembly

It has been estimated that about 70% of the life cycle cost of a product is determined by basic decisions made during product design [12]. These design decisions include the choice of part material, part geometry, tolerances, surface finish, how parts are organized into subassemblies, and the assembly methods to be used. Once these decisions are made, the potential to reduce the manufacturing cost of the product is limited. For example, if the product designer decides that a part is to be made of an aluminum sand casting but the part possesses features that can be achieved only by machining (such as threaded holes and close tolerances), the manufacturing engineer has no alternative except to plan a process sequence that starts with sand casting followed by the sequence of machining operations needed to achieve the specified features. In this example, a better decision might be to use a plastic molded part that can be made in a single step. It is important for the manufacturing engineer to have the opportunity to advise the design engineer as the product design is evolving, to favorably influence the manufacturability of the product.

Terms used to describe such attempts to favorably influence the manufacturability of a new product are *design for manufacturing* (DFM) and *design for assembly* (DFA). Of course, DFM and DFA are inextricably linked, so the term *design for manufacturing and*

assembly (DFM/A) is used here. It involves the systematic consideration of manufacturability and assemblability in the development of a new product design. This includes (1) organizational changes and (2) design principles and guidelines.

Organizational Changes in DFM/A. Effective implementation of DFM/A involves making changes in a company's organizational structure, either formally or informally, so that closer interaction and better communication occurs between design and manufacturing personnel. This can be accomplished in several ways: (1) by creating project teams consisting of product designers, manufacturing engineers, and other specialties (e.g., quality engineers, material scientists) to develop the new product design; (2) by requiring design engineers to spend some career time in manufacturing to witness firsthand how manufacturability and assemblability are impacted by a product's design; and (3) by assigning manufacturing engineers to the product design department on either a temporary or full-time basis to serve as producibility consultants.

Design Principles and Guidelines. DFM/A also relies on the use of design principles and guidelines to maximize manufacturability and assemblability. Some of these are universal design guidelines that can be applied to nearly any product design situation, such as those presented in Table 24.3. In other cases, there are design principles that apply to specific processes, for example, the use of drafts or tapers in casted and molded parts to facilitate removal of the part from the mold. These process-specific guidelines are covered in texts on manufacturing processes, such as you will find in reference [8].

The guidelines sometimes conflict with one another. For example, one of the guidelines in Table 24.3 is to "simplify part geometry; avoid unnecessary features." But another guideline in the same table states that "special geometric features must sometimes be added to components" to design the product for foolproof assembly. And it may also be desirable to combine features of several assembled parts into one component to minimize the number of parts in the product. In these instances, a suitable compromise must be found between design for part manufacture and design for assembly.

24.3.2 Other Concurrent Engineering Objectives

To complete the coverage of concurrent engineering, other design objectives are briefly described: design for quality, cost, and life cycle.

Design for Quality. It might be argued that DFM/A is the most important component of concurrent engineering because it has the potential for the greatest impact on product cost and development time. However, the importance of quality in international competition cannot be minimized. High quality does not just happen. Procedures for achieving it must be devised during product design and process planning. Design for quality (DFQ) refers to the principles and procedures employed to ensure that the highest possible quality is designed into the product. The general objectives of DFQ are [1]: (1) to design the product to meet or exceed customer requirements; (2) to design the product to be "robust," in the sense of Taguchi (Section 20.6.1), that is, to design the product so that its function and performance are relatively insensitive to variations in manufacturing and subsequent application; and (3) to continuously improve the performance, functionality, reliability, safety, and other quality aspects of the product to provide superior value to the customer. The discussion of quality in Part V is certainly consistent with design for quality, but the emphasis in those chapters was directed more at the operational aspects of quality during production.

TABLE 24.3 General Principles and Guidelines in DFM/A

Guideline	Interpretation and Advantages
Minimize number of components	Reduced assembly costs. Greater reliability in final product. Easier disassembly in maintenance and field service. Automation is often easier with reduced part count. Reduced work-in-process and inventory control problems. Fewer parts to purchase; reduced ordering costs.
Use standard commercially available components	Reduced design effort. Fewer part numbers. Better inventory control possible. Avoids design of custom-engineered components. Quantity discounts are possible.
Use common parts across product lines	Group technology (Chapter 18) can be applied. Quantity discounts are possible. Permits development of manufacturing cells.
Design for ease of part fabrication	Use net shape and near-net shape processes where possible. Simplify part geometry; avoid unnecessary features. Avoid making surface smoother than necessary since additional processing may be needed.
Design parts with tolerances that are within process capability	Avoid tolerances less than process capability (Section 20.3.2). Specify bilateral tolerances. Otherwise, additional processing or sortation and scrap are required.
Design the product to be foolproof during assembly	Assembly should be unambiguous. Components should be designed so they can be assembled only one way. Special geometric features must sometimes be added to components.
Minimize flexible components	These include components made of rubber, belts, gaskets, electrical cables, etc. Flexible components are generally more difficult to handle.
Design for ease of assembly	Include part features such as chamfers and tapers on mating parts. Use base part to which other components are added. Use modular design (see following guideline). Design assembly for addition of components from one direction, usually vertically; in mass production this rule can be violated because fixed automation can be designed for multiple direction assembly. Avoid threaded fasteners (screws, bolts, nuts) where possible, especially when automated assembly is used; use fast assembly techniques such as snap fits and adhesive bonding. Minimize number of distinct fasteners.
Use modular design	Each subassembly should consists of 5–15 parts. Easier maintenance and field service. Facilitates automated (and manual) assembly. Reduces inventory requirements. Reduces final assembly time.
Shape parts and products for ease of packaging	Compatible with automated packaging equipment. Facilitates shipment to customer. Can use standard packaging cartons.
Eliminate or reduce adjustments	Many assembled products require adjustments and calibrations. During product design, the need for adjustments and calibrations should be minimized because they are often time consuming in assembly.

Source: Groover [8].

Design for Product Cost. The cost of a product is a major factor in determining its commercial success. Cost affects the price charged for the product and the profit made by the company producing it. Design for product cost (DFC) refers to the efforts of a company to specifically identify how design decisions affect product costs and to develop ways to reduce cost through design. Although the objectives of DFC and DFM/A overlap to some degree, because improved manufacturability usually results in lower cost, the scope of design for product cost extends beyond manufacturing in its pursuit of cost savings. It includes costs of inspection, purchasing, distribution, inventory control, and overhead.

Design for Life Cycle. To the customer, the price paid for the product may be a small portion of its total cost when life cycle costs are considered. Design for life cycle refers to the product after it has been manufactured and includes factors ranging from product delivery to product disposal. Other life cycle factors include installability, reliability, maintainability, serviceability, and upgradeability. Some customers (e.g., the federal government) include consideration of these costs in their purchasing decisions. The producer of the product is often obliged to offer service contracts that limit customer liability for out-of-control maintenance and service costs. In these cases, accurate estimates of these life cycle costs must be included in the total product cost.

24.4 ADVANCED MANUFACTURING PLANNING

Advanced manufacturing planning emphasizes planning for the future. It is a corporate-level activity that is distinct from process planning because it is concerned with products being contemplated in the company's long-term plans (2- to 10-year future), rather than products currently being designed and released. Advanced manufacturing planning involves working with sales, marketing, and design engineering to forecast the future products that will be introduced and determine what production resources will be needed to make those products. The future products may require manufacturing technologies and facilities not currently available in the firm. In advanced manufacturing planning, the current equipment and facilities are compared with the processing needs of future planned products to determine what new technologies and facilities should be installed. The general planning cycle is portrayed in Figure 24.5. The feedback loop at the top of the diagram is intended to indicate that the firm's future manufacturing capabilities may motivate new product ideas not previously considered.

Activities in advanced manufacturing planning include (1) new technology evaluation, (2) investment project management, (3) facilities planning, and (4) manufacturing research.

New Technology Evaluation. One of the reasons a company may consider installing new technologies is because future product lines require processing methods not currently used by the company. To introduce the new products, the company must either implement new processing technologies in-house or purchase the components made by the new technologies from vendors. For strategic reasons, it may be in the company's interest to implement a new technology internally and develop staff expertise in that technology as a distinctive competitive advantage for the company. The pros and cons must be analyzed, and the technology itself must be evaluated to assess its merits and demerits.

A good example of the need for technology evaluation has occurred in the microelectronics industry, whose history spans only the past several decades. The technology of

Chapter 8

Industrial Robotics

CHAPTER CONTENTS

8.1 Robot Anatomy and Related Attributes
- 8.1.1 Joints and Links
- 8.1.2 Common Robot Configurations
- 8.1.3 Joint Drive Systems
- 8.1.4 Sensors in Robotics

8.2 Robot Control Systems

8.3 End Effectors
- 8.3.1 Grippers
- 8.3.2 Tools

8.4 Applications of Industrial Robots
- 8.4.1 Material Handling Applications
- 8.4.2 Processing Operations
- 8.4.3 Assembly and Inspection
- 8.4.4 Economic Justification of Industrial Robots

8.5 Robot Programming
- 8.5.1 Leadthrough Programming
- 8.5.2 Robot Programming Languages
- 8.5.3 Simulation and Off-Line Programming

8.6 Robot Accuracy and Repeatability

An ***industrial robot*** is defined as "an automatically controlled, reprogrammable, multipurpose manipulator programmable in three or more axes, which may be either fixed in

place or mobile for use in industrial automation applications."[1] It is a general-purpose machine possessing certain anthropomorphic characteristics, the most obvious of which is its mechanical arm. Other human-like characteristics are the robot's capabilities to respond to sensory inputs, communicate with other machines, and make decisions. These capabilities permit robots to perform a variety of industrial tasks. The development of robotics technology followed the development of numerical control (Historical Note 8.1), and the two technologies are quite similar. They both involve coordinated control of multiple axes (the axes are called *joints* in robotics), and they both use dedicated digital computers as controllers. Whereas NC (numerical control) machines are designed to perform specific processes (e.g., machining, sheet metal hole punching, and thermal cutting), robots are designed for a wider variety of tasks. Typical production applications of industrial robots include spot welding, material transfer, machine loading, spray painting, and assembly.

Some of the qualities that make industrial robots commercially and technologically important are the following:

- Robots can be substituted for humans in hazardous or uncomfortable work environments.
- A robot performs its work cycle with a consistency and repeatability that cannot be attained by humans.
- Robots can be reprogrammed. When the production run of the current task is completed, a robot can be reprogrammed and equipped with the necessary tooling to perform an altogether different task.
- Robots are controlled by computers and can therefore be connected to other computer systems to achieve computer integrated manufacturing.

Historical Note 8.1 A Short History of Industrial Robots [5] [12]

The word "robot" entered the English language through a Czechoslovakian play titled *Rossum's Universal Robots*, written by Karel Capek in the early 1920s. The Czech word *robota* means forced worker. In the English translation, the word was converted to "robot." The story line of the play centers around a scientist named Rossum who invents a chemical substance similar to protoplasm and uses it to produce robots. The scientist's goal is for robots to serve humans and perform physical labor. Rossum continues to make improvements in his invention, ultimately perfecting it. These "perfect beings" begin to resent their subservient role in society and turn against their masters, killing off all human life.

Capek's play was pure science fiction. This brief history must include two real inventors who made original contributions to the technology of industrial robotics. The first was Cyril W. Kenward, a British inventor who devised a manipulator that moved on an x–y–z axis system. In 1954, Kenward applied for a British patent for his robotic device, and in 1957 the patent was issued.

The second inventor was an American named George C. Devol. Devol is credited with two inventions related to robotics. The first was a device for magnetically recording

[1]Industrial robot as defined in the ISO 8373 standard [16].

electrical signals so that the signals could be played back to control the operation of machinery. This device was invented around 1946, and a U.S. patent was issued in 1952. The second invention was a robotic device developed in the 1950s, which Devol called "programmed article transfer." This device was intended for parts handling. The U.S. patent was finally issued in 1961. It was a prototype for the hydraulically driven robots that were later built by Unimation, Inc.

Although Kenward's robot was chronologically the first (at least in terms of patent date), Devol's proved ultimately to be far more important in the development and commercialization of robotics technology. The reason for this was a catalyst in the person of Joseph Engelberger. Engelberger had graduated from Columbia University with degrees in physics in 1946 and 1949. As a student, he had read science fiction novels about robots. By the mid-1950s, he was working for a company that made control systems for jet engines. Hence, by the time a chance meeting occurred between Engelberger and Devol in 1956, Engelberger was "predisposed by education, avocation, and occupation toward the notion of robotics."[2] The meeting took place at a cocktail party in Fairfield, Connecticut. Devol described his programmed article transfer invention to Engelberger, and they subsequently began considering how to develop the device as a commercial product for industry. In 1962, Unimation, Inc., was founded, with Engelberger as president. The name of the company's first product was "Unimate," a polar configuration robot. The first application of a Unimate robot was unloading a die casting machine at a General Motors plant in New Jersey in 1961.

8.1 ROBOT ANATOMY AND RELATED ATTRIBUTES

The arm or manipulator of an industrial robot consists of a series of joints and links. Robot anatomy is concerned with the types and sizes of these joints and links and other aspects of the manipulator's physical construction. The robot's anatomy affects its capabilities and the tasks for which it is best suited.

8.1.1 Joints and Links

A robot's joint, or *axis* as it is also called in robotics, is similar to a joint in the human body: It provides relative motion between two parts of the body. Robots are often classified according to the total number of axes they possess. Connected to each joint are two links, 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.

Most robots are mounted on a stationary base on the floor. Let this base and its connection to the first joint be referred to as link 0. It is the input link to joint 1, the first in the series of joints used in the construction of the robot. The output link of joint 1 is link 1. Link 1 is the input link to joint 2, whose output link is link 2, and so forth. This joint-link numbering scheme is illustrated in Figure 8.1.

Nearly all industrial robots have mechanical joints that can be classified into one of five types: two types that provide translational motion and three types that provide

[2]This quote was borrowed from Groover et al., *Industrial Robotics: Technology, Programming, and Applications* [5].

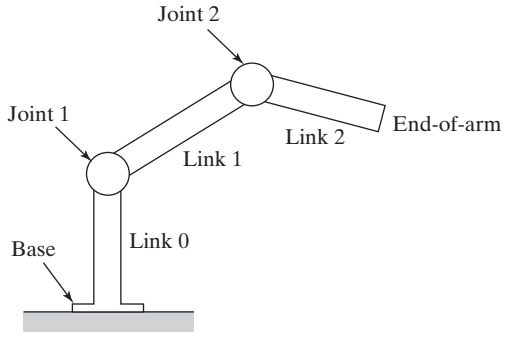


Figure 8.1 Diagram of robot construction showing how a robot is made up of a series of joint-link combinations.

rotary motion. These joint types are illustrated in Figure 8.2 and are based on a scheme described in [5]. The five joint types are

1. *Linear joint* (type L joint). The relative movement between the input link and the output link is a translational telescoping motion, with the axes of the two links being parallel.
2. *Orthogonal joint* (type O joint). This is also a translational sliding motion, but the input and output links are perpendicular to each other.
3. *Rotational joint* (type R joint). This type provides rotational relative motion, with the axis of rotation perpendicular to the axes of the input and output links.
4. *Twisting joint* (type T joint). This joint also involves rotary motion, but the axis of rotation is parallel to the axes of the two links.
5. *Revolving joint* (type V joint, V from the "v" in revolving). In this joint type, 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.

Each of these joint types has a range over which it can be moved. The range for a translational joint is usually less than a meter, but for large gantry robots, the range may be several meters. The three types of rotary joints may have a range as small as a few degrees or as large as several complete revolutions.

8.1.2 Common Robot Configurations

A robot manipulator can be divided into two sections: a body-and-arm assembly and a wrist assembly. There are usually three axes associated with the body-and-arm, and either two or three axes associated with the wrist. At the end of the manipulator's wrist is a device related to the task that must be accomplished by the robot. The device, called an *end effector* (Section 8.3), is usually either (1) a gripper for holding a work part or (2) a tool for performing some process. The body-and-arm of the robot is used to position the end effector, and the robot's wrist is used to orient the end effector.

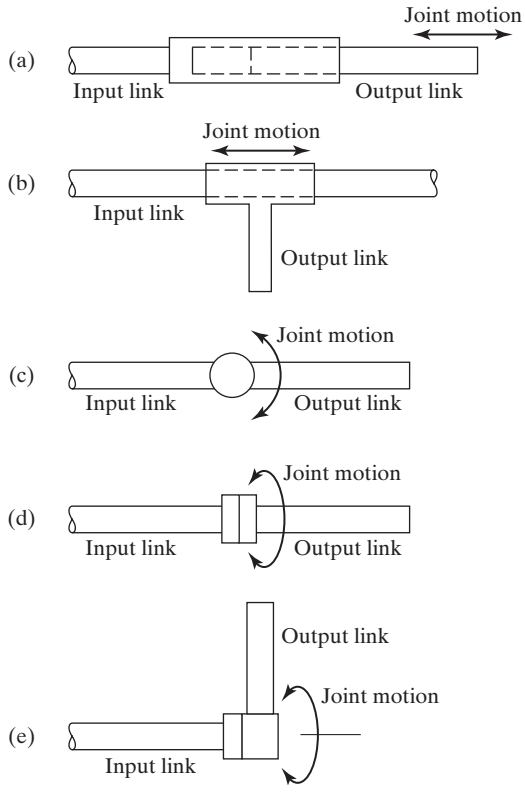


Figure 8.2 Five types of joints commonly used in industrial robot construction: (a) linear joint (type L joint), (b) orthogonal joint (type O joint), (c) rotational joint (type R joint), (d) twisting joint (type T joint), and (e) revolving joint (type V joint).

Body-and-Arm Configurations. Given the five types of joints defined earlier, there are $5 \times 5 \times 5 = 125$ possible combinations of joints that could be used to design the body-and-arm assembly for a three-axis manipulator. In addition, there are design variations within the individual joint types (e.g., physical size of the joint and range of motion). It is somewhat remarkable, therefore, that only a few configurations are commonly available in commercial industrial robots. These configurations are:

1. *Articulated robot.* Also known as a *jointed-arm robot* (Figure 8.3), it has the general configuration of a human shoulder and arm. It consists of an upright body that swivels about the base using a T joint. At the top of the body is a shoulder joint (shown as an R joint in the figure), whose output link connects to an elbow joint (another R joint).
2. *Polar configuration.* This configuration (Figure 8.4) consists of a sliding arm (L joint) actuated relative to the body, which can rotate about both a vertical axis (T joint) and a horizontal axis (R joint).[3]

[3]The polar configuration was the design used for the first commercial robot, the Unimate, produced by Unimation, Inc., in the 1960s. Once the most widely used robot configuration, for machine loading and automobile spot welding, it is no longer favored by robot designers today.

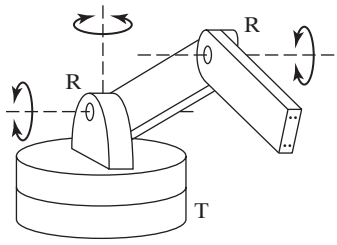


Figure 8.3 Articulated robot (jointed-arm robot).

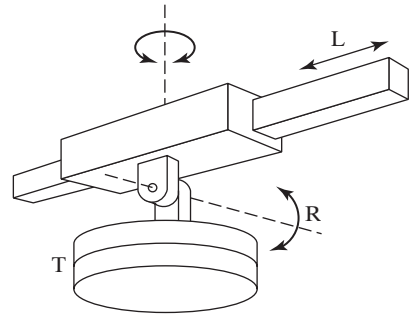


Figure 8.4 Polar configuration.

3. *SCARA*. SCARA is an acronym for *S*electively *C*ompliant *A*rm for *R*obotic *A*ssembly. This configuration (Figure 8.5) is similar to the jointed-arm robot except that the shoulder and elbow rotational axes are vertical, which means that the arm is very rigid in the vertical direction, but compliant in the horizontal direction. This permits the robot to perform insertion tasks (for assembly) in a vertical direction, where some side-to-side alignment may be needed to mate the two parts properly.
4. *Cartesian coordinate robot.* Other names for this configuration include *gantry robot*, *rectilinear robot*, and *x–y–z robot*. As shown in Figure 8.6, it consists of three orthogonal joints (type O) to achieve linear motions in a three-dimensional rectangular work space. It is commonly used for overhead access to load and unload production machines.
5. *Delta robot*. This unusual design, depicted in Figure 8.7, consists of three arms attached to an overhead base. Each arm is articulated and consists of two rotational joints (type R), the first of which is powered and the second is unpowered. All three arms are connected to a small platform below, to which the end effector is attached. The platform and end effector can be manipulated in three dimensions. The delta robot is used for high-speed movement of small objects, as in product packaging.

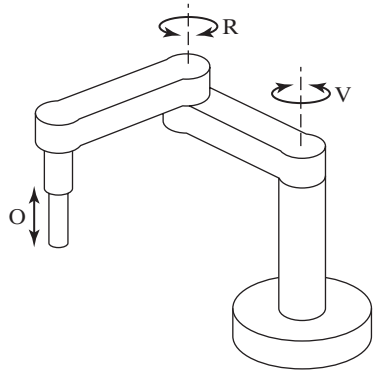


Figure 8.5 SCARA configuration.

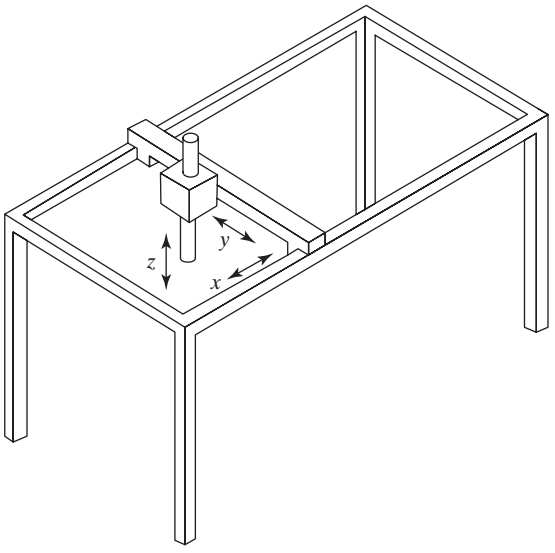


Figure 8.6 Cartesian coordinate robot.

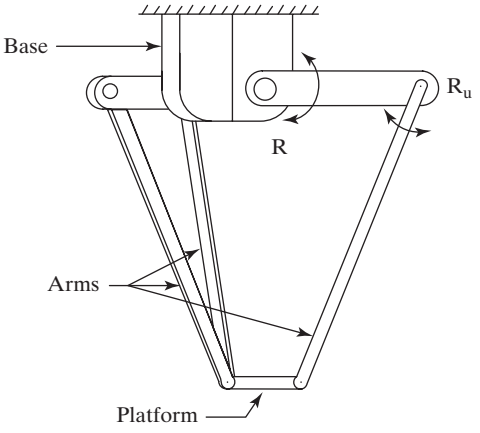


Figure 8.7 Delta robot.

The first three configurations follow the serial joint-link configuration mounted on the floor, as pictured in Figure 8.1. The Cartesian coordinate and delta configurations are exceptions to this convention. The usual Cartesian coordinate robot is suspended from a gantry structure. The first and second axes permit x–y movement over a rectangular area above the floor. The third axis permits movement in the z direction to reach downward.[4] Depending on the application requirements, a wrist assembly can be attached to the end of the arm (link 3).

The delta robot is also suspended from an overhead base rather than floor mounted. Its most unique feature is its three articulated arms that are all connected to the platform below. Each arm has one powered joint and one follower joint. The overhead base is link 0 for all three arms. In each arm, joint 1 is rotational and powered. Its output link 1 is connected to joint 2, which is unpowered, and output link 2 is connected to the platform. By coordinating the actuations of the three powered joints, the position of the platform can be controlled in three dimensions while maintaining the orientation of the end effector. A separate wrist assembly is usually not included in the delta robot. The end effector is attached directly to the underside of the platform.

Like the delta robot, the SCARA configuration typically does not have a separate wrist assembly. As indicated in the description, it is used for insertion-type assembly operations in which the insertion is made from above. Accordingly, orientation requirements are minimal, and the wrist is not needed. Orientation of the object to be inserted is sometimes required, and an additional rotary joint added to link 3 can be provided for this purpose.

[4]A recently introduced robot from ABB Robotics combines the overhead gantry design with an articulated robot arm [14]. The gantry structure allows a large x–y work envelope, and the jointed arm provides orientation capability for equipment access.

There are more manipulator configurations than those described, and they come in many different sizes. The interested reader can peruse the websites of some of the robot manufacturers in [14], [15], and [17].

Wrist Configurations. The robot's wrist is used to establish the orientation of the end effector. Robot wrists usually consist of two or three joints that almost always consist of R and T type rotary joints. Figure 8.8 illustrates one possible configuration for a three-axis wrist assembly. The three joints are defined as follows: (1) ***roll***, using a T joint to accomplish rotation about the robot's arm axis; (2) ***pitch***, which involves up-and-down rotation, typically using an R joint; and (3) ***yaw***, which involves right-and-left rotation, also accomplished by means of an R-joint. A two-axis wrist typically includes only roll and pitch joints (T and R joints).

To avoid confusion in the pitch and yaw definitions, the wrist roll should be assumed in its center position, as shown in the figure. To demonstrate the possible confusion, consider a two-jointed wrist assembly. With the roll joint in its center position, the second joint (R joint) provides up-and-down rotation (pitch). However, if the roll position were 90 degrees from center (either clockwise or counterclockwise), the second joint would provide a right-left rotation (yaw).

Joint Notation System. The letter symbols for the five joint types (L, O, R, T, and V) can be used in a joint notation system for the robot manipulator and wrist. In this notation system, the manipulator is described by the joint types that make up the body-and-arm assembly, followed by the joint symbols that make up the wrist. For example, the notation TLR:RT represents a five-axis manipulator whose body-and-arm is made up of a twisting joint (joint 1 = T), a linear joint (joint 2 = L), and a rotational joint (joint 3 = R). The wrist consists of two joints, a rotational joint (joint 4 = R) and a twisting joint (joint 5 = T). A colon separates the body-and-arm notation from the wrist notation. Common wrist joint notations are TRR, TR, and RT.

Typical joint notations for the five body-and-arm configurations are presented in Table 8.1. The notation for the delta robot indicates that there are three arms, and each arm has two rotational joints, the second of which is unpowered (R_u).

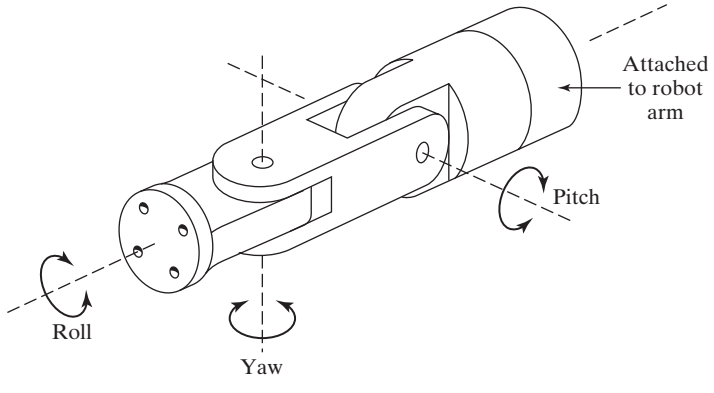


Figure 8.8 Typical configuration of a three-axis wrist assembly showing roll, pitch, and yaw.

TABLE 8.1 Joint Notations for the Five Common Robot Body-and-Arm Configurations

Body-and-Arm	Joint Notation	Work Volume
Articulated	TRR (Figure 8.3)	Partial sphere
Polar	TRL (Figure 8.4)	Partial sphere
SCARA	VRO (Figure 8.5)	Cylindrical
Cartesian coordinate	OOO (Figure 8.6)	Rectangular solid
Delta	$3(RR_u)$ (Figure 8.7)	Hemisphere

Work Volume. The work volume (also known as *work envelope*) of the manipulator is defined as the three-dimensional space within which the robot can manipulate the end of its wrist. Work volume is determined by the number and types of joints in the manipulator (body-and-arm and wrist), the ranges of the various joints, and the physical sizes of the links. The shape of the work volume depends largely on the robot's configuration, as indicated in Table 8.1.

8.1.3 Joint Drive Systems

Robot joints are actuated using any of three types of drive systems: (1) electric, (2) hydraulic, or (3) pneumatic. Electric drive systems use electric motors as joint actuators (e.g., servomotors or stepper motors, Sections 6.2.1 and 7.4). The motors are connected to the joints either using no gear reduction (called *direct drive*) or with a gear reduction to increase torque or force. Hydraulic and pneumatic drive systems use devices such as linear pistons and rotary vane actuators (Section 6.2.2) to move the joint.

Pneumatic drive is typically limited to smaller robots used in simple part transfer applications. Electric drive and hydraulic drive are used on more sophisticated industrial robots. Electric drive has become the preferred drive system in commercially available robots, as electric motor technology has advanced in recent years. It is more readily adaptable to computer control, which is the dominant technology used today for robot controllers. Electric drive robots are relatively accurate compared with hydraulically powered robots. By contrast, hydraulic drive robots can be designed with greater lift capacity.

The drive system, position sensors (and speed sensors if used), and feedback control systems for the joints determine the dynamic response characteristics of the manipulator. The speed with which the robot can move to a programmed position and the stability of its motion are important characteristics of dynamic response in robotics. ***Motion speed*** refers to the absolute velocity of the manipulator at its end-of-arm. The maximum speed of a large robot is around 2 m/sec (6 ft/sec). Speed can be programmed into the work cycle so that different portions of the cycle are carried out at different velocities. What is sometimes more important than speed is the robot's capability to accelerate and decelerate in a controlled manner. In many work cycles, much of the robot's movement is performed in a confined region of the work volume, so the robot never achieves its top-rated velocity. In these cases, nearly all of the motion cycle is engaged in acceleration and deceleration rather than in constant speed. Other factors that influence speed of motion are the weight (mass) of the object that is being manipulated and the precision with which the object must be located at the end of a given move. All of these factors are included in the ***speed of response***, which is the time required for the

manipulator to move from one point in space to the next. Speed of response is important because it influences the robot's cycle time, which in turn affects the production rate in the application. ***Motion stability*** refers to the amount of overshoot and oscillation that occurs in the robot motion at the end-of-arm as it attempts to move to the next programmed location. More oscillation in the motion is an indication of less stability. The problem is that robots with greater stability are inherently slower in their response, whereas faster robots are generally less stable.

Load carrying capacity depends on the robot's physical size and construction as well as the force and power that can be transmitted to the end of the wrist. The weight carrying capacity of commercial robots ranges from less than 1 kg up to approximately 1,200 kg (2,600 lb) [15]. Medium sized robots designed for typical industrial applications have capacities in the range of 10–60 kg (22–130 lb). One factor that should be kept in mind when considering load carrying capacity is that a robot usually works with a tool or gripper attached to its wrist. Grippers are designed to grasp and move objects about the work cell. The net load-carrying capacity of the robot is obviously reduced by the weight of the gripper. If the robot is rated at 10 kg (22 lb) and the weight of the gripper is 4 kg (9 lbs), then the net weight carrying capacity is reduced to 6 kg (13 lb).

8.1.4 Sensors in Robotics

The general topic of sensors as components in control systems is discussed in Section 6.1. The discussion here is on how sensors are applied in robotics. Sensors used in industrial robotics can be classified into two categories: (1) internal and (2) external. Internal sensors are components of the robot and are used to control the positions and velocities of the robot joints. These sensors form a feedback control loop with the robot controller. Typical sensors used to control the position of the robot arm include potentiometers and optical encoders. Tachometers of various types are used to control the speed of the robot arm.

External sensors are external to the robot and are used to coordinate the operation of the robot with other equipment in the cell. In many cases, these external sensors are relatively simple devices, such as limit switches that determine whether a part has been positioned properly in a fixture or that a part is ready to be picked up at a conveyor. Other situations require more advanced sensor technologies, including the following:

- *Tactile sensors.* These are used to determine whether contact is made between the sensor and another object. Tactile sensors can be divided into two types in robot applications: (1) touch sensors and (2) force sensors. Touch sensors indicate simply that contact has been made with the object. Force sensors indicate the magnitude of the force with the object. This might be useful in a gripper to measure and control the force being applied to grasp a delicate object.
- *Proximity sensors.* These indicate when an object is close to the sensor. When this type of sensor is used to indicate the actual distance of the object, it is called a *range sensor*.
- *Optical sensors.* Photocells and other photometric devices can be utilized to detect the presence or absence of objects and are often used for proximity detection.
- *Machine vision.* Machine vision is used in robotics for inspection, parts identification, guidance, and other uses. Section 22.5 provides a more complete discussion of machine vision in automated inspection. Improvements in programming of

vision-guided robot (VGR) systems have made implementations of this technology easier and faster [20], and machine vision is being implemented as an integral feature in more and more robot installations, especially in the automotive industry [13].

- *Other sensors.* A miscellaneous category includes other types of sensors that might be used in robotics, such as devices for measuring temperature, fluid pressure, fluid flow, electrical voltage, current, and various other physical properties (Table 6.2).

8.2 ROBOT CONTROL SYSTEMS

The actuations of the individual joints must be controlled in a coordinated fashion for the manipulator to perform a desired motion cycle. Microprocessor-based controllers are commonly used today in robotics as the control system hardware. The controller is organized in a hierarchical structure as indicated in Figure 8.9 so that each joint has its own feedback control system, and a supervisory controller coordinates the combined actuations of the joints according to the sequence of the robot program. Different types of control are required for different applications. Robot controllers can be classified into four categories [5]: (1) limited-sequence control, (2) playback with point-to-point control, (3) playback with continuous path control, and (4) intelligent control.

Limited-Sequence Control. This is the most elementary control type. It can be utilized only for simple motion cycles, such as pick-and-place operations (i.e., picking an object up at one location and placing it at another location). It is usually implemented by setting limits or mechanical stops for each joint and sequencing the actuation of the joints to accomplish the cycle. Interlocks (Section 5.3.2) are sometimes used to indicate that the particular joint actuation has been accomplished so that the next step in the sequence can be initiated. However, there is no servo-control to accomplish precise positioning of the joint. Many pneumatically driven robots are limited-sequence robots.

Playback with Point-to-Point Control. Playback robots represent a more sophisticated form of control than limited-sequence robots. Playback control means that the controller has a memory to record the sequence of motions in a given work cycle, as well as the locations and other parameters (such as speed) associated with each motion, and then to subsequently play back the work cycle during execution of the program.

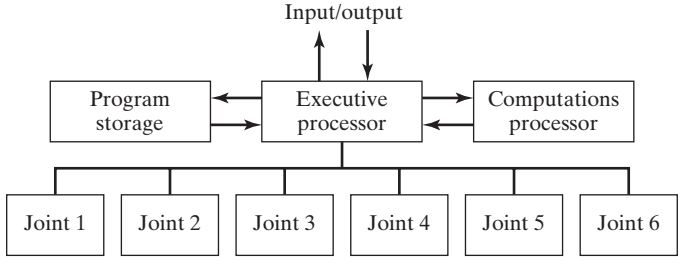


Figure 8.9 Hierarchical control structure of a robot microcomputer controller.

In point-to-point (PTP) control, individual positions of the robot arm are recorded into memory. These positions are not limited to mechanical stops for each joint as in limited-sequence robots. Instead, each position in the robot program consists of a set of values representing locations in the range of each joint of the manipulator. Thus, each "point" consists of five or six values corresponding to the positions of each of the five or six joints of the manipulator. For each position defined in the program, the joints are thus directed to actuate to their respective specified locations. Feedback control is used during the motion cycle to confirm that the individual joints achieve the specified locations in the program. Interlocks are used to coordinate the actions of the robot with the actions of other equipment in the work cell.

Playback with Continuous Path Control. Continuous path robots have the same playback capability as the previous type. The difference between continuous path and point-to-point is the same in robotics as it is in NC (Section 7.1.3). A playback robot with continuous path control is capable of one or both of the following:

1. *Greater storage capacity.* The controller has a far greater storage capacity than its point-to-point counterpart, so the number of locations that can be recorded into memory is far greater than for point-to-point. Thus, the points constituting the motion cycle can be spaced very closely together to permit the robot to accomplish a smooth continuous motion. In PTP, only the final location of the individual motion elements are controlled, so the path taken by the arm to reach the final location is not controlled. In a continuous path motion, the movement of the arm and wrist is controlled during the motion.
2. *Interpolation calculations.* The controller computes the path between the starting point and the ending point of each move using interpolation routines similar to those used in NC. These routines generally include linear and circular interpolation (Table 7.1).

The difference between PTP and continuous path control can be explained mathematically as follows. Consider a three-axis Cartesian coordinate manipulator in which the end-of-arm is moved in x–y–z space. In point-to-point systems, the x-, y-, and z-axes are controlled to achieve a specified point location within the robot's work volume. In continuous path systems, not only are the x-, y-, and z-axes controlled, but the velocities dx/dt, dy/dt, and dz/dt are controlled simultaneously to achieve the specified linear or curvilinear path. Servo-control is used to continuously regulate the position and speed of the manipulator. It should be mentioned that a playback robot with continuous path control has the capacity for PTP control.

Intelligent Control. Industrial robots are becoming increasingly intelligent. In this context, an intelligent robot is one that exhibits behavior that makes it seem intelligent. Some of the characteristics that make a robot appear intelligent include the capacities to interact with its environment, make decisions when things go wrong during the work cycle, communicate with humans, make computations during the motion cycle, and respond to advanced sensor inputs such as machine vision.

In addition, robots with intelligent control possess playback capability for both PTP and continuous path control. All of these features require (1) a relatively high level of computer control and (2) an advanced programming language to input the decision-making logic and other "intelligence" into memory.

8.3 END EFFECTORS

As mentioned in Section 8.1.2 on robot configurations, an end effector is usually attached to the robot's wrist. The end effector enables the robot to accomplish a specific task. Because there is a wide variety of tasks performed by industrial robots, the end effector is usually custom-engineered and fabricated for each different application. The two categories of end effectors are grippers and tools.

8.3.1 Grippers

Grippers are end effectors used to grasp and manipulate objects during the work cycle. The objects are usually work parts 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, most grippers must 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 close on the work part (Figure 8.10 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*, which use an adhesive substance to hold a flexible material such as a fabric
- *Simple mechanical devices,* such as hooks and scoops.

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 for machine loading and unloading. With a single gripper, the robot must reach into the production machine twice, once to unload the finished part and position it in a location external to the machine, and the second time to pick up the next part and load it into the machine. With a dual gripper, the robot picks up the next work part while the

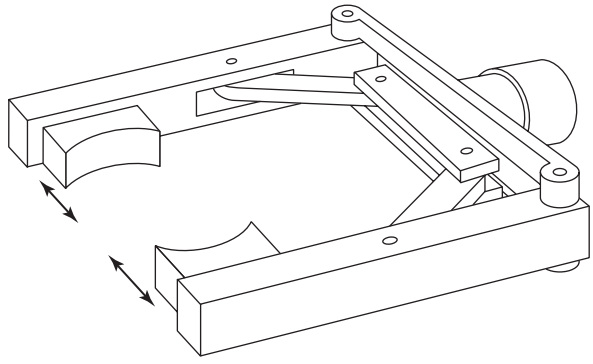


Figure 8.10 Robot mechanical gripper.

machine is still processing the previous part. When the machine cycle is finished, the robot reaches into the machine only 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 work part or (2) applying a specified limited force to the work part during gripping (for fragile work parts).
- *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.

8.3.2 Tools

The robot uses tools to perform processing operations on the work part. The robot manipulates the tool relative to a stationary or slowly moving object (e.g., work part or subassembly). Examples of 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 similar operations; assembly tool (e.g., automatic screwdriver); heating torch; ladle (for metal die casting); and water jet cutting tool. In each case, the robot must not only control the relative position of the tool with respect to the work as a function of time, it must also control the operation of the tool. For this purpose, the robot must be able to transmit control signals to the tool for starting, stopping, and otherwise regulating its actions.

In some applications, the robot may use multiple tools during the work cycle. For example, several sizes of routing or drilling bits must be applied to the work part. Thus, the robot must have a means of rapidly changing the tools. The end effector in this case takes the form of a fast-change tool holder for quickly fastening and unfastening the various tools used during the work cycle.

8.4 APPLICATIONS OF INDUSTRIAL ROBOTS

Robots are used in a wide field of applications in industry. Most of the current applications are in manufacturing. The applications can usually be classified into one of the following categories: (1) material handling, (2) processing operations, and (3) assembly and inspection. Section 8.4.4 lists some of the work characteristics that must be present in the application to make the installation of a robot technically and economically feasible.

8.4.1 Material Handling Applications

In material handling applications, the robot moves materials or parts from one place to another. To accomplish the transfer, the robot is equipped with a gripper that must be designed to handle the specific part or parts to be moved. Included within this application category are (1) material transfer and (2) machine loading and/or unloading. In many material handling applications, the parts must be presented to the robot in a known position and orientation. This requires some form of material handling device to deliver the parts into the work cell in this position and orientation.

Chapter 7

Computer Numerical Control

CHAPTER CONTENTS

7.1 Fundamentals of NC Technology
- 7.1.1 Basic Components of an NC System
- 7.1.2 NC Coordinate Systems
- 7.1.3 Motion Control Systems

7.2 Computers and Numerical Control
- 7.2.1 The CNC Machine Control Unit
- 7.2.2 CNC Software
- 7.2.3 Distributed Numerical Control

7.3 Applications of NC
- 7.3.1 Machine Tool Applications
- 7.3.2 Other NC Applications
- 7.3.3 Advantages and Disadvantages of NC

7.4 Analysis of Positioning Systems
- 7.4.1 Open-Loop Positioning Systems
- 7.4.2 Closed-Loop Positioning Systems
- 7.4.3 Precision in Positioning Systems

7.5 NC Part Programming
- 7.5.1 Manual Part Programming
- 7.5.2 Computer-Assisted Part Programming
- 7.5.3 CAD/CAM Part Programming
- 7.5.4 Manual Data Input

Appendix 7A: Coding for Manual Part Programming

Numerical control (NC) is a form of programmable automation in which the mechanical actions of a machine tool or other equipment are controlled by a program containing coded alphanumeric data. The alphanumeric data represent relative positions between a work head and a work part as well as other instructions needed to operate the machine. The work head is a cutting tool or other processing apparatus, and the work part is the object being processed. When the current job is completed, the program of instructions can be changed to process a new job. The capability to change the program makes NC suitable for low and medium production. It is much easier to write new programs than to make major alterations in the processing equipment.

Numerical control can be applied to a wide variety of processes. The applications divide into two categories: (1) machine tool applications, such as drilling, milling, turning, and other metal working; and (2) other applications, such as assembly, rapid prototyping, and inspection. The common operating feature of NC in all of these applications is control of the work head movement relative to the work part. The concept for NC dates from the late 1940s. The first NC machine was developed in 1952 (Historical Note 7.1).

Historical Note 7.1 The First NC Machines [1], [4], [7], [9]

The development of NC owes much to the U.S. Air Force and the early aerospace industry. The first work in the area of NC is attributed to John Parsons and his associate Frank Stulen at Parsons Corporation in Traverse City, Michigan. Parsons was a contractor for the Air Force during the 1940s and had experimented with the concept of using coordinate position data contained on punched cards to define and machine the surface contours of airfoil shapes. He had named his system the *Cardamatic* milling machine, since the numerical data was stored on punched cards. Parsons and his colleagues presented the idea to the Wright-Patterson Air Force Base in 1948. The initial Air Force contract was awarded to Parsons in June 1949. A subcontract was awarded by Parsons in July 1949 to the Servomechanism Laboratories at the Massachusetts Institute of Technology to (1) perform a systems engineering study on machine tool controls and (2) develop a prototype machine tool based on the Cardamatic principle. Research commenced on the basis of this subcontract, which continued until April 1951, when a contract was signed by MIT and the Air Force to complete the development work.

Early in the project, it became clear that the required data transfer rates between the controller and the machine tool could not be achieved using punched cards, so it was proposed to use either punched paper tape or magnetic tape to store the numerical data. These and other technical details of the control system for machine tool control had been defined by June 1950. The name *numerical control* was adopted in March 1951 based on a contest sponsored by John Parsons among "MIT personnel working on the project." The first NC machine was developed by retrofitting a Cincinnati Milling Machine Company vertical Hydro-Tel milling machine (a 24-in $\times$ 60-in conventional tracer mill) that had been donated by the Air Force from surplus equipment. The controller combined analog and digital components, consisted of 292 vacuum tubes, and occupied a floor area greater than the machine tool itself. The prototype successfully performed simultaneous control of three-axis motion based on coordinate-axis data on punched binary tape. This experimental machine was in operation by March 1952.

A patent for the machine tool system entitled *Numerical Control Servo System* was filed in August 1952, and awarded in December 1962. Inventors were listed as Jay Forrester, William Pease, James McDonough, and Alfred Susskind, all Servomechanisms Lab staff

during the project. It is of interest to note that a patent was also filed by John Parsons and Frank Stulen in May 1952 for a *Motor Controlled Apparatus for Positioning Machine Tool* based on the idea of using punched cards and a mechanical rather than electronic controller. This patent was issued in January 1958. In hindsight, it is clear that the MIT research provided the prototype for subsequent developments in NC technology. As far as is known, no commercial machines were ever introduced using the Parsons–Stulen configuration.

Once the NC machine was operational in March 1952, trial parts were solicited from aircraft companies across the country to learn about the operating features and economics of NC. Several potential advantages of NC were apparent from these trials. These included good accuracy and repeatability, reduction of noncutting time in the machining cycle, and the capability to machine complex geometries. Part programming was recognized as a difficulty with the new technology. A public demonstration of the machine was held in September 1952 for machine tool builders (anticipated to be the companies that would subsequently develop products in the new technology), aircraft component producers (expected to be the principal users of NC), and other interested parties.

Reactions of the machine tool companies following the demonstrations "ranged from guarded optimism to outright negativism" [9, p. 61]. Most of the companies were concerned about a system that relied on vacuum tubes, not realizing that tubes would soon be displaced by transistors and integrated circuits. They were also worried about their staff's qualifications to maintain such equipment and were generally skeptical of the NC concept. Anticipating this reaction, the Air Force sponsored two additional tasks: (1) information dissemination to industry and (2) an economic study. The information dissemination task included many visits by Servo Lab personnel to companies in the machine tool industry as well as visits to the Lab by industry personnel to observe demonstrations of the prototype machine. The economic study showed clearly that the applications of general-purpose NC machine tools were in low- and medium-quantity production, as opposed to Detroit-type transfer lines, which could be justified only for very large quantities.

In 1956, the Air Force decided to sponsor the development of NC machine tools at several aircraft companies, and these machines were placed in operation between 1958 and 1960. The advantages of NC soon became apparent, and the aerospace companies began placing orders for new NC machines. In some cases, they even built their own units. This served as a stimulus to the remaining machine tool companies that had not yet embraced NC. Advances in computer technology also stimulated further development. The first application of the digital computer for NC was part programming. In 1956, MIT demonstrated the feasibility of a computer-aided part programming system using an early digital computer prototype that had been developed at MIT. Based on this demonstration, the Air Force sponsored development of a part programming language. This research resulted in the development of the APT language in 1958.

The automatically programmed tool system (APT) was the brainchild of mathematician Douglas Ross, who worked in the MIT Servomechanisms Lab at the time. Recall that this project was started in the 1950s, a time when digital computer technology was in its infancy, as were the associated computer programming languages and methods. The APT project was a pioneering effort, not only in the development of NC technology, but also in computer programming concepts, computer graphics, and computer-aided design (CAD). Ross envisioned a part programming system in which (1) the user would prepare instructions for operating the machine tool using English-like words, (2) the digital computer would translate these instructions into a language that the computer could understand and process, (3) the computer would carry out the arithmetic and geometric calculations needed to execute the instructions, and (4) the computer would further process (post-process) the instructions so that they could be interpreted by the machine tool controller. He further recognized that the programming system should be expandable for applications beyond those considered in the immediate research at MIT (milling applications).

Ross's work at MIT became a focal point for NC programming, and a project was initiated to develop a two-dimensional version of APT, with nine aircraft companies plus IBM Corporation participating in the joint effort and MIT as project coordinator. The 2D-APT system was ready for field evaluation at plants of participating companies in April 1958. Testing, debugging, and refining the programming system took approximately three years. In 1961, the Illinois Institute of Technology Research Institute (IITRI) was selected to become responsible for long-range maintenance and upgrading of APT. In 1962, IITRI announced the completion of APT-III, a commercial version of APT for three-dimensional part programming. In 1974, APT was accepted as the U.S. standard for programming NC metal cutting machine tools. In 1978, it was accepted by the ISO as the international standard.

Numerical control technology was in its second decade before computers were employed to actually control machine tool motions. In the mid-1960s, the concept of *direct numerical control* (DNC) was developed, in which individual machine tools were controlled by a mainframe computer located remotely from the machines. The computer bypassed the punched tape reader, instead transmitting instructions to the machine control unit (MCU) in real time, one block at a time. The first prototype system was demonstrated in 1966 [4]. Two companies that pioneered the development of DNC were General Electric Company and Cincinnati Milling Machine Company (which changed its name to Cincinnati Milacron in 1970). Several DNC systems were demonstrated at the National Machine Tool Show in 1970.

Mainframe computers represented the state of the technology in the mid-1960s. There were no personal computers or microcomputers at that time. But the trend in computer technology was toward the use of integrated circuits of increasing levels of integration, which resulted in dramatic increases in computational performance at the same time that the size and cost of the computer were reduced. At the beginning of the 1970s, the economics were right for using a dedicated computer as the MCU. This application came to be known as *computer numerical control* (CNC). At first, minicomputers were used as the controllers; subsequently, microcomputers were used as the performance/size trend continued.

7.1 FUNDAMENTALS OF NC TECHNOLOGY

This section identifies the basic components of an NC system. Then, NC coordinate systems in common use and types of motion controls are described.

7.1.1 Basic Components of an NC System

An NC system consists of three basic components: (1) a part program of instructions, (2) a machine control unit, and (3) processing equipment. The general relationship among the three components is illustrated in Figure 7.1.

The ***part program*** is the set of detailed step-by-step commands that direct the actions of the processing equipment. In machine tool applications, the person who prepares the program is called a *part programmer*. In these applications, the individual commands refer to positions of a cutting tool relative to the worktable on which the work part is fixtured. Additional instructions are usually included, such as spindle speed, feed rate, cutting tool selection, and other functions. The program is coded on a suitable medium for submission to the machine control unit. For many years, the common medium was 1-in wide punched tape, using a standard format that could be interpreted by the machine control unit. Today, punched tape has largely been replaced by newer storage technologies in modern machine shops. These technologies include magnetic tape, diskettes, and electronic transfer of part programs from a computer.

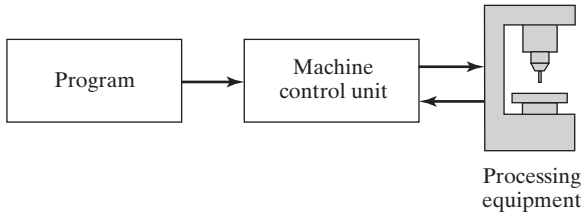


Figure 7.1 Basic components of an NC system.

In modern NC technology, the ***machine control unit*** (MCU) is a microcomputer and related control hardware that stores the program of instructions and executes it by converting each command into mechanical actions of the processing equipment, one command at a time. The related hardware of the MCU includes components to interface with the processing equipment and feedback control elements. The MCU also includes one or more reading devices for entering part programs into memory. Software residing in the MCU includes control system software, calculation algorithms, and translation software to convert the NC part program into a usable format for the MCU. Because the MCU is a computer, the term *computer numerical control* (CNC) is used to distinguish this type of NC from its technological ancestors that were based entirely on hardwired electronics. Today, virtually all new MCUs are based on computer technology.

The third basic component of an NC system is the processing equipment that performs the actual productive work (e.g., machining). It accomplishes the processing steps to transform the starting workpiece into a completed part. Its operation is directed by the MCU, which in turn is driven by instructions contained in the part program. In the most common example of NC, machining, the processing equipment consists of the worktable and spindle as well as the motors and controls to drive them.

7.1.2 NC Coordinate Systems

To program the NC processing equipment, a part programmer must define a standard axis system by which the position of the work head relative to the work part can be specified. There are two axis systems used in NC, one for flat and prismatic work parts and the other for rotational parts. Both systems are based on the Cartesian coordinates.

The axis system for flat and block-like parts consists of the three linear axes (x, y, z) in the Cartesian coordinate system, plus three rotational axes (a, b, c), as shown in Figure 7.2(a). In most machine tool applications, the x- and y-axes are used to move and position the worktable to which the part is attached, and the z-axis is used to control the vertical position of the cutting tool. Such a positioning scheme is adequate for simple NC applications such as drilling and punching of flat sheet metal. Programming these machine tools consists of little more than specifying a sequence of x–y coordinates.

The a-, b-, and c-rotational axes specify angular positions about the x-, y-, and z-axes, respectively. To distinguish positive from negative angles, the *right-hand rule* is used: Using the right hand with the thumb pointing in the positive linear axis direction ($+x$, $+y$, or $+z$), the fingers of the hand are curled in the positive rotational direction. The rotational axes can be used for one or both of the following: (1) orientation of the work part to present different surfaces for machining or (2) orientation of the tool or work head at some angle relative to the part. These additional axes permit machining of

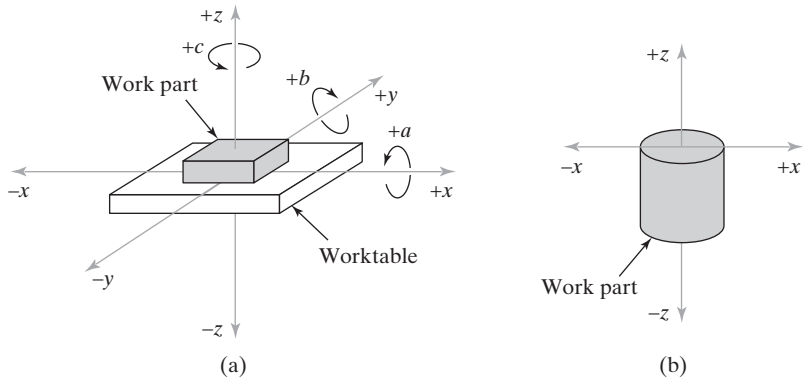


Figure 7.2 Coordinate systems used in NC (a) for flat and prismatic work and (b) for rotational work. (On most turning machines, the z-axis is horizontal rather than vertical as shown here.)

complex work part geometries. Machine tools with rotational axis capability generally have either four or five axes: three linear axes plus one or two rotational axes.

The coordinate axes for a rotational NC system are illustrated in Figure 7.2(b). These systems are associated with NC lathes and turning machines. Although the workpiece rotates, this is not one of the controlled axes on most turning machines. Consequently, the y-axis is not used. The path of the cutting tool relative to the rotating workpiece is defined in the x–z plane, where the x-axis is the radial location of the tool and the z-axis is parallel to the axis of rotation of the part.

Some machine tools are equipped with more than the number of axes described above. The additional axes are usually included to control more than one tool or spindle. Examples of these machine tools are mill-turn centers and multitasking machines (Section 14.2.3).

The part programmer must decide where the origin of the coordinate axis system should be located. This decision is usually based on programming convenience. For example, the origin might be located at one of the corners of the part. If the work part is symmetrical, the zero point might be most conveniently defined at the center of symmetry. Wherever the location, this zero point is communicated to the machine tool operator. At the beginning of the job, the operator must move the cutting tool under manual control to some *target point* on the worktable, where the tool can be easily and accurately positioned. The target point has been previously referenced to the origin of the coordinate axis system by the part programmer. When the tool has been accurately positioned at the target point, the operator indicates to the MCU where the origin is located for subsequent tool movements.

7.1.3 Motion Control Systems

Some NC processes are performed at discrete locations on the work part (e.g., drilling and spot welding). Others are carried out while the work head is moving (e.g., turning, milling, and continuous arc welding). If the work head is moving, it may be necessary to follow a straight line path or a circular or other curvilinear path. These different types of movement are accomplished by the motion control system, whose features are explained below.

Point-to-Point Versus Continuous Path Control. Motion control systems for NC (and robotics, Chapter 8) can be divided into two types: (1) point-to-point and

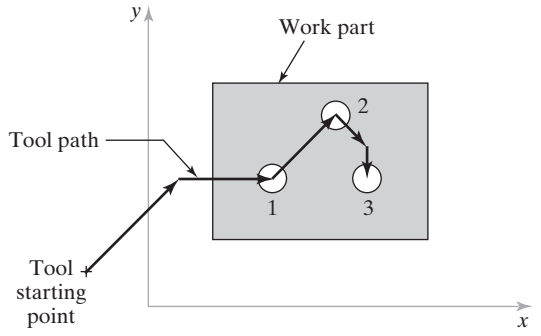


Figure 7.3 Point-to-point (positioning) control in NC. At each *x*–*y* position, table movement stops to perform the hole-drilling operation.

(2) continuous path. ***Point-to-point*** systems, also called *positioning systems*, move the worktable to a programmed location without regard for the path taken to get to that location. Once the move has been completed, some processing action is accomplished by the work head at the location, such as drilling or punching a hole. Thus, the program consists of a series of point locations at which operations are performed, as depicted in Figure 7.3.

Continuous path systems are capable of continuous simultaneous control of two or more axes. This provides control of the tool trajectory relative to the work part. In this case, the tool performs the process while the worktable is moving, thus enabling the system to generate angular surfaces, two-dimensional curves, or three-dimensional contours in the work part. This control mode is required in many milling and turning operations. A simple two-dimensional profile milling operation is shown in Figure 7.4 to illustrate continuous path control. When continuous path control is utilized to move the tool parallel to only one of the major axes of the machine tool worktable, this is called *straight-cut NC*. When continuous path control is used for simultaneous control of two or more axes in machining operations, the term *contouring* is used.

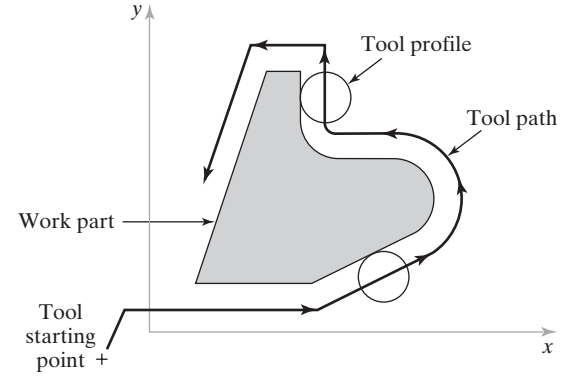


Figure 7.4 Continuous path (contouring) control in NC (*x*–*y* plane only). Note that cutting tool path must be offset from the part outline by a distance equal to its radius.

Interpolation Methods. One of the important aspects of contouring is interpolation. The paths that a contouring-type NC system is required to generate often consist of circular arcs and other smooth nonlinear shapes. Some of these shapes can be defined mathematically by relatively simple geometric formulas (e.g., the equation for a circle is $x^2 + y^2 = R^2$, where R = the radius of the circle and the center of the circle is at the origin), whereas others cannot be mathematically defined except by approximation. In any case, a fundamental problem in generating these shapes using NC equipment is that they are continuous, whereas NC is digital. To cut along a circular path, the circle must be divided into a series of straight line segments that approximate the curve. The tool is commanded to machine each line segment in succession so that the machined surface closely matches the desired shape. The maximum error between the nominal (desired) surface and the actual (machined) surface can be controlled by the lengths of the individual line segments, as shown in Figure 7.5.

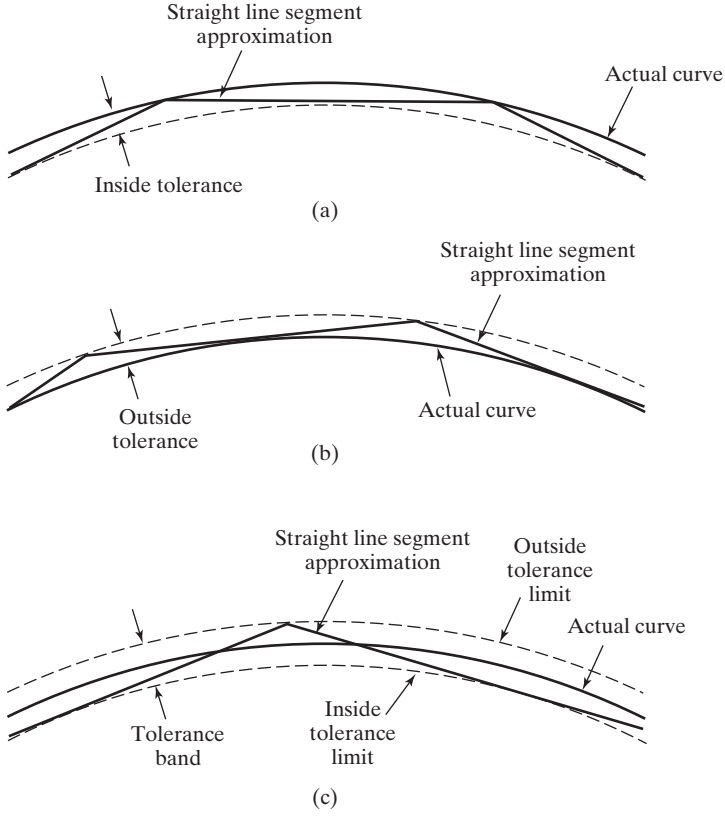


Figure 7.5 Approximation of a curved path in NC by a series of straight line segments. The accuracy of the approximation is controlled by the maximum deviation (called the tolerance) between the nominal (desired) curve and the straight line segments that are machined by the NC system. In (a), the tolerance is defined on only the inside of the nominal curve. In (b), the tolerance is defined on only the outside of the desired curve. In (c), the tolerance is defined on both the inside and outside of the desired curve.

TABLE 7.1 Numerical Control Interpolation Methods for Continuous Path Control

Linear interpolation. This is the most basic method and is used when a straight line path is to be generated in continuous path NC. Two-axis and three-axis linear interpolation routines are sometimes distinguished in practice, but conceptually they are the same. The programmer specifies the beginning point and endpoint of the straight line and the feed rate to be used along the straight line. The interpolator computes the feed rates for each of the two (or three) axes to achieve the specified feed rate.
Circular interpolation. This method permits programming of a circular arc by specifying the following parameters: (1) the coordinates of the starting point, (2) the coordinates of the endpoint, (3) either the center or radius of the arc, and (4) the direction of the cutter along the arc. The generated tool path consists of a series of small straight line segments (see Figure 7.5) calculated by the interpolation module. The cutter is directed to move along each line segment one by one to generate the smooth circular path. A limitation of circular interpolation is that the plane in which the circular arc exists must be a plane defined by two axes of the NC system ($x - y$, $x - z$, or $y - z$).
Helical interpolation. This method combines the circular interpolation scheme for two axes with linear movement of a third axis. This permits the definition of a helical path in three-dimensional space. Applications include the machining of large internal threads, either straight or tapered.
Parabolic and *cubic interpolations.* These routines provide approximations of free-form curves using higher order equations. They generally require considerable computational power and are not as common as linear and circular interpolation. Most applications are in the aerospace and automotive industries for free-form designs that cannot accurately and conveniently be approximated by combining linear and circular interpolations.

If the programmer were required to specify the endpoints for each of the line segments, the programming task would be extremely arduous and fraught with errors. Also, the part program would be extremely long because of the large number of points. To ease the burden, interpolation routines have been developed that calculate the intermediate points to be followed by the cutter to generate a particular mathematically defined or approximated path.

A number of interpolation methods are available to deal with the problems encountered in generating a smooth continuous path in contouring. They include (1) linear interpolation, (2) circular interpolation, (3) helical interpolation, (4) parabolic interpolation, and (5) cubic interpolation. Each of these procedures, briefly described in Table 7.1, permits the programmer to generate machine instructions for linear or curvilinear paths using relatively few input parameters. The interpolation module in the MCU performs the calculations and directs the tool along the path. In CNC systems, the interpolator is generally accomplished by software. Linear and circular interpolators are almost always included in modern CNC systems, whereas helical interpolation is a common option. Parabolic and cubic interpolations are less common because they are only needed by machine shops that produce complex surface contours.

Absolute Versus Incremental Positioning. Another aspect of motion control is concerned with whether positions are defined relative to the origin of the coordinate system (absolute positioning) or relative to the previous location of the tool (incremental positioning). In absolute positioning, the work head locations are always defined with respect to the origin of the axis system. In incremental positioning, the next work head position is defined relative to the present location. The difference is illustrated in Figure 7.6.

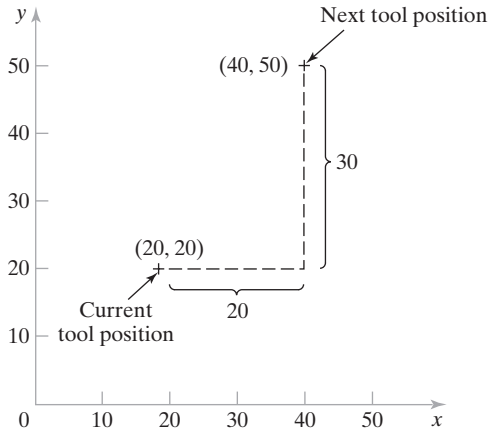


Figure 7.6 Absolute versus incremental positioning. The work head is presently at point (20, 20) and is to be moved to point (40, 50). In absolute positioning, the move is specified by $x = 40, y = 50$; whereas in incremental positioning, the move is specified by $x = 20, y = 30$.

7.2 COMPUTERS AND NUMERICAL CONTROL

Since the introduction of NC in 1952, there have been dramatic advances in digital computer technology. The physical size and cost of a digital computer have been significantly reduced at the same time that its computational capabilities have been substantially increased. The makers of NC equipment incorporated these advances in computer technology into their products, starting with large mainframe computers in the 1960s and followed by minicomputers in the 1970s and microcomputers in the 1980s. Today, NC means *computer numerical control* (CNC), which is defined as an NC system whose MCU consists of a dedicated microcomputer rather than a hardwired controller. The latest computer controllers for CNC feature high-speed processors, large memories, solid-state memory, improved servos, and bus architectures [12].

Computer NC systems include additional features beyond what is feasible with conventional hardwired NC. A list of many of these features is compiled in Table 7.2.

7.2.1 The CNC Machine Control Unit

The MCU is the hardware that distinguishes CNC from conventional NC. The general configuration of the MCU in a CNC system is illustrated in Figure 7.7. The MCU consists of the following components and subsystems: (1) central processing unit, (2) memory, (3) I/O interface, (4) controls for machine tool axes and spindle speed, and (5) sequence controls for other machine tool functions. These subsystems are interconnected by means of a system bus, which communicates data and signals among the components of the network.

TABLE 7.2 Features of Computer Numerical Control that Distinguish It from Conventional NC

Storage of more than one part program. With improvements in storage technology, newer CNC controllers have sufficient capacity to store multiple programs. Controller manufacturers generally offer one or more memory expansions as options to the MCU.

Program editing at the machine tool. CNC permits a part program to be edited while it resides in the MCU computer memory. Hence, a program can be tested and corrected entirely at the machine site. Editing also permits cutting conditions in the machining cycle to be optimized. After the program has been corrected and optimized, the revised version can be stored for future use.

Fixed cycles and programming subroutines. The increased memory capacity and the ability to program the control computer provide the opportunity to store frequently used machining cycles as *macros* that can be called by the part program. Instead of writing the full instructions for the particular cycle into every program, a programmer includes a call statement in the part program to indicate that the macro cycle should be executed. These cycles often require that certain parameters be defined, for example, a bolt hole circle, in which the diameter of the bolt circle, the spacing of the bolt holes, and other parameters must be specified.

Adaptive control. In this feature, the MCU measures and analyzes machining variables, such as spindle torque, power, and tool-tip temperature, and adjusts cutting speed and/or feed rate to maximize machining performance. Benefits include reduced cycle time and improved surface finish.

Interpolation. Some of the interpolation schemes described in Table 7.1 are normally executed on a CNC system because of the computational requirements. Linear and circular interpolations are sometimes hardwired into the control unit, but helical, parabolic, and cubic interpolations are usually executed by a stored program algorithm.

Positioning features for setup. Setting up the machine tool for a given work part involves installing and aligning a fixture on the machine tool table. This must be accomplished so that the machine axes are established with respect to the work part. The alignment task can be facilitated using certain features made possible by software options in a CNC system, such as *position set*. With position set, the operator is not required to locate the fixture on the machine table with extreme accuracy. Instead, the machine tool axes are referenced to the location of the fixture using a target point or set of target points on the work or fixture.

Acceleration and deceleration calculations. This feature is applicable when the cutter moves at high feed rates. It is designed to avoid tool marks on the work surface that would be generated due to machine tool dynamics when the cutter path changes abruptly. Instead, the feed rate is smoothly decelerated in anticipation of a tool path change and then accelerated back up to the programmed feed rate after the direction change.

Communications interface. With the trend toward interfacing and networking in plants today, modern CNC controllers are equipped with a standard communications interface to link the machine to other computers and computer-driven devices. This is useful for applications such as (1) downloading part programs from a central data file; (2) collecting operational data such as workpiece counts, cycle times, and machine utilization; and (3) interfacing with peripheral equipment, such as robots that load and unload parts.

Diagnostics. Many modern CNC systems possess a diagnostics capability that monitors certain aspects of the machine tool to detect malfunctions or signs of impending malfunctions or to diagnose system breakdowns.

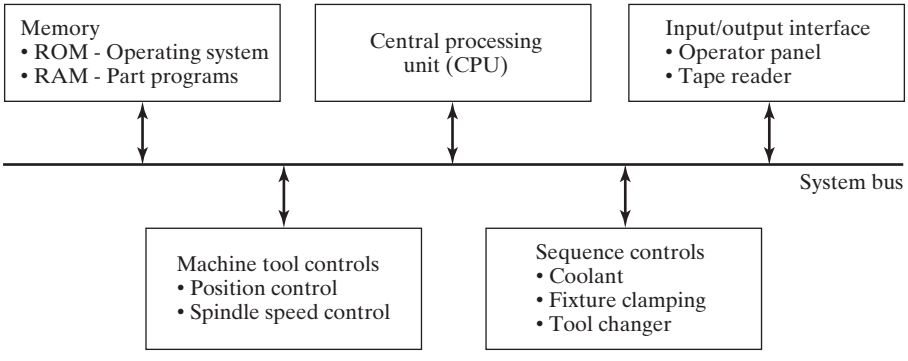


Figure 7.7 Configuration of CNC machine control unit.

Central Processing Unit. The central processing unit (CPU) is the brain of the MCU. It manages the other components in the MCU based on software contained in main memory. The CPU can be divided into three sections: (1) control section, (2) arithmetic-logic unit, and (3) immediate access memory. The control section retrieves commands and data from memory and generates signals to activate other components in the MCU. In short, it sequences, coordinates, and regulates the activities of the MCU computer. The arithmetic-logic unit (ALU) consists of the circuitry to perform various calculations (addition, subtraction, multiplication), counting, and logical functions required by software residing in memory. The immediate access memory provides a temporary storage for data being processed by the CPU. It is connected to main memory by means of the system data bus.

Memory. The immediate access memory in the CPU is not intended for storing CNC software. A much greater storage capacity is required for the various programs and data needed to operate the CNC system. As with most other computer systems, CNC memory can be divided into two categories: (1) main memory and (2) secondary memory. Main memory consists of ROM (read-only memory) and RAM (random access memory) devices. Operating system software and machine interface programs (Section 7.2.2) are generally stored in ROM. These programs are usually installed by the manufacturer of the MCU. NC part programs are stored in RAM devices. Current programs in RAM can be erased and replaced by new programs as jobs are changed.

High-capacity secondary memory devices are used to store large programs and data files, which are transferred to main memory as needed. Common among the secondary memory devices are hard disks and solid-state memory devices to store part programs, macros, and other software. These high-capacity storage devices are permanently installed in the CNC machine control unit and have replaced most of the punched paper tape traditionally used to store part programs.

Input/Output Interface. The I/O interface provides communication between the various components of the CNC system, other computer systems, and the machine operator. As its name suggests, the I/O interface transmits and receives data and signals to and from external devices, several of which are indicated in Figure 7.7. The operator control panel is the basic interface by which the machine operator communicates to the CNC system. This is used to enter commands related to part program editing, MCU operating mode (e.g., program control vs. manual control), speeds and feeds, cutting fluid pump on/off, and similar functions. Either an alphanumeric keypad or keyboard is usually included in the operator control panel. The I/O interface also includes a display to communicate data and information from the MCU to the machine operator. The display is used to indicate current status of the program as it is being executed and to warn the operator of any malfunctions in the system.

Also included in the I/O interface are one or more means of entering part programs into storage. Programs can be entered manually by the machine operator or stored at a central computer site and transmitted via local area network (LAN) to the CNC system. Whichever means is employed by the plant, a suitable device must be included in the I/O interface to allow input of the programs into MCU memory.

Controls for Machine Tool Axes and Spindle Speed. These are hardware components that control the position and velocity (feed rate) of each machine axis as well as the rotational speed of the machine tool spindle. Control signals generated by the MCU

must be converted to a form and power level suited to the particular position control systems used to drive the machine axes. Positioning systems can be classified as open loop or closed loop, and different hardware components are required in each case. A more detailed discussion of these hardware elements is presented in Section 7.4, together with an analysis of how they operate to achieve position and feed rate control. Some of the hardware components are resident in the MCU.

Depending on the type of machine tool, the spindle is used to drive either (1) the workpiece, as in turning, or (2) a rotating cutter, as in milling and drilling. Spindle speed is a programmed parameter. Components for spindle speed control in the MCU usually consist of a drive control circuit and a feedback sensor interface.

Sequence Controls for Other Machine Tool Functions. In addition to control of table position, feed rate, and spindle speed, several additional functions are accomplished under part program control. These auxiliary functions generally involve on/off (binary) actuations, interlocks, and discrete numerical data. The functions include cutting fluid control, fixture clamping, emergency warnings, and interlock communications for robot loading and unloading of the machine tool.

7.2.2 CNC Software

The NC computer operates by means of software. There are three types of software programs used in CNC systems: (1) operating system software, (2) machine interface software, and (3) application software.

The principal function of the operating system software is to interpret the NC part programs and generate the corresponding control signals to drive the machine tool axes. It is installed by the controller manufacturer and is stored in ROM in the MCU. The operating system software consists of the following: (1) an editor, which permits the machine operator to input and edit NC part programs and perform other file management functions; (2) a control program, which decodes the part program instructions, performs interpolation and acceleration/deceleration calculations, and accomplishes other related functions to produce the coordinate control signals for each axis; and (3) an executive program, which manages the execution of the CNC software as well as the I/O operations of the MCU. The operating system software also includes any diagnostic routines that are available in the CNC system.

Machine interface software is used to operate the communication link between the CPU and the machine tool to accomplish the CNC auxiliary functions. The I/O signals associated with the auxiliary functions are sometimes implemented by means of a programmable logic controller interfaced to the MCU, so the machine interface software is often written in the form of ladder logic diagrams (Section 9.2).

Finally, the application software consists of the NC part programs that are written for machining (or other) applications in the user's plant. The topic of part programming is postponed to Section 7.5.

7.2.3 Distributed Numerical Control

Historical Note 7.1 describes several ways in which digital computers have been used to implement NC. This section describes two approaches: (1) direct numerical control and (2) distributed numerical control.

Direct numerical control (DNC) was the first attempt to use a digital computer to control NC machines. It was in the late 1960s, before the advent of CNC. As initially implemented, direct numerical control involved the control of a number of machine tools by a single (mainframe) computer through direct connection and in real time. Instead of using a punched tape reader to enter the part program into the MCU, the program was transmitted to the MCU directly from the computer, one block of instructions at a time. An instruction block provides the commands for one complete move of the machine tool, including location coordinates, speeds, feeds, and other data (Section 7.5.1). This mode of operation was referred to by the term *behind the tape reader* (BTR). The DNC computer provided instruction blocks to the machine tool on demand; when a machine needed control commands, they were communicated to it immediately. As each block was executed by the machine, the next block was transmitted. As far as the machine tool was concerned, the operation was no different from that of a conventional NC controller. In theory, DNC relieved the NC system of its least reliable components: the punched tape and tape reader.

The general configuration of a DNC system is depicted in Figure 7.8. The system consisted of four components: (1) central computer, (2) bulk memory at the central computer site, (3) set of controlled machines, and (4) telecommunications lines to connect the machines to the central computer. In operation, the computer called the required part program from bulk memory and sent it (one block at a time) to the designated machine tool. This procedure was replicated for all machine tools under direct control of the computer.

In addition to transmitting data to the machines, the central computer also received data back from the machines to indicate operating performance in the shop (e.g., number of machining cycles completed, machine utilization, and breakdowns). A central objective of DNC was to achieve two-way communication between the machines and the central computer.

As the installed base of CNC machines grew during the 1970s and 1980s, a new form of DNC emerged, called *distributed numerical control* (DNC). The configuration of the new DNC is very similar to that shown in Figure 7.8 except that the central computer is connected to MCUs, which are themselves computers; basically this is a distributed control system (Section 5.3.3). Complete part programs are sent to the machine tools, not

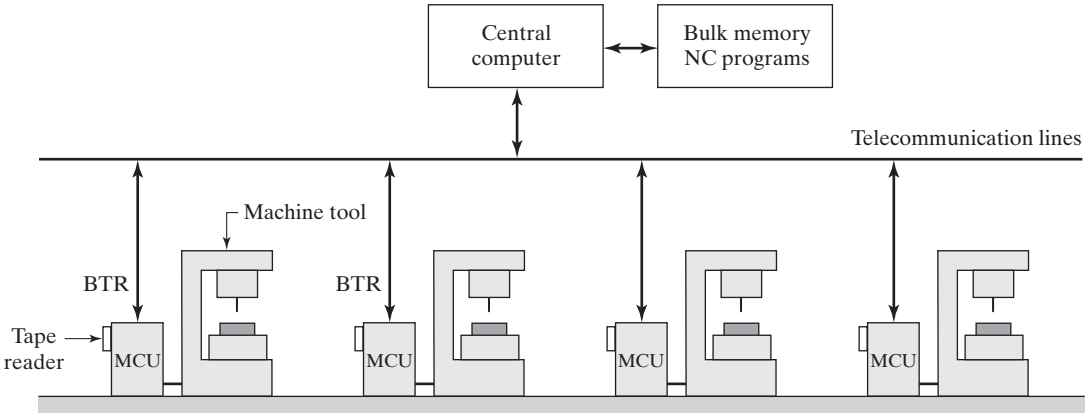


Figure 7.8 General configuration of a DNC system. Connection to MCU is behind the tape reader. Key: BTR = behind the tape reader, MCU = machine control unit.

TABLE 7.3 Flow of Data and Information Between Central Computer and Machine Tools in DNC

Data and Information Downloaded from the Central Computer to the Machine Tools	Data and Information Uploaded from the Machine Tools to the Central Computer
NC part programs	Piece counts
List of tools needed for job	Actual machining cycle times
Machine tool setup instructions	Tool life statistics
Machine operator instructions	Machine uptime and downtime statistics
Machining cycle time for part program	Product quality data
Data about when program was last used	Machine utilization
Production schedule information	

one block at a time. The distributed NC approach permits easier and less costly installation of the overall system, because the individual CNC machines can be put into service and distributed NC can be added later. Redundant computers improve system reliability compared with the original DNC. The new DNC permits two-way communication of data between the shop floor and the central computer, which was one of the important features included in the old DNC. However, improvements in data collection devices as well as advances in computer and communications technologies have expanded the range and flexibility of the information that can be gathered and disseminated. Some of the data and information sets included in the two-way communication flow are itemized in Table 7.3. This flow of information in DNC is similar to the information flow in shop floor control, discussed in Chapter 25.

7.3 APPLICATIONS OF NC

The operating principle of NC has many applications. There are many industrial operations in which the position of a work head must be controlled relative to a part or product being processed. The applications divide into two categories: (1) machine tool applications and (2) other applications. Most machine tool applications are associated with the metalworking industry. The other applications comprise a diverse group of operations in other industries. It should be noted that the applications are not always identified by the name "numerical control"; this term is used principally in the machine tool industry.

7.3.1 Machine Tool Applications

The most common applications of NC are in machine tool control. Machining was the first application of NC, and it is still one of the most important commercially.

Machining Operations and NC Machine Tools. Machining is a manufacturing process in which the geometry of the work is produced by removing excess material (Section 2.2.1). Control of the relative motion between a cutting tool and the workpiece creates the desired geometry. Machining is considered one of the most versatile processes because it can be used to create a wide variety of shapes and surface finishes. It can be performed at relatively high production rates to yield highly accurate parts at relatively low cost.

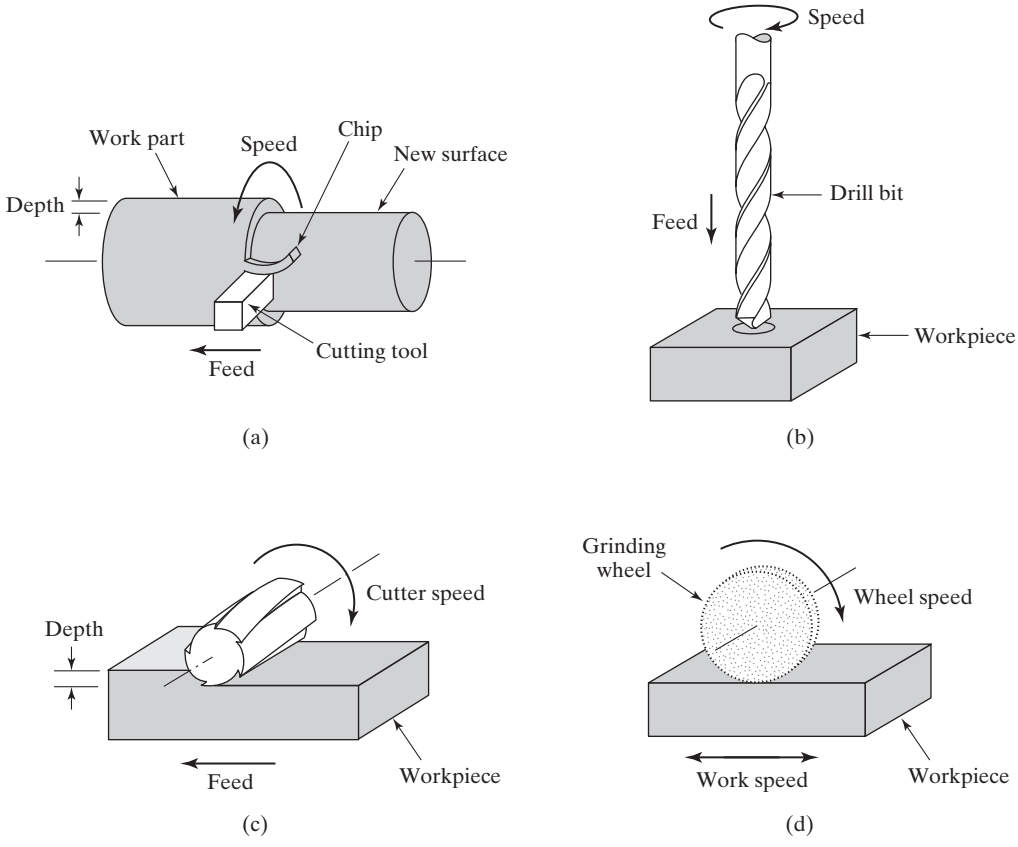


Figure 7.9 The four common machining operations are (a) turning, (b) drilling, (c) peripheral milling, and (d) surface grinding.

There are four common types of machining operations: (a) turning, (b) drilling, (c) milling, and (d) grinding, shown in Figure 7.9. Each of the machining operations is carried out at a certain combination of speed, feed, and depth of cut, collectively called the *cutting conditions*. The terminology varies somewhat for grinding. These cutting conditions are illustrated in Figure 7.9 for turning, drilling, and milling. Consider milling. The ***cutting speed*** is the velocity of the milling cutter relative to the work surface, m/min (ft/min). This is usually programmed into the machine as a spindle rotation speed, rev/min. Cutting speed can be converted into spindle rotation speed by means of the equation

$$N = \frac{v}{\pi D} \tag{7.1}$$

where N = spindle rotation speed, rev/min; v = cutting speed, m/min (ft/min); and D = milling cutter diameter, m (ft). In milling, the ***feed*** usually means the size of the chip formed by each tooth in the milling cutter, often referred to as the *chip load* per tooth. This must normally be programmed into the NC machine as the feed rate (the travel rate of the machine tool table). Therefore, feed must be converted to feed rate as

$$f_r = Nn_t f \tag{7.2}$$

where f_r = feed rate, mm/min (in/min); N = spindle rotational speed, rev/min; n_t = number of teeth on the milling cutter; and f = feed, mm/tooth (in/tooth). For a turning operation, feed is defined as the lateral movement of the cutting tool per revolution of the workpiece, mm/rev (in/rev). ***Depth of cut*** is the distance the tool penetrates below the original surface of the work, mm (in). For drilling, depth of cut refers to the depth of the hole. These are the parameters that must be controlled during the operation of an NC machine through motion or position commands in the part program.

Each of the four machining processes is traditionally carried out on a machine tool designed to perform that process. Turning is performed on a lathe, drilling is done on a drill press, milling on a milling machine, and so on. The following is a list of the common material-removal CNC machine tools along with their typical features:

- *NC lathe*, either horizontal or vertical axis. Turning requires two-axis, continuous path control, either to produce a straight cylindrical geometry (straight turning) or to create a profile (contour turning).
- *NC boring mill*, horizontal or vertical spindle. Boring is similar to turning, except that an internal cylinder is created instead of an external cylinder. The operation requires continuous path, two-axis control.
- *NC drill press*. This machine uses point-to-point control of a work head (spindle containing the drill bit) and two axis (x–y) control of a worktable. Some NC drill presses have turrets containing six or eight drill bits. The turret position is programmed under NC control, allowing different drill bits to be applied to the same work part during the machine cycle without requiring the machine operator to manually change the tool.
- *NC milling machine*. A milling machine requires continuous path control to perform straight cut or contouring operations. Figure 7.10 illustrates the features of a CNC four-axis milling machine.
- *NC cylindrical grinder*. This machine operates like a turning machine, except that the tool is a grinding wheel. It has continuous path two-axis control, similar to an NC lathe.

Numerical control has had a profound influence on the design and operation of machine tools. One of the effects is that the proportion of time spent by the machine cutting metal is significantly greater than with manually operated machines. This causes certain components such as the spindle, drive gears, and feed screws to wear more rapidly. These components must be designed to last longer on NC machines. Secondly, the addition of the electronic control unit has increased the cost of the machine, requiring higher equipment utilization. Instead of running the machine during only one shift, which is the typical schedule with manually operated machines, NC machines are often operated during two or even three shifts to obtain the required economic payback. Third, the increasing cost of labor has altered the relative roles of the human operator and the machine tool. Instead of being the highly skilled worker who controlled every aspect of part production, the NC machine operator performs only part loading and unloading, tool-changing, chip clearing, and the like. With these reduced responsibilities, one operator can often run two or three NC machines.

The functions performed by the machine tool have also changed. NC machines are designed to be highly automatic and capable of combining several operations in one setup that formerly required several different machines. They are also designed to reduce the time consumed by the noncutting elements in the operation cycle, such as changing tools

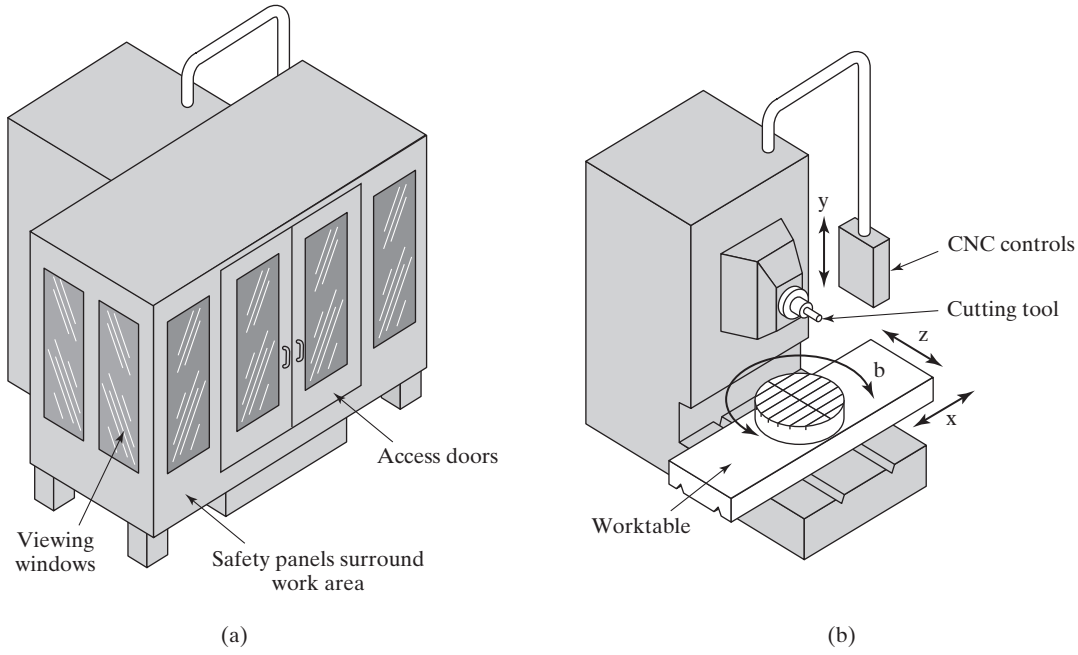


Figure 7.10 (a) Four-axis CNC horizontal milling machine with safety panels installed and (b) with safety panels removed to show typical axis configuration for the horizontal spindle.

and loading and unloading the work part. These changes are best exemplified by a new type of machine that did not exist prior to the development of NC: the ***machining center***, which is a machine tool capable of performing multiple machining operations on a single workpiece in one setup. The operations involve rotating cutters, such as milling and drilling, and the feature that enables more than one operation to be performed in one setup is automatic tool-changing. Machining centers and related machine tools are discussed in Chapter 14 on single-station manufacturing cells (Section 14.2.3).

NC Application Characteristics. In general, NC technology is appropriate for low-to-medium production of medium-to-high variety product. Using the terminology of Section 2.4.1, the product is low-to-medium Q, medium-to-high P. Over many years of machine shop practice, the following part characteristics have been identified as most suited to the application of NC:

1. *Batch production.* NC is most appropriate for parts produced in small or medium lot sizes (batch sizes ranging from one unit up to several hundred units). Dedicated automation would not be economical for these quantities because of the high fixed cost. Manual production would require many separate machine setups and would result in higher labor cost, longer lead time, and higher scrap rate.
2. *Repeat orders.* Batches of the same parts are produced at random or periodic intervals. Once the NC part program has been prepared, parts can be economically produced in subsequent batches using the same part program.

3. *Complex part geometry.* The part geometry includes complex curved surfaces such as those found on airfoils and turbine blades. Mathematically defined surfaces such as circles and helixes can also be accomplished with NC. Some of these geometries would be difficult if not impossible to achieve accurately using conventional machine tools.
4. *Much metal needs to be removed from the work part.* This condition is often associated with a complex part geometry. The volume and weight of the final machined part is a relatively small fraction of the starting block. Such parts are common in the aircraft industry to fabricate large structural sections with low weights.
5. *Many separate machining operations on the part.* This applies to parts consisting of many machined features requiring different cutting tools, such as drilled and/or tapped holes, slots, flats, and so on. If these operations were machined by a series of manual operations, many setups would be needed. The number of setups can be reduced significantly using NC.
6. *The part is expensive.* This factor is often a consequence of one or more of preceding factors 3, 4, and 5. It can also result from using a high-cost starting work material. When the part is expensive, and mistakes in processing would be costly, the use of NC helps to reduce rework and scrap losses.

Although these characteristics pertain mainly to machining, they are adaptable to other production applications as well.

NC for Other Metalworking Processes. NC machine tools have been developed for other metalworking processes besides machining. These machines include the following:

- *Punch presses* for sheet metal hole punching. The two-axis NC operation is similar to that of a drill press except that holes are produced by punching rather than drilling. Different hole sizes and shapes are implemented using a tool turret.
- *Presses* for sheet metal bending. Instead of cutting sheet metal, these systems bend sheet metal according to programmed commands.
- *Welding machines*. Both spot welding and continuous arc welding machines are available with automatic controls based on NC.
- *Thermal cutting machines*, such as oxy-fuel cutting, laser cutting, and plasma arc cutting. The stock is usually flat; thus, two-axis control is adequate. Some laser cutting machines can cut holes in preformed sheet metal stock, requiring four- or five-axis control.
- *Tube bending* and *wire bending machines*. Automatic tube and wire bending machines are programmed to control the location (along the length of the stock) and the angle of the bend. Important tube bending applications include frames for bicycles and motorcycles. Wire bending applications include springs and paper clips.
- *Wire EDM*. Electric discharge wire cutting operates in a manner similar to a band saw, except that the saw is a small diameter wire that uses sparks to cut metal stock that is positioned by an x–y positioning table.

7.3.2 Other NC Applications

The operating principle of NC has a host of other applications besides metalworking. Some of the machines with NC-type controls that position a work head relative to an object being processed are the following:

- *Rapid prototyping and additive manufacturing*. These include a number of processes that add material, one thin layer at a time, to construct a part. Many of them operate by means of a work head that is manipulated by NC over the partially constructed part. Some processes use lasers to cure photosensitive liquid polymers (stereolithography) or fuse solid powders (selective laser sintering); others use extruder heads that add material (fused deposition modeling).
- *Water jet cutters* and *abrasive water jet cutters*. These machines are used to cut various materials, including metals and nonmetals (e.g., plastic, cloth), by means of a fine, high-pressure, high-velocity stream of water. Abrasive particles are added to the stream in the case of abrasive water jet cutting to facilitate cutting of more difficult materials (e.g., metals). The work head is manipulated relative to the work material by means of numerical control.
- *Component placement machines.* This equipment is used to position components on an x–y plane, usually a printed circuit board. The program specifies the x- and y-axis positions in the plane where the components are to be located. Component placement machines find extensive applications for placing electronic components on printed circuit boards. Machines are available for either through-hole or surface-mount applications as well as similar insertion-type mechanical assembly operations.
- *Coordinate measuring machines.* A coordinate measuring machine (CMM) is an inspection machine used for measuring or checking dimensions of a part. A CMM has a probe that can be manipulated in three axes and that identifies when contact is made against a part surface. The location of the probe tip is determined by the CMM control unit, thereby indicating some dimension on the part. Many coordinate measuring machines are programmed to perform automated inspections under NC. Coordinate measuring machines are discussed in Section 22.3.
- *Wood routers and granite cutters*. These machines perform operations similar to NC milling for metal machining, except the work materials are not metals. Many wood cutting lathes are also NC machines.
- *Tape laying machines for polymer composites.* The work head of this machine is a dispenser of uncured polymer matrix composite tape. The machine is programmed to lay the tape onto the surface of a contoured mold, following a back-and-forth and crisscross pattern to build up a required thickness. The result is a multilayered panel of the same shape as the mold.
- *Filament winding machines for polymer composites.* These are similar to the preceding machine except that a filament is dipped in uncured polymer and wrapped around a rotating pattern of roughly cylindrical shape.

7.3.3 Advantages and Disadvantages of NC

When the production application satisfies the characteristics identified in Section 7.3.1, NC yields many advantages over manual production methods. These advantages translate into economic savings for the user company. However, NC involves more sophisticated technology than conventional methods, and there are costs that must be considered to apply the technology effectively. This section examines the advantages and disadvantages of NC.

Advantages of NC. The advantages generally attributed to NC, with emphasis on machine tool applications, are the following:

- *Nonproductive time is reduced*. NC reduces the proportion of time the machine is not cutting metal. This is achieved through fewer setups, less setup time, reduced workpiece handling time, and automatic tool changes on some NC machines, all of which translate into labor cost savings and lower elapsed times to produce parts.
- *Greater accuracy and repeatability*. Compared with manual production methods, NC reduces or eliminates variations due to operator skill differences, fatigue, and other factors attributed to inherent human variabilities. Parts are made closer to nominal dimensions, and there is less dimensional variation among parts in the batch.
- *Lower scrap rates*. Because greater accuracy and repeatability are achieved, and because human errors are reduced, more parts are produced within tolerance. The ultimate goal in NC is zero defects.
- *Inspection requirements are reduced*. Less inspection is needed when NC is used because parts produced from the same NC part program are virtually identical. Once the program has been verified, there is no need for the high level of sampling inspection that is required when parts are produced by conventional manual methods. Except for tool wear and equipment malfunctions, NC produces exact replicates of the part each cycle.
- *More complex part geometries are possible*. NC technology has extended the range of possible part geometries beyond what is practical with manual machining methods. This is an advantage for product design in several ways: (1) More functional features can be designed into a single part, thus reducing the total number of parts in the product and the associated cost of assembly, (2) mathematically defined surfaces can be fabricated with high precision, and (3) the limits within which the designer's imagination can wander to create new part and product geometries are expanded.
- *Engineering changes can be accommodated more gracefully*. Instead of making alterations in a complex fixture so that the part can be machined to the engineering change, revisions are made in the NC part program to accomplish the change.
- *Simpler fixtures*. NC requires simpler fixtures because accurate positioning of the tool is accomplished by the NC machine tool. Tool positioning does not have to be designed into the jig.
- *Shorter manufacturing lead times*. Jobs can be set up more quickly and fewer setups are required per part when NC is used. This results in shorter elapsed time between order release and completion.
- *Reduced parts inventory*. Because fewer setups are required and job changeovers are easier and faster, NC permits production of parts in smaller lot sizes. The economic lot size is lower in NC than in conventional batch production. Average parts inventory is therefore reduced.
- *Less floor space*. This results from the fact that fewer NC machines are required to perform the same amount of work compared to the number of conventional machine tools needed. Reduced parts inventory also contributes to less floor space requirements.
- *Operator skill requirements are reduced*. Workers need fewer skills to operate an NC machine than to operate a conventional machine tool. Tending an NC machine tool usually consists only of loading and unloading parts and periodically changing tools. The machining cycle is carried out under program control. Performing a comparable machining cycle on a conventional machine requires much more participation by the operator and a higher level of training and skill.

Disadvantages of NC. There are certain commitments to NC technology that must be made by the machine shop that installs NC equipment, and these commitments, most of which involve additional cost to the company, might be seen as disadvantages. These include the following:

- *Higher investment cost.* An NC machine tool has a higher first cost than a comparable conventional machine tool. There are several reasons why: (1) NC machines include CNC controls and electronics hardware; (2) software development costs of the CNC controls manufacturer must be included in the cost of the machine; (3) more reliable mechanical components are generally used in NC machines; and (4) NC machine tools often possess additional features not included on conventional machines, such as automatic tool changers and part changers (Section 14.2.3).
- *Higher maintenance effort.* In general, NC equipment requires more maintenance than conventional equipment, which translates to higher maintenance and repair costs. This is due largely to the computer and other electronics that are included in a modern NC system. The maintenance staff must include personnel who are trained in maintaining and repairing this type of equipment.
- *Part programming.* NC equipment must be programmed. To be fair, it should be mentioned that process planning must be accomplished for any part, whether or not it is produced by NC. However, NC part programming is a special preparation step in batch production that is absent in conventional machine shop operations.
- *Higher utilization of NC equipment.* To maximize the economic benefits of NC, some companies operate multiple shifts. This might mean adding one or two extra shifts to the plant's normal operations, with the requirement for supervision and other staff support.

7.4 ANALYSIS OF POSITIONING SYSTEMS

An NC positioning system converts the coordinate axis values in the NC part program into relative positions of the tool and work part during processing. Consider the simple positioning system shown in Figure 7.11. The system consists of a cutting tool and a worktable on which a work part is fixtured. The table is designed to move the part relative to the tool. The worktable moves linearly by means of a rotating leadscrew or ball screw, which is driven by a stepper motor or servomotor (Section 6.2.1). For simplicity, only one

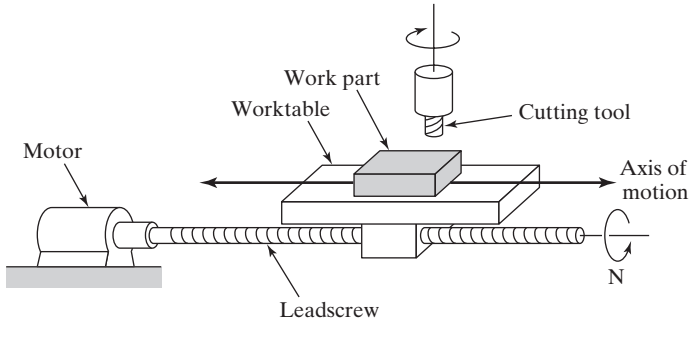


Figure 7.11 Motor and leadscrew arrangement in an NC positioning system.

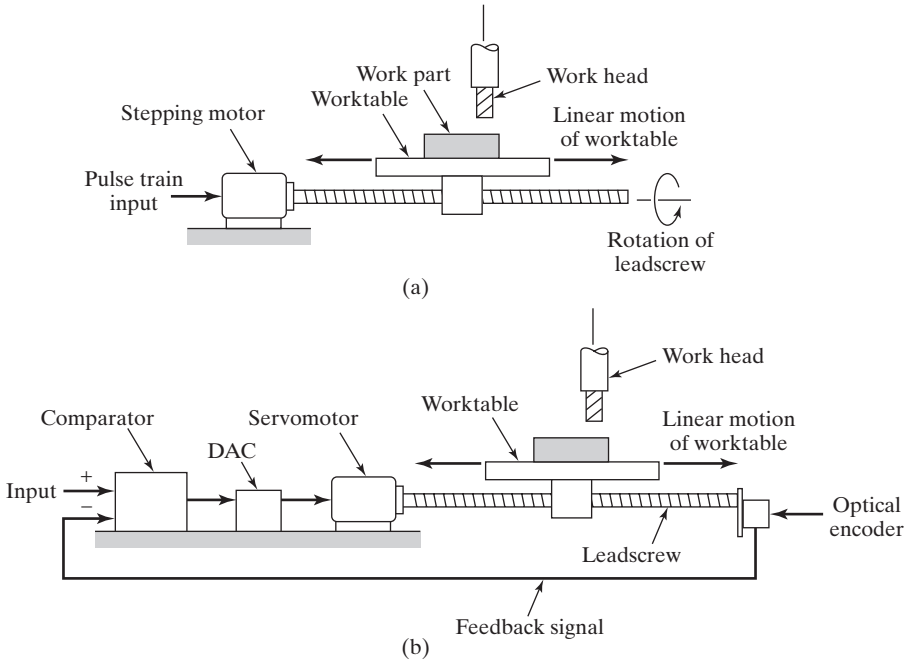


Figure 7.12 Two types of motion control in NC: (a) open loop and (b) closed loop.

axis is shown in the sketch. To provide *x*–*y* capability, the system shown would be piggybacked on top of a second axis perpendicular to the first. The screw has a certain pitch p, mm/thread (in/thread). Thus, the table moves a distance equal to the pitch for each revolution. The velocity of the worktable, which corresponds to the feed rate in a machining operation, is determined by the rotational speed of the screw.

Two types of control systems are used in positioning systems: (a) open loop and (b) closed loop, as shown in Figure 7.12. An open-loop system operates without verifying that the actual position achieved in the move is the same as the desired position. A closed-loop system uses feedback measurements to confirm that the final position of the worktable is the location specified in the program. Open-loop systems cost less than closed-loop systems and are appropriate when the force resisting the actuating motion is minimal. Closed-loop systems are normally specified for machines that perform continuous path operations such as milling or turning, in which there are significant forces resisting the forward motion of the cutting tool.

7.4.1 Open-Loop Positioning Systems

An open-loop positioning system typically uses a stepper motor to rotate the leadscrew or ball screw. A stepper motor is driven by a series of electrical pulses, which are generated by the MCU in an NC system. Each pulse causes the motor to rotate a fraction of one revolution, called the step angle. The possible step angles must be consistent with the relationship

$$\alpha = \frac{360}{n_s} \tag{7.3}$$

where α = step angle, °; and n_s = the number of step angles for the motor, which must be an integer. The angle through which the motor shaft rotates is given by

$$A_m = n_p \alpha \tag{7.4}$$

where A_m = angle of motor shaft rotation, °; n_p = number of pulses received by the motor; and α = step angle, °/pulse. The motor shaft is generally connected to the screw through a gearbox, which reduces the angular rotation of the screw. The angle of the screw rotation must take the gear ratio into account as

$$A_s = \frac{A_m}{r_g} = \frac{n_p \alpha}{r_g} \tag{7.5}$$

where A_s = angle of screw rotation, °; and r_g = gear ratio, defined as the number of turns of the motor for each single turn of the screw. That is,

$$r_g = \frac{A_m}{A_s} = \frac{N_m}{N_s} \tag{7.6}$$

where N_m = rotational speed of the motor, rev/min; and N_s = rotational speed of the screw, rev/min.

The linear movement of the worktable is given by the number of full and partial rotations of the screw multiplied by its pitch,

$$x = \frac{pA_s}{360} \tag{7.7}$$

where x = x-axis position relative to the starting position, mm (in); p = pitch of the screw mm/rev (in/rev); and $A_s/360$ = number of screw revolutions. The number of pulses required to achieve a specified x-position increment in a point-to-point system can be found using combinations of Equations (7.3), (7.5), and (7.7):

$$n_p = \frac{360 x r_g}{p\alpha} = \frac{n_s x r_g}{p} \tag{7.8}$$

Control pulses are transmitted from the pulse generator at a certain frequency, which drives the worktable at a corresponding velocity or feed rate in the direction of the screw axis. The rotational speed of the screw depends on the frequency of the pulse train as

$$N_s = \frac{60 f_p}{n_s r_g} \tag{7.9}$$

where N_s = screw rotational speed, rev/min; f_p = pulse train frequency, Hz; and n_s = steps per revolution or pulses per revolution. For a two-axis table with continuous path control, the relative velocities of the axes are coordinated to achieve the desired travel direction.

The table travel speed in the direction of screw axis is determined by the rotational speed as

$$v_t = f_r = N_s p \tag{7.10}$$

where v_t = table travel speed, mm/min (in/min); f_r = table feed rate, mm/min (in/min); N_s = screw rotational speed, rev/min; and p = screw pitch, mm/rev (in/rev).

The required pulse train frequency to drive the table at a specified linear travel rate can be obtained by combining Equations (7.9) and (7.10) and rearranging to solve for f_p:

$$f_p = \frac{v_t n_s r_g}{60p} = \frac{f_r n_s r_g}{60p} = \frac{N_m n_s}{60} = \frac{N_s n_s r_g}{60} \tag{7.11}$$

EXAMPLE 7.1 NC Open-Loop Positioning

The worktable of a positioning system is driven by a ball screw whose pitch = 6.0 mm. The screw is connected to the output shaft of a stepper motor through a gearbox whose ratio is 5:1 (five turns of the motor to one turn of the screw). The stepper motor has 48 step angles. The table must move a distance of 250 mm from its present position at a linear velocity = 500 mm/min. Determine (a) how many pulses are required to move the table the specified distance and (b) the required motor speed and pulse rate to achieve the desired table velocity.

Solution: (a) Rearranging Equation (7.7) to find the screw rotation angle A_s corresponding to a distance $x = 250$ mm,

$$A_s = \frac{360x}{p} = \frac{360(250)}{6.0} = 15{,}000^\circ$$

With 48 step angles, each step angle is

$$\alpha = \frac{360}{48} = 7.5^\circ$$

Thus, the number of pulses to move the table 250 mm is

$$n_p = \frac{360xr_g}{p\alpha} = \frac{A_s r_g}{\alpha} = \frac{15{,}000(5)}{7.5} = \mathbf{10{,}000\ pulses}$$

(b) The rotational speed of the screw corresponding to a table speed of 500 mm/min is determined from Equation (7.10):

$$N_s = \frac{v_t}{p} = \frac{500}{6} = \mathbf{83.333\ rev/min}$$

Equation (7.6) is used to find the motor speed:

$$N_m = r_g N_s = 5(83.333) = \mathbf{416.667\ rev/min}$$

The applied pulse rate to drive the table is given by Equation (7.11):

$$f_p = \frac{v_t n_s r_g}{60p} = \frac{500(48)(5)}{60(6)} = \mathbf{333.333\ Hz}$$

7.4.2 Closed-Loop Positioning Systems

A closed-loop NC system, illustrated in Figure 7.12(b), uses servomotors and feedback measurements to ensure that the worktable is moved to the desired position. A common feedback sensor used for NC (and also for industrial robots) is an optical encoder, depicted in Figure 7.13. An ***optical encoder*** is a device for measuring rotational speed that consists of a light source and a photodetector on either side of a disk. The disk contains slots uniformly spaced around the outside of its face. These slots allow the light source to shine through and energize the photodetector. The disk is connected to a rotating shaft whose angular position and velocity are to be measured. As the shaft rotates, the slots cause the light source to be seen by the photocell as a series of flashes. The flashes are converted into an equal number of electrical pulses. The optical encoder is connected directly to the leadscrew or ball screw, which drives the worktable. By counting the pulses and computing the frequency of the pulse train, the worktable position and velocity can be determined. There is usually a gear reduction between the servomotor and the screw driving the worktable.

The equations that define the operation of a closed-loop NC positioning system are similar to those for an open-loop system. In the basic optical encoder, the angle between slots in the disk must satisfy the following requirement:

$$\alpha = \frac{360}{n_s} \tag{7.12}$$

where α = angle between slots, °/slot; and n_s = number of slots in the disk, slots/rev. For a certain angular rotation of the encoder shaft, the number of pulses sensed by the encoder is given by

$$n_p = \frac{A_s}{\alpha} = \frac{A_s n_s}{360} \tag{7.13}$$

where n_p = pulse count emitted by the encoder; A_s = angle of rotation of the encoder shaft, °; and α = angle between slots, which converts to °/pulse. The pulse count can be used to determine the distance moved by the worktable along the x-axis (or y-axis):

$$\Delta x = \frac{p n_p}{n_s} = \frac{p A_s}{n_s \alpha} = \frac{p A_s}{360} \tag{7.14}$$

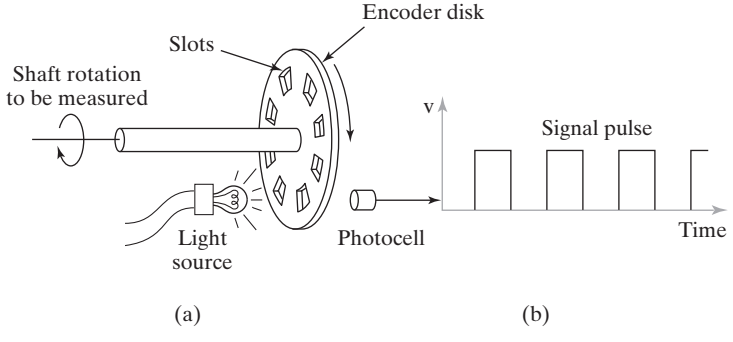


Figure 7.13 Optical encoder: (a) apparatus and (b) series of pulses emitted to measure rotation of disk.

where Δx = distance moved along the axis, mm (in); n_p and n_s are defined above; and p = screw pitch, mm/rev (in/rev).

The velocity of the worktable, which is normally the feed rate in a machining operation, is determined by the rotational speed of the screw, which in turn is driven by the servomotor:

$$v_t = f_r = N_s p = \frac{N_m p}{r_g} \tag{7.15}$$

where v_t = worktable velocity, mm/min (in/min); and f_r = feed rate, mm/min (in/min); N_s = screw rotational speed, rev/min; N_m = motor rotational speed, rev/min; r_g = *gear* reduction ratio.

At the worktable velocity or feed rate given by Equation (7.15), the pulse frequency emitted by the encoder is the following:

$$f_p = \frac{v_t n_s}{60p} = \frac{f_r n_s}{60p} \tag{7.16}$$

where f_p = frequency of the pulse train, Hz; and the constant 60 converts worktable velocity or feed rate from mm/sec (in/sec) to mm/min (in/min).

The pulse train generated by the encoder is compared with the coordinate position and feed rate specified in the part program, and the difference is used by the MCU to drive a servomotor, which in turn drives the worktable. A digital-to-analog converter (Section 6.3.2) is used to convert the digital signals used by the MCU into a continuous analog current that powers the drive motor. Closed-loop NC systems of the type described here are appropriate when a reactionary force resists the movement of the table. Metal cutting machine tools that perform continuous path cutting operations, such as milling and turning, fall into this category.

EXAMPLE 7.2 NC Closed-Loop Positioning

An NC worktable operates by closed-loop positioning. The system consists of a servomotor, ball screw, and optical encoder. The screw has a pitch of 6.0 mm and is coupled to the motor shaft with a gear ratio of 5:1 (five turns of the drive motor for each turn of the screw). The optical encoder generates 48 pulses/rev of its output shaft. The table has been programmed to move a distance of 250 mm at a feed rate = 500 mm/min. Determine (a) how many pulses should be received by the control system to verify that the table has moved exactly 250 mm, (b) the pulse rate of the encoder, and (c) the drive motor speed that corresponds to the specified feed rate.

Solution: (a) Rearranging Equation (7.14) to find n_p,

$$n_p = \frac{\Delta x n_s}{p} = \frac{250(48)}{6.0} = \mathbf{2{,}000\ pulses}$$

(b) The pulse rate corresponding to 500 mm/min is obtained by Equation (7.16):

$$f_p = \frac{f_r n_s}{60p} = \frac{500(48)}{60(6.0)} = \mathbf{66.667\ Hz}$$

(c) Motor speed = table velocity (feed rate) divided by screw pitch, corrected for gear ratio:

$$N_m = \frac{r_g f_r}{p} = \frac{5(500)}{6.0} = \textbf{416.667 rev/min}$$

Comment: Note that motor speed has the same numerical value as in Example 7.1 because the table velocity and motor gear ratio are the same.

7.4.3 Precision in Positioning Systems

To accurately machine or otherwise process a work part, an NC positioning system must possess a high degree of precision. Three measures of precision can be defined for an NC positioning system: (1) control resolution, (2) accuracy, and (3) repeatability. These terms are most readily explained by considering a single axis of the positioning system, as depicted in Figure 7.14. Control resolution refers to the control system's ability to divide the total range of the axis movement into closely spaced points that can be distinguished by the MCU. ***Control resolution*** is defined as the distance separating two adjacent addressable points in the axis movement. ***Addressable points*** are locations along the axis to which the worktable can be specifically directed to go. It is desirable for control resolution to be as small as possible. This depends on limitations imposed by (1) the electromechanical components of the positioning system and/or (2) the number of bits used by the controller to define the axis coordinate location.

A number of electromechanical factors affect control resolution, including screw pitch, gear ratio in the drive system, and the step angle in a stepper motor for an open-loop system or the angle between slots in an encoder disk for a closed-loop system. For an open-loop positioning system driven by a stepper motor, these factors can be combined into an expression that defines control resolution as

$$CR_1 = \frac{p}{n_s r_g} \tag{7.17}$$

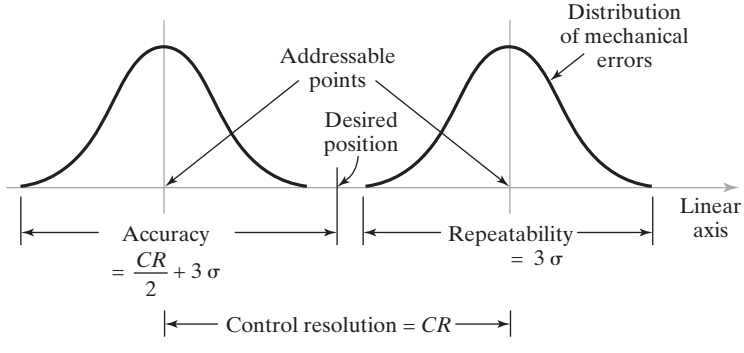


Figure 7.14 A portion of a linear positioning system axis, with definition of control resolution, accuracy, and repeatability.

where CR_1 = control resolution of the electromechanical components, mm (in); p = leadscrew pitch, mm/rev (in/rev); n_s = number of steps per revolution; and r_g = gear ratio between the motor shaft and the screw as defined in Equation (7.6). The same expression can be used for a closed-loop positioning system.

The second factor that limits control resolution is the number of bits used by the MCU to specify the axis coordinate value. For example, this limitation may be imposed by the bit storage capacity of the controller. If B = the number of bits in the storage register for the axis, then the number of control points into which the axis range can be divided = 2^B. Assuming that the control points are separated equally within the range, then

$$CR_2 = \frac{L}{2^B - 1} \tag{7.18}$$

where CR_2 = control resolution of the computer control system, mm (in); and L = axis range, mm (in). The control resolution of the positioning system is the maximum of the two values; that is,

$$CR = \text{Max}\{CR_1, CR_2\} \tag{7.19}$$

A desirable criterion is $CR_2 \leq CR_1$, meaning that the electromechanical system is the limiting factor that determines control resolution. The bit storage capacity of a modern computer controller is sufficient to satisfy this criterion except in unusual situations. Resolutions of 0.0025 mm (0.0001 in) are within the current state of CNC technology.

The ability of a positioning system to move the worktable to the exact location defined by a given addressable point is limited by mechanical errors that are due to various imperfections in the mechanical system. These imperfections include play between the screw and the worktable, backlash in the gears, and deflection of machine components. The mechanical errors are assumed to form an unbiased normal statistical distribution about the control point whose mean $\mu = 0$. It is further assumed that the standard deviation σ of the distribution is constant over the range of the axis under consideration. Given these assumptions, nearly all of the mechanical errors (99.73%) are contained within $\pm 3\sigma$ of the control point. This is pictured in Figure 7.14 for a portion of the axis range that includes two control points.

These definitions of control resolution and mechanical error distribution can now be used to define accuracy and repeatability of a positioning system. Accuracy is defined under worst case conditions in which the desired target point lies in the middle between two adjacent addressable points. Since the table can only be moved to one or the other of the addressable points, there will be an error in the final position of the worktable. This is the maximum possible positioning error, because if the target were closer to either one of the addressable points, then the table would be moved to the closer point and the error would be smaller. It is appropriate to define accuracy under this worst-case scenario. The ***accuracy*** of any given axis of a positioning system is the maximum possible error that can occur between the desired target point and the actual position taken by the system. In equation form,

$$Ac = \frac{CR}{2} + 3\sigma \tag{7.20}$$

where Ac = accuracy, mm (in); CR = control resolution, mm (in); and σ = standard deviation of the error distribution. Accuracies in machine tools are generally expressed for a certain range of table travel, for example, ± 0.01 mm for 250 mm (± 0.0004 in. for 10 in) of table travel.

Repeatability refers to the ability of the positioning system to return to a given addressable point that has been previously programmed. This capability can be measured in terms of the location errors encountered when the system attempts to position itself at the addressable point. Location errors are a manifestation of the mechanical errors of the positioning system, which follow a normal distribution, as assumed previously. Thus, the ***repeatability*** of any given axis of a positioning system is ± 3 standard deviations of the mechanical error distribution associated with the axis. This can be written as

$$Re = \pm 3\sigma \tag{7.21}$$

where Re = repeatability, mm (in).

EXAMPLE 7.3 Control Resolution, Accuracy, and Repeatability in NC

Suppose the mechanical inaccuracies in the open-loop positioning system of Example 7.1 are described by a normal distribution with standard deviation = 0.005 mm. The range of the worktable axis is 1,000 mm, and there are 16 bits in the binary register used by the digital controller to store the programmed position. Other relevant parameters from Example 7.1 are the following: pitch = 6.0 mm, gear ratio between motor shaft and screw = 5.0, and number of step angles in the stepper motor = 48. Determine the (a) control resolution, (b) accuracy, and (c) repeatability of the positioning system.

Solution: (a) Control resolution is the greater of CR_1 and CR_2 as defined by Equations (7.17) and (7.18).

$$CR_1 = \frac{p}{n_s r_g} = \frac{6.0}{48(5.0)} = 0.025 \text{ mm}$$

$$CR_2 = \frac{1{,}000}{2^{16} - 1} = \frac{1{,}000}{65{,}535} = 0.01526 \text{ mm}$$

$$CR = \text{Max}\{0.025, 0.01526\} = \mathbf{0.025\ mm}$$

(b) Accuracy is given by Equation (7.20):

$$Ac = 0.5(0.025) + 3(0.005) = \mathbf{0.0275\ mm}$$

(c) Repeatability $Re = \pm 3(0.005) = \pm \mathbf{0.015\ mm}$

7.5 NC PART PROGRAMMING

NC part programming consists of planning and documenting the sequence of processing steps to be performed by an NC machine. The part programmer must have a knowledge of machining (or other processing technology for which the NC machine is designed), as well as geometry and trigonometry. The documentation portion of part programming involves the input medium used to transmit the program of instructions to the NC machine control unit. The traditional input medium dating back to the first NC machines in the 1950s is 1-in wide punched tape. More recently, magnetic tape, floppy disks, and portable solid-state memory

devices have been used for NC due to their much higher data density. Distributed numerical control is also commonly used to transmit part programs from a central storage unit.

Part programming can be accomplished using a variety of procedures ranging from highly manual to highly automated methods. The methods are (1) manual part programming, (2) computer-assisted part programming, (3) CAD/CAM part programming, and (4) manual data input.

7.5.1 Manual Part Programming

In manual part programming, the programmer prepares the NC code using a low-level machine language that is described briefly in this section and more thoroughly in Appendix 7A. The coding system is based on binary numbers. This coding is the low-level machine language that can be understood by the MCU. When higher level languages are used, such as APT (Section 7.5.2) and CAD/CAM (Section 7.5.3), the statements in these respective programs are converted to this basic code. NC uses a combination of the binary and decimal number systems, called the *binary-coded decimal* (BCD) system. In this coding scheme, each of the ten digits (0–9) in the decimal system is coded as a four-digit binary number, and these binary numbers are added in sequence as in the decimal number system. Conversion of the ten digits in the decimal number system into binary numbers is shown in Table 7.4. Example 7.4 illustrates the conversion process.

In addition to numerical values, the NC coding system must also provide for alphabetical characters and other symbols. Eight binary digits are used to represent all of the characters required for NC part programming. Out of a sequence of characters, a word

EXAMPLE 7.4 Binary Coded Decimal

Convert the decimal value 1,258 to binary coded decimal.

Solution: The conversion of the four digits is shown in the following table:

Number Sequence	Binary Number	Decimal Value
First	0001	1,000
Second	0010	200
Third	0101	50
Fourth	0000	8
Sum		1,258

TABLE 7.4 Comparison of Binary and Decimal Numbers

Binary	Decimal	Binary	Decimal
0000	0	0101	5
0001	1	0110	6
0010	2	0111	7
0011	3	1000	8
0100	4	1001	9

is formed. A ***word*** specifies a detail about the operation, such as *x*-position, *y*-position, feed rate, or spindle speed. Out of a collection of words, a block is formed. A ***block*** is one complete NC instruction. It specifies the destination for the move, the speed and feed of the cutting operation, and other commands that determine explicitly what the machine tool will do. For example, an instruction block for a two-axis NC milling machine would likely include the *x*- and *y*-coordinates to which the machine table should be moved, the type of motion to be performed (linear or circular interpolation), the rotational speed of the milling cutter, and the feed rate at which the milling operation should be performed.

The organization of words within a block is known as a *block format.* Although a number of different block formats have been developed over the years, all modern controllers use the word address format, which uses a letter prefix to identify each type of word, and spaces to separate words within the block. This format also allows for variations in the order of words within the block, and omission of words from the block if their values do not change from the previous block. For example, the two commands in word address format to perform the two drilling operations illustrated in Figure 7.15 are

N001 G00 X07000 Y03000 M03
N002 Y06000

where N is the sequence number prefix, and X and Y are the prefixes for the *x*- and *y*-axes, respectively. G-words and M-words require some elaboration. G-words are called preparatory words. They consist of two numerical digits (following the "G" prefix) that prepare the MCU for the instructions and data contained in the block. For example, G00 prepares the controller for a point-to-point rapid traverse move between the present location and the endpoint defined in the current command. M-words are used to specify miscellaneous or auxiliary functions that are available on the machine tool. The M03 in the example is used to start the spindle rotation. Other examples include stopping the spindle for a tool change, and turning the cutting fluid on or off. Of course, the particular machine tool must possess the function that is being called.

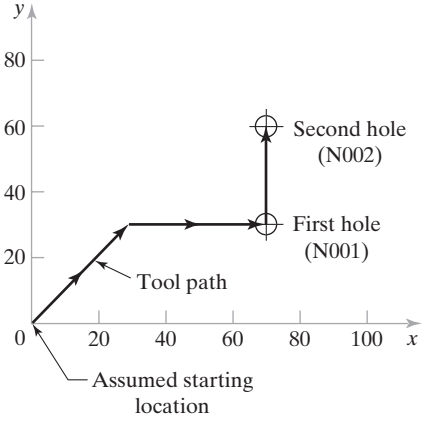


Figure 7.15 Drilling sequence for word address format example. Dimensions are in millimeters.

Words in an instruction block are intended to convey all of the commands and data needed for the machine tool to execute the move defined in the block. The words required for one machine tool type may differ from those required for a different type; for example, turning requires a different set of commands than milling. The words in a block are usually given in the following order (although the word address format allows variations in the order):

- sequence number (N-word)
- preparatory word (G-word)
- coordinates (X-, Y-, Z-words for linear axes, A-, B-, C-words for rotational axes)
- feed rate (F-word)
- spindle speed (S-word)
- tool selection (T-word)
- miscellaneous command (M-word)

For the interested reader, Appendix 7A has been prepared. This describes the details of the coding system used in manual part program. Examples of programming commands are provided and the various G-words and M-words are defined.

Manual part programming can be used for both point-to-point and contouring jobs. It is most suited for point-to-point machining operations such as drilling. It can also be used for simple contouring jobs, such as milling and turning when only two axes are involved. However, for complex three-dimensional machining operations, there is an advantage in using a more powerful part programming technique such as CAD/CAM.

7.5.2 Computer-Assisted Part Programming

Manual part programming can be time consuming, tedious, and subject to errors for parts possessing complex geometries or requiring many machining operations. A number of NC part programming language systems have been developed to accomplish many of the calculations that the programmer would otherwise have to do. The program is written in English-like statements that are subsequently converted to the low-level machine language described in Section 7.5.1 and Appendix 7A. Using this programming arrangement, the various tasks are divided between the human part programmer and the computer.

The Part Programmer's Job. In computer-assisted part programming, the machining instructions are written in English-like statements that are subsequently translated by the computer into the low-level machine code that can be interpreted and executed by the machine tool controller. The two main tasks of the programmer are (1) defining the geometry of the part and (2) specifying the tool path and operation sequence.

No matter how complicated the work part may appear, it is composed of basic geometric elements and mathematically defined surfaces. Consider the sample part in Figure 7.16. Although its appearance is somewhat irregular, the outline of the part consists of intersecting straight lines and a partial circle. The hole locations in the part can be defined in terms of the x- and y-coordinates of their centers. Nearly any component that can be conceived by a designer can be described by points, straight lines, planes, circles, cylinders, and other mathematically defined surfaces. It is the part programmer's task in computer-assisted part programming to identify and enumerate the geometric elements

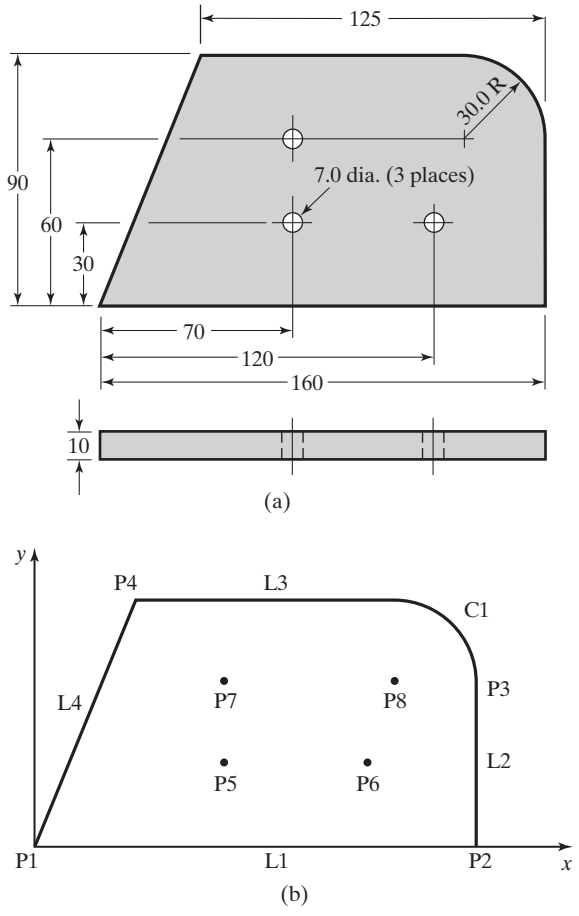


Figure 7.16 Sample part with geometry elements (points, lines, and circle) labeled for computer-assisted part programming.

of which the part is constructed. Each element must be defined in terms of its dimensions and location relative to other elements. A few examples will be instructive here to show how geometric elements are defined. The sample part will be used to illustrate, with labels of geometry elements added in Figure 7.16(b). The statements are taken from APT, which stands for *automatically programmed tooling*.[1]

The simplest geometric element is a point, and the simplest way to define a point is by means of its coordinates; for example,

$$\text{P4} = \text{POINT}/35, 90, 0$$

[1]Readers familiar with previous editions of this book will note that the appendix on APT is no longer included in the current edition. The author's impression is that APT is not widely used, especially in the United States, because it has largely been replaced by CAD/CAM part programming (Section 7.5.3). In addition to CAD/CAM part programming, manual part programming (G-codes and M-codes, Section 7.5.1 and Appendix 7A) and manual data input (Section 7.5.4) are also common in machine shops.

where the point is identified by a symbol (P4), and its coordinates are given in the order x, y, z in millimeters ($x = 35$ mm, $y = 90$ mm, and $z = 0$). A line can be defined by two points, as in the following:

$$\text{L1} = \text{LINE/P1, P2}$$

where L1 is the line defined in the statement, and P1 and P2 are two previously defined points. And finally, a circle can be defined by its center location and radius,

$$\text{C1} = \text{CIRCLE/CENTER, P8, RADIUS, 30}$$

where C1 is the newly defined circle, with center at previously defined point P8 and radius = 30 mm. The APT language offers many alternative ways to define points, lines, circles, and other geometric elements.

After the part geometry has been defined, the part programmer must next specify the tool path that the cutter will follow to machine the part. The tool path consists of a sequence of connected line and arc segments, using the previously defined geometry elements to guide the cutter. Consider how the outline of the sample part in Figure 7.16 would be machined in a profile milling operation (contouring). A cut has just been finished along surface L1 in a counterclockwise direction around the part, and the tool is presently located at the intersection of surfaces L1 and L2. The following APT statement could be used to command the tool to make a left turn from L1 onto L2 and to cut along L2:

$$\text{GOLFT/L2, TANTO, C1}$$

The tool proceeds along surface L2 until it is tangent to (TANTO) circle C1. This is a continuous path motion command. Point-to-point commands tend to be simpler; for example, the following statement directs the tool to go to a previously defined point P5:

$$\text{GOTO/P5}$$

In addition to defining part geometry and specifying tool path, the programmer must enter other programming functions, such as naming the program, identifying the machine tool on which the job will be performed, specifying cutting speeds and feed rates, designating the cutter size (cutter radius, tool length, etc.), and specifying tolerances in circular interpolation.

Computer Tasks in Computer-Assisted Part Programming. The computer's role in computer-assisted part programming consists of the following steps, performed more or less in the sequence given: (1) input translation, (2) arithmetic and cutter offset computations, (3) editing, and (4) post-processing. The first three steps are carried out under the supervision of the language processing program. For example, the APT language uses a processor designed to interpret and process the words, symbols, and numbers written in APT. Other high-level languages require their own processors, but they work similarly to APT. The fourth step, post-processing, requires a separate computer program. The sequence and relationship of the steps of the part programmer and the computer are portrayed in Figure 7.17.

The part programmer enters the program using APT or some other high-level part programming language. The input translation module converts the coded instructions contained in the program into computer-usable form, preparatory to further processing. In APT, input translation accomplishes the following tasks: (1) syntax check of the input code to identify errors in format, punctuation, spelling, and statement sequence; (2) assigning

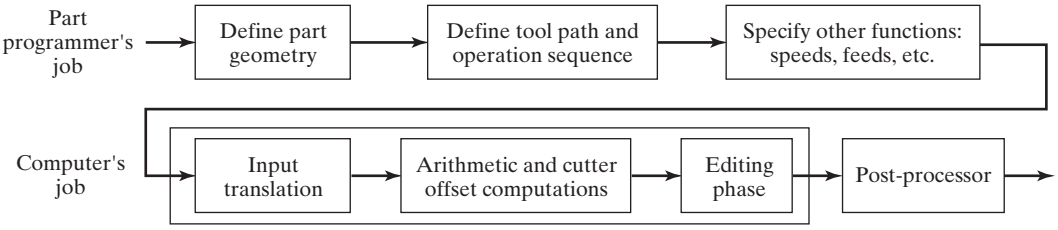


Figure 7.17 Steps in computer-assisted part programming.

a sequence number to each APT statement in the program; (3) converting geometry elements into a suitable form for computer processing; and (4) generating an intermediate file called PROFIL that is utilized in subsequent arithmetic calculations.

The arithmetic module consists of a set of subroutines to perform the mathematical computations required to define the part surface and generate the tool path, including compensation for cutter offset. The individual subroutines are called by the various statements used in the part programming language. The arithmetic computations are performed on the PROFIL file. The arithmetic module frees the programmer from the time-consuming and error-prone geometry and trigonometry calculations to concentrate on issues related to work part processing. The output of this module is a file called CLFILE, which stands for "cutter location file." As its name suggests, this file consists mainly of tool path data.

During the editing stage, the computer edits the CLFILE and generates a new file called CLDATA. When printed, CLDATA provides readable data on cutter locations and machine tool operating commands. The machine tool commands can be converted to specific instructions during post-processing. The output of the editing phase is a part program in a format that can be post-processed for the given machine tool on which the job will be accomplished.

NC machine tool systems are different. They have different features and capabilities. High-level part programming languages, such as APT, are generally not intended for only one machine tool type. They are designed to be general purpose. Accordingly, the final task of the computer in computer-assisted part programming is ***post-processing***, in which the cutter location data and machining commands in the CLDATA file are converted into low-level code that can be interpreted by the NC controller for a specific machine tool. The output of post-processing is a part program consisting of G-codes, *x*-, *y*-, and *z*-coordinates, S, F, M, and other functions in word address format. The post-processor is separate from the high-level part programming language. A unique post-processor must be written for each machine tool system.

7.5.3 CAD/CAM[2] Part Programming

A CAD/CAM system is a computer interactive graphics system equipped with software to accomplish certain tasks in design and manufacturing and to integrate the design and manufacturing functions. CAD/CAM is discussed in Chapter 23. One of the important tasks performed on a CAD/CAM system is NC part programming. In this method of part programming, portions of the procedure usually done by the part programmer are

[2]CAD/CAM stands for computer-aided design/computer-aided manufacturing.

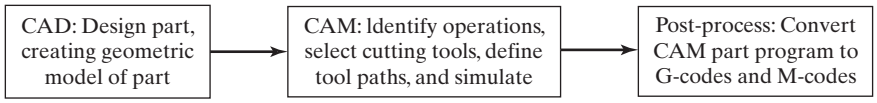


Figure 7.18 Steps in CAD/CAM part programming.

instead done by the computer. Advantages of NC part programming using CAD/CAM include the following [11]: (1) the part program can be simulated off-line on the CAD/CAM system to verify its accuracy; (2) the time and cost of the machining operation can be determined by the CAD/CAM system; (3) the most appropriate tooling can be automatically selected for the operation; and (4) the CAD/CAM system can automatically insert the optimum values for speeds and feeds for the work material and operations.

Other advantages are described below. Recall that the two main tasks of the part programmer in computer-assisted part programming are defining the part geometry and specifying the tool path. CAD/CAM systems automate portions of both of these tasks. The procedure in CAD/CAM part programming can be summarized in three steps, illustrated in Figure 7.18: (1) CAD: create geometric model of part; (2) CAM: define tool paths, select cutting tools, and simulate tool paths; and (3) post-processing to generate a part program in word address format.

CAD/CAM Part Geometry Definition. A fundamental objective of CAD/CAM is to integrate the design engineering and manufacturing engineering functions. Certainly one of the important design functions is to design the individual components of the product. If a CAD/CAM system is used, a computer graphics model of each part is developed by the designer and stored in the CAD/CAM database. That model contains all the geometric, dimensional, and material specifications for the part.

When the same CAD/CAM system is used to perform NC part programming, the programmer can retrieve the part geometry model from the CAD database and use that model to construct the appropriate cutter path. The significant advantage of using CAD/CAM in this way is that it eliminates one of the time-consuming steps in computer-assisted part programming: geometry definition. After the part geometry has been retrieved, the usual procedure is to label the geometric elements that will be used during part programming. These labels are the variable names (symbols) given to the lines, circles, and surfaces that comprise the part. CAM systems automatically label the geometry elements of the part and display the labels on the monitor. The programmer can then refer to those labeled elements during tool path construction.

An NC programmer who does not have access to the database must define the geometry of the part, using similar interactive graphics techniques that the product designer would use to design the part. Points are defined in a coordinate system using the computer graphics system, lines and circles are defined from the points, surfaces are defined, and so forth, to construct a geometric model of the part. The advantage of the interactive graphics system over conventional computer-assisted part programming is that the programmer receives immediate visual verification of the geometric elements being created. This tends to improve the speed and accuracy of the geometry definition process.

CAD/CAM Tool Path Generation and Simulation. The second task of the NC programmer in CAD/CAM part programming is tool path specification for the various operations to be performed. For each operation, the programmer selects a cutting tool

from a tool library listing the tools available in the tool crib. The programmer must decide which of the available tools is most appropriate for the operation under consideration and then specify it for the tool path. This permits the tool diameter and other dimensions to be entered automatically for tool offset calculations. If the desired cutting tool is not available in the library, the programmer can specify an appropriate tool. It then becomes part of the library for future use.

The next step is tool path definition. There are differences among CAM systems that result in different approaches for generating the tool path. The most basic approach involves the use of the interactive graphics system to enter the motion commands one by one, similar to computer-assisted part programming. Individual statements in APT or other part programming language are entered, and the CAD/CAM system provides an immediate graphic display of the action resulting from the command, thereby validating the statement.

A more advanced approach for generating tool path commands is to use one of the automatic software modules available on the CAD/CAM system. This can most readily be done for certain NC processes that involve well-defined, relatively simple part geometries. Examples are point-to-point operations such as NC drilling and electronic component assembly machines. In these processes, the program consists basically of a series of locations in an x–y coordinate system where work is to be performed (e.g., holes are to be drilled or components are to be inserted). These locations are defined by data generated during product design. Special algorithms are used to process the design data and generate the NC statements.

Additional modules have been developed to accomplish a number of common machining cycles for milling and turning. They are subroutines in the NC programming package that can be called and the required parameters entered to execute the machining cycle. Several of these modules are identified in Table 7.5 and Figure 7.19. NC contouring systems are capable of a similar level of automation.

When the complete part program has been developed, the CAD/CAM system can provide an animated simulation of the program for validation purposes. Any corrections to the tool path are made at this time.

TABLE 7.5 Some Common NC Modules for Automatic Programming of Machining Cycles

Module Type	Brief Description
Profile milling	Generates cutter path around the periphery of a part, usually a two-dimensional contour where depth remains constant.
Pocket milling	Generates the tool path to machine a cavity, as in Figure 7.19(a). A series of cuts is usually required to complete the bottom of the cavity to the desired depth.
Engraving	Generates tool path to engrave (mill) alphanumeric characters and other symbols to specified font and size.
Contour turning	Generates tool path for a series of turning cuts to provide a defined contour on a rotational part, as in Figure 7.19(b).
Facing	Generates tool path for a series of facing cuts to remove excess stock from the end of a rotational part or to create a shoulder on the part by a series of facing operations, as in Figure 7.19(c).
Threading	Generates tool path for a series of threading cuts to cut external, internal, or tapered threads on a rotational part, as in Figure 7.19(d) for external threads.

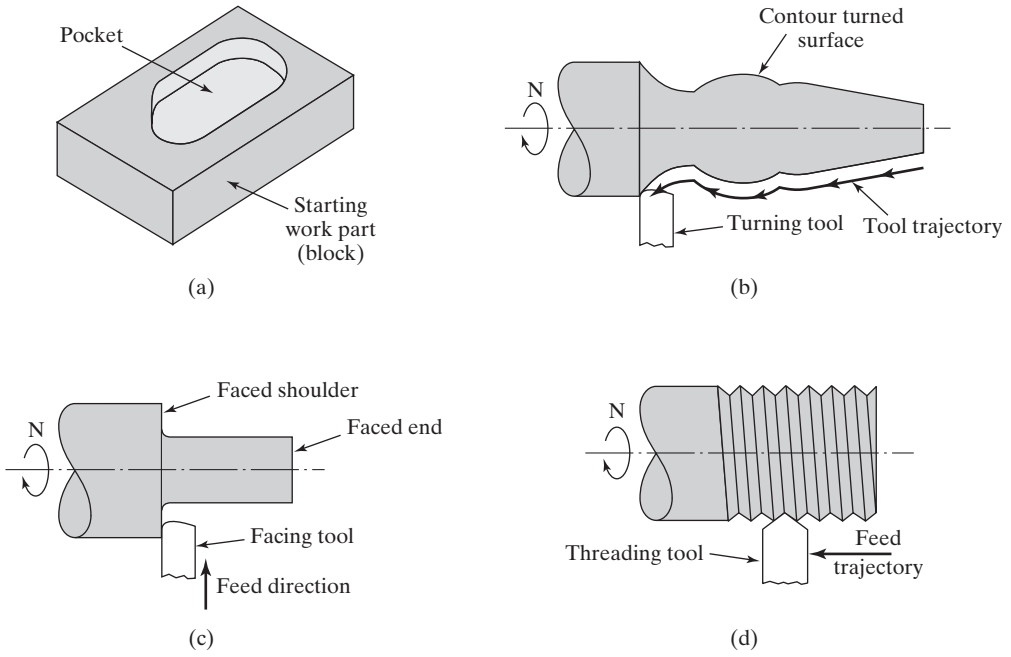


Figure 7.19 Examples of machining cycles available in automatic programming modules. (a) pocket milling, (b) contour turning, (c) facing and shoulder facing, and (d) threading (external).

Finally, the part program that has been developed and verified using CAD/CAM is post-processed to create the machine-language part program in word address format for the particular machine tool that will be used for the job.

Mastercam. Mastercam is the leading commercial CAD/CAM software package for CNC part programming. It is available from CNC Software, Inc. [16]. The package includes a CAD capability for designing parts in addition to its CAM features for part programming. If an alternative computer-aided design package is used for design, files from these other packages can be translated for use within Mastercam. Processes to which Mastercam can be applied include milling and drilling, turning, plasma cutting, and laser cutting. The typical steps that a programmer uses in Mastercam to accomplish a part-programming job are listed in Table 7.6. The output of the Mastercam program is a part program in word address format.

STEP-NC. In CAD/CAM part programming, several aspects of the procedure are automated, as indicated above. Given the geometric model of a part that has been defined during product design on a CAD system, a future CAM system would possess sufficient logic and decision-making capability to accomplish NC part programming for the entire part without human assistance. Research and development is proceeding on a new machine tool control language that would eliminate the need for machine-level part programming using G-codes and M-codes. In effect, it would result in the automatic generation of NC part programs without the participation of human part programmers.

TABLE 7.6 Typical Sequence of Steps in CNC Part Programming Using Mastercam for a Sequence of Milling and Drilling Operations

Step	Description
1	Develop a CAD model of the part to be machined using Mastercam, or import the CAD model from a compatible CAD package.
2	Orient the starting workpiece relative to the axis system of the machine.
3	Identify the workpiece material and specified grade (e.g., Aluminum 2024, for selection of cutting conditions).
4	Select the operation to be performed (e.g., drilling, pocket milling, contouring) and the surface to be machined.
5	Select the cutting tool (e.g., 0.250-in drill) from the tool library.
6	Enter applicable cutting parameters such as hole depth.
7	Repeat steps 4 through 6 for each additional machining operation to be performed on the part.
8	Select appropriate post-processor to generate the part program in word address format for the machine tool on which the machining job will be accomplished.
9	Verify the part program by animated simulation of the sequence of machining operations to be performed on the part.

Called STEP-NC, the research is part of a larger international project to develop standards to define and exchange product data in a format that can be interpreted by the computer. The larger project is called STEP, which is an acronym for *Standard for the Exchange of Product Model Data*. The international standard is ISO 10303 [18], and the application protocol that deals with NC part programming is ISO 10303-238 (also known as AP 238), which is titled *Application Interpreted Model for Computer Numeric Controllers*.

The limitation of part programs based on G-codes and M-codes is that they consist of instructions that only direct the actions of the cutting tool, without any related information content about the part being machined. STEP-NC would replace G-codes and M-codes with a more advanced language that directly associates the CNC processing instructions to the geometric model contained in the CAD database. The CNC machine control unit would receive a STEP-NC file and be capable of converting that file into tooling and machining commands to the machine tool without any additional part programming.

7.5.4 Manual Data Input

Manual and computer-assisted part programming require a high degree of formal documentation. There is lead time required to write and validate the programs. CAD/CAM part programming automates a substantial portion of the procedure, but a significant commitment in equipment, software, and training is required. One method of simplifying the procedure is to have the machine operator perform the part-programming task at the machine tool. This is called *manual data input* (MDI) because the operator manually enters the part geometry data and motion commands directly into the MCU prior to running the job. Also known as *conversational programming* [5], [10], MDI was conceived as a way for the small machine shop to introduce NC into its operations without needing to acquire special NC part-programming equipment and hiring a part programmer. MDI

permits the shop to make a minimal initial investment to begin the transition to modern CNC technology. The limitation of manual data input is the risk of programming errors as jobs become more complicated. For this reason, MDI is usually applied for relatively simple parts.

Communication between the machine operator-programmer and the MDI system is accomplished using a graphical user interface (GUI), consisting of a display monitor and alphanumeric keyboard. Entering the programming commands into the controller is typically done using a menu-driven procedure in which the operator responds to prompts and questions posed by the NC system about the job to be machined. The sequence of questions is designed so that the operator inputs the part geometry and machining commands in a logical and consistent manner. A computer graphics capability is included in modern MDI programming systems to permit the operator to visualize the machining operations and verify the program. Typical verification features include tool path display and animation of the tool path sequence.

A minimum of training in NC part programming is required of the machine operator. The operator must have the ability to read an engineering drawing of the part and must be familiar with the machining process. An important caveat in the use of MDI is to make certain that the NC system does not become an expensive toy that stands idle while the operator is entering the programming instructions. Efficient use of the system requires that programming for the next part be accomplished while the current part is being machined. Most MDI systems permit these two functions to be performed simultaneously to reduce changeover time between jobs.

REFERENCES

[1] Chang, C. H., and M. Melkanoff, *NC Machine Programming and Software Design*, Prentice Hall, Englewood Cliffs, NJ, 1989.

[2] Groover, M. P., and E. W. Zimmers, Jr., *CAD/CAM: Computer-Aided Design and Manufacturing*, Prentice Hall, Englewood Cliffs, NJ, 1984.

[3] Illinois Institute of Technology Research Institute, *APT Part Programming*, McGraw-Hill Book Company, New York, 1967.

[4] Lin, S. C., *Computer Numerical Control: Essentials of Programming and Networking*, Delmar Publishers Inc., Albany, NY, 1994.

[5] Lynch, M., *Computer Numerical Control for Machining*, McGraw-Hill, New York, 1992.

[6] Mattson, M., *CNC Programming: Principles and Applications*, Delmar, Thomson Learning, Albany, NY, 2002.

[7] Noble, D. F., *Forces of Production*, Alfred A. Knopf, New York, 1984.

[8] Quesada, R., *Computer Numerical Control, Machining and Turning Centers,* Pearson/Prentice Hall, Upper Saddle River, NJ, 2005.

[9] Reintjes, J. F., *Numerical Control: Making a New Technology*, Oxford University Press, New York, 1991.

[10] Stenerson, J., and K. Curran, *Computer Numerical Control*: *Operation and Programming*, 3rd ed., Pearson/Prentice Hall, Upper Saddle River, NJ, 2007.

[11] Valentino, J. V., and J. Goldenberg, *Introduction to Computer Numerical Control,* 5th ed., Pearson/Prentice Hall, Upper Saddle River, NJ, 2012.

[12] Waurzyniak, P., "Machine Controllers: Smarter and Faster," *Manufacturing Engineering*, June 2005, pp. 61–73.

machine is still processing the previous part. When the machine cycle is finished, the robot reaches into the machine only 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 work part or (2) applying a specified limited force to the work part during gripping (for fragile work parts).
- *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.

8.3.2 Tools

The robot uses tools to perform processing operations on the work part. The robot manipulates the tool relative to a stationary or slowly moving object (e.g., work part or subassembly). Examples of 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 similar operations; assembly tool (e.g., automatic screwdriver); heating torch; ladle (for metal die casting); and water jet cutting tool. In each case, the robot must not only control the relative position of the tool with respect to the work as a function of time, it must also control the operation of the tool. For this purpose, the robot must be able to transmit control signals to the tool for starting, stopping, and otherwise regulating its actions.

In some applications, the robot may use multiple tools during the work cycle. For example, several sizes of routing or drilling bits must be applied to the work part. Thus, the robot must have a means of rapidly changing the tools. The end effector in this case takes the form of a fast-change tool holder for quickly fastening and unfastening the various tools used during the work cycle.

8.4 APPLICATIONS OF INDUSTRIAL ROBOTS

Robots are used in a wide field of applications in industry. Most of the current applications are in manufacturing. The applications can usually be classified into one of the following categories: (1) material handling, (2) processing operations, and (3) assembly and inspection. Section 8.4.4 lists some of the work characteristics that must be present in the application to make the installation of a robot technically and economically feasible.

8.4.1 Material Handling Applications

In material handling applications, the robot moves materials or parts from one place to another. To accomplish the transfer, the robot is equipped with a gripper that must be designed to handle the specific part or parts to be moved. Included within this application category are (1) material transfer and (2) machine loading and/or unloading. In many material handling applications, the parts must be presented to the robot in a known position and orientation. This requires some form of material handling device to deliver the parts into the work cell in this position and orientation.

Material Transfer. These applications are ones in which the primary purpose of the robot is to move parts from one location to another. In many cases, reorientation of the part is accomplished during the move. The basic application in this category is called a *pick-and-place* operation, in which the robot picks up a part and deposits it at a new location. Transferring parts from one conveyor to another is an example. The requirements of the application are modest; a low-technology robot (e.g., limited-sequence type) is often sufficient. Only two or three joints are required for many of the applications, and pneumatically powered robots are often used. Also, delta robots are used for many high-speed picking and packaging operations.

A more complex example of material transfer is ***palletizing***, in which the robot retrieves parts, cartons, or other objects from one location and deposits them onto a pallet or other container at multiple positions on the pallet. The problem is illustrated in Figure 8.11. Although the pickup point is the same for every cycle, the deposit location on the pallet is different for each carton. This adds to the degree of difficulty of the task. Either the robot must be taught each position on the pallet using the powered-leadthrough method (Section 8.5.1), or it must compute the location based on the dimensions of the pallet and the center distances between the cartons in both x- and y-directions, and in the z-direction if the pallet is stacked.

Other applications similar to palletizing include ***depalletizing***, which consists of removing parts from an ordered arrangement in a pallet and placing them at another location (e.g., onto a moving conveyor); ***stacking*** operations, which involve placing flat parts on top of each other, such that the vertical location of the drop-off position is continuously changing with each cycle; and ***insertion*** operations, in which the robot inserts parts into the compartments of a divided carton.

Machine Loading and/or Unloading. In machine loading and/or unloading applications, the robot transfers parts into and/or from a production machine. The three possible cases are (1) machine loading, in which the robot loads parts into the production machine, but the parts are unloaded from the machine by some other means; (2) machine unloading, in which the raw materials are fed into the machine without using the robot, and

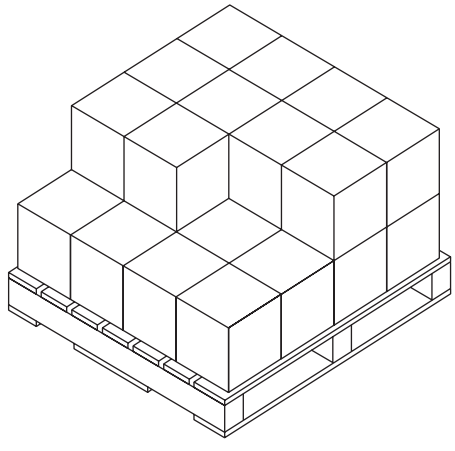


Figure 8.11 Typical part arrangement for a robot palletizing operation.

the robot unloads the finished parts; and (3) machine loading and unloading, which involves both loading of the raw work part and unloading of the finished part by the robot. Industrial robot applications of machine loading and/or unloading include the following processes:

- *Die casting.* The robot unloads parts from the die casting machine. Peripheral operations sometimes performed by the robot include dipping the parts into a water bath for cooling.
- *Plastic molding.* Plastic molding is similar to die casting. The robot unloads molded parts from the injection molding machine.
- *Metal machining operations.* The robot loads raw blanks into the machine tool and unloads finished parts from the machine. The change in shape and size of the part before and after machining often presents a problem in end effector design, and dual grippers (Section 8.3.1) are often used to deal with this issue.
- *Forging.* The robot typically loads the raw hot billet into the die, holds it during the forging strikes, and removes it from the forge hammer. The hammering action and the risk of damage to the die or end effector are significant technical problems.
- *Pressworking.* Human operators work at considerable risk in sheetmetal pressworking operations because of the action of the press. Robots are used to substitute for the workers to reduce the danger. In these applications, the robot loads the blank into the press, then the stamping operation is performed, and the part falls out of the machine into a container.
- *Heat-treating.* These are often relatively simple operations in which the robot loads and/or unloads parts from a furnace.

8.4.2 Processing Operations

In processing applications, the robot performs some operation on a work part, such as grinding or spray painting. A distinguishing feature of this category is that the robot is equipped with some type of tool as its end effector (Section 8.3.2). To perform the process, the robot must manipulate the tool relative to the part. Examples of industrial robot applications in the processing category include spot welding, arc welding, spray painting, and various machining and other rotating spindle processes.

Spot Welding. Spot welding is a metal joining process in which two sheet metal parts are fused together at localized points of contact. Two electrodes squeeze the metal parts together and then a large electrical current is applied across the contact point to cause fusion to occur. The electrodes, together with the mechanism that actuates them, constitute the welding gun in spot welding. Because of its widespread use in the automobile industry for car body fabrication, spot welding represents one of the most common applications of industrial robots today. The end effector is the spot welding gun used to pinch the car panels together and perform the resistance welding process. The welding gun used for automobile spot welding is typically heavy. Prior to the application of robots, human workers performed this operation, and the heavy welding tools were difficult for humans to manipulate accurately. As a consequence, there were many instances of missed welds, poorly located welds, and other defects, resulting in overall low quality of the finished product. The use of industrial robots in this application has dramatically improved the consistency of the welds.

Robots used for spot welding are usually large, with sufficient payload capacity to wield the heavy welding gun. Five or six axes are generally required to achieve the

required position and orientation of the welding gun. Playback robots with point-to-point control are used. Jointed-arm robots are the most common type in automobile spot-welding lines, which may consist of several dozen robots.

Arc Welding. Arc welding is used to provide continuous welds rather than individual spot welds at specific contact points. The resulting arc-welded joint is substantially stronger than in spot welding. Because the weld is continuous, it can be used in airtight pressure vessels and other weldments in which strength and continuity are required. There are various forms of arc welding, but they all follow the general description given here.

The working conditions for humans who perform arc welding are not good. The welder must wear a face helmet for eye protection against the ultraviolet light emitted by the arc welding process. The helmet window must be dark enough to mask the UV radiation. High electrical current is used in the welding process, and this creates a hazard for the welder. Finally, there is the obvious danger from the high temperatures in the process, high enough to melt the steel, aluminum, or other metal that is being welded. Significant hand–eye coordination is required by human welders to make sure that the arc follows the desired path with sufficient accuracy to make a good weld. This, together with the conditions described above, results in worker fatigue. Consequently, the welder is only accomplishing the welding process for perhaps 20–30% of the time. This ***arc-on time*** is defined as the proportion of time during the shift when the welding arc is on and performing the process. To assist the welder, a second worker is usually present at the work site, called a *fitter*, whose job is to set up the parts to be welded and to perform other similar chores in support of the welder.

Because of these conditions in manual arc welding, automation is used where technically and economically feasible. For welding jobs involving long continuous joints that are accomplished repetitively, mechanized welding machines have been designed to perform the process. These machines are used for long straight sections and regular round parts, such as pressure vessels, tanks, and pipes.

Industrial robots can also be used to automate the arc welding process. The cell consists of the robot, the welding apparatus (power unit, controller, welding tool, and wire feed mechanism), and a fixture that positions the components for the robot. The fixture might be mechanized with one or two axes so that it can present different portions of the work to the robot for welding (the term *positioner* is used for this type of fixture). For greater productivity, two fixtures are often used so that a human helper or another robot can unload the completed job and load the components for the next work cycle while the welding robot is simultaneously welding the present job. Figure 8.12 illustrates this kind of workplace arrangement.

The robot used in arc welding must be capable of continuous path control. Jointed-arm robots consisting of six joints are frequently used. Some robot vendors provide manipulators that have hollow upper arms, so that the cables connected to the welding torch can be contained in the arm for protection, rather than attached to the exterior. Also, programming improvements for arc welding based on CAD/CAM have made it much easier and faster to implement a robot welding cell. The weld path can be developed directly from the CAD model of the assembly [9].

Spray Coating. Spray coating directs a spray gun at the object to be coated. Fluid (e.g., paint) flows through the nozzle of the spray gun to be dispersed and applied over the surface of the object. Spray painting is the most common application in the category, but spray coating refers to a broader range of applications that includes painting.

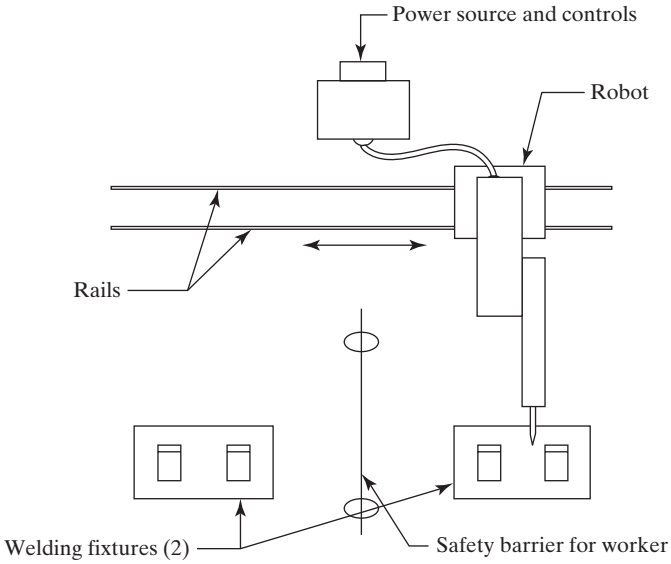


Figure 8.12 Robot arc welding cell in which the welding robot moves between welding fixtures on rails. A worker unloads the completed weldment and loads parts into one fixture while the robot performs the welding cycle at the other fixture.

The work environment for humans who perform this process is filled with health hazards. These hazards include harmful and noxious fumes in the air and noise from the spray gun nozzle. To mitigate these hazards, robots are being used more and more for spray coating tasks, particularly in high-production operations.

Robot applications include spray coating of automobile car bodies, appliances, engines, and other parts; spray staining of wood products; and spraying of porcelain coatings on bathroom fixtures. The robot must be capable of continuous path control to accomplish the smooth motion sequences required in spray painting. The most convenient programming method is manual leadthrough (Section 8.5.1). Jointed-arm robots seem to be the most common anatomy for this application. The robot must possess a work volume sufficient to access all areas of the work part to be coated in the application.

The use of industrial robots for spray coating offers a number of benefits in addition to protecting workers from a hazardous environment. These other benefits include greater uniformity in applying the coating than humans can accomplish, reduced waste of paint, lower needs for ventilating the work area because humans are not present during the process, and greater productivity.

Other Processing Applications. Spot welding, arc welding, and spray coating are common processing applications of industrial robots. The list of industrial processes that are being performed by robots is continually growing. Among these are the following:

- *Drilling, routing, and other machining processes.* These applications use a rotating spindle as the end effector. The cutting tool is mounted in the spindle chuck. One of the problems with this application is the high cutting forces encountered in machining. The robot must be strong enough to withstand these cutting forces and maintain the required accuracy of the cut.

- *Grinding, wire brushing, and similar operations*. Most of these operations use a rotating spindle as the end effector to drive a grinding wheel, wire brush, polishing wheel, or similar tool at high speed to accomplish finishing and deburring operations on the workpiece. In an alternative approach described in [13], the robot is equipped with a gripper to hold and manipulate the workpiece against a rotating deburring head.
- *Waterjet cutting*. This is a process in which a high-pressure stream of water is forced through a small nozzle at high speed to cut plastic sheets, fabrics, cardboard, and other materials with precision. The end effector is the waterjet nozzle that is directed to follow the desired cutting path by the robot.
- *Laser cutting*. The function of the robot in this application is similar to its function in waterjet cutting. Laser beam welding is a similar application. The laser gun is attached to the robot as its end effector. In an application described in [6], robots are used to trim excess sheet metal from parts produced in hot stamping operations. The hot-stamped sheet metal is too hard to trim with conventional cutting dies, so laser cutting must be used.

8.4.3 Assembly and Inspection

In some respects, assembly and inspection are hybrids of the previous two categories: material handling and processing. Assembly and inspection can involve either the handling of materials or the manipulation of a tool. For example, assembly operations typically involve the addition of components to build a product. This requires the movement of parts from a supply location in the workplace to the product being assembled, which is material handling. In some cases, the fastening of the components requires a tool to be used by the robot (e.g., driving a screw). Similarly, some robot inspection operations require that parts be manipulated, while other applications require that an inspection tool be manipulated.

Traditionally, assembly and inspection are labor-intensive activities. They are also highly repetitive and usually boring. For these reasons, they are logical candidates for robotic applications. However, assembly work typically involves diverse and sometimes difficult tasks, often requiring adjustments to be made in parts that don't quite fit together. A sense of feel is often required to achieve a close fitting of parts. Inspection work requires high precision and patience, and human judgment is often needed to determine whether a product is within quality specifications or not. Because of these complications in both types of work, the application of robots has not been easy. Nevertheless, the potential rewards are so great that substantial efforts have been made to develop the necessary technologies to achieve success in these applications.

Assembly. Assembly involves the combining of two or more parts to form a new entity, called a subassembly or assembly. The new entity is made secure by fastening the parts together using mechanical fastening techniques (e.g., screws, bolts and nuts, rivets) or joining processes (e.g., welding, brazing, soldering, or adhesive bonding). Welding applications have already been discussed.

Because of the economic importance of assembly, automated methods are often applied. Fixed automation is appropriate in mass production of relatively simple products, such as pens, mechanical pencils, cigarette lighters, and garden hose nozzles. Robots are usually at a disadvantage in these high-production situations because they cannot operate at the high speeds that fixed-automated equipment can. The most appealing application

of industrial robots for assembly involves situations in which a mix of similar models are produced in the same work cell or assembly line. Examples of these kinds of products include electric motors, small appliances, and various other small mechanical and electrical products. In these instances, the basic configuration of the different models is the same, but there are variations in size, geometry, options, and other features. Such products are often made in batches on manual assembly lines. However, the pressure to reduce inventories makes mixed-model assembly lines (Appendix 15A.2) more attractive. Robots can be used to substitute for some or all of the manual stations on these lines. What makes robots viable in mixed-model assembly is their capability to execute programmed variations in the work cycle to accommodate different product configurations.

Industrial robots used for the types of assembly operations described here are typically small, with light load capacities. The most common configurations are jointed arm, SCARA, and Cartesian coordinate. Accuracy and repeatability requirements in assembly work are often more demanding than in other robot applications, and the more precise robots in this category have repeatabilities of ± 0.05 mm (± 0.002 in). In addition, the requirements of the end effector are sometimes difficult. It may have to perform multiple functions at a single workstation to reduce the number of robots required in the cell. These functions may include handling more than one part geometry and performing both as a gripper and an assembly tool.

Inspection. There is often a need in automated production to inspect the work that is done. Inspections accomplish the following functions: (1) making sure that a given process has been completed, (2) ensuring that parts have been assembled as specified, and (3) identifying flaws in raw materials and finished parts. The topic of automated inspection is considered in more detail in Chapter 21. The purpose here is to identify the role played by industrial robots in inspection. Inspection tasks performed by robots can be divided into the following two cases:

1. The robot performs loading and unloading to support an inspection or testing machine. This case is really machine loading and unloading, where the machine is an inspection machine. The robot picks parts (or assemblies) that enter the cell, loads and unloads them to carry out the inspection process, and places them at the cell output. In some cases, the inspection may result in sorting of parts that must be accomplished by the robot. Depending on the quality level of the parts, the robot places them in different containers or on different exit conveyors.
2. The robot manipulates an inspection device, such as a mechanical probe or vision sensor, to inspect the product. This case is similar to a processing operation in which the end effector attached to the robot's wrist is the inspection probe. To perform the process, the part is delivered to the workstation in the correct position and orientation, and the robot must manipulate the inspection device as required.

8.4.4 Economic Justification of Industrial Robots

One of the earliest installations of an industrial robot was in 1961 in a die casting operation (Historical Note 8.1). The robot was used to unload castings from the die casting machine. The typical environment in die casting is not pleasant for humans due to the heat and fumes emitted by the casting process. It seemed desirable to use a robot in this type of work environment instead of a human operator.

Characteristics of Robot Applications. The general characteristics of industrial work situations that tend to promote the substitution of robots for human labor are the following:

1. *Hazardous work for humans.* When the work and the environment in which it is performed are hazardous, unsafe, unhealthful, uncomfortable, or otherwise unpleasant for humans, an industrial robot should be considered for the task. In addition to die casting, there are many other work situations that are hazardous or unpleasant for humans, including spray painting, arc welding, and spot welding. Industrial robots are applied in all of these processes.
2. *Repetitive work cycle.* A second characteristic that tends to promote the use of robotics is a repetitive work cycle. If the sequence of motion elements in the work cycle is the same, or nearly the same, a robot is usually capable of performing the cycle with greater consistency and repeatability than a human worker. Greater consistency and repeatability are manifested as higher product quality than what can be achieved in a manual operation.
3. *Difficult handling for humans.* If the task involves the handling of parts or tools that are heavy or otherwise difficult to manipulate, an industrial robot may be available that can perform the operation. Parts or tools that are too heavy for humans to handle conveniently are well within the load-carrying capacity of a large robot.
4. *Multishift operation.* In manual operations requiring second and third shifts, substitution of a robot provides a much faster financial payback than a single shift operation. Instead of replacing one worker, the robot replaces two or three workers.
5. *Infrequent changeovers.* Most batch or job shop operations require a changeover of the physical workplace between one job and the next. The time required to make the changeover is nonproductive time because parts are not being made. Consequently, robots have traditionally been easier to justify for relatively long production runs where changeovers are infrequent. Advances have been made in robot technology to reduce programming time, and shorter production runs have become more economical.
6. *Part position and orientation are established in the work cell.* Most robots in today's industrial applications do not possess vision capability. Their capacity to pick up a part or manipulate a tool during each work cycle relies on the work unit being in a known position and orientation. The work unit must be presented to the robot at the same location each cycle.[5]

Cycle Time and Cost Analysis. The cycle time and cost of a proposed robotic application can be analyzed using the methods of Chapter 3. Restating the basic cycle time equation, Equation (3.1):

$$T_c = T_o + T_h + T_t \tag{8.1}$$

where T_c = cycle time, min/pc; T_o = time of the actual processing or assembly operation, min/pc; T_h = work part handling time, min/pc; and T_t = average tool handling time,

[5]As mentioned in Section 8.1.4, many of the robots installed today are equipped with vision capability or are vision-compatible, meaning that their controllers have the software to readily integrate vision into the work cycle. Vision capability reduces the need for the work unit to be in a known position and orientation.

min/pc, if such an activity is applicable. Most robot applications involve either some form of material handling (Section 8.4.1) or a processing operation (Section 8.4.2). Assembly and inspection applications (Section 8.4.3) can be included within these two categories.

As indicated in Section 8.4.1, material handling applications include (1) material transfer, in which case Equation (8.1) reduces to $T_c = T_h$, or (2) machine loading and/or unloading, in which the robot is used to support a principal production machine performing the actual processing operation. In the second case, the robot's participation in the work cycle is T_h, and the production machine consists of T_o and possibly T_t, depending on the type of process.

If the robot application consists of a processing operation, in which the robot manipulates some tool as its end effector, then the robot is the principal production machine, and Equation (8.1) is probably a representative model for the work cycle. Work part handling is performed by support equipment, perhaps another robot or a human worker as a last resort.

The production rate of a robotic cell is based on the average production time, which must include the time to set up the cell.

$$T_p = \frac{T_{su} + QT_c}{Q} \tag{8.2}$$

where T_p = average production time per work unit, min; T_{su} = setup time, min; and Q = quantity of work units produced in the production run. The setup time must include the on-site time to program the robot in addition to the other physical setup activities prior to the actual production run. Production rate is the reciprocal of average production time:

$$R_p = \frac{60}{T_p} \tag{8.3}$$

where R_p is expressed in work units per hour, pc/hr. For long-running jobs, R_p approaches the cycle rate R_c, which is the reciprocal of T_c. That is, as Q becomes very large $(T_{su}/Q) \rightarrow 0$ and

$$R_p \rightarrow R_c = \frac{60}{T_c} \tag{8.4}$$

where R_c = operation cycle rate of the machine, pc/hr; and T_c = operation cycle time, min/pc, from Equation (8.1).

EXAMPLE 8.1 Robot Cycle Time Analysis

An articulated robot loads and unloads parts in a CNC (computer numerical control) machine cell in a high production run (assume setup time can be neglected). The machine tool operates on semiautomatic cycle which is coordinated with the robot using interlocks. The programmed machining cycle takes 2.25 min. Cutting tools wear out and must be periodically changed, which takes 5.0 min every 25 cycles and is performed by a human worker. At the end of each machining cycle, the robot reaches into the machine and removes the just-completed part, places it in a tote pan, then reaches for a starting work part from another tote pan and places it in the machine tool

chuck. This sequence of handling activities takes 30 sec. Tote pans are exchanged every 20 work cycles by the same worker who changes tools, but there is no lost production time. Determine (a) the average production time and (b) production rate of the cell.

Solution: (a) Changing cutting tools involves lost production time. On a per cycle basis, $T_t = 5.0/25 = 0.20$ min.

Average production time $T_p = T_c = T_o + T_h + T_t = 2.25 + 30/60 + 0.20 =$ **2.95 min/pc**

(b) Production rate $R_p = 60/2.95 =$ **20.34 pc/hr**

The costs of operating a robot cell divide into fixed and variable costs, where fixed costs are associated with equipment and variable costs include labor, raw materials, and power to operate the equipment. Total cost of the cell is the sum of the two categories:

$$TC = C_f + C_v Q \tag{8.5}$$

where TC = total annual cost, \$/yr; C_f = fixed annual cost, \$/yr; C_v = variable cost, \$/pc; and Q = annual quantity produced, pc/yr. A break-even analysis can be used to compare alternatives such as a manually operated work cell versus a robotic cell, similar to Example 3.5. The cost of the robot and other equipment in the cell can be reduced to an hourly rate using the methods of Section 3.2.3.

8.5 ROBOT PROGRAMMING

To accomplish useful work, a robot must be programmed to perform a motion cycle. A ***robot program*** can be defined as a path in space to be followed by the manipulator, combined with peripheral actions that support the work cycle. Examples of peripheral actions include opening and closing a gripper, performing logical decision making, and communicating with other pieces of equipment in the cell. A robot is programmed by entering the programming commands into its controller memory. Different robots use different methods of entering the commands.

In the case of limited-sequence robots, programming is accomplished by setting limit switches and mechanical stops to control the endpoints of its motions. The sequence in which the motions occur is regulated by a sequencing device. This device determines the order in which each joint is actuated to form the complete motion cycle. Setting the stops and switches and wiring the sequencer is more akin to a manual setup than programming.

Today, nearly all industrial robots have digital computers as their controllers, and compatible storage devices as their memory units. For these robots, three programming methods can be distinguished: (1) leadthrough programming, (2) computer-like robot programming languages, and (3) off-line programming.

8.5.1 Leadthrough Programming

Leadthrough programming dates from the early 1960s before computer control was prevalent. The same basic methods are used today for many computer-controlled robots. In leadthrough programming, the task is taught to the robot by moving the manipulator

through the required motion cycle, simultaneously entering the program into the controller memory for subsequent playback.

Powered Leadthrough and Manual Leadthrough. There are two methods of performing the leadthrough teach procedure: (1) powered leadthrough and (2) manual leadthrough. The difference between the two is in the manner in which the manipulator is moved through the motion cycle during programming. Powered leadthrough is commonly used as the programming method for playback robots with point-to-point control. It involves the use of a teach pendant (handheld control box) that has toggle switches and/or contact buttons for controlling the movement of the manipulator joints. Using the toggle switches or buttons, the programmer power-drives the robot arm to the desired positions, in sequence, and records the positions into memory. During subsequent playback, the robot moves through the sequence of positions under its own power.

Manual leadthrough is convenient for programming playback robots with continuous path control where the continuous path is an irregular motion pattern such as in spray painting. This programming method requires the operator to physically grasp the tool attached to the end of the arm and move it through the motion sequence, recording the path into memory. Because the robot arm itself may have significant mass and would therefore be difficult to move, a special programming device often substitutes for the actual robot during the teach procedure. The programming device has the same joint configuration as the robot and is equipped with a trigger handle (or other control switch), which the operator activates when recording motions into memory. The motions are recorded as a series of closely spaced points. During playback, the path is recreated by controlling the actual robot arm through the same sequence of points.

Motion Programming. The leadthrough methods provide a very natural way to program motion commands into the robot controller. In manual leadthrough, the operator simply moves the arm through the required path to create the program. In powered leadthrough, the operator uses a handheld teach pendant to drive the manipulator. The programmer moves the various joints of the manipulator to the required positions in the work space by activating the switches or buttons of the teach pendant in a coordinated fashion.

Coordinating the individual joints with the teach pendant is an awkward and tedious way to enter motion commands to the robot. For example, it is difficult to coordinate the individual joints of an articulated robot (TRR configuration) to drive the end-of-arm in a straight-line motion. Therefore, many robots using powered leadthrough provide two alternative methods for controlling movement of the entire manipulator during programming, in addition to controls for individual joints. With these methods, the programmer can move the robot's wrist end in straight line paths. The names given to these alternatives are (1) world-coordinate system and (2) tool-coordinate system. Both systems make use of Cartesian coordinates. In a world-coordinate system, the origin and axes are defined relative to the robot base, as illustrated in Figure 8.13(a). In a tool-coordinate system, Figure 8.13(b), the alignment of the axis system is defined relative to the orientation of the wrist faceplate (to which the end effector is attached). In this way, the programmer can orient the tool in a desired way and then control the robot to make linear moves in directions parallel or perpendicular to the tool.

The world- and tool-coordinate systems are useful only if the robot has the capacity to move its wrist end in a straight line motion, parallel to one of the axes of the coordinate system. Straight line motion is quite natural for a Cartesian coordinate robot (LOO configuration) but unnatural for robots with any combination of rotational

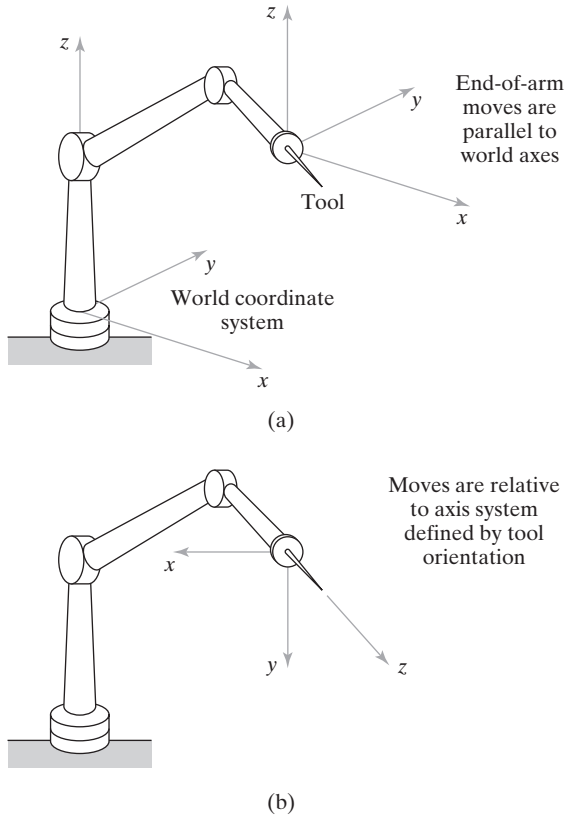


Figure 8.13 (a) World-coordinate system. (b) Tool-coordinate system.

joints (types R, T, and V). Accomplishing straight line motion requires manipulators with these types of joints to carry out a linear interpolation process. In ***straight line interpolation***, the control computer calculates the sequence of addressable points in space through which the wrist end must move to achieve a straight line path between two points.

Other types of interpolation are available. More common than straight line interpolation is joint interpolation. When a robot is commanded to move its wrist end between two points using ***joint interpolation***, it actuates each of the joints simultaneously at its own constant speed such that all of the joints start and stop at the same time. The advantage of joint interpolation over straight line interpolation is that usually less total motion energy is required to make the move. This may mean that the move could be made in slightly less time. It should be noted that in the case of a Cartesian coordinate robot, joint interpolation and straight line interpolation result in the same motion path.

Still another form of interpolation is used in manual-leadthrough programming. In this case, the robot must follow the sequence of closely spaced points that are defined during the programming procedure. In effect, this is an interpolation process for a path that usually consists of irregular smooth motions, such as in spray painting.

The speed of the robot is controlled by means of a dial or other input device, located on the teach pendant and/or the main control panel. Certain motions in the work cycle should be performed at high speed (e.g., moving parts over substantial distances in the cell), while other motions require low speed (e.g., motions that require high precision in positioning the work part). Speed control also permits a given program to be tried out at a safe slow speed and then used at a higher speed during production.

Advantages and Disadvantages. The advantage offered by the leadthrough methods is that they can be readily learned by shop personnel. Programming the robot by moving its arm through the required motion path is a logical way for someone to teach the work cycle. It is not necessary for the robot programmer to possess knowledge of computer programming. The robot languages described in the next section, especially the more advanced languages, are more easily learned by someone whose background includes computer programming.

There are several inherent disadvantages of the leadthrough programming methods. First, regular production must be interrupted during the leadthrough programming procedures. In other words, leadthrough programming results in downtime of the robot cell or production line. The economic consequence of this is that the leadthrough methods are most appropriate for relatively long production runs and are less appropriate for small batch sizes.

Second, the teach pendant used with powered leadthrough and the programming devices used with manual leadthrough are limited in terms of the decision-making logic that can be incorporated into the program. It is much easier to write logical instructions using a computer-like robot language than a leadthrough method.

Third, because the leadthrough methods were developed before computer control became common for robots, these methods are not readily compatible with modern computer-based technologies such as CAD/CAM, manufacturing databases, and local communications networks. The capability to readily interface the various computer-automated subsystems in the factory for transfer of data is considered a requirement for achieving computer integrated manufacturing.

8.5.2 Robot Programming Languages

The use of textual programming languages became an appropriate programming method as digital computers took over the control function in robotics. Their use has been stimulated by the increasing complexity of the tasks that robots are called on to perform, with the concomitant need to imbed logical decisions into the robot work cycle. These computer-like programming languages are really a combination of on-line and off-line methods, because the robot must still be taught its locations using the leadthrough method. Textual programming languages for robots provide the opportunity to perform the following functions that leadthrough programming cannot readily accomplish:

- Enhanced sensor capabilities, including the use of analog as well as digital inputs and outputs
- Improved output capabilities for controlling external equipment
- Program logic that is beyond the capabilities of leadthrough methods
- Computations and data processing similar to computer programming languages
- Communications with other computer systems.

This section reviews some of the capabilities of the robot programming languages. Many of the language statements are taken from commercially available robot languages.

Motion Programming. Motion programming with robot languages usually requires a combination of textual statements and leadthrough techniques. Accordingly, this method of programming is sometimes referred to as *on-line/off-line programming*. The textual statements are used to describe the motion, and the leadthrough methods are used to define the position and orientation of the robot during and/or at the end of the motion. To illustrate, the basic motion statement is

MOVE P1

which commands the robot to move from its current position to a position and orientation defined by the variable name P1. The point P1 must be defined, and the most convenient way to define P1 is to use either powered leadthrough or manual leadthrough to place the robot at the desired point and record that point into memory. Statements such as

HERE P1

or

LEARN P1

are used in the leadthrough procedure to indicate the variable name for the point. What is recorded into the robot's control memory is the set of joint positions or coordinates used by the controller to define the point. For example, the aggregate

(236, 158, 65, 0, 0, 0)

could be utilized to represent the joint positions for a six-axis manipulator. The first three values (236, 158, 65) give the joint positions of the body-and-arm, and the last three values (0, 0, 0) define the wrist joint positions. The values are specified in millimeters or degrees, depending on the joint types.

There are variants of the MOVE statement. These include the definition of straight line interpolation motions, incremental moves, approach and depart moves, and paths. For example, the statement

MOVES P1

denotes a move that is to be made using straight line interpolation. The suffix S on MOVE designates straight line motion.

An incremental move is one whose endpoint is defined relative to the current position of the manipulator rather than to the absolute coordinate system of the robot. For example, suppose the robot is presently at a point defined by the joint coordinates (236, 158, 65, 0, 0, 0), and it is desired to move joint 4 (corresponding to a twisting motion of the wrist) from 0 to 125. The following form of statement might be used to accomplish this move:

DMOVE (4, 125)

The new joint coordinates of the robot would therefore be given by (236, 158, 65, 125, 0, 0). The prefix D is interpreted as delta, so DMOVE represents a delta move, or incremental move.

Approach and depart statements are useful in material handling operations. The APPROACH statement moves the gripper from its current position to within a certain

distance of the pickup (or drop-off) point, and then a MOVE statement positions the end effector at the pickup point. After the pickup is made, a DEPART statement moves the gripper away from the point. The following statements illustrate the sequence:

```
APPROACH P1, 40 MM
MOVE P1
(command to actuate gripper)
DEPART 40 MM
```

The destination is point P1, but the APPROACH command moves the gripper to a safe distance (40 mm) above the point. This might be useful to avoid obstacles such as other parts in a tote pan. The orientation of the gripper at the end of the APPROACH move is the same as that defined for point P1, so that the final MOVE P1 is really a spatial translation of the gripper. This permits the gripper to be moved directly to the part for grasping.

A path in a robot program is a series of points connected together in a single move. The path is given a variable name, as illustrated in the following statement:

```
DEFINE PATH123 = PATH(P1,P2,P3)
```

This is a path that consists of points P1, P2, and P3. The points are defined in the manner described above using HERE or LEARN statements. A MOVE statement is used to drive the robot through the path.

```
MOVE PATH123
```

The speed of the robot is controlled by defining either a relative velocity or an absolute velocity. The following statement represents the case of relative velocity definition:

```
SPEED 75
```

When this statement appears within the program, it is typically interpreted to mean that the manipulator should operate at 75% of the initially commanded velocity in the statements that follow in the program. The initial speed is given in a command that precedes the execution of the robot program. For example,

```
SPEED 0.5 MPS
EXECUTE PROGRAM1
```

indicates that the program named PROGRAM1 is to be executed by the robot at a speed of 0.5 m/sec.

Interlock and Sensor Commands. The two basic interlock commands (Section 5.3.2) used for industrial robots are WAIT and SIGNAL. The WAIT command is used to implement an input interlock. For example,

```
WAIT 20, ON
```

would cause program execution to stop at this statement until the input signal coming into the robot controller at port 20 was in an "on" condition. This might be used in a situation where the robot needed to wait for the completion of an automatic machine cycle in a loading and unloading application.

The SIGNAL statement is used to implement an output interlock. This is used to communicate to some external piece of equipment. For example,

SIGNAL 21, ON

would switch on the signal at output port 21, perhaps to actuate the start of an automatic machine cycle.

The above interlock commands represent situations where the execution of the statement occurs at the point in the program where the statement appears. There are other situations in which it is desirable for an external device to be continuously monitored for any change that might occur. This would be useful, for example, in safety monitoring where a sensor is set up to detect the presence of humans who might wander into the robot's work volume. The sensor reacts to the presence of the humans by signaling the robot controller. The following type of statement might be used for this case:

REACT 25, SAFESTOP

This command would be written to continuously monitor input port 25 for any changes in the incoming signal. If and when a change in the signal occurs, regular program execution is interrupted, and control is transferred to a subroutine called SAFESTOP. This subroutine would stop the robot from further motion and/or cause some other safety action to be taken.

Although end effectors are attached to the wrist of the manipulator, they are actuated very much like external devices. Special commands are usually written for controlling the end effector. In the case of grippers, the basic commands are

OPEN

and

CLOSE

which cause the gripper to actuate to fully open and fully closed positions, respectively, where fully closed is the position for grasping the object in the application. Greater control over the gripper is available in some sensored and servo-controlled hands. For grippers with force sensors that can be regulated through the robot controller, a command such as

CLOSE 2.0 N

controls the closing of the gripper until a 2.0-N force is encountered by the gripper fingers. A similar command used to close the gripper to a given opening width is

CLOSE 25 MM

A special set of statements is often required to control the operation of tool-type end effectors, such as spot welding guns, arc welding tools, spray painting guns, and powered spindles (e.g., for drilling or grinding). Spot welding and spray painting controls are typically simple binary commands (e.g., open/close and on/off), and these commands would be similar to those used for gripper control. In the case of arc welding and powered spindles, a greater variety of control statements is needed to control feed rates and other parameters of the operation.

Computations and Program Logic. Many robot languages possess capabilities for performing computations and data processing operations that are similar to

computer programming languages. Most robot applications do not require a high level of computational power. As the sophistication of applications increases in the future, the computing and data processing duties of the controller will also increase for functions such as calculating complex motion paths, decision making, and integrating with other computer systems.

Many of today's robot applications require the use of branches and subroutines in the program. Statements such as

GO TO 150

and

IF (logical expression) GO TO 150

cause the program to branch to some other statement in the program (e.g., to statement number 150 in the above illustrations).

A subroutine in a robot program is a group of statements that are to be executed separately when called from the main program. In a preceding example, the subroutine SAFESTOP was named in the REACT statement for use in safety monitoring. Other uses of subroutines include making calculations or performing repetitive motion sequences at a number of different places in the program. Using a subroutine is more efficient than writing the same steps several places in the program.

8.5.3 Simulation and Off-Line Programming

The trouble with leadthrough methods and textual programming techniques is that the robot must be taken out of production for a certain length of time to accomplish the programming. Off-line programming permits the robot program to be prepared at a remote computer terminal and downloaded to the robot controller for execution without interrupting production. In true off-line programming, there is no need to physically locate the positions in the workspace for the robot as required with present textual programming languages. Some form of graphical computer simulation is required to validate the programs developed off-line, similar to the off-line procedures used in NC part programming.

The off-line programming procedures that are commercially available use graphical simulation to construct a three-dimensional model of the robot cell for evaluation and off-line programming. The cell might consist of the robot, machine tools, conveyors, and other hardware. The simulator displays these cell components on the graphics monitor and shows the robot performing its work cycle in animated computer graphics. After the program has been developed using the simulation procedure, it is then converted into the textual language corresponding to the particular robot employed in the cell. This is a step in off-line robot programming that is equivalent to post-processing in NC part programming.

In off-line programming, some adjustment must be performed to account for geometric differences between the three-dimensional model in the computer and the actual physical cell. For example, the position of a machine tool chuck in the physical layout might be slightly different than in the model used to do the off-line programming. For the robot to reliably load and unload the machine, it must have an accurate location of the load/unload point recorded in its control memory. A calibration procedure is used to correct the three-dimensional computer model by substituting actual location data from the cell for the approximate values developed in the original model. The disadvantage of calibrating the cell is that some production time is lost in performing this procedure.

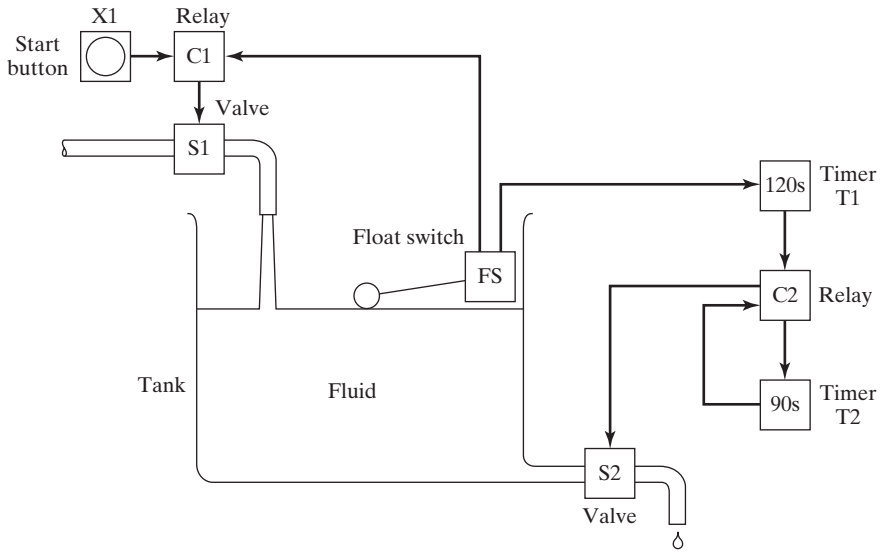


Figure 9.10 Fluid filling operation of Example 9.6.

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 initiates timer T2, which waits 90 sec to allow the contents of the tank to be drained, and (2) it energizes solenoid S2, which opens a valve to allow the fluid to flow out of the tank. At the end of the 90 sec, the timer breaks the current and de-energizes solenoid S2, thus closing the out-flow valve. Depressing the start button X1 resets the timers and opens their respective contacts. Construct the ladder logic diagram for the system.

Solution: The ladder logic diagram is constructed as shown in Figure 9.5.

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 9.6, 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 programmable logic controllers.

9.3 PROGRAMMABLE LOGIC CONTROLLERS

A programmable logic controller (PLC) can be defined as a microcomputer-based controller that uses stored instructions in programmable memory to implement logic, sequencing, timing, counting, and arithmetic functions through digital or analog input/output (I/O) modules, for controlling machines and processes. PLC applications are found in both the process industries and discrete manufacturing. Examples of applications in process industries include chemical processing, paper mill operations, and

food production. PLCs are primarily associated with discrete manufacturing industries to control individual machines, machine cells, transfer lines, material handling equipment, and automated storage systems. Before the PLC was introduced around 1970, hardwired controllers composed of relays, coils, counters, timers, and similar components were used to implement this type of industrial control (Historical Note 9.1). Today, many older pieces of equipment have been retrofitted with PLCs to replace the original hardwired controllers, often making the equipment more productive and reliable than when it was new.

Historical Note 9.1 Programmable Logic Controllers [2], [6], [8], [9].

In the mid-1960s, Richard Morley was a partner in Bedford Associates, a New England consulting firm specializing in control systems for machine tool companies. Most of the firm's work involved replacing relays with minicomputers in machine tool controls. In January 1968, Morley devised the notion and wrote the specifications for the first programmable controller.[2] It would overcome some of the limitations of conventional computers used for process control at the time; namely, it would be a real-time processor (Section 5.3.1), it would be predictable and reliable, and it would be modular and rugged. Programming would be based on ladder logic, which was widely used for industrial controls. The controller that emerged was named the Modicon Model 084. MODICON was an abbreviation of MOdular DIgital CONtroller. Model 084 was derived from the fact that this was the 84th product developed by Bedford Associates. Morley and his associates elected to start up a new company to produce the controllers, and Modicon was incorporated in October 1968. In 1977, Modicon was sold to Gould and became Gould's PLC division.

In the same year that Morley invented the PLC, the Hydramatic Division of General Motors Corporation developed a set of specifications for a PLC. The specifications were motivated by the high cost and lack of flexibility of electromechanical relay-based controllers used extensively in the automotive industry to control transfer lines and other mechanized and automated systems. The requirements for the device were that it must (1) be programmable and reprogrammable, (2) be designed to operate in an industrial environment, (3) accept 120 V AC signals from standard push-buttons and limit switches, (4) have outputs designed to switch and continuously operate loads such as motors and relays of 2-A rating, and (5) have a price and installation cost competitive with relay and solid-state logic devices then in use. In addition to Modicon, several other companies saw a commercial opportunity in the GM specifications and developed various versions of the PLC.

Capabilities of the first PLCs were similar to those of the relay controls they replaced. They were limited to on/off control. Within five years, product enhancements included better operator interfaces, arithmetic capability, data manipulation, and computer communications. Improvements over the next five years included larger memory, analog and positioning control, and remote I/O (permitting remote devices to be connected to a satellite I/O subsystem that was multiplexed to the PLC using twisted pair). Much of the progress was based on advancements taking place in microprocessor technology. By the mid-1980s, the micro PLC had been introduced. This was a down-sized PLC with much lower size and cost (typical size = 75 mm by 75 mm by 125 mm, and typical cost less than $500). By the mid-1990s, the nano PLC had arrived, which was even smaller and less expensive.

[2]Morley used the abbreviation PC to refer to the programmable controller. This term was used for many years until IBM began to call its personal computers by the same abbreviation in the early 1980s. The term *PLC*, widely used today for programmable logic controller, was coined by Allen-Bradley, a leading PLC supplier.

There are significant advantages to using a PLC rather than conventional relays, timers, counters, and other hardwired control components. These advantages include (1) programming the PLC is easier than wiring the relay control panel; (2) the PLC can be reprogrammed, whereas conventional controls must be rewired and are often scrapped instead; (3) PLCs take less floor space than relay control panels; (4) reliability is greater, and maintenance is easier; (5) the PLC can be connected to computer systems more easily than relays; and (6) PLCs can perform a greater variety of control functions than relay-based controls.

This section describes the components, operation, and programming of the PLC. Although its principal applications are in logic and sequence control, many PLCs also perform additional functions (Section 9.4).

9.3.1 Components of the PLC

A schematic diagram of a PLC is presented in Figure 9.11. The basic components are the following: (1) processor, (2) memory unit, (3) power supply, and (4) I/O module. These components are housed in a suitable cabinet designed for the industrial environment. In addition, there is (5) a programming device that can be disconnected from the PLC when not required.

The processor is the central processing unit (CPU) of the PLC. It executes the various logic and sequence control functions by operating on the PLC inputs to determine the appropriate output signals. The typical CPU operating cycle is described in Section 9.3.2. The CPU consists of one or more microprocessors similar to those used in personal computers and other data processing equipment. The difference is that they have a real-time operating system and are designed to facilitate I/O transactions and execute ladder logic functions. In addition, PLCs are built so that the CPU and other electronic components will operate in the electrically noisy environment of the factory.

Connected to the CPU is the memory unit, which contains the programs of logic, sequencing, and I/O operations. It also holds data files associated with these programs, including I/O status bits, counter and timer constants, and other variable and parameter values. The memory unit is referred to as the user memory because its contents are

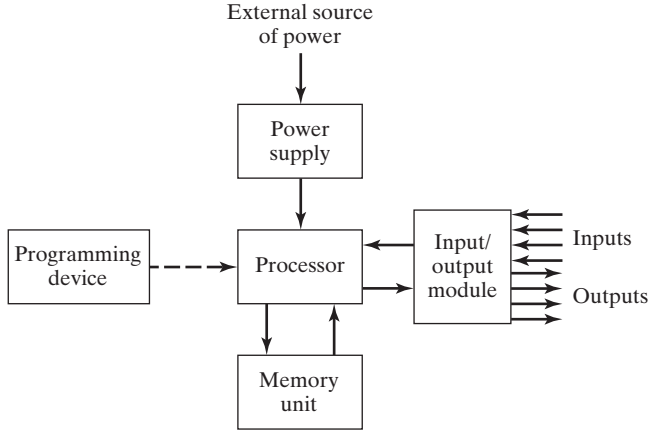


Figure 9.11 Components of a PLC.

entered by the user. In addition, the processor also contains the operating system memory, which directs the execution of the control program and coordinates I/O operations. The operating system is entered by the PLC manufacturer and cannot be accessed or altered by the user.

A power supply of 120 VAC is typically used to drive the PLC (some PLCs operate on 240 VAC). The power supply converts the 120 VAC into direct current (DC) voltages of ± 5 V. These low voltages are used to operate equipment that may have much higher voltage and power ratings than the PLC itself. The power supply often includes a battery backup that switches on automatically in the event of an external power source failure.

The input/output module provides the connections to the industrial equipment or process that is to be controlled. Inputs to the controller are signals from limit switches, push-buttons, sensors, and other on/off devices. Outputs from the controller are on/off signals to operate motors, valves, and other devices required to actuate the process. In addition, many PLCs are capable of accepting continuous signals from analog sensors and generating signals suitable for analog actuators. The size of a PLC is usually rated in terms of the number of its I/O terminals.

The PLC is programmed by means of a programming device, which is usually detachable from the PLC cabinet so that it can be shared among multiple controllers. Different PLC manufacturers provide different devices, ranging from simple teach-pendant type devices, similar to those used in robotics, to special PLC programming keyboards and displays. Personal computers can also be used to program PLCs. A PC used for this purpose sometimes remains connected to the PLC to serve a process monitoring or supervisory function and for conventional data processing applications related to the process.

9.3.2 PLC Operating Cycle

As far as the PLC user is concerned, the steps in the control program are executed simultaneously and continuously. In truth, a certain amount of time is required for the PLC processor to execute the user program during one cycle of operation. The typical operating cycle of the PLC, called a *scan*, consists of three parts: (1) input scan, (2) program scan, and (3) output scan. During the input scan, the inputs to the PLC are read by the processor and the status of each input is stored in memory. Next, the control program is executed during the program scan. The input values stored in memory are used in the control logic calculations to determine the values of the outputs. Finally, during the output scan, the outputs are updated to agree with the calculated values. The time to perform the scan is called the *scan time*, and this time depends on the number of inputs that must be read, the complexity of control functions to be performed, and the number of outputs that must be changed. Typical scan times are measured in milliseconds [20].

One of the potential problems that can occur during the scan cycle is that the value of an input can change immediately after it has been sampled. Since the program uses the input value stored in memory, any output values that are dependent on that input are determined incorrectly. There is obviously a potential risk involved in this mode of operation. However, the risk is minimized because the time between updates is so short that it is unlikely that the output value being incorrect for such a short time will have a serious effect on process operation. The risk becomes most significant in processes in which the response times are very fast and where hazards can occur during the scan time. Some PLCs have special features for making "immediate" updates of output signals when input variables are known to cycle back and forth at frequencies faster than the scan time.

9.3.3 Programming the PLC

Programming is the means by which the user enters the control instructions to the PLC through the programming device. The most basic control instructions consist of switching, logic, sequencing, counting, and timing. Virtually all PLC programming methods provide instruction sets that include these functions. Many control applications require additional instructions to accomplish analog control of continuous processes, complex control logic, data processing and reporting, and other advanced functions not readily performed by the basic instruction set. Owing to these differences in requirements, various PLC programming languages have been developed. A standard for PLC programming was published by the International Electrotechnical Commission in 1992, entitled *International Standard for Programmable Controllers* (IEC 61131–3). This standard specifies three graphical languages and two text-based languages for programming PLCs, respectively: (1) ladder logic diagrams, (2) function block diagrams, (3) sequential functions charts, (4) instruction list, and (5) structured text. Table 9.6 lists the five languages along with the most suitable application of each. IEC 61131–3 also states that the five languages must be able to interact with each other to allow for all possible levels of control sophistication in any given application.

Ladder Logic Diagram. The most widely used PLC programming language today involves ladder diagrams (LDs), examples of which are shown in several previous figures. As indicated in Section 9.2, ladder diagrams are very convenient for shop personnel who are familiar with ladder and circuit diagrams but may not be familiar with computers and computer programming. To use ladder logic diagrams, they do not need to learn an entirely new programming language.

Direct entry of the ladder logic diagram into the PLC memory requires the use of a keyboard and monitor with graphics capability to display symbols representing the components and their interrelationships in the ladder logic diagram. The symbols are similar to those presented in Figure 9.6. The PLC keyboard is often designed with keys for each of the individual symbols. Programming is accomplished by inserting the appropriate components into the rungs of the ladder diagram. The components are of two basic types: contacts and coils. Contacts represent input switches, relay contacts, and similar components. Coils represent loads such as motors, solenoids, relays, timers, and counters. In effect, the programmer inputs the ladder logic circuit diagram rung by rung into the PLC memory with the monitor displaying the results for verification.

TABLE 9.6 Features of the Five PLC Languages Specified in the IEC 61131–3 Standard

Language	Abbreviation	Type	Applications Best Suited for
Ladder logic diagram	(LD)	Graphical	Discrete control
Function block diagram	(FBD)	Graphical	Continuous control
Sequential function chart	(SFC)	Graphical	Sequence control
Instruction list	(IL)	Textual	Discrete control
Structured text	(ST)	Textual	Complex logic, computations, etc.

Function Block Diagrams. A function block diagram (FBD) provides a means of inputting high-level instructions. Instructions are composed of operational blocks. Each block has one or more inputs and one or more outputs. Within a block, certain operations take place on the inputs to transform the signals into the desired outputs. The function blocks include operations such as timers and counters, control computations using equations (e.g., proportional-integral-derivative control), data manipulation, and data transfer to other computer-based systems. Function blocks are described in Hughes [4].

Sequential Function Charts. The sequential function chart (SFC, also called the *Grafcet* method) graphically displays the sequential functions of an automated system as a series of steps and transitions from one state of the system to the next. The sequential function chart is described in Boucher [1]. It has become a standard method for documenting logic control and sequencing in much of Europe. However, its use in the United States is more limited, and the reader is referred to the cited reference for more details on the method.

Instruction List. Instruction list (IL) programming also provides a way of entering the ladder logic diagram into PLC memory. In this method, the programmer uses a low-level computer language to construct the ladder logic diagram by entering statements that specify the various components and their relationships for each rung of the ladder diagram. This approach can be demonstrated by introducing a hypothetical PLC instruction set, which is a composite of various manufacturers' languages. It contains fewer features than most commercially available PLCs. The programming device typically consists of a special keyboard for entering the individual components on each rung of the ladder logic diagram. A monitor capable of displaying each ladder rung (and perhaps several rungs that precede it) is useful to verify the program. The instruction set for the PLC is presented in Table 9.7 with a concise explanation of each instruction.

TABLE 9.7 Typical Low-Level Language Instruction Set for a PLC

STR	Store a new input and start a new rung of the ladder.
AND	Logical AND referenced with the previously entered component. This is interpreted as a series circuit relative to the previously entered component.
OR	Logical OR referenced with the previously entered component. This is interpreted as a parallel circuit relative to the previously entered component.
NOT	Logical NOT or inverse of the previously entered component.
OUT	Output component for the rung of the ladder diagram.
TMR	Timer component. Requires one input signal to initiate timing sequence. Output is delayed relative to input by a duration specified by the programmer in seconds. Resetting the timer is accomplished by interrupting (stopping) the input signal.
CTR	Counter component. Requires two inputs: One is the incoming pulse train that is counted by the CTR component, the other is the reset signal indicating a restart of the counting procedure.

EXAMPLE 9.7 Language Commands for AND, OR, and NOT Circuits

Using the command set in Table 9.7, write the PLC programs for the three ladder diagrams from Figure 9.7, depicting the logic gates (a) AND, (b) OR, and (c) NOT.

Solution: Commands for the control relay are listed below, with explanatory comments.

Command		Comment
(a) AND	STR X1	Store input X1
	AND X2	Input X2 in series with X1
	OUT Y	Output Y
(b) OR	STR X1	Store input X1
	OR X2	Input X2 parallel with X1
	OUT Y	Output Y
(c) NOT	STR NOT X1	Store inverse of X1
	OUT Y	Output Y

EXAMPLE 9.8 Language Commands for Control Relay

Using the command set in Table 9.7, write the PLC program for the control relay depicted in the ladder logic diagram of Figure 9.9.

Solution: Commands for the control relay are listed below, with explanatory comments.

Command	Comment
STR X	Store input X
OUT C	Output contact relay C
STR NOT C	Store inverse of C output
OUT Y1	Output load Y1
STR C	Store C output
OUT Y2	Output load Y2

The low-level languages are generally limited to the kinds of logic and sequence control functions that can be defined in a ladder logic diagram. Although timers and counters have not been illustrated in the two preceding examples, some of the exercise problems at the end of the chapter require the reader to use them.

Structured Text. Structured text (ST) is a high-level computer-type language likely to become more common in the future to program PLCs and PCs for automation and control applications. The principal advantage of a high-level language is its capability to perform data processing and calculations on values other than binary. Ladder diagrams and low-level PLC languages are usually quite limited in their ability to operate

on signals that are other than on/off types. The capability to perform data processing and computation permits the use of more complex control algorithms, communication with other computer-based systems, display of data on a monitor, and input of data by a human operator. Another advantage is the relative ease with which a complicated control program can be interpreted by a user. Explanatory comments can be inserted into the program to facilitate interpretation.

9.4 PERSONAL COMPUTERS AND PROGRAMMABLE AUTOMATION CONTROLLERS

Programmable logic controllers were originally designed to execute the logic and sequence control functions described in Section 9.1, and these are the functions for which the PLC is best suited. However, modern industrial control applications have evolved to include requirements in addition to logic and sequence control. These additional requirements include the following:

- *Analog control.* Proportional-integral-derivative (PID) control is available on some programmable controllers for regulation of continuous variables like temperature and force. These control algorithms have traditionally been implemented using analog controllers. Today the analog control schemes are approximated using the digital computer, with either a PLC or a computer process controller.
- *Motion and servomotor control.* This function is basically the same task as performed by the machine control unit in computer numerical control. Many industrial applications require precise control of motor actuators, and PLCs are being used to implement this type of control.
- *Arithmetic functions.* Use of functions such as basic addition, subtraction, multiplication, and division permit more complex control algorithms to be developed than what is possible with conventional logic and sequence control or PID control.
- *Matrix functions.* The capability to perform matrix operations on stored values can be used to compare the actual values of a set of inputs and outputs with the values stored in memory to determine if some error has occurred.
- *Data processing and reporting.* These functions are typically associated with business applications of personal computers. Controller manufacturers have found it necessary to include these PC capabilities in some of their controller products.
- *Network connectivity and enterprise data integration.* Again, the company's business computer systems are usually associated with these functions, which have evolved to include the need for factory data to be integrated into the corporate information system.

This section considers two approaches that address the need for these additional requirements for industrial control: (1) personal computers and (2) programmable automation controllers.

9.4.1 Personal Computers for Industrial Control

In the early 1990s, personal computers began to be used in industrial control applications normally reserved for programmable logic controllers. PLCs were traditionally favored for use in factories because they were designed to operate in harsh environments, while PCs

were designed for office environments. In addition, with their built-in input/output (I/O) interfaces and real-time operating systems, PLCs could be readily connected to external equipment for process control, whereas PCs required special I/O cards and programs to enable such functions. Finally, personal computers sometimes locked up for no apparent cause, and usually lockups cannot be tolerated in industrial control applications. PLCs are not prone to such malfunctions.

These PLC advantages notwithstanding, the technological evolution of programmable logic controllers has not kept pace with the development of personal computers, new generations of which are introduced with much greater frequency than PLCs. There is much more proprietary software and architecture in PLCs than in PCs, making it difficult to mix and match components from different vendors. Programming a PLC is usually accomplished using ladder logic, but there are differences in the ladder logic coding from different PLC manufacturers, and ladder logic has its limitations for most of the industrial control requirements listed in the introduction to this section. Over time, these factors have resulted in a performance disadvantage for PLCs. PC speeds are typically doubling every 18 months or so, while improvements in PLC technology occur much more slowly and require that individual companies redesign their proprietary software and architectures for each new generation of microprocessors.

PCs can now be purchased in more sturdy enclosures for the dirty and noisy plant environment. They can be equipped with membrane-type keyboards for protection against factory moisture, oil, and dirt, as well as electrical noise. They can be ordered with I/O cards and related hardware to provide the necessary devices to connect to the plants' equipment and processes. Operating systems designed to implement real-time control applications can be installed in addition to traditional office software. And the traditional advantages of personal computers, such as graphics capability for the human-machine interface (HMI), data logging and storage for maintaining process records, and easier network connectivity, are becoming more and more relevant in industrial control applications. In addition, PCs can be readily integrated with peripheral devices such as scanners and printers for automatic identification and data capture (Chapter 12).

There are two basic approaches used in PC-based control systems [10]: soft logic and hard real-time control. In the soft logic configuration, the PC's operating system is Windows, and control algorithms are installed as high-priority programs under the operating system. However, it is possible to interrupt the control tasks in order to service certain system functions in Windows, such as network communications and disk access. When this happens, the control function is delayed, with possible negative consequences to the process. Thus, a soft logic control system cannot be considered a real-time controller in the sense of a PLC. In high-speed control applications or volatile processes, lack of real-time control is a potential hazard. In less critical processes, soft logic works well.

In a hard real-time control system, the PC operates like a PLC, using a real-time operating system, and the control software takes priority over all other software. Windows tasks are executed at a lower priority under the real-time operating system. Windows cannot interrupt the execution of the real-time controller. If Windows locks up, it does not affect the controller operation. Also, the real-time operating system resides in the PC's active memory, so a failure of the hard disk has no effect in a hard real-time control system.

PART III
Material Handling and Identification

Chapter 10

Material Transport Systems

CHAPTER CONTENTS

10.1 Overview of Material Handling
  10.1.1 Material Handling Equipment
  10.1.2 Design Considerations in Material Handling
10.2 Material Transport Equipment
  10.2.1 Industrial Trucks
  10.2.2 Automated Guided Vehicles
  10.2.3 Rail-Guided Vehicles
  10.2.4 Conveyors
  10.2.5 Cranes and Hoists
10.3 Analysis of Material Transport Systems
  10.3.1 Analysis of Vehicle-Based Systems
  10.3.2 Conveyor Analysis

Material handling is defined by the Material Handling Industry of America[1] as "the movement, protection, storage and control of materials and products throughout the process of manufacture and distribution, consumption and disposal" [22]. The handling of materials must be performed safely, efficiently, at low cost, in a timely manner, accurately (the right materials in the right quantities to the right locations), and without damage to the materials. Material handling is an important yet often overlooked issue in production. The cost of material handling is a significant portion of total production cost, estimates

[1]The Material Handling Industry of America (MHIA) is the trade association for material handling companies that do business in North America.

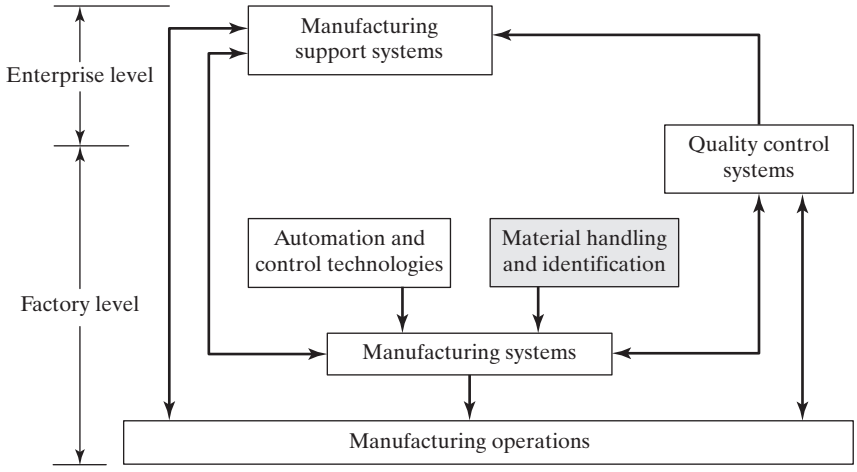


Figure 10.1 Material handling and identification in the production system.

averaging around 20–25% of total manufacturing labor cost in the United States [3]. This proportion varies, depending on type of production and degree of automation in material handling.

This part of the book covers the material handling and identification systems used in production. The position of material handling in the larger production system is shown in Figure 10.1. The coverage is divided into three major categories: (1) material transport systems, discussed in the present chapter, (2) storage systems (Chapter 11), and (3) automatic identification and data capture (Chapter 12). In addition, several material handling devices are discussed in other chapters of the text, including industrial robots (Chapter 8), pallet shuttles in NC machining centers (Chapter 14), conveyors in manual assembly lines (Chapter 15), transfer mechanisms in automated transfer lines (Chapter 16), and parts feeding devices in automated assembly (Chapter 17).

10.1 OVERVIEW OF MATERIAL HANDLING

Material handling is one of the activities in the larger distribution system by which materials, parts, and products are moved, stored, and tracked in the world's commercial infrastructure. The term commonly used for the larger system is *logistics*, which is concerned with the acquisition, movement, storage, and distribution of materials and products, as well as the planning and control of these operations in order to satisfy customer demand. Logistics operations can be divided into two basic categories: external logistics and internal logistics. ***External logistics*** is concerned with transportation and related activities that occur outside of a facility. In general, these activities involve the movement of materials between different geographical locations. The five traditional modes of transportation are rail, truck, air, ship, and pipeline. ***Internal logistics***, more popularly known as material handling, involves the movement and storage of materials inside a given facility. The interest in this book is on internal logistics. This section describes the various types of equipment used in material handling, and then identifies several considerations in the design of material handling systems.

10.1.1 Material Handling Equipment

A great variety of material handling equipment is available commercially. The equipment can be classified into five categories [21]: (1) transport equipment, (2) positioning equipment, (3) unit load formation equipment, (4) storage equipment, and (5) identification and control equipment.

Transport Equipment. Material transport equipment is used to move materials inside a factory, warehouse, or other facility. The five main types of equipment are (1) industrial trucks, (2) automated guided vehicles, (3) rail-guided vehicles, (4) conveyors, and (5) hoists and cranes. These equipment types are described in Section 10.2.

Positioning Equipment. This category consists of equipment used to handle parts and other materials at a single location: for example, loading and unloading parts from a production machine in a work cell. Positioning is accomplished by industrial robots that perform material handling (Section 8.4.1) and parts feeders in automated assembly (Section 17.1.2). Hoists used at a single location can also be included in this category. The general role of positioning at a workstation is discussed in Section 13.1.2.

Unit Load Formation Equipment. The term *unitizing equipment* refers to (1) containers used to hold individual items during handling and (2) equipment used to load and package the containers. Containers include pallets, tote pans, boxes, baskets, barrels, and drums, some of which are shown in Figure 10.2. Although seemingly mundane, containers are very important for moving materials efficiently as a unit load, rather than as individual items. Pallets and other containers that can be handled by forklift equipment are widely used in production and distribution operations. Most factories, warehouses, and distribution centers use forklift trucks to move unit loads on pallets. A given facility must often standardize on a specific type and size of container if it utilizes automatic transport and/or storage equipment to handle the loads.

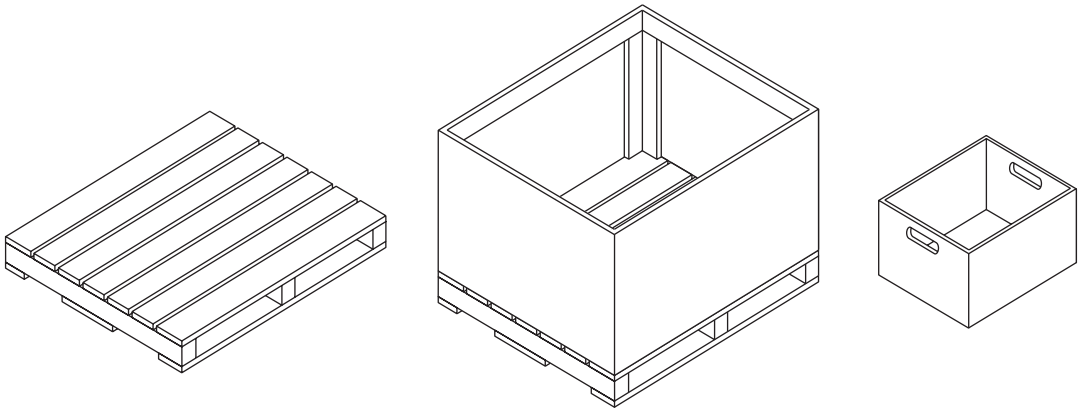


Figure 10.2 Examples of unit load containers for material handling: (a) wooden pallet, (b) pallet box, and (c) tote box.

The second category of unitizing equipment includes ***palletizers***, which are designed to automatically load cartons onto pallets and shrink-wrap plastic film around them for shipping, and ***depalletizers***, which are designed to unload cartons from pallets. Other wrapping and packaging machines are also included in this equipment category.

Storage Equipment. Although it is generally desirable to reduce the storage of materials in manufacturing, it seems unavoidable that raw materials and work-in-process spend some time in storage, even if only temporarily. And finished products are likely to spend time in a warehouse or distribution center before being delivered to the final customer. Accordingly, companies must give consideration to the most appropriate methods for storing materials and products prior to, during, and after manufacture.

Storage methods and equipment can be classified into two major categories: (1) conventional storage methods and (2) automated storage systems. Conventional storage methods include bulk storage (storing items in an open floor area), rack systems (for pallets), shelving and bins, and drawer storage. In general, conventional storage methods are labor-intensive. Human workers put materials into storage and retrieve them from storage. Automated storage systems are designed to reduce or eliminate the manual labor involved in these functions. Both conventional and automated storage methods are described in detail in Chapter 11. Mathematical models are developed to predict throughput and other performance characteristics of automated storage systems.

Identification and Control Equipment. The scope of material handling includes keeping track of the materials being moved and stored. This is usually done by affixing some kind of label to the item, carton, or unit load that uniquely identifies it. The most common label used today is a bar code that can be read quickly and automatically by bar code readers. This is the same basic technology used by grocery stores and retail merchandisers. An alternative identification technology that is growing in importance is RFID (for radio frequency identification). Bar codes, RFID, and other automatic identification techniques are discussed in Chapter 12.

10.1.2 Design Considerations in Material Handling

Material handling equipment is usually assembled into a system. The system must be specified and configured to satisfy the requirements of a particular application. Design of the system depends on the materials to be handled, quantities and distances to be moved, type of production facility served by the handling system, and other factors, including available budget. This section considers these factors that influence the design of the material handling system.

Material Characteristics. For handling purposes, materials can be classified by the physical characteristics presented in Table 10.1, suggested by a classification scheme of Muther and Haganas [15]. Design of the material handling system must take these factors into account. For example, if the material is a liquid that is to be moved over long distances in great volumes, then a pipeline is the appropriate transport means. But this handling method would be infeasible for moving a liquid contained in barrels or other containers. Materials in a factory usually consist of solid items: raw materials, parts, and finished or semifinished products.

TABLE 10.1 Characteristics of Materials in Material Handling

Category	Measures or Descriptors
Physical state	Solid, liquid, or gas
Size	Volume, length, width, height
Weight	Weight per piece, weight per unit volume
Shape	Long and flat, round, square, etc.
Condition	Hot, cold, wet, dirty, sticky
Risk of damage	Fragile, brittle, sturdy
Safety risk	Explosive, flammable, toxic, corrosive, etc.

Flow Rate, Routing, and Scheduling. In addition to material characteristics, other factors must be considered in determining which type of equipment is most appropriate for the application. These other factors include (1) quantities and flow rates of materials to be moved, (2) routing factors, and (3) scheduling of the moves.

The amount or quantity of material to be moved affects the type of handling system that should be installed. If large quantities of material must be handled, then a dedicated handling system is appropriate. If the quantity of a particular material type is small but there are many different material types to be moved, then the handling system must be designed to be shared by the various materials moved. The amount of material moved must be considered in the context of time, that is, how much material is moved within a given time period. The amount of material moved per unit time is referred to as the *flow rate*. Depending on the form of the material, flow rate is measured in pieces per hour, pallet loads per hour, tons per hour, or similar units. Whether the material must be moved as individual units, in batches, or continuously has an effect on the selection of handling method.

Routing factors include pickup and drop-off locations, move distances, routing variations, and conditions that exist along the routes. Given that other factors remain constant, handling cost is directly related to the distance of the move: The longer the move distance, the greater the cost. Routing variations occur because different materials follow different flow patterns in the factory or warehouse. If these differences exist, the material handling system must be flexible enough to deal with them. Conditions along the route include floor surface condition, traffic congestion, whether a portion of the move is outdoors, whether the path is straight line or involves turns and changes in elevation, and the presence or absence of people along the path. All of these factors affect the design of the material transport system.

Scheduling relates to the timing of each individual delivery. In production as well as in many other material handling applications, the material must be picked up and delivered promptly to its proper destination to maintain peak performance and efficiency of the overall system. To the extent required by the application, the handling system must be responsive to this need for timely pickup and delivery of the items. Rush jobs increase material handling cost. Scheduling urgency is often mitigated by providing space for buffer stocks of materials at pickup and drop-off points. This allows a "float" of materials to exist in the system, thus reducing the pressure on the handling system for immediate response to a delivery request.

Plant Layout. The material handling system is an important factor in plant layout design. When a new facility is being planned, the handling system should be considered

part of the layout. In this way, there is greater opportunity to create a layout that optimizes material flow in the building and utilizes the most appropriate type of handling system. In the case of an existing facility, there are more constraints on the design of the handling system. The present arrangement of departments and equipment in the building often limits the attainment of optimum flow patterns.

Section 2.3 describes the conventional plant layouts used in manufacturing: (1) process layout, (2) product layout, and (3) fixed-position layout. Different material handling systems are generally required for the three layout types. Table 10.2 summarizes the characteristics of the three conventional layout types and the kinds of material handling equipment usually associated with each layout type.

In process layouts, a variety of parts and/or products are manufactured in small or medium batch sizes. The handling system must be flexible to deal with the variations. Considerable work-in-process is usually one of the characteristics of batch production, and the material handling system must be capable of accommodating this inventory. Hand trucks and forklift trucks (for moving pallet loads of parts) are commonly used in process layouts. Factory applications of automated guided vehicle systems are growing because they represent a versatile means of handling the different load configurations in medium and low volume production. Work-in-progress is often stored on the factory floor near the next scheduled machines. More systematic ways of managing in-process inventory include automated storage systems (Section 11.3).

A product layout involves production of a standard or nearly identical types of product in relatively high quantities. Final assembly plants for cars, trucks, and appliances are usually designed as product layouts. The transport system that moves the product is typically characterized as fixed route, mechanized, and capable of large flow rates. It sometimes serves as a storage area for work-in-process to reduce effects of downtime between production areas along the line of product flow. Conveyor systems are common in product layouts. Delivery of component parts to the various assembly workstations along the flow path is accomplished by trucks and similar unit load vehicles.

Finally, in a fixed-position layout, the product is large and heavy and therefore remains in a single location during most of its fabrication. Heavy components and subassemblies must be moved to the product. Handling systems used for these moves in fixed-position layouts are large and often mobile. Cranes, hoists, and trucks are common in this situation.

Unit Load Principle. The Unit Load Principle stands as an important and widely applied principle in material handling. A ***unit load*** is simply the mass that is to be moved

TABLE 10.2 Types of Material Handling Equipment Associated with Three Layout Types

Layout Type	Characteristics	Typical Material Handling Equipment
Process	Variations in product and processing, low and medium production rates	Hand trucks, forklift trucks, automated guided vehicle systems
Product	Limited product variety, high production rate	Conveyors for product flow, industrial trucks and automated guided vehicles to deliver components to stations
Fixed-position	Large product size, low production rate	Cranes, hoists, industrial trucks

TABLE 10.3 Standard Pallet Sizes Commonly Used in Factories and Warehouses

Depth = *x* Dimension	Width = *y* Dimension
800 mm (32 in)	1,000 mm (40 in)
900 mm (36 in)	1,200 mm (48 in)
1,000 mm (40 in)	1,200 mm (48 in)
1,060 mm (42 in)	1,060 mm (42 in)
1,200 mm (48 in)	1,200 mm (48 in)

Sources: [6], [16].

or otherwise handled at one time. The unit load may consist of only one part, a container loaded with multiple parts, or a pallet loaded with multiple containers of parts. In general, the unit load should be designed to be as large as is practical for the material handling system that will move or store it, subject to considerations of safety, convenience, and access to the materials making up the unit load. This principle is widely applied in the truck, rail, and ship industries. Palletized unit loads are collected into truck loads, which then become larger unit loads themselves. Then these truck loads are aggregated once again on freight trains or ships, in effect becoming even larger unit loads.

There are good reasons for using unit loads in material handling [16]: (1) multiple items can be handled simultaneously, (2) the required number of trips is reduced, (3) loading and unloading times are reduced, and (4) product damage is decreased. Using unit loads results in lower cost and higher operating efficiency.

Included in the definition of unit load is the container that holds or supports the materials to be moved. To the extent possible, these containers are standardized in size and configuration to be compatible with the material handling system. Examples of containers used to form unit loads in material handling are illustrated in Figure 10.2. Of the available containers, pallets are probably the most widely used, owing to their versatility, low cost, and compatibility with various types of material handling equipment. Most factories and warehouses use forklift trucks to move materials on pallets. Table 10.3 lists some of the most popular standard pallet sizes in use today. These standard pallet sizes are used in the analysis of automated storage/retrieval systems in Chapter 11.

10.2 MATERIAL TRANSPORT EQUIPMENT

This section covers the five categories of material transport equipment commonly used to move parts and other materials in manufacturing and warehouse facilities: (1) industrial trucks, manual and powered; (2) automated guided vehicles; (3) rail-guided vehicles; (4) conveyors; and (5) cranes and hoists. Table 10.4 summarizes the principal features and kinds of applications for each equipment category. Section 10.3 considers quantitative techniques by which material transport systems consisting of this equipment can be analyzed.

10.2.1 Industrial Trucks

Industrial trucks are divided into two categories: nonpowered and powered. The nonpowered types are often referred to as hand trucks because they are pushed or pulled by human workers. Quantities of material moved and distances traveled are relatively

TABLE 10.4 Summary of Features and Applications of Five Categories of Material Handling Equipment

Handling Equipment	Features	Typical Applications
Industrial trucks, manual	Low cost Low rate of deliveries	Moving light loads in a factory
Industrial trucks, powered	Medium cost	Movement of pallet loads and palletized containers in a factory or warehouse
Automated guided vehicle systems	High cost Battery-powered vehicles Flexible routing Non-obstructive pathways	Moving pallet loads in factory or warehouse Moving work-in-process along variable routes in low and medium production
Rail-guided vehicles	High cost Flexible routing On-the-floor or overhead types	Moving assemblies, products, or pallet loads along variable routes in factory or warehouse Moving large quantities of items over fixed routes in a factory or warehouse
Conveyors, powered	Great variety of equipment In-floor, on-the-floor, or overhead Mechanical power to move loads resides in pathway	Sortation of items in a distribution center Moving products along a manual assembly line
Cranes and hoists	Lift capacities of more than 100 tons possible	Moving large, heavy items in factories, mills, warehouses, etc.

low when this type of equipment is used to transport materials. Hand trucks are classified as either two-wheel or multiple-wheel. Two-wheel hand trucks, Figure 10.3(a), are generally easier to manipulate by the worker but are limited to lighter loads. Multiple-wheeled hand trucks are available in several types and sizes. Two common types are dollies and pallet trucks. Dollies are simple frames or platforms as shown in Figure 10.3(b). Various wheel configurations are possible, including fixed wheels and caster-type wheels. Pallet trucks, shown in Figure 10.3(c), have two forks that can be inserted through the openings in a pallet. A lift mechanism is actuated by the worker to lift and lower the pallet off the ground using small diameter wheels near the end of the forks. In operation, the worker inserts the forks into the pallet, elevates the load, pulls the truck to its destination, lowers the pallet, and removes the forks.

Powered trucks are self-propelled and guided by a worker. Three common types are used in factories and warehouses: (a) walkie trucks, (b) forklift rider trucks, and

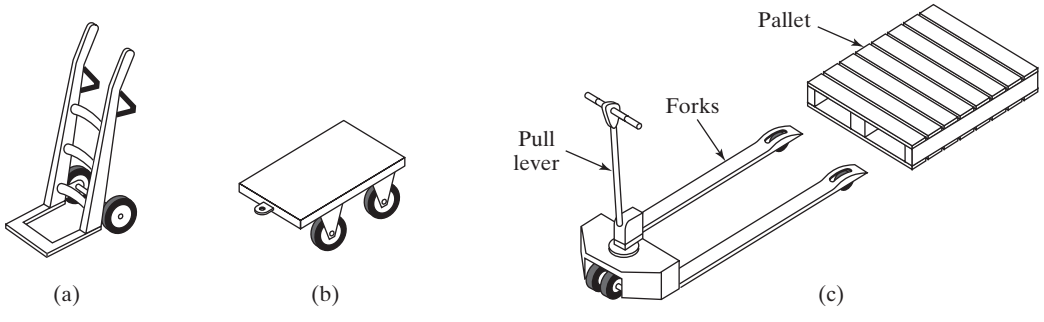


Figure 10.3 Examples of nonpowered industrial trucks (hand trucks): (a) two-wheel hand truck, (b) four-wheel dolly, and (c) hand-operated low-lift pallet truck.

(c) towing tractors. Walkie trucks, Figure 10.4(a), are battery-powered vehicles equipped with wheeled forks for insertion into pallet openings but with no provision for a worker to ride on the vehicle. The truck is steered by a worker using a control handle at the front of the vehicle. The forward speed of a walkie truck is limited to around 3 mi/hr (5 km/hr), about the normal walking speed of a human.

Forklift rider trucks, Figure 10.4(b), are distinguished from walkie trucks by the presence of a cab for the worker to sit in and drive the vehicle. Forklift trucks range in load carrying capacity from about 450 kg (1,000 lb) up to more than 4,500 kg (10,000 lb). Forklift trucks have been modified to suit various applications. Some trucks have high reach capacities for accessing pallet loads on high rack systems, while others are capable of operating in the narrow aisles of high-density storage racks. Power sources for forklift trucks are either internal combustion engines (gasoline, liquefied petroleum gas, or compressed natural gas) or electric motors (using on-board batteries).

Industrial towing tractors, Figure 10.4(c), are designed to pull one or more trailing carts over the relatively smooth surfaces found in factories and warehouses. They are generally used for moving large amounts of materials between major collection and distribution areas. The runs between origination and destination points are usually fairly long. Power is supplied either by electric motor (battery-powered) or internal combustion engine. Tow tractors also find significant applications in air transport operations for moving baggage and air freight in airports.

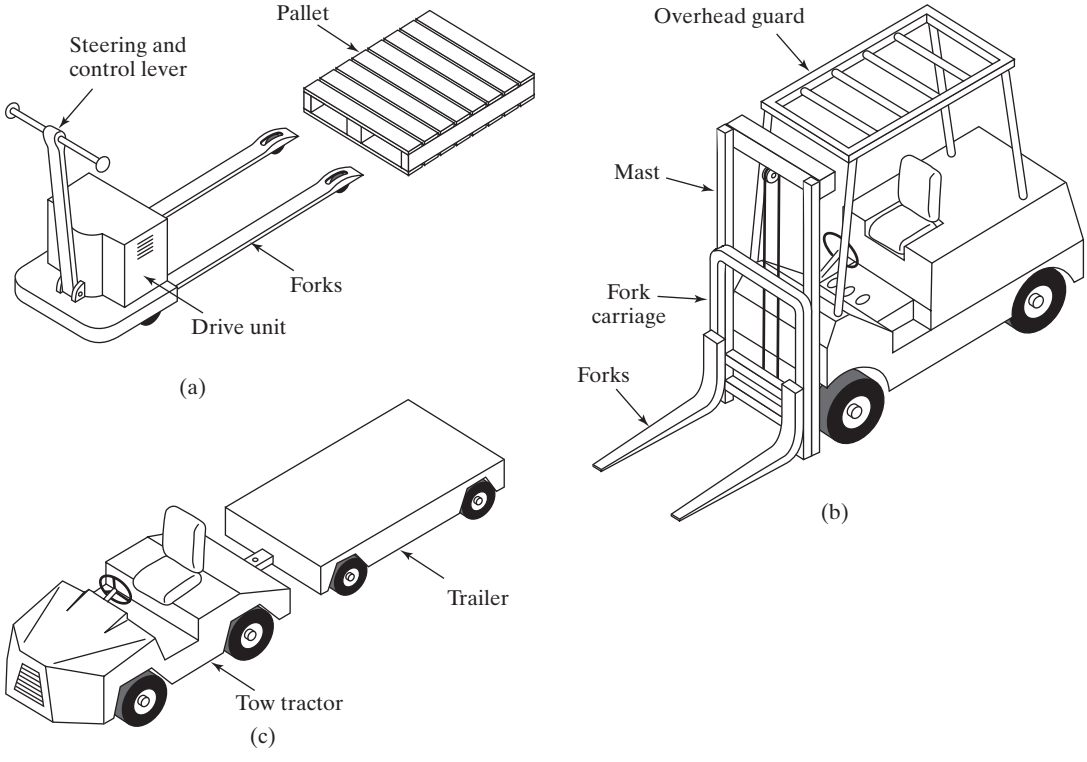


Figure 10.4 Three principal types of powered trucks: (a) walkie truck, (b) forklift truck, and (c) towing tractor.

10.2.2 Automated Guided Vehicles

An automated guided vehicle system (AGVS) is a material handling system that uses independently operated, self-propelled vehicles guided along defined pathways.[2] The AGVs are powered by on-board batteries that allow many hours of operation (8–16 hr is typical) before needing to be recharged. A distinguishing feature of an AGVS, compared to rail-guided vehicle systems and most conveyor systems, is that the pathways are unobtrusive.

Types of Vehicles. Automated guided vehicles can be divided into the following categories: (1) towing vehicles for driverless trains, (2) pallet trucks, and (3) unit load carriers, illustrated in Figure 10.5. A driverless train consists of a towing vehicle (the AGV) pulling one or more trailers to form a train, as in Figure 10.5(a). It was the first type of AGVS to be introduced and is still widely used today. A common application is moving heavy payloads over long distances in warehouses or factories with or without intermediate pickup and drop-off points along the route. For trains consisting of 5–10 trailers, this is an efficient transport system.

Automated guided pallet trucks, Figure 10.5(b), are used to move palletized loads along predetermined routes. In the typical application the vehicle is backed into the

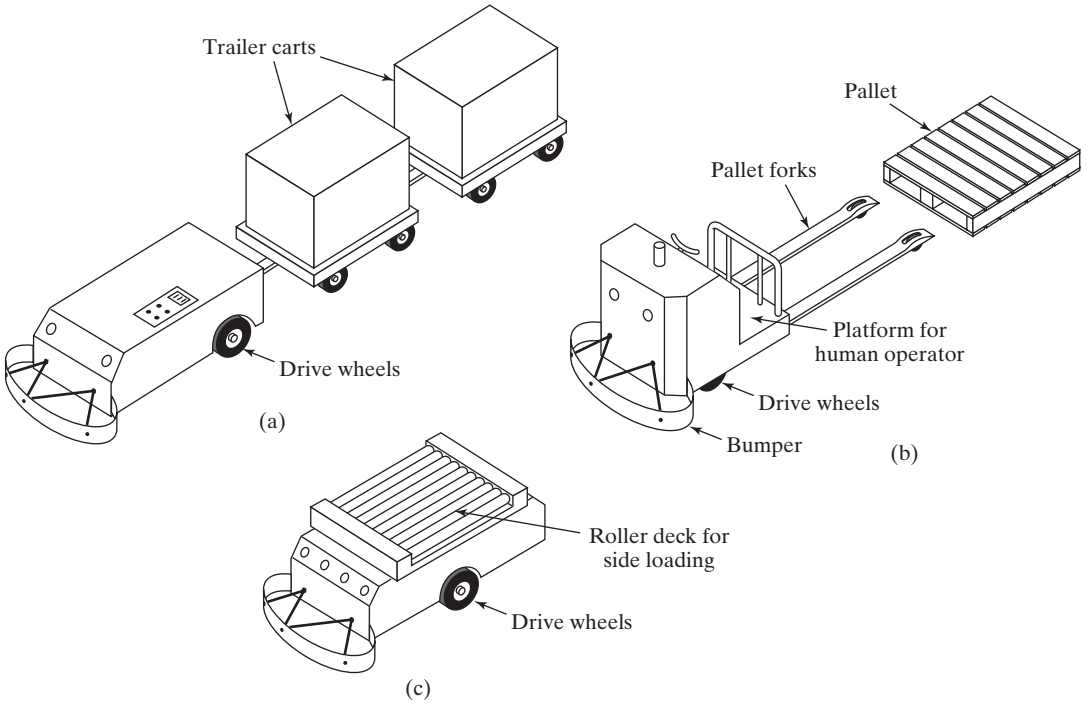


Figure 10.5 Three types of automated guided vehicles: (a) driverless automated guided train, (b) AGV pallet truck, and (c) unit load carrier.

[2]The term *automated guided cart* (AGC) is used by some AGV vendors to identify vehicles that are smaller and lighter weight than conventional AGVs and are available at significantly lower prices.

loaded pallet by a human worker who steers the truck and uses its forks to elevate the load slightly. Then the worker drives the pallet truck to the guide path and programs its destination, and the vehicle proceeds automatically to the destination for unloading. The load capacity of an AGVS pallet truck ranges up to several thousand kilograms, and some trucks are capable of handling two pallets rather than one. A special type of pallet truck is the forklift AGV, which uses forks that are similar to those of a forklift truck to engage pallets. This vehicle can achieve significant vertical movement of its forks to reach loads on racks and shelves.

AGV unit load carriers are used to move unit loads from one station to another. They are often equipped for automatic loading and unloading of pallets or tote pans by means of powered rollers, moving belts, mechanized lift platforms, or other devices built into the vehicle deck. A typical unit load AGV is illustrated in Figure 10.5(c).

Variations of unit load carriers include light load AGVs, assembly line AGVs, and heavy-duty AGVs. The light load AGV is a relatively small vehicle with corresponding light load capacity (typically 250 kg or less). It does not require the same large aisle width as a conventional AGV. Light load guided vehicles are designed to move small loads (single parts, small baskets, or tote pans of parts) through plants of limited size engaged in light manufacturing. An assembly line AGV is designed to carry a partially completed subassembly through a sequence of assembly workstations to build the product. Heavy-duty unit load AGVs are used for loads up to 125 tons [20]. Applications include moving large paper rolls in printing companies, heavy steel coils in stamping plants, and cargo containers in seaport docking operations.

AGVS Applications. In general, an AGVS is appropriate when different materials are moved from various load points to various unload points. The principal AGVS applications in production and logistics are (1) driverless train operations, (2) storage and distribution, (3) assembly line applications, and (4) flexible manufacturing systems. Driverless train operations have already been described; they involve the movement of large quantities of material over relatively long distances.

The second application area is storage and distribution operations. Unit load carriers and pallet trucks are typically used in these applications, which involve movement of material in unit loads. The applications often interface the AGVS with some other automated handling or storage system, such as an automated storage/retrieval system (AS/RS) in a distribution center. The AGVS delivers incoming unit loads contained on pallets from the receiving dock to the AS/RS, which places the items into storage, and the AS/RS retrieves individual pallet loads from storage and transfers them to vehicles for delivery to the shipping dock. Storage/distribution operations also include light manufacturing and assembly plants in which work-in-process is stored in a central storage area and distributed to individual workstations for processing. Electronics assembly is an example of these kinds of applications. Components are "kitted" at the storage area and delivered in tote pans or trays to the assembly workstations in the plant. Light load AGVs are the appropriate vehicles in these applications.

AGV systems are used in assembly line operations, based on a trend that began in Europe in the automotive industry. Unit load carriers and light load guided vehicles are used in these lines. Station-to-station movement of car bodies and engines in final assembly plants is a typical application. In these situations, the AGVs remain with the work units during assembly, rather than serving a pickup and drop-off function.

Another application area for AGVS technology is flexible manufacturing systems (FMSs, Chapter 19). In the typical operation, starting work parts are placed onto pallet

fixtures by human workers in a staging area, and the AGVs deliver the parts to the individual workstations in the system. When the AGV arrives at the station, the pallet is transferred from the vehicle platform to the station (such as the worktable of a machine tool) for processing. At the completion of processing, a vehicle returns to pick up the work and transport it to the next assigned station. An AGVS provides a versatile material handling system to complement the flexibility of the FMS.

AGVS technology is still developing, and the industry is continually working to design new systems to respond to new application requirements. An interesting example that combines two technologies involves the use of a robotic manipulator mounted on an automated guided vehicle to provide a mobile robot for performing complex handling tasks at various locations in a plant.

Vehicle Guidance Technologies. The guidance system is the method by which AGVS pathways are defined and vehicles are controlled to follow the pathways. The technologies used in commercial AGV systems for vehicle guidance include (1) imbedded guide wires, (2) paint strips, (3) magnetic tape, (4) laser-guided vehicles (LGVs), and (5) inertial navigation.

In the imbedded guide wire method, electrical wires are placed in a shallow channel that has been cut into the surface of the floor. After the guide wire is installed, the channel is filled with cement to eliminate the discontinuity in the floor surface. The guide wire is connected to a frequency generator, which emits a low-voltage, low-frequency signal in the range 1–15 kHz. This induces a magnetic field along the pathway that can be followed by sensors on board each vehicle. The operation of a typical system is illustrated in Figure 10.6. Two sensors are mounted on the vehicle on either side of the guide wire. When the vehicle is located such that the guide wire is directly between the two coils, the intensity of the magnetic field measured by each sensor is equal. If the vehicle strays to one side or the other, or if the guide wire path changes direction, the magnetic field intensity at the two sensors becomes unequal. This difference is used to control the steering motor, which makes the required changes in vehicle direction to equalize the two sensor signals, thereby tracking the guide wire.

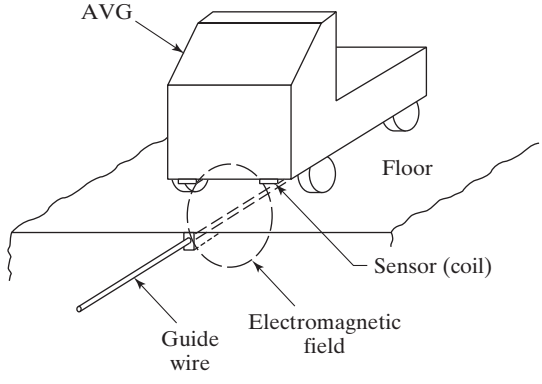


Figure 10.6 Operation of the on-board sensor system that uses two coils to track the magnetic field in the guide wire.

A typical AGVS layout contains multiple loops, branches, side tracks, and spurs, as well as pickup and drop-off stations. The most appropriate route must be selected from the alternative pathways available to a vehicle as it moves to a specified destination in the system. When a vehicle approaches a branching point where the guide path forks into two (or more) pathways, the vehicle must have a means of deciding which path to take. The two principal methods of making this decision in commercial wire-guided systems are (1) the frequency select method and (2) the path switch select method. In the frequency select method, the guide wires leading into the two separate paths at the switch have different frequencies. As the vehicle enters the switch, it reads an identification code on the floor to determine its location. Depending on its programmed destination, the vehicle selects the correct guide path by following only one of the frequencies. This method requires a separate frequency generator for each different frequency used in the guide-path layout.

The path switch select method operates with a single frequency throughout the guide-path layout. To control the path of a vehicle at a switch, the power is turned off in all other branches except the one that the vehicle is to travel on. To accomplish routing by the path switch select method, the guide-path layout is divided into blocks that are electrically insulated from each other. The blocks can be turned on and off either by the vehicles themselves or by a central control computer.

When paint strips are used to define the pathway, the vehicle uses an optical sensor capable of tracking the paint. The strips can be taped, sprayed, or painted on the floor. One system uses a 1-in-wide paint strip containing fluorescent particles that reflect an ultraviolet (UV) light source from the vehicle. The on-board sensor detects the reflected light in the strip and controls the steering mechanism to follow it. Paint strip guidance is useful in environments where electrical noise renders the guide wire system unreliable or when the installation of guide wires in the floor surface is not practical. One problem with this guidance method is that the paint strip deteriorates with time. It must be kept clean and periodically replaced.

Magnetic tape is installed on the floor surface to define the pathways. It avoids the cutting of the floor surface that is required when imbedded guide wires are used. It also allows the pathways to be conveniently redefined as the needs of the facility change over time. Unlike imbedded wire guidance, which emits an active powered signal, magnetic tape is a passive guidance technology.

Unlike the previous guidance methods, laser-guided vehicles (LGVs) operate without continuously defined pathways. Instead, they use a combination of dead reckoning and reflective beacons located throughout the plant that can be identified by on-board laser scanners. ***Dead reckoning*** refers to the capability of a vehicle to follow a given route in the absence of a defined pathway in the floor. Movement of the vehicle along the route is accomplished by computing the required number of wheel rotations in a sequence of specified steering angles. The computations are performed by the vehicle's on-board computer. As one would expect, positioning accuracy of dead reckoning decreases over long distances. Accordingly, the location of the laser-guided vehicle must be periodically verified by comparing the calculated position with one or more known positions. These known positions are established using the reflective beacons located strategically throughout the plant on columns, walls, and machines. These beacons can be sensed by the laser scanner on the vehicle. Based on the positions of the beacons, the on-board navigation computer uses triangulation to update the positions calculated by dead reckoning.

It should be noted that dead reckoning can be used by AGV systems that are normally guided by in-floor guide wires, paint strips, or magnetic tape. This capability allows

the vehicle to cross steel plates in the factory floor where guide wires cannot be installed or to depart from the guide path for positioning at a load/unload station. At the completion of the dead reckoning maneuver, the vehicle is programmed to return to the guide path to resume normal guidance control.

Inertial navigation, also known as *inertial guidance*, involves the use of on-board gyroscopes and/or other motion sensors to determine the position of the vehicle by detecting changes in its speed and acceleration. It is the same basic navigation technology used for guided missiles, aircraft, and submarines. When used in AGVS installations, magnetic transponders imbedded in the floor along the desired pathway are detected by the AGV to correct any errors in its position.

The advantage of laser-guided vehicle technology and inertial navigation over fixed pathways (guide wires, paint strips, and magnetic tape) is its flexibility. The LGV pathways are defined in software. The path network can be changed by entering the required data into the navigation computer. New docking points can be defined. The pathway network can be expanded by installing new beacons. These changes can be made quickly and without major alterations to the facility.

Vehicle Management. For the AGVS to operate efficiently, the vehicles must be well managed. Delivery tasks must be allocated to vehicles to minimize waiting times at load/unload stations. Traffic congestion in the guide-path network must be minimized. Two aspects of vehicle management are considered here: (1) traffic control and (2) vehicle dispatching.

The purpose of traffic control in an automated guided vehicle system is to minimize interference between vehicles and to prevent collisions. Two methods of traffic control used in commercial AGV systems are (1) on-board vehicle sensing and (2) zone control. The two techniques are often used in combination. On-board vehicle sensing, also called *forward sensing*, uses one or more sensors on each vehicle to detect the presence of other vehicles and obstacles ahead on the guide path. Sensor technologies include optical and ultrasonic devices. When the on-board sensor detects an obstacle in front of it, the vehicle stops. When the obstacle is removed, the vehicle proceeds. If the sensor system is 100% effective, collisions between vehicles are avoided. The effectiveness of forward sensing is limited by the capability of the sensor to detect obstacles that are in front of it on the guide path. These systems are most effective on straight pathways. They are less effective at turns and convergence points where forward vehicles may not be directly in front of the sensor.

In zone control, the AGVS layout is divided into separate zones, and the operating rule is that no vehicle is permitted to enter a zone that is already occupied by another vehicle. The length of a zone is at least sufficient to hold one vehicle plus allowances for safety and other considerations. Other considerations include number of vehicles in the system, size and complexity of the layout, and the objective of minimizing the number of separate zones. For these reasons, the zones are normally much longer than a vehicle length. Zone control is illustrated in Figure 10.7 in its simplest form. When one vehicle occupies a given zone, any trailing vehicle is not allowed to enter that zone. The leading vehicle must proceed into the next zone before the trailing vehicle can occupy the current zone. When the forward movement of vehicles in the separate zones is controlled, collisions are prevented, and traffic in the overall system is controlled. One method to implement zone control is to use a central computer, which monitors the location of each vehicle and attempts to optimize the movement of all vehicles in the system.

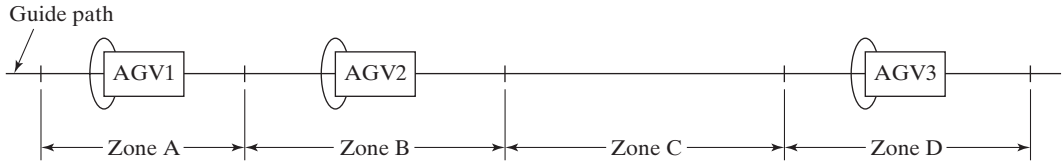


Figure 10.7 Zone control to implement blocking system. Zones A, B, and D are blocked. Zone C is free. Vehicle 2 is blocked from entering Zone A by Vehicle 1. Vehicle 3 is free to enter Zone C.

For an AGVS to serve its function, vehicles must be dispatched in a timely and efficient manner to the points in the system where they are needed. Several methods are used in AGV systems to dispatch vehicles: (1) on-board control panels, (2) remote call stations, and (3) central computer control. These dispatching methods are generally used in combination to maximize responsiveness and efficiency.

Each guided vehicle is equipped with an on-board control panel for the purpose of manual vehicle control, vehicle programming, and other functions. Most commercial vehicles can be dispatched by means of this control panel to a given station in the AGVS layout. Dispatching with an on-board control panel provides the AGVS with flexibility and timeliness to cope with changes and variations in delivery requirements.

Remote call stations represent another method for an AGVS to satisfy delivery requirements. The simplest call station is a push-button mounted at the load/unload station. This transmits a hailing signal for any available vehicle in the neighborhood to dock at the station and either pick up or drop off a load. The on-board control panel might then be used to dispatch the vehicle to the desired destination point.

In a large factory or warehouse involving a high degree of automation, the AGVS servicing the facility must also be highly automated to achieve efficient operation of the entire production-storage-handling system. Central computer control is used to dispatch vehicles according to a preplanned schedule of pickups and deliveries in the layout and/or in response to calls from the various load/unload stations. In this dispatching method, the central computer issues commands to the vehicles in the system concerning their destinations and the operations they must perform. To perform the dispatching function, the central computer must have current information on the location of each vehicle so it can make appropriate decisions about which vehicles to dispatch to what locations. Hence, the vehicles must continually communicate their whereabouts to the central controller. Radio frequency (RF) is commonly used to achieve the required communication links.

Vehicle Safety. The safety of humans located along the pathway is an important objective in AGVS operations. An inherent safety feature of an AGV is that its traveling speed is slower than the normal walking pace of a human. This minimizes the danger that it will overtake a human walking along the path in front of the vehicle.

In addition, AGVs are usually provided with several other features specifically for safety reasons. A safety feature included in most guidance systems is automatic stopping of the vehicle if it strays more than a short distance, typically 50–150 mm (2–6 in), from the guide path; the distance is referred to as the vehicle's *acquisition distance*. This automatic stopping feature prevents a vehicle from running wild in the building. Alternatively, in the event that the vehicle is off the guide path (e.g., for loading), its sensor system is capable of locking onto the guide path when the vehicle is moved to within the acquisition distance.

Another safety device is an obstacle detection sensor located on each vehicle. This is the same on-board sensor used for traffic control. The sensor can detect obstacles along the path ahead, including humans. The vehicles are programmed either to stop when an obstacle is sensed ahead or to slow down. The reason for slowing down is that the sensed object may be located off to the side of the vehicle path or directly ahead but beyond a turn in the guide path, or the obstacle may be a person who will move out of the way as the AGV approaches. In any of these cases, the vehicle is permitted to proceed at a slower (safer) speed until it has passed the obstacle. The disadvantage of programming a vehicle to stop when it encounters an obstacle is that this delays the delivery and degrades system performance.

A safety device included on virtually all commercial AGVs is an emergency bumper. The bumpers are prominent in the illustrations shown in Figure 10.5. The bumper surrounds the front of the vehicle and protrudes ahead of it by a distance of 300 mm (12 in) or more. When the bumper makes contact with an object, the vehicle is programmed to brake immediately. Depending on the speed of the vehicle, its load, and other conditions, the distance the vehicle needs to come to a complete stop will vary from several inches to several feet. Most vehicles are programmed to require manual restarting after an obstacle has been encountered by the emergency bumper. Other safety devices on a typical vehicle include warning lights (blinking or rotating lights) and/or warning bells, which alert humans that the vehicle is present.

10.2.3 Rail-Guided Vehicles

The third category of material transport equipment consists of motorized vehicles that are guided by a fixed rail system. The rail system consists of either one rail, called a *monorail*, or two parallel rails. Monorails in factories and warehouses are typically suspended overhead from the ceiling. In rail-guided vehicle systems using parallel fixed rails, the tracks generally protrude up from the floor. In either case, the presence of a fixed rail pathway distinguishes these systems from automated guided vehicle systems. As with AGVs, the vehicles operate asynchronously and are driven by an on-board electric motor. Unlike AGVs, which are powered by their own on-board batteries, rail-guided vehicles pick up electrical power from an electrified rail (similar to an urban rapid transit rail system). This relieves the vehicle from periodic recharging of its battery; however, the electrified rail system introduces a safety hazard not present in an AGVS.

Routing variations are possible in rail-guided vehicle systems through the use of switches, turntables, and other specialized track sections. This permits different loads to travel different routes, similar to an AGVS. Rail-guided systems are generally considered to be more versatile than conveyor systems but less versatile than automated guided vehicle systems. One of the original applications of nonpowered monorails was in the meat-processing industry before 1900. The slaughtered animals were hung from meat hooks attached to overhead monorail trolleys. The trolleys were moved through the different departments of the plant manually by the workers. It is likely that Henry Ford got the idea for the assembly line from observing these meat packing operations. Today, the automotive industry makes considerable use of electrified overhead monorails to move large components and subassemblies in its manufacturing operations.

10.2.4 Conveyors

A conveyor is a mechanical apparatus for moving items or bulk materials, usually inside a facility. Conveyors are generally used when material must be moved in relatively large quantities between specific locations over a fixed path, which may be in the floor, above

the floor, or overhead. Conveyors are either powered or nonpowered. In powered conveyors, the power mechanism is contained in the fixed path, using chains, belts, rotating rolls, or other devices to propel loads along the path. Powered conveyors are commonly used in automated material transport systems in manufacturing plants, warehouses, and distribution centers. In nonpowered conveyors, materials are moved either manually by human workers who push the loads along the fixed path or by gravity from one elevation to a lower elevation.

Types of Conveyors. A variety of conveyor equipment is commercially available. The primary interest here is in powered conveyors. Most of the major types of powered conveyors, organized according to the type of mechanical power provided in the fixed path, are briefly described in the following:

- *Roller conveyors.* In roller conveyors, the pathway consists of a series of tubes (rollers) that are perpendicular to the direction of travel, as in Figure 10.8(a). Loads must possess a flat bottom surface of sufficient area to span several adjacent rollers. Pallets, tote pans, or cartons serve this purpose well. The rollers are contained in a fixed frame that elevates the pathway above floor level from several inches to several feet. The loads move forward as the rollers rotate. Roller conveyors can either be powered or nonpowered. Powered roller conveyors are driven by belts or chains. Nonpowered roller conveyors are often driven by gravity so that the pathway has a

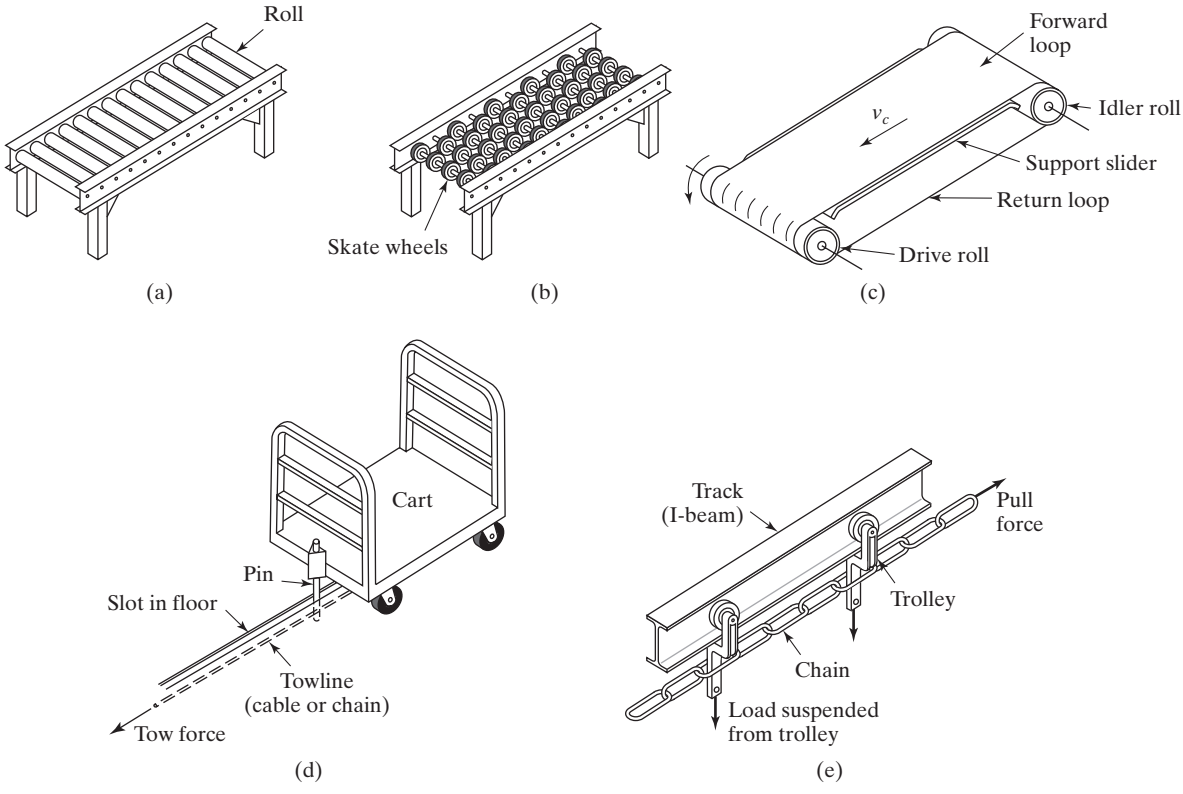


Figure 10.8 Types of Conveyors: (a) Roller conveyor, (b) skate-wheel conveyor, (c) belt (flat) conveyor (support frame not shown), (d) in-floor towline conveyor, and (e) overhead trolley conveyor.

downward slope sufficient to overcome rolling friction. Roller conveyors are used in a wide variety of applications, including manufacturing, assembly, packaging, sortation, and distribution.

- *Skate-wheel conveyors.* These are similar in operation to roller conveyors. Instead of rollers, they use skate wheels rotating on shafts connected to a frame to roll pallets, tote pans, or other containers along the pathway, as in Figure 10.8(b). Skate-wheel conveyors are lighter weight than roller conveyors. Applications of skate-wheel conveyors are similar to those of roller conveyors, except that the loads must generally be lighter since the contacts between the loads and the conveyor are much more concentrated. Because of their lightweight, skate-wheel conveyors are sometimes built as portable units that can be used for loading and unloading truck trailers at shipping and receiving docks at factories and warehouses.
- *Belt conveyors.* Belt conveyors consist of a continuous loop, with half its length used for delivering materials and the other half for the return run, as in Figure 10.8(c). The belt is made of reinforced elastomer (rubber), so that it possesses high flexibility but low extensibility. At one end of the conveyor is a drive roll that powers the belt. The flexible belt is supported by a frame that has rollers or support sliders along its forward loop. Belt conveyors are available in two common forms: (1) flat belts for pallets, cartons, and individual parts; and (2) troughed belts for bulk materials. Materials placed on the belt surface travel along the moving pathway. In the case of troughed belt conveyors, the rollers and supports give the flexible belt a V shape on the forward (delivery) loop to contain bulk materials such as coal, gravel, grain, or similar particulate materials.
- *Chain conveyors.* The typical equipment in this category consists of chain loops in an over-and-under configuration around powered sprockets at the ends of the pathway. The conveyor may consist of one or more chains operating in parallel. The chains travel along channels in the floor that provide support for the flexible chain sections. Either the chains slide along the channel or they ride on rollers in the channel. The loads are generally dragged along the pathway using bars that project up from the moving chain.
- *In-floor towline conveyor.* These conveyors use four-wheel carts powered by moving chains or cables located in trenches in the floor, as in Figure 10.8(d). The chain or cable is the towline. Pathways for the conveyor system are defined by the trench and towline, and the towline is driven as a powered pulley system. It is possible to switch between powered pathways to achieve flexibility in routing. The carts use steel pins that project below floor level into the trench to engage the chain for towing. (Gripper devices are substituted for pins when cable is used for the pulley system, as in the San Francisco trolleys.) The pin can be pulled out of the chain (or the gripper releases the cable) to disengage the cart for loading, unloading, switching, accumulating parts, and manually pushing a cart off the main pathway. Towline conveyor systems are used in manufacturing plants and warehouses.
- *Overhead trolley conveyor.* A ***trolley*** in material handling is a wheeled carriage running on an overhead rail from which loads can be suspended. An overhead trolley conveyor, Figure 10.8(e), consists of multiple trolleys, usually equally spaced along a fixed track. The trolleys are connected together and moved along the track by means of a chain or cable that forms a complete loop. Suspended from the trolleys are hooks, baskets, or other containers to carry loads. The chain (or cable) is attached to a drive pulley that pulls the chain at a constant velocity. The conveyor

path is determined by the configuration of the track system, which has turns and possible changes in elevation. Overhead trolley conveyors are often used in factories to move parts and assemblies between major production departments. They can be used for both delivery and storage.

- *Power-and-free overhead trolley conveyor.* This conveyor is similar to the overhead trolley conveyor, except that the trolleys can be disconnected from the drive chain, providing the conveyor with an asynchronous capability. This is usually accomplished by using two tracks, one just above the other. The upper track contains the continuously moving endless chain, and the trolleys that carry loads ride on the lower track. Each trolley includes a mechanism by which it can be connected to the drive chain and disconnected from it. When connected, the trolley is pulled along its track by the moving chain in the upper track. When disconnected, the trolley is idle.
- *Cart-on-track conveyor.* This equipment consists of individual carts riding on a track a few feet above floor level. The carts are driven by means of a spinning tube, as illustrated in Figure 10.9. A drive wheel, attached to the bottom of the cart and set at an angle to the rotating tube, rests against it and drives the cart forward. The cart speed is controlled by regulating the angle of contact between the drive wheel and the spinning tube. When the axis of the drive wheel is 45°, as in the figure, the cart is propelled forward. When the axis of the drive wheel is parallel to the tube, the cart does not move. Thus, control of the drive wheel angle on the cart allows power-and-free operation of the conveyor. One of the advantages of cart-on-track systems relative to many other conveyors is that the carts can be positioned with high accuracy. This permits their use for positioning work during production. Applications of cart-on-track systems include robotic spot welding lines in automobile body plants and mechanical assembly systems.
- *Other conveyor types.* Other powered conveyors include vibration-based systems and vertical lift conveyors. Screw conveyors are powered versions of the Archimedes screw, the water-raising device invented in ancient times, consisting of a large screw inside a tube, turned by hand to pump water uphill for irrigation purposes. Vibration-based conveyors use a flat track connected to an electromagnet that imparts an angular vibratory motion to the track to propel items in the desired direction. This same principle is used in vibratory bowl feeders to deliver components in automated assembly systems (Section 17.1.2). Vertical lift conveyors include a variety of mechanical elevators designed to provide vertical motion, such as between floors or to link floor-based conveyors with overhead conveyors. Other conveyor types include nonpowered chutes, ramps, and tubes, which are driven by gravity.

Conveyor Operations and Features. As indicated by the preceding discussion, conveyor equipment covers a wide variety of operations and features. The discussion here is limited to powered conveyors. Conveyor systems divide into two basic types in terms of the characteristic motion of the materials moved by the system: (1) continuous and (2) asynchronous. Continuous motion conveyors move at a constant velocity v_c along the path. They include belt, roller, skate-wheel, and overhead trolley.

Asynchronous conveyors operate with a stop-and-go motion in which loads move between stations and then stop and remain at the station until released. Asynchronous handling allows independent movement of each carrier in the system. Examples of this type include overhead power-and-free trolley, in-floor towline, and cart-on-track conveyors. Some roller and skate-wheel conveyors can also be operated asynchronously.

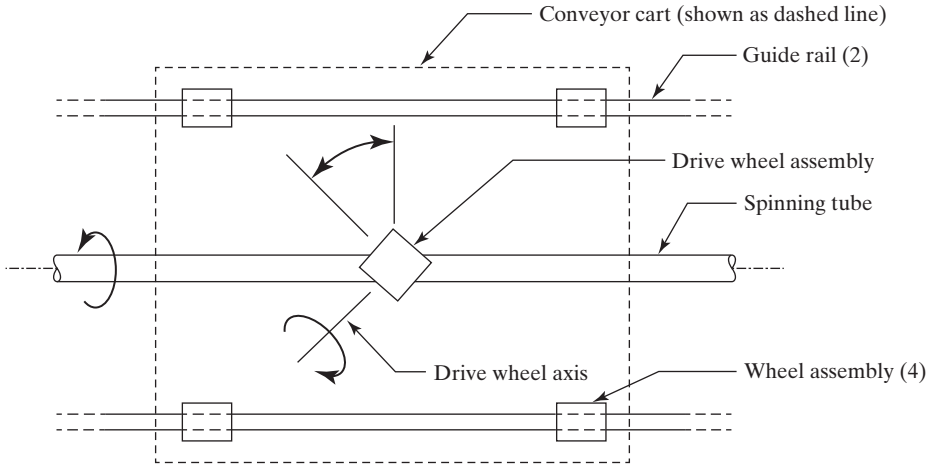


Figure 10.9 Cart-on-track conveyor. Drive wheel at 45° angle resting on spinning tube provides forward motion of cart.

Reasons for using asynchronous conveyors include (1) to accumulate loads, (2) to temporarily store items, (3) to allow for differences in production rates between adjacent processing areas, (4) to smooth production when cycle times are variable at stations along the conveyor, and (5) to accommodate different conveyor speeds along the pathway.

Conveyors can also be classified as (1) single direction, (2) continuous loop, and (3) recirculating. Section 10.3.2 presents equations and techniques for analyzing these conveyor systems. Single direction conveyors are used to transport loads one way from origination point to destination point, as depicted in Figure 10.10(a). These systems are appropriate when there is no need to move loads in both directions or to return containers or carriers from the unloading stations back to the loading stations. Single direction powered conveyors include roller, skate-wheel, belt, and chain-in-floor types. In addition, all gravity conveyors operate in one direction.

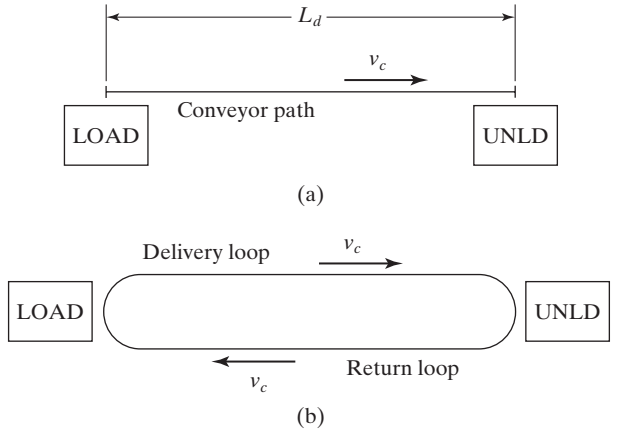


Figure 10.10 (a) Single direction conveyor and (b) continuous loop conveyor.

Continuous loop conveyors form a complete circuit, as in Figure 10.10(b). An overhead trolley conveyor is an example of this conveyor type. However, any conveyor type can be configured as a loop, even those previously identified as single direction conveyors, simply by connecting several single direction conveyor sections into a closed loop. Continuous loop conveyors are used when loads are moved in carriers (e.g., hooks, baskets) between load and unload stations and the carriers are affixed to the conveyor loop. In this design, the empty carriers are automatically returned from the unload station back to the load station.

The preceding description of a continuous loop conveyor assumes that items loaded at the load station(s) are unloaded at the unload station(s). There are no loads in the return loop; the purpose of the return loop is simply to send the empty carriers back for reloading. This method of operation overlooks an important opportunity offered by a closed-loop conveyor: to store as well as deliver items. Conveyor systems that allow parts or products to remain on the return loop for one or more revolutions are called *recirculating conveyors*. In providing a storage function, the conveyor system can be used to accumulate parts to smooth out effects of loading and unloading variations at stations in the conveyor. Two problems can plague the operation of a recirculating conveyor system. One is that there may be times during the operation of the conveyor when no empty carriers are immediately available at the loading station when needed. The other problem is that there may be times when no loaded carriers are immediately available at the unloading station when needed.

It is possible to construct branching and merging points into a conveyor track to permit different routing of different loads moving in the system. In nearly all conveyor systems, it is possible to build switches, shuttles, or other mechanisms to achieve these alternate routings. In some systems, a push-pull mechanism or lift-and-carry device is required to actively move the load from the current pathway onto the new pathway.

10.2.5 Cranes and Hoists

The fifth category of transport equipment in material handling is cranes and hoists. Cranes are used for horizontal movement of materials in a facility, and hoists are used for vertical lifting. A crane invariably includes a hoist; thus, the hoist component of the crane lifts the load, and the crane transports the load horizontally to the desired destination. This class of material handling equipment includes cranes capable of lifting and moving very heavy loads, in some cases over 100 tons.

A ***hoist*** is a mechanical device used to raise and lower loads. As seen in Figure 10.11, a hoist consists of one or more fixed pulleys, one or more moving pulleys, and a rope, cable, or chain strung between the pulleys. A hook or other means for attaching the load is connected to the moving pulley(s). The number of pulleys in the hoist determines its mechanical advantage, which is the ratio of the load weight to the driving force required to lift the weight. The mechanical advantage of the hoist in Figure 10.11 is 4. The driving force to operate the hoist is usually applied by electric or pneumatic motor.

Cranes include a variety of material handling equipment designed for lifting and moving heavy loads using one or more overhead beams for support. Principal types of cranes found in factories include (a) bridge cranes, (b) gantry cranes, and (c) jib cranes. In all three types, at least one hoist is mounted to a trolley that rides on the overhead beam of the crane. A bridge crane consists of one or two horizontal girders or beams suspended between fixed rails on either end which are connected to the structure of the building, as

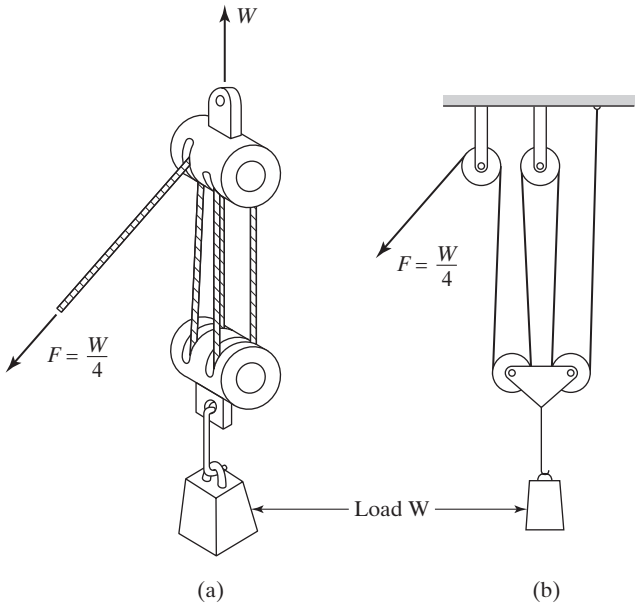


Figure 10.11 A hoist with a mechanical advantage of 4.0: (a) sketch of the hoist and (b) diagram to illustrate mechanical advantage.

shown in Figure 10.12(a). The hoist trolley can be moved along the length of the bridge, and the bridge can be moved the length of the rails in the building. These two drive capabilities provide motion in the x- and y-axes of the building, and the hoist provides motion in the z-axis direction. Thus, the bridge crane achieves vertical lifting due to its hoist and horizontal movement due to its orthogonal rail system. Large bridge cranes have girders that span up to 36.5 m (120 ft) and are capable of carrying loads up to 90,000 kg (100 tons). Large bridge cranes are controlled by operators riding in cabs on the bridge. Applications include heavy machinery fabrication, steel and other metal mills, and power-generating stations.

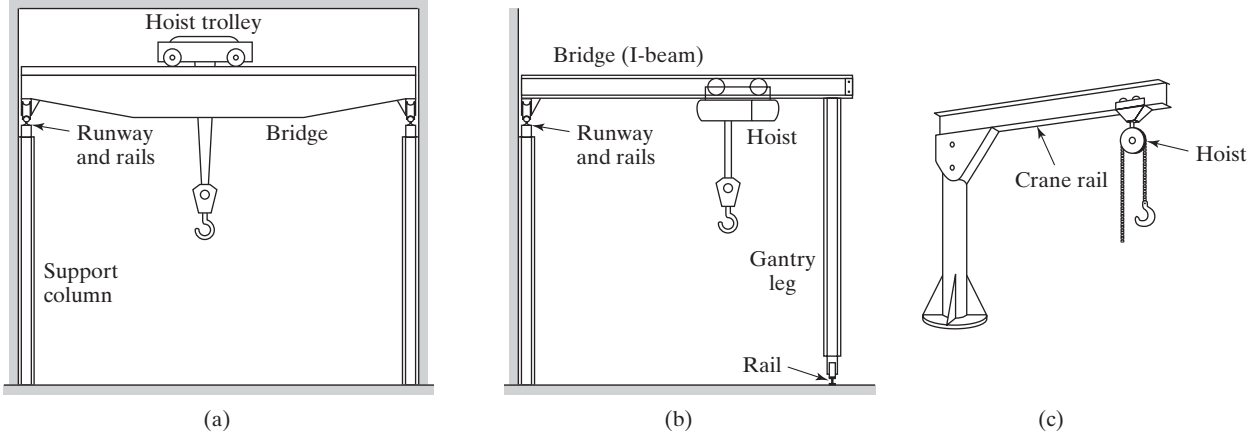


Figure 10.12 Three types of cranes: (a) bridge crane, (b) gantry crane (a half gantry crane is shown), and (c) jib crane.

A gantry crane is distinguished from a bridge crane by the presence of one or two vertical legs that support the horizontal bridge. As with the bridge crane, a gantry crane includes one or more hoists that accomplish vertical lifting. Gantries are available in a variety of sizes and capacities, the largest possessing spans of about 46 m (150 ft) and load capacities of 136,000 kg (150 tons). A double gantry crane has two legs. A half gantry crane, Figure 10.12(b), has a single leg on one end of the bridge, and the other end is supported by a rail mounted on the wall or other structural member of a building. A cantilever gantry crane has a bridge that extends beyond the span created by the support legs.

A jib crane consists of a hoist supported on a horizontal beam that is cantilevered from a vertical column or wall support, as in Figure 10.12(c). The horizontal beam pivots about the vertical axis formed by the column or wall to provide a horizontal sweep for the crane. The beam also serves as the track for the hoist trolley to provide radial travel along the length of the beam. Thus, the horizontal area included by a jib crane is circular or semicircular. As with other cranes, the hoist provides vertical lift-and-lower motions. Standard capacities of jib cranes range up to about 5,000 kg (11,000 lb). Wall-mounted jib cranes can achieve a swing of about 180°, while a floor-mounted jib crane using a column or post as its vertical support can sweep a full 360°.

10.3 ANALYSIS OF MATERIAL TRANSPORT SYSTEMS

Quantitative models are useful for analyzing material flow rates, delivery cycle times, and other aspects of system performance. The analysis may be useful in determining equipment requirements—for example, how many forklift trucks will be required to satisfy a specified flow rate. Material transport systems can be classified as vehicle-based systems or conveyor systems.[3] The following coverage of the quantitative models is organized along these lines.

10.3.1 Analysis of Vehicle-Based Systems

Equipment used in vehicle-based material transport systems includes industrial trucks (both hand trucks and powered trucks), automated guided vehicles, rail-guided vehicles, and certain types of conveyor systems. These systems are commonly used to deliver individual loads between origination and destination points. Two graphical tools that are useful for displaying and analyzing data in these deliveries are the from-to chart and the network diagram. The ***from-to chart*** is a table that can be used to indicate material flow data and/or distances between multiple locations. Table 10.5 illustrates a from-to chart that lists flow rates and distances between five workstations in a manufacturing system. The left-hand vertical column lists the origination points (loading stations), while the horizontal row at the top identifies the destination locations (unloading stations).

Network diagrams can also be used to indicate the same type of information. A ***network diagram*** consists of nodes and arrows, and the arrows indicate relationships among the nodes. In material handling, the nodes represent locations (e.g., load and unload stations), and the arrows represent material flows and/or distances between the stations. Figure 10.13 shows a network diagram containing the same information as Table 10.5.

[3]Exceptions exist. Some conveyor systems use vehicles to carry loads. Examples include the in-floor tow line conveyor and the cart-on-track conveyor.

TABLE 10.5 From-To Chart Showing Flow Rates, loads/hr (Value Before the Slash), and Travel Distances, m (Value After the Slash), Between Stations in a Layout

	To	1	2	3	4	5
From	1	0	9/50	5/120	6/205	0
	2	0	0	0	0	9/80
	3	0	0	0	2/85	3/170
	4	0	0	0	0	8/85
	5	0	0	0	0	0

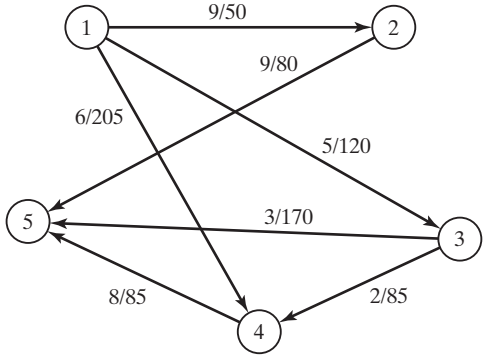


Figure 10.13 Network diagram showing material deliveries between load/unload stations. Nodes represent the load/unload stations, and arrows are labeled with flow rates, loads/hr, and distances m.

Mathematical equations can be developed to describe the operation of vehicle-based material transport systems. It is assumed that the vehicle moves at a constant velocity throughout its operation and that effects of acceleration, deceleration, and other speed differences are ignored. The time for a typical delivery cycle in the operation of a vehicle-based transport system consists of (1) loading at the pickup station, (2) travel time to the drop-off station, (3) unloading at the drop-off station, and (4) empty travel time of the vehicle between deliveries. The total cycle time per delivery per vehicle is given by

$$T_c = T_L + \frac{L_d}{v_c} + T_U + \frac{L_e}{v_c} \tag{10.1}$$

where T_c = delivery cycle time, min/del; T_L = time to load at load station, min; L_d = distance the vehicle travels between load and unload station, m (ft); v_c = carrier velocity, m/min (ft/min); T_U = time to unload at unload station, min; and L_e = distance the vehicle travels empty until the start of the next delivery cycle, m (ft).

The T_c calculated by Equation (10.1) must be considered an ideal value, because it ignores any time losses due to reliability problems, traffic congestion, and other factors that may slow down a delivery. In addition, not all delivery cycles are the same.

Originations and destinations may be different from one delivery to the next, which affect the L_d and L_e terms in the equation. Accordingly, these terms are considered to be average values for the loaded and empty distances traveled by the vehicle during a shift or other period of analysis.

The delivery cycle time T_c can be used to determine two values of interest in a vehicle-based transport system: (1) rate of deliveries per vehicle and (2) number of vehicles required to satisfy a specified total delivery requirement. The analysis is based on hourly rates and requirements, but the equations can readily be adapted for other time periods.

The hourly rate of deliveries per vehicle is 60 min divided by the delivery cycle time T_c, adjusting for any time losses during the hour. The possible time losses include (1) availability, (2) traffic congestion, and (3) efficiency of manual drivers in the case of manually operated trucks. ***Availability*** A is a reliability factor (Section 3.1.1) defined as the proportion of total shift time that the vehicle is operational and not broken down or being repaired.

To deal with the time losses due to traffic congestion, the ***traffic factor*** F_t is defined as a parameter for estimating the effect of these losses on system performance. Sources of inefficiency accounted for by the traffic factor include waiting at intersections, blocking of vehicles (as in an AGVS), and waiting in a queue at load/unload stations. If these situations do not occur, then $F_t = 1.0$. As blocking increases, the value of F_t decreases. F_t is affected by the number of vehicles in the system relative to the size of the layout. If there is only one vehicle in the system, no blocking should occur, and the traffic factor will be 1.0. For systems with many vehicles, there will be more instances of blocking and congestion, and the traffic factor will take a lower value. Typical values of traffic factor for an AGVS range between 0.85 and 1.0 [4].

For systems based on industrial trucks, including both hand trucks and powered trucks that are operated by human workers, traffic congestion is probably not the main cause of low operating performance. Instead, performance depends primarily on the work efficiency of the operators who drive the trucks. ***Worker efficiency*** is defined as the actual work rate of the human operator relative to the work rate expected under standard or normal performance. Let E_w symbolize worker efficiency.

With these factors defined, the available time per hour per vehicle can now be expressed as 60 min adjusted by A, F_t, and E_w. That is,

$$AT = 60AF_tE_w \tag{10.2}$$

where AT = available time, min/hr per vehicle; A = availability; F_t = traffic factor, and E_w = worker efficiency. The parameters A, F_t, and E_w do not take into account poor vehicle routing, poor guide-path layout, or poor management of the vehicles in the system. These factors should be minimized, but if present they are accounted for in the values of L_d, L_e, T_L, and T_u.

Equations for the two performance parameters of interest can now be written. The rate of deliveries per vehicle is given by

$$R_{dv} = \frac{AT}{T_c} \tag{10.3}$$

where R_{dv} = hourly delivery rate per vehicle, deliveries/hr per vehicle; T_c = delivery cycle time computed by Equation (10.1), min/del; and AT = the available time in 1 hour, adjusted for time losses, min/hr.

The total number of vehicles (trucks, AGVs, trolleys, carts, etc.) needed to satisfy a specified total delivery schedule R_f in the system can be estimated by first calculating the total workload required and then dividing by the available time per vehicle. ***Workload*** is defined as the total amount of work, expressed in terms of time, that must be accomplished by the material transport system in 1 hr. This can be expressed as

$$WL = R_f T_c \tag{10.4}$$

where WL = workload, min/hr; R_f = specified flow rate of total deliveries per hour for the system, deliveries/hr; and T_c = delivery cycle time, min/del. Now the number of vehicles required to accomplish this workload can be written as

$$n_c = \frac{WL}{AT} \tag{10.5}$$

where n_c = number of carriers (vehicles) required, WL = workload, min/hr; and AT = available time per vehicle, min/hr per vehicle. Substituting Equations (10.3) and (10.4) into Equation (10.5) provides an alternative way to determine n_c:

$$n_c = \frac{R_f}{R_{dv}} \tag{10.6}$$

where n_c = number of carriers required, R_f = total delivery requirements in the system, deliveries/hr; and R_{dv} = delivery rate per vehicle, deliveries/hr per vehicle. Although the traffic factor accounts for delays experienced by the vehicles, it does not include delays encountered by a load/unload station that must wait for the arrival of a vehicle. Because of the random nature of the load/unload demands, workstations are likely to experience waiting time while vehicles are busy with other deliveries. The preceding equations do not consider this idle time or its impact on operating cost. If station idle time is to be minimized, then more vehicles may be needed than the number indicated by Equations (10.5) or (10.6). Mathematical models based on queueing theory are appropriate to analyze this more complex stochastic situation.

EXAMPLE 10.1 Determining Number of Vehicles in an AGVS

Consider the AGVS layout in Figure 10.14. Vehicles travel counterclockwise around the loop to deliver loads from the load station to the unload station. Loading time at the load station = 0.75 min, and unloading time at the unload station = 0.50 min. The following performance parameters are given: vehicle speed = 50 m/min, availability = 0.95, and traffic factor = 0.90. Operator efficiency does not apply, so $E_w = 1.0$. Determine (a) travel distances loaded and empty, (b) ideal delivery cycle time, and (c) number of vehicles required to satisfy the delivery demand if a total of 40 deliveries per hour must be completed by the AGVS.

Solution: (a) Ignoring effects of slightly shorter distances around the curves at corners of the loop, the values of L_d and L_e are readily determined from the layout to be **110 m** and **80 m**, respectively.

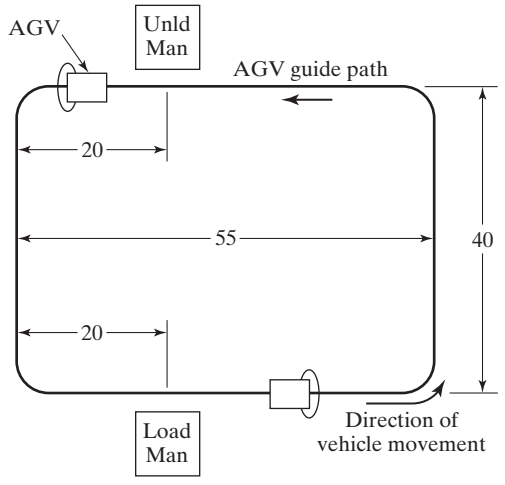


Figure 10.14 AGVS loop layout for Example 10.1. Key: Unld = unload, Man = manual operation, dimensions in meters (m).

(b) Ideal cycle time per delivery per vehicle is given by Equation (10.1):

$$T_c = 0.75 + \frac{110}{50} + 0.50 + \frac{80}{50} = \textbf{5.05 min}$$

(c) To determine the number of vehicles required to make 40 deliveries/hr, compute the workload of the AGVS and the available time per hour per vehicle:

$$WL = 40(5.05) = 202 \text{ min/hr}$$

$$AT = 60(0.95)(0.90)(1.0) = 51.3 \text{ min/hr per vehicle}$$

Therefore, the number of vehicles required is

$$n_c = \frac{202}{51.3} = \textbf{3.94 vehicles}$$

This value should be rounded up to **4 vehicles**, since the number of vehicles must be an integer.

Determining the average travel distances, L_d and L_e, requires analysis of the particular AGVS layout and how the vehicles are managed. For a simple loop layout such as Figure 10.14, determining these values is straightforward. For a complex AGVS layout, the problem is more difficult. The following example illustrates the issue.

EXAMPLE 10.2 Determining L_d for a More-Complex AGVS Layout

The layout for this example is shown in Figure 10.15, and the from-to chart is presented in Table 10.5. The AGVS includes load station 1 where raw parts enter the system for delivery to any of three production stations 2, 3, and 4. Unload station 5 receives finished parts from the production stations. Load and unload times at stations 1 and 5 are each 0.5 min. Production rates for each workstation are indicated by the delivery requirements in Table 10.5. A complicating factor is that some parts must be transshipped between stations 3 and 4. Vehicles move in the direction indicated by the arrows in the figure. Determine the average delivery distance, L_d.

Solution: Table 10.5 shows the number of deliveries and corresponding distances between the stations. The distance values are taken from the layout drawing in Figure 10.15. To determine the value of L_d, a weighted average must be calculated based on the number of trips and corresponding distances shown in the from-to chart for the problem:

$$L_d = \frac{9(50) + 5(120) + 6(205) + 9(80) + 2(85) + 3(170) + 8(85)}{9 + 5 + 6 + 9 + 2 + 3 + 8} = \frac{4{,}360}{42} = \mathbf{103.8\ m}$$

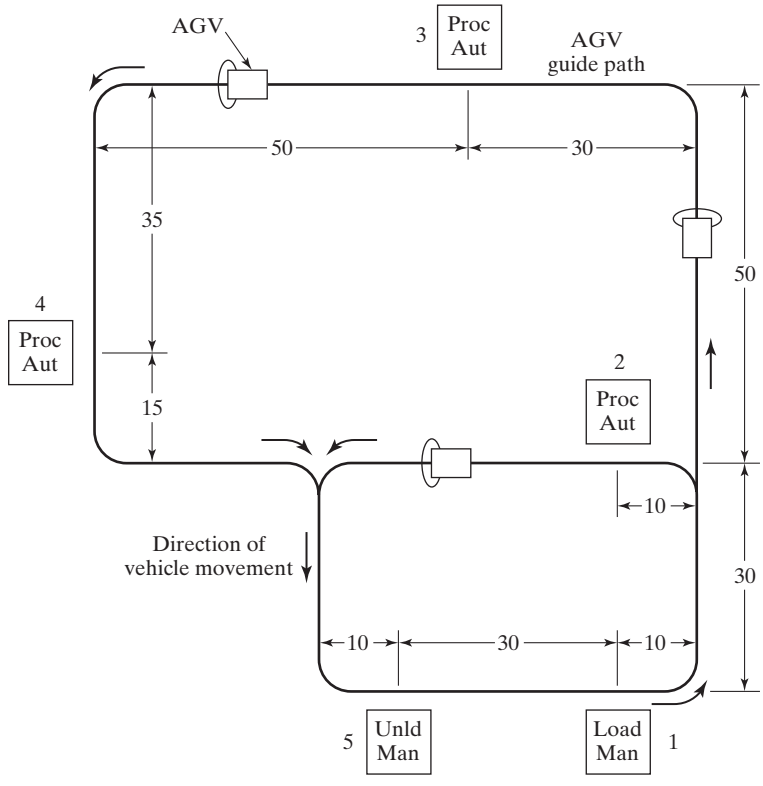


Figure 10.15 AGVS layout for production system of Example 10.2. Key: Proc = processing operation, Aut = automated, Unld = unload, Man = manual operation, dimensions in meters (m).

Determining L_e, the average distance a vehicle travels empty during a delivery cycle, is more complicated. It depends on the dispatching and scheduling methods used to decide how a vehicle should proceed from its last drop-off to its next pickup. In Figure 10.15, if each vehicle must travel back to station 1 after each drop-off at stations 2, 3, and 4, then the empty distance between pick-ups would be very long indeed. L_e would be greater than L_d. On the other hand, if a vehicle could exchange a raw work part for a finished part while stopped at a given workstation, then empty travel time for the vehicle would be minimized. However, this would require a two-position platform at each station to enable the exchange. So this issue must be considered in the initial design of the AGVS. Ideally, L_e should be reduced to zero. It is highly desirable to minimize the average distance a vehicle travels empty through good design of the AGVS and good scheduling of the vehicles. The mathematical model of vehicle-based systems indicates that the delivery cycle time will be reduced if L_e is minimized, and this will have a beneficial effect on the vehicle delivery rate and the number of vehicles required to operate the system. Two of the exercise problems at the end of the chapter ask the reader to determine L_e under different operating scenarios.

10.3.2 Conveyor Analysis

Conveyor operations have been analyzed in the research literature [8], [9], [11], [12], [13], and [14]). In the discussion here, the three basic types of conveyor operations discussed in Section 10.2.4 are considered: (1) single direction conveyors, (2) continuous loop conveyors, and (3) recirculating conveyors.

Single Direction Conveyors. Consider the case of a single direction powered conveyor with one load station at the upstream end and one unload station at the downstream end, as in Figure 10.10(a). Materials are loaded at one end and unloaded at the other. The materials may be parts, cartons, pallet loads, or other unit loads. Assuming the conveyor operates at a constant speed, the time required to move materials from load station to unload station is given by

$$T_d = \frac{L_d}{v_c} \tag{10.7}$$

where T_d = delivery time, min; L_d = length of conveyor between load and unload stations, m (ft), and v_c = conveyor velocity, m/min (ft/min).

The flow rate of materials on the conveyor is determined by the rate of loading at the load station. The loading rate is limited by the reciprocal of the time required to load the materials. Given the conveyor speed, the loading rate establishes the spacing of materials on the conveyor. Summarizing these relationships,

$$R_f = R_L = \frac{v_c}{s_c} \leq \frac{1}{T_L} \tag{10.8}$$

where R_f = material flow rate, parts/min; R_L = loading rate, parts/min; s_c = center-to-center spacing of materials on the conveyor, m/part (ft/part); and T_L = loading time, min/part. One might be tempted to think that the loading rate R_L is the reciprocal of the loading time T_L. However, R_L is set by the flow rate requirement R_f, while T_L is determined by ergonomic factors. The worker who loads the conveyor may be capable of

performing the loading task at a rate that is faster than the required flow rate. On the other hand, the flow rate requirement cannot be set faster than it is humanly possible to perform the loading task.

An additional requirement for loading and unloading is that the time required to unload the conveyor must be equal to or less than the reciprocal of material flow rate. That is,

$$T_U \leq \frac{1}{R_f} \tag{10.9}$$

where T_U = unloading time, min/part. If unloading requires more time than the time interval between arriving loads, then loads may accumulate or be dumped onto the floor at the downstream end of the conveyor.

Parts are being used as the material in Equations (10.8) and (10.9), but the relationships apply to other unit loads as well. The advantage of the Unit Load Principle (Section 10.1.2) can be demonstrated by transporting n_p parts in a container rather than a single part. Recasting Equation (10.8) to reflect this advantage,

$$R_f = \frac{n_p v_c}{s_c} \leq \frac{1}{T_L} \tag{10.10}$$

where R_f = flow rate, parts/min; n_p = number of parts per container; s_c = center-to-center spacing of containers on the conveyor, m/container (ft/container); and T_L = loading time per container, min/container. The flow rate of parts transported by the conveyor is potentially much greater in this case. However, loading time is still a limitation, and T_L may consist of not only the time to load the container onto the conveyor but also the time to load parts into the container. The preceding equations must be interpreted and perhaps adjusted for the given application.

EXAMPLE 10.3 Single Direction Conveyor

A roller conveyor follows a pathway 35 m long between a parts production department and an assembly department. Velocity of the conveyor is 40 m/min. Parts are loaded into large tote pans, which are placed onto the conveyor at the load station in the production department. Two operators work at the loading station. The first worker loads parts into tote pans, which takes 25 sec. Each tote pan holds 20 parts. Parts enter the loading station from production at a rate that is in balance with this 25-sec cycle. The second worker loads tote pans onto the conveyor, which takes only 10 sec. Determine (a) spacing between tote pans along the conveyor, (b) maximum possible flow rate in parts/min, and (c) the maximum time allowed to unload the tote pan in the assembly department.

Solution: (a) Spacing between tote pans on the conveyor is determined by the loading time. It takes only 10 sec to load a tote pan onto the conveyor, but 25 sec are required to load parts into the tote pan. Therefore, the loading cycle is limited by this 25 sec. At a conveyor speed of 40 m/min, the spacing will be

$$s_c = (25/60 \text{ min})(40 \text{ m/min}) = \mathbf{16.67\ m}$$

(b) Flow rate is given by Equation (10.10):

$$R_f = \frac{20(40)}{16.67} = \mathbf{48\ parts/min}$$

This is consistent with the parts loading rate of 20 parts in 25 sec, which is 0.8 parts/sec or 48 parts/min.

(c) The maximum allowable time to unload a tote pan must be consistent with the flow rate of tote pans on the conveyor. This flow rate is one tote pan every 25 sec, so

$$T_U \leq \mathbf{25\ sec}$$

Continuous Loop Conveyors. Consider a continuous loop conveyor such as an overhead trolley in which the pathway is formed by an endless chain moving in a track loop, and carriers are suspended from the track and pulled by the chain. The conveyor moves parts in the carriers between a load station and an unload station. The complete loop is divided into two sections: a delivery (forward) loop in which the carriers are loaded and a return loop in which the carriers travel empty, as shown in Figure 10.10(b). The length of the delivery loop is L_d, and the length of the return loop is L_e. Total length of the conveyor is therefore $L = L_d + L_e$. The total time required to travel the complete loop is

$$T_c = \frac{L}{v_c} \tag{10.11}$$

where T_c = total cycle time, min; and v_c = speed of the conveyor chain, m/min (ft/min). The time a load spends in the forward loop is

$$T_d = \frac{L_d}{v_c} \tag{10.12}$$

where T_d = delivery time on the forward loop, min.

Carriers are equally spaced along the chain at a distance s_c apart. Thus, the total number of carriers in the loop is given by

$$n_c = \frac{L}{s_c} \tag{10.13}$$

where n_c = number of carriers; L = total length of the conveyor loop, m (ft); and s_c = center-to-center distance between carriers, m/carrier (ft/carrier). The value of n_c must be an integer, and so the values of L and s_c must be consistent with that requirement.

Each carrier is capable of holding parts on the delivery loop, and it holds no parts on the return trip. Since only those carriers on the forward loop contain parts, the maximum number of parts in the system at any one time is given by

$$\text{Total parts in system} = \frac{n_p n_c L_d}{L} \tag{10.14}$$

As in the single direction conveyor, the maximum flow rate between load and unload stations is

$$R_f = \frac{n_p v_c}{s_c}$$

where R_f = parts per minute, pc/min. Again, this rate must be consistent with limitations on the time it takes to load and unload the conveyor, as defined in Equations (10.8) through (10.10).

Recirculating Conveyors. Recall the two problems complicating the operation of a recirculating conveyor system (Section 10.2.4): (1) the possibility that no empty carriers are immediately available at the loading station when needed and (2) the possibility that no loaded carriers are immediately available at the unloading station when needed. The case of a recirculating conveyor with one load station and one unload station was analyzed by Kwo [8], [9]. According to his analysis, three basic principles must be obeyed in designing such a conveyor system:

1. *Speed Rule.* The operating speed of the conveyor must be within a certain range. The lower limit of the range is determined by the required loading and unloading rates at the respective stations. These rates are dictated by the external systems served by the conveyor. Let R_L and R_U represent the required loading and unloading rates at the two stations, respectively. Then the conveyor speed must satisfy the relationship

$$\frac{n_p v_c}{s_c} \geq \text{Max}\{R_L, R_U\} \tag{10.15}$$

   where R_L = required loading rate, pc/min; and R_U = the corresponding unloading rate. The upper speed limit is determined by the physical capabilities of the material handlers who perform the loading and unloading tasks. Their capabilities are defined by the time required to load and unload the carriers, so that

$$\frac{v_c}{s_c} \leq \text{Min}\left\{\frac{1}{T_L}, \frac{1}{T_U}\right\} \tag{10.16}$$

   where T_L = time required to load a carrier, min/carrier; and T_U = time required to unload a carrier. In addition to Equations (10.15) and (10.16), another limitation is of course that the speed must not exceed the physical limits of the mechanical conveyor itself.
2. *Capacity Constraint.* The flow rate capacity of the conveyor system must be at least equal to the flow rate requirement to accommodate reserve stock and allow for the time elapsed between loading and unloading due to delivery distance. This can be expressed as follows:

$$\frac{n_p v_c}{s_c} \geq R_f \tag{10.17}$$

   In this case, R_f must be interpreted as a system specification required of the recirculating conveyor.
3. *Uniformity Principle.* This principle states that parts (loads) should be uniformly distributed throughout the length of the conveyor, so that there will be no sections of the conveyor in which every carrier is full while other sections are virtually empty. The reason for the uniformity principle is to avoid unusually long waiting times at the load or unload stations for empty or full carriers (respectively) to arrive.

EXAMPLE 10.4 Recirculating Conveyor Analysis: Kwo

A recirculating conveyor has a total length of 300 m. Its speed is 60 m/min, and the spacing of part carriers along its length is 12 m. Each carrier can hold two parts. The task time required to load two parts into each carrier is 0.20 min and the unload time is the same. The required loading and unloading rates are both defined by the specified flow rate, which is 4 parts/min. Evaluate the conveyor system design with respect to Kwo's three principles.

Solution: *Speed Rule:* The lower limit on speed is set by the required loading and unloading rates, both 4 parts/min. Checking this against Equation (10.15),

$$\frac{n_p v_c}{s_c} \geq \text{Max}\{R_L, R_U\}$$

$$\frac{(2 \text{ parts/carrier})(60 \text{ m/min})}{12 \text{ m/carrier}} = 10 \text{ parts/min} > \mathbf{4 \text{ parts/min}}$$

Checking the lower limit,

$$\frac{60 \text{ m/min}}{12 \text{ m/carrier}} = 5 \text{ carriers/min} \leq \text{Min}\left\{\frac{1}{0.2}, \frac{1}{0.2}\right\} = \text{Min}\{5, 5\} = 5$$

The Speed Rule is satisfied.

Capacity Constraint: The conveyor flow rate capacity = 10 parts/min as computed above. Since this is substantially greater than the required delivery rate of 4 parts/min, the capacity constraint is satisfied. Kwo provides guidelines for determining the flow rate requirement that should be compared to the conveyor capacity.

Uniformity Principle: The conveyor is assumed to be uniformly loaded throughout its length, since the loading and unloading rates are equal and the flow rate capacity is substantially greater than the load/unload rate. Conditions for checking the uniformity principle are available in the original papers by Kwo [8], [9].

REFERENCES

[1] Bose, P. P., "Basics of AGV Systems," Special Report 784, *American Machinist and Automated Manufacturing*, March 1986, pp. 105–122.

[2] Castelberry, G., *The AGV Handbook*, AGV Decisions, Inc., published by Braun-Brumfield, Inc., Ann Arbor, MI, 1991.

[3] Eastman, R. M., *Materials Handling*, Marcel Dekker, Inc., New York, 1987.

[4] Fitzgerald, K. R., "How to Estimate the Number of AGVs You Need," *Modern Materials Handling*, October 1985, p. 79.

[5] Kulwiec, R. A., *Basics of Material Handling*, Material Handling Institute, Pittsburgh, PA, 1981.

[6] Kulwiec, R. A., Editor, *Materials Handling Handbook*, 2nd ed., John Wiley & Sons, Inc., New York, 1985.

[7] Kulwiec, R., "Cranes for Overhead Handling," *Modern Materials Handling*, July 1998, pp. 43–47.

[8] Kwo, T. T., "A Theory of Conveyors," *Management Science*, Vol. 5, No. 1, 1958, pp. 51–71.

[9] Kwo, T. T., "A Method for Designing Irreversible Overhead Loop Conveyors," *Journal of Industrial Engineering*, Vol. 11, No. 6, 1960, pp. 459–466.

[10] Miller, R. K., *Automated Guided Vehicle Systems*, Co-published by SEAI Institute, Madison, GA and Technical Insights, Fort Lee, NJ, 1983.

[11] Muth, E. J., "Analysis of Closed-Loop Conveyor Systems," *AIIE Transactions*, Vol. 4, No. 2, 1972, pp. 134–143.

[12] Muth, E. J., "Analysis of Closed-Loop Conveyor Systems: The Discrete Flow Case," *AIIE Transactions*, Vol. 6, No. 1, 1974, pp. 73–83.

[13] Muth, E. J., "Modelling and Analysis of Multistation Closed-Loop Conveyors," *International Journal of Production Research*, Vol. 13, No. 6, 1975, pp. 559–566.

[14] Muth, E. J., and J. A. White, "Conveyor Theory: A Survey," *AIIE Transactions*, Vol. 11, No. 4, 1979, pp. 270–277.

[15] Muther, R., and K. Haganas, *Systematic Handling Analysis*, Management and Industrial Research Publications, Kansas City, MO, 1969.

[16] Tompkins, J. A., J. A. White, Y. A. Bozer, E. H. Frazelle, J. M. Tanchoco, and J. Trevino, *Facilities Planning*, 4th ed., John Wiley & Sons, Inc., New York, 2010.

[17] Witt, C. E., "Palletizing Unit Loads: Many Options," *Material Handling Engineering*, January, 1999, pp. 99–106.

[18] Zollinger, H. A., "Methodology to Concept Horizontal Transportation Problem Solutions," paper presented at the *MHI 1994 International Research Colloquium*, Grand Rapids, MI, June 1994.

[19] www.agvsystems.com

[20] www.jervisbwebb.com

[21] www.mhi.org/cicmhe/resources/taxonomy

[22] www.mhia.org

[23] www.wikipedia.org/wiki/Automated_guided_vehicle

REVIEW QUESTIONS

10.1 Provide a definition of material handling.

10.2 How does material handling fit within the scope of logistics?

10.3 Name the five major categories of material handling equipment.

10.4 What is included within the term *unitizing equipment*?

10.5 What is the Unit Load Principle?

10.6 What are the five categories of material transport equipment commonly used to move parts and materials inside a facility?

10.7 Give some examples of industrial trucks used in material handling.

10.8 What is an automated guided vehicle system (AGVS)?

10.9 Name three categories of automated guided vehicles.

10.10 What features distinguish laser-guided vehicles from conventional AGVs?

10.11 What is forward sensing in AGVS terminology?

10.12 What are some of the differences between rail-guided vehicles and automated guided vehicles?

10.13 What is a conveyor?

10.14 Name some of the different types of conveyors used in industry.

10.15 What is a recirculating conveyor?

10.16 What is the difference between a hoist and a crane?

PROBLEMS

Answers to problems labeled **(A)** are listed in the appendix.

Analysis of Vehicle-based Systems

10.1 **(A)** An automated guided vehicle system has an average travel distance per delivery = 220 m and an average empty travel distance = 160 m. Load and unload times are each 24 sec and the speed of the AGV = 1 m/sec. Traffic factor = 0.9 and availability = 0.94. How many vehicles are needed to satisfy a delivery requirement of 35 deliveries/hr?

10.2 In Example 10.2, suppose that the vehicles operate according to the following scheduling rules: (1) vehicles delivering raw work parts from station 1 to stations 2, 3, and 4 must return empty to station 5; and (2) vehicles picking up finished parts at stations 2, 3, and 4 for delivery to station 5 must travel empty from station 1. (a) Determine the empty travel distances associated with each delivery and develop a from-to chart in the format of Table 10.5. (b) The AGVs travel at a speed of 50 m/min and the traffic factor = 0.90. Assume reliability = 100%. From Example 10.2, the delivery distance L_d = 103.8 m. Determine the value of L_e. (c) How many automated guided vehicles will be required to operate the system?

10.3 In Example 10.2, suppose that the vehicles operate according to the following scheduling rule in order to minimize the distances the vehicles travel empty: vehicles delivering raw work parts from station 1 to stations 2, 3, and 4 must pick up finished parts at these respective stations for delivery to station 5. (a) Determine the empty travel distances associated with each delivery and develop a from-to chart in the format of Table 10.5. (b) The AGVs travel at a speed of 50 m/min and the traffic factor = 0.90. Assume reliability = 100%. From Example 10.2, the delivery distance L_d = 103.8 m. Determine the value of L_e. (c) How many automated guided vehicles will be required to operate the system?

10.4 A planned manufacturing system will have the layout pictured in Figure P10.4 and will use an automated guided vehicle system to move parts between stations in the layout. All work parts are loaded into the system at station 1, moved to one of three processing stations (2, 3, or 4), and then brought back to station 1 for unloading. Once loaded onto its AGV, each work part stays onboard the vehicle throughout its time in the manufacturing system. Load and unload times at station 1 are each 0.5 min. Processing times at the processing stations are 6.5 min at station 2, 8.0 min at station 3, and 9.5 min at station 4. Vehicle speed = 50 m/min. Assume that the traffic factor = 1.0 and vehicle availability = 100%. (a) Construct the from-to chart for distances. (b) Determine the maximum hourly production rate for each of the three processing stations, assuming that 15 sec will be lost between successive vehicles at each station; this is the time for the vehicle presently at the station to move out and the next vehicle to move into the station for processing. (c) Find the total number of AGVs that will be needed to achieve these production rates.

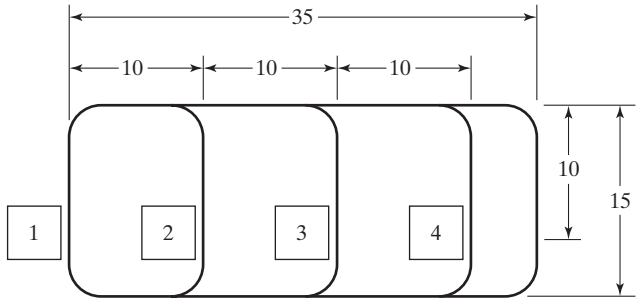


Figure P10.4 FMS layout for Problem 10.4.

10.5 In the previous problem, it is unrealistic to assume that the traffic factor will be 1.0 and that vehicle availability will be 100%. It is also unrealistic to believe that the processing stations will operate at 100% reliability. Solve the problem except the traffic factor = 90% and availability of the vehicles and processing workstations = 95%.

10.6 **(A)** A fleet of forklift trucks is being planned for a new warehouse. The average travel distance per delivery will be 500 ft loaded and the average empty travel distance will be 400 ft. The fleet must make a total of 50 deliveries/hr. Load and unload times are each 0.75 min and the speed of the vehicles = 350 ft/min. Assume the traffic factor for the system = 0.85, availability = 0.95, and worker efficiency = 90%. Determine (a) ideal cycle time per delivery, (b) the resulting average number of deliveries/hr that a forklift truck can make, and (c) how many trucks are required to accomplish the 50 deliveries/hr.

10.7 Three forklift trucks are used to deliver pallet loads of parts between work cells in a factory. Average travel distance loaded is 350 ft and the travel distance empty is estimated to be the same. The trucks are driven at an average speed of 3 miles/hr when loaded and 4 miles/hr when empty. Terminal time per delivery averages 1.0 min (load = 0.5 min and unload = 0.5 min). If the traffic factor is assumed to be 0.90, availability = 100%, and worker efficiency = 0.95, what is the maximum hourly delivery rate of the three trucks?

10.8 An AGVS has an average loaded travel distance per delivery = 300 ft. The average empty travel distance is not known. Required number of deliveries/hr = 50. Load and unload times are each 0.5 min and the AGV speed = 200 ft/min. Anticipated traffic factor = 0.85 and availability = 0.95. Develop an equation that relates the number of vehicles required to operate the system as a function of the average empty travel distance L_e.

10.9 A rail-guided vehicle system is being planned as part of an assembly cell consisting of two parallel lines, as in Figure P10.9. In operation, a base part is loaded at station 1 and delivered to either station 2 or 4, where components are added to the base part. The RGV then goes to either station 3 or 5, respectively, where further assembly of components is accomplished. From stations 3 or 5, the completed product moves to station 6 for removal from the system. Vehicles remain with the products as they move through the station sequence; thus, there is no loading and unloading of parts at stations 2, 3, 4, and 5. After unloading parts at station 6, the vehicles then travel empty back to station 1 for reloading. The hourly moves (parts/hr, above the slash) and distances (ft, below the slash) are listed in the following table. Moves indicated by "L" are trips in which the vehicle is loaded, while "E" indicates moves in which the vehicle is empty. RGV speed = 150 ft/min. Assembly cycle times at stations 2 and 3 = 4.0 min each and at stations 4 and 5 = 6.0 min each. Load and unload times at stations 1 and 6, respectively, are each 0.75 min. Traffic factor = 1.0 and availability = 1.0. How many vehicles are required to operate the system?

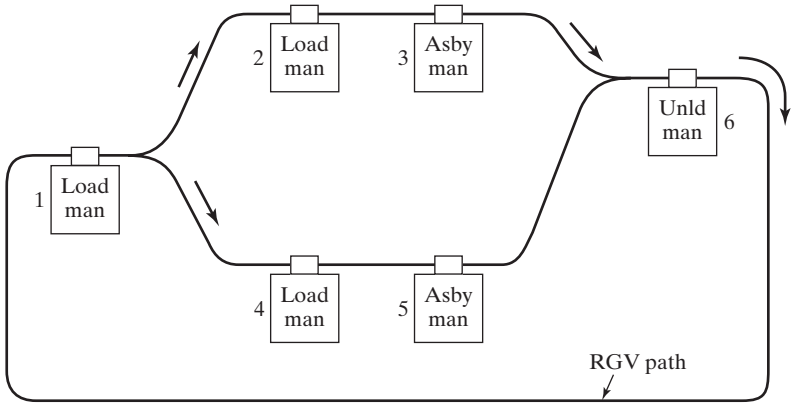


Figure P10.9 Layout for Problem 10.9. Key: Man = manual, Asby = assembly, Unld = unload.

	To	1	2	3	4	5	6
From	1	0/0	13L/100	—	9L/80	—	—
	2	—	0/0	13L/30	—	—	—
	3	—	—	0/0	—	—	13L/50
	4	—	—	—	0/0	9L/30	—
	5	—	—	—	—	0/0	9L/70
	6	22E/300	—	—	—	—	0/0

10.10 An AGVS will be used to satisfy the material flows in the from-to chart below, which shows delivery rates between stations (pc/hr, above the slash) and distances between stations (m, below the slash). Moves indicated by "L" are trips in which the vehicle is loaded, while "E" indicates moves in which the vehicle is empty. It is assumed that availability = 0.90, traffic factor = 0.85, and efficiency = 1.0. Speed of an AGV = 0.9 m/sec. If load handling time per delivery cycle = 1.0 min (load = 0.5 min and unload = 0.5 min), how many vehicles are needed to satisfy the indicated deliveries per hour?

	To	1	2	3	4
From	1	0/0	9L/90	7L/120	5L/75
	2	5E/90	0/0	—	4L/80
	3	7E/120	—	0/0	—
	4	9E/75	—	—	0/0

10.11 An automated guided vehicle system is being proposed to deliver parts between 40 workstations in a factory. Loads must be moved from each station about once every hour; thus, the delivery rate = 40 loads/hr. Average travel distance loaded is estimated to be 250 ft and travel distance empty is estimated to be 300 ft. Vehicles move at a speed = 200 ft/min. Total handling time per delivery = 1.5 min (load = 0.75 min and unload = 0.75 min). Traffic factor F_t becomes increasingly significant as the number of vehicles n_c increases; this can be modeled as $F_t = 1.0 - 0.05(n_c - 1)$, for n_c = Integer > 0. Determine the minimum number of vehicles needed in the factory to meet the flow rate requirement. Assume that availability = 1.0 and worker efficiency = 1.0.

10.12 A driverless train AGVS is being planned for a warehouse complex. Each train will consist of a towing vehicle plus four carts. Speed of the trains = 160 ft/min. Only the pulled carts carry loads. Average loaded travel distance per delivery cycle is 2,000 ft and empty travel distance is the same. Assume the travel factor = 0.95 and availability = 1.0. The load handling time per train per delivery is expected to be 10 min. If the requirements on the AGVS are 25 cart loads/hr, determine the number of trains required.

10.13 The from-to chart below indicates the number of loads moved per 8-hr day (above the slash) and the distances in ft (below the slash) between departments in a particular factory. Fork lift trucks are used to transport the materials. They move at an average speed = 275 ft/min (loaded) and 350 ft/min (empty). Load handling time (loading plus unloading) per delivery is 1.5 min and anticipated traffic factor = 0.9. Availability = 95% and worker efficiency = 110%. Determine the number of trucks required under each of the following assumptions: (a) the trucks never travel empty; and (b) the trucks travel empty a distance equal to their loaded distance.

	To Dept.	A	B	C	D	E
From Dept	A	0/0	62/500	51/450	45/350	—
	B	—	0/0	—	22/400	—
	C	—	—	0/0	—	76/200
	D	—	—	—	0/0	65/150
	E	—	—	—	—	0/0

10.14 A warehouse consists of five aisles of racks (racks on both sides of each aisle) and a loading dock. The rack system is four levels high. Forklift trucks are used to transport loads between the loading dock and the storage compartments of the rack system in each aisle. The trucks move at an average speed = 120 m/min (loaded) and 150 m/min (empty). Load handling time (loading plus unloading) per delivery totals 1.0 min per storage/retrieval delivery on average, and the anticipated traffic factor = 0.90. Worker efficiency = 100% and vehicle availability = 95%. The average distance between the loading dock and the centers of aisles 1 through 5 are 150 m, 250 m, 350 m, 450 m, and 550 m, respectively. These values are to be used to compute travel times. The required rate of storage/retrieval deliveries is 75/hr, distributed evenly among the five aisles, and the trucks perform either storage or retrieval deliveries, but not both in one delivery cycle. Determine the number of forklift trucks required to achieve the 75 deliveries per hour.

10.15 Suppose the warehouse in the preceding problem were organized according to a class-based dedicated storage strategy based on activity level of the pallet loads in storage, so that aisles 1 and 2 accounted for 70% of the deliveries (class A) and aisles 3, 4, and 5 accounted for the remaining 30% (class B). Assume that deliveries in class A are evenly divided between aisles 1 and 2, and that deliveries in class B are evenly divided between aisles 3, 4, and 5. How many forklift trucks would be required to achieve 75 storage/retrieval deliveries per hour?

10.16 Major appliances are assembled on a production line at the rate of 50/hr. The products are moved along the line on work pallets (one product per pallet). At the final workstation the finished products are removed from the pallets. The pallets are then removed from the line and delivered back to the front of the line for reuse. Automated guided vehicles are used to transport the pallets to the front of the line, a distance of 500 ft. Return trip distance (empty) to the end of the line is also 500 ft. Each AGV carries three pallets and travels at a speed of 200 ft/min (loaded or empty). The pallets form queues at each end of the line, so that neither the production line nor the AGVs are ever starved for pallets. Time required to load each pallet onto an AGV = 15 sec; time to release a loaded AGV and move an empty AGV into position for loading at the end of the line = 12 sec. The same times apply

for pallet handling and release/positioning at the unload station located at the front of the production line. Assume availability = 100% and traffic factor = 1.0 because the route is a simple loop. How many vehicles are needed to operate the AGV system?

10.17 For the production line in the previous problem, assume that a single AGV train consisting of a tractor and multiple trailers is used to make deliveries rather than separate vehicles. Time required to load a pallet onto a trailer = 15 sec; and the time to release a loaded train and move an empty train into position for loading at the end of the production line = 30 sec. The same times apply for pallet handling and release/positioning at the unload station located at the front of the production line. The velocity of the AGV train = 175 ft/min (loaded or empty). Assume availability = 100% and traffic factor = 1.0 because there is only one train. If each trailer carries three pallets, how many trailers should be included in the train?

10.18 **(A)** An AGVS will be installed to deliver loads between four workstations: A, B, C, and D. Hourly flow rates (loads/hr, above the slash) and distances (m, below the slash) within the system are given in the table below (travel loaded denoted by "L" and travel empty denoted by "E"). Load and unload times are each 0.45 min, and travel speed of each vehicle is 1.4 m/sec. A total of 43 loads enter the system at station A, and 30 loads exit the system at station A. In addition, during each hour, six loads exit the system from station B and seven loads exit the system from station D. This is why there are a total of 13 empty trips made by the vehicles within the AGVS. How many vehicles are required to satisfy these delivery requirements, assuming the traffic factor = 0.85 and availability = 95%?

	To	A	B	C	D
From	A	—	18L/95	10L/80	15L/150
	B	6E/95	—	12L/65	
	C			—	22L/80
	D	30L/150 7E/150			—

Analysis of Conveyor Systems

10.19 **(A)** An overhead trolley conveyor is configured as a closed loop. The delivery loop has a length of 100 m and the return loop is 60 m. All parts loaded at the load station are unloaded at the unload station. Each hook on the conveyor can hold one part and the hooks are separated by 2 m. Conveyor speed = 0.5 m/sec. Determine (a) number of parts in the conveyor system under normal operations, (b) parts flow rate; and (c) maximum loading and unloading times that are compatible with the operation of the conveyor system?

10.20 A 400-ft long roller conveyor operates at a velocity = 50 ft/min and is used to move parts in containers between load and unload stations. Each container holds 15 parts. One worker at the load station is able to load parts into containers and place the containers onto the conveyor in 45 sec. It takes 30 sec to unload at the unload station. Determine (a) center-to-center distance between containers, (b) number of containers on the conveyor at one time, and (c) hourly flow rate of parts. (d) By how much must conveyor speed be increased in order to increase flow rate to 1,500 parts/hr?

10.21 **(A)** A roller conveyor moves tote pans in one direction at 200 ft/min between a load station and an unload station, a distance of 350 ft. With one worker, the time to load parts into a tote pan at the load station is 3 sec per part. Each tote pan holds 10 parts. In addition, it takes 15 sec to load a tote pan of parts onto the conveyor. Determine (a) spacing between

tote pan centers flowing in the conveyor system and (b) flow rate of parts on the conveyor system. (c) Consider the effect of the Unit Load Principle. Suppose the tote pans were smaller and could hold only one part instead of 10. Determine the flow rate of parts in this case if it takes 7 sec to load a tote pan onto the conveyor (instead of 15 sec for the larger tote pan), and it takes the same 3 sec to load the part into the tote pan.

10.22 A closed loop overhead conveyor must be designed to deliver parts from one load station to one unload station. The specified flow rate of parts that must be delivered between the two stations is 300 parts/hr. The conveyor has carriers spaced at a center-to-center distance that is to be determined. Each carrier holds one part. Forward and return loops will each be 90 m long. Conveyor speed = 0.5 m/sec. Times to load and unload parts at the respective stations are each = 12 sec. Is the system feasible and if so, what is the appropriate number of carriers and spacing between carriers that will achieve the specified flow rate?

10.23 Consider the previous problem, only the carriers are larger and capable of holding up to four parts ($n_p = 2, 3,$ or 4). The loading time $T_L = 9 + 3n_p$, where T_L is in seconds. With other parameters defined in the previous problem, determine which of the three values of n_p are feasible. For those values that are feasible, specify the appropriate design parameters for (a) spacing between carriers and (b) number of carriers that will achieve this flow rate.

10.24 A recirculating conveyor has a total length of 700 ft and a speed of 90 ft/min. Spacing of part carriers = 14 ft. Each carrier holds one part. Automated machines load and unload the conveyor at the load and unload stations. Time to load a part is 0.10 min and unload time is the same. To satisfy production requirements, the loading and unloading rates are each 2.0 parts per min. Evaluate the conveyor system design with respect to the three principles developed by Kwo.

10.25 A recirculating conveyor has a total length of 200 m and a speed of 50 m/min. Spacing of part carriers = 5 m. Each carrier holds two parts. Time needed to load a part carrier = 0.15 min. Unloading time is the same. The required loading and unloading rates are 6 parts per min. Evaluate the conveyor system design with respect to the three Kwo principles.

10.26 There is a plan to install a continuous loop conveyor system with a total length of 1,000 ft and a speed of 50 ft/min. The conveyor will have carriers separated by 25 ft. Each carrier can hold one part. A load station and an unload station are to be located 500 ft apart along the conveyor loop. Each day, the conveyor system is planned to operate as follows, starting empty at the beginning of the day. The load station will load parts at the rate of one part every 30 sec, continuing this loading operation for 10 min, then resting for 10 min during which no loading occurs. It will repeat this 20 min cycle of loading and then resting throughout the 8-hr shift. The unload station will wait until loaded carriers begin to arrive, then will unload parts at the rate of one part every minute during the 8 hr, continuing until all carriers are empty. Will the planned conveyor system work? Present calculations and arguments to justify your answer.

Chapter 11

Storage Systems

CHAPTER CONTENTS

11.1 Introduction to Storage Systems
  11.1.1 Storage System Performance
  11.1.2 Storage Location Strategies
11.2 Conventional Storage Methods and Equipment
11.3 Automated Storage Systems
  11.3.1 Fixed-Aisle Automated Storage/Retrieval Systems
  11.3.2 Carousel Storage Systems
11.4 Analysis of Storage Systems
  11.4.1 Fixed-Aisle Automated Storage/Retrieval Systems
  11.4.2 Carousel Storage Systems

The function of a material storage system is to store materials for a period of time and to permit access to those materials when required. Some production plants and storage facilities use manual methods for storing and retrieving materials. The storage function is often accomplished inefficiently in terms of human resources, factory floor space, and material control. More effective approaches and automated methods are available to improve the efficiency of the storage function.

This chapter provides an overview of material storage systems and describes the types of methods and systems used in a company's manufacturing and distribution operations. When used in manufacturing, storage systems are sometimes physically integrated with the production equipment. These cases are covered in several other chapters in this book, including parts storage in single-station automated cells (Chapter 14), buffer

storage in transfer lines (Chapter 16), storage of component parts in automated assembly systems (Chapter 17), and temporary storage of work-in-process in flexible manufacturing systems (Chapter 19). The present chapter emphasizes the storage system itself and the approaches and equipment related to it. Coverage of storage equipment is divided into two major categories: (1) conventional storage methods and (2) automated storage systems. The chapter begins with an overview of storage systems, focusing on performance measures and strategies. The final section of the chapter presents a quantitative analysis of automated storage systems, with emphasis on two important performance measures: storage capacity and throughput.

11.1 INTRODUCTION TO STORAGE SYSTEMS

Materials stored by a manufacturing firm include a variety of types, as indicated in Table 11.1. Categories (1) through (5) relate directly to the product, (6) through (8) relate to the process, and (9) and (10) relate to overall support of factory operations. The different categories require different storage methods and controls, as discussed in the coverage of equipment in Sections 11.2 and 11.3.

Whatever the stored materials, certain metrics and approaches can be used to design and operate the storage system so that it is as efficient as possible in fulfilling its function for the company. The coverage addresses the following topics: (1) measures of storage system performance and (2) storage location strategies.

TABLE 11.1 Types of Materials Typically Stored in a Factory

Type	Description
1. Raw materials	Raw stock to be processed (e.g., bar stock, sheet metal, plastic molding compound)
2. Purchased parts	Parts from vendors to be processed or assembled (e.g., castings, purchased components)
3. Work-in-process	Partially completed parts between processing operations and parts awaiting assembly
4. Finished product	Completed product ready for shipment
5. Rework and scrap	Parts that do not meet specifications, either to be reworked or scrapped
6. Refuse	Chips, swarf, oils, other waste products left over after processing; these materials must be disposed of, sometimes using special precautions
7. Tooling and supplies	Cutting tools, jigs, fixtures, molds, dies, welding wire, and other tools used in production; supplies such as helmets and gloves
8. Spare parts	Parts needed for maintenance and repair of factory equipment
9. Office supplies	Paper, paper forms, writing instruments, and other items used in support of plant office
10. Plant records	Records on product, production, equipment, personnel, etc. (paper documents and electronic media)

11.1.1 Storage System Performance

The performance of a storage system in accomplishing its function must be sufficient to justify its investment and operating expense. Various measures used to assess the performance of a storage system include (1) storage capacity, (2) storage density, (3) accessibility, and (4) throughput. In addition, standard measures used for mechanized and automated systems include (5) utilization and (6) reliability.[1]

Storage capacity can be defined and measured in two ways: (1) as the total volumetric space available or (2) as the total number of storage compartments in the system available to hold items or loads. In many storage systems, materials are stored as unit loads that are held in standard-size containers (pallets, tote pans, or other containers). The standard container can readily be handled, transported, and stored by the storage system and by the material transport system that may be connected to it. Hence, storage capacity is conveniently measured as the number of unit loads that can be stored in the system. The physical capacity of the storage system should be greater than the maximum number of loads anticipated to be stored, to provide available empty spaces for materials entering the system, and to allow for variations in maximum storage requirements.

Storage density is defined as the volumetric space available for actual storage relative to the total volumetric space in the storage facility. In many warehouses, aisle space and wasted overhead space account for more volume than the volume available for actual storage of materials. Floor area is sometimes used to assess storage density, because it is convenient to measure this on a floor plan of the facility. However, volumetric density is usually a more appropriate measure than area density.

For efficient use of space, the storage system should be designed to achieve a high density. However, as storage density is increased, accessibility, another important measure of storage performance, is adversely affected. ***Accessibility*** refers to the capability to access any desired item or load stored in the system. In the design of a given storage system, appropriate trade-offs must be made between storage density and accessibility.

System throughput is defined as the hourly rate at which the storage system (1) receives and puts loads into storage and/or (2) retrieves and delivers loads to the output station. In many factory and warehouse operations, there are certain periods of the day when the required rate of storage and/or retrieval transactions is greater than at other times. The storage system must be designed for the maximum throughput that will be required during the day.

System throughput is limited by the time to perform a storage or retrieval (S/R) transaction. A typical storage transaction consists of the following elements: (1) pick up load at input station, (2) travel to storage location, (3) place load in storage location, and (4) travel back to input station. A retrieval transaction consists of: (1) travel to storage location, (2) pick up item from storage, (3) travel to output station, and (4) unload at output station. Each element takes time. The sum of the element times is the transaction time that determines throughput of the storage system. Throughput can sometimes be increased by combining storage and retrieval transactions in one cycle, thus reducing travel time; this is called a *dual-command cycle*. When either a storage or a retrieval transaction alone is performed in the cycle, it is called a *single-command cycle*. The ability to perform dual-command cycles rather than single-command cycles depends on demand and

[1]The discussion here is limited to physical storage and not electronic storage media, although analogous performance metrics would apply to electronic storage.

scheduling issues. If, during a certain portion of the day, there is demand for only storage transactions and no retrievals, then it is not possible to include both types of transactions in the same cycle. If both transaction types are required, then greater throughput will be achieved by scheduling dual-command cycles. This scheduling is more readily done by a computerized (automated) storage system than by one controlled manually.

In manually operated systems, time is often lost looking up the storage location of the item being stored or retrieved. Computer-generated pick lists can be used to reduce such losses. Also, greater efficiency can be achieved in manual systems by combining multiple storage and/or retrieval transactions in one cycle, thus reducing time traveling to and from the input/output station. However, the drawback of manual systems is that they are subject to the variations and motivations of the human workers in these systems, and there is a lack of management control over the operations.

Two additional performance measures applicable to mechanized and automated storage systems are utilization and availability. ***Utilization*** is defined as the proportion of time that the system is actually being used for performing S/R operations compared with the time it is available. Utilization varies throughout the day, as requirements change from hour to hour. It is desirable to design an automated storage system for relatively high utilization, in the range 80–90%. If utilization is too low, then the system is probably overdesigned. If utilization is too high, then there is no allowance for rush periods or system breakdowns.

Availability is a measure of system reliability, defined as the proportion of time that the system is capable of operating (not broken down) compared with the normally scheduled shift hours (Section 3.1.1). Malfunctions and failures of the equipment cause downtime. Reasons for downtime include computer failures, mechanical breakdowns, load jams, improper maintenance, and incorrect procedures by personnel using the system. The reliability of an existing system can be improved by following good preventive maintenance procedures and by having repair parts on hand for critical components. Backup procedures should be devised to mitigate the effects of system downtime.

11.1.2 Storage Location Strategies

Several strategies can be used to organize stock in a storage system. These storage location strategies affect the performance measures discussed above. The two basic strategies applied in warehousing operations are (1) randomized storage and (2) dedicated storage. Each item type stored in a warehouse is known as a *stock-keeping-unit* (SKU). The SKU uniquely identifies that item type. The inventory records of the storage facility maintain a count of the quantities of each SKU that are in storage.

In randomized storage, items are stored in any available location in the storage system. In the usual implementation of randomized storage, incoming items are placed into storage in the nearest available open location. When an order is received for a given SKU, the stock is retrieved from storage according to a first-in-first-out policy so that the items held in storage the longest are used to make up the order.

In dedicated storage, SKUs are assigned to specific locations in the storage facility. This means that locations are reserved for all SKUs stored in the system, and so the number of storage locations for each SKU must be sufficient to accommodate its maximum inventory level. The basis for specifying the storage locations is usually one of the following: (1) items are stored in part number or product number sequence; (2) items are stored according to activity level, the more active SKUs being located closer to the input/output

station; or (3) items are stored according to their activity-to-space ratios, the higher ratios being located closer to the input/output station.

When comparing the benefits of the two strategies, it is generally found that less total space is required in a storage system that uses randomized storage, but higher throughput rates can usually be achieved when a dedicated storage strategy is implemented based on activity level. Example 11.1 illustrates the storage density advantage of randomized storage.

EXAMPLE 11.1 Comparison of Storage Strategies

Suppose that a total of 50 SKUs must be stored in a storage system. For each SKU, average order quantity = 100 cartons, average depletion rate = 2 cartons/day, and safety stock level = 10 cartons. Each carton requires one storage location in the system. Based on this data, each SKU has an inventory cycle that lasts 50 days. Since there are 50 SKUs in all, management has scheduled incoming orders so that a different SKU arrives each day. Determine the number of storage locations required in the system under two alternative strategies: (a) randomized storage and (b) dedicated storage.

Solution: The inventory for each SKU varies over time as shown in Figure 11.1. The maximum inventory level, which occurs just after an order has been received, is the sum of the order quantity and safety stock level:

$$\text{Maximum inventory level} = 100 + 10 = 110 \text{ cartons}$$

The average inventory is the average of the maximum and minimum inventory levels under the assumption of uniform depletion rate. The minimum value occurs just before an order is received when the inventory is depleted to the safety stock level:

$$\text{Minimum inventory level} = 10 \text{ cartons}$$

$$\text{Average inventory level} = (110 + 10)/2 = 60 \text{ cartons}$$

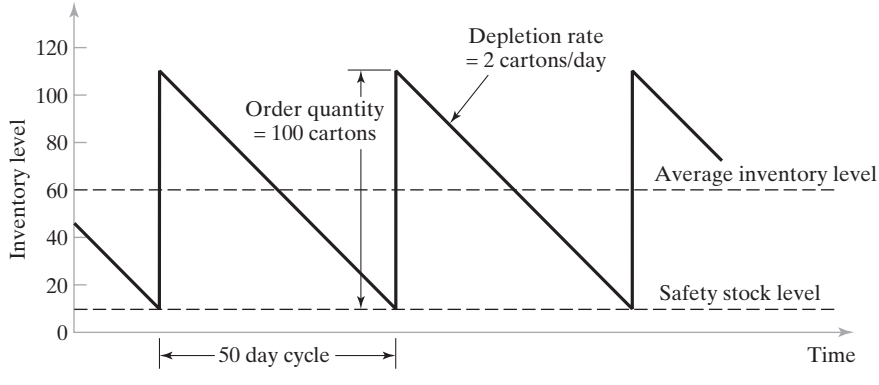


Figure 11.1 Inventory level as a function of time for each SKU in Example 11.1.

(a) Under a randomized storage strategy, the number of locations required for each SKU is equal to the average inventory level of the item, since incoming orders are scheduled each day throughout the 50-day cycle. This means that when the inventory level of one SKU near the beginning of its cycle is high, the level for another SKU near the end of its cycle is low. Thus, the number of storage locations required in the system $= (50\ \text{SKUs})(60\ \text{cartons}) =$ **3,000 locations**.

(b) Under a dedicated storage strategy, the number of locations required for each SKU must equal its maximum inventory level. Thus, the number of storage locations required in the storage system $= (50\ \text{SKUs})(110\ \text{cartons}) =$ **5,500 locations**.

Some of the advantages of both storage strategies can be obtained in a class-based dedicated storage allocation, in which the storage system is divided into several classes according to activity level, and a randomized storage strategy is used within each class. The classes containing more active SKUs are located closer to the input/output point of the storage system for increased throughput, and the randomized locations within the classes reduce the total number of storage compartments required. The effect of class-based dedicated storage on throughput is considered in several end-of-chapter problems.

11.2 CONVENTIONAL STORAGE METHODS AND EQUIPMENT

A variety of storage methods and equipment are available to store the various materials listed in Table 11.1. The choice of method and equipment depends largely on the materials to be stored, the operating philosophy of the personnel managing the storage facility, and budgetary limitations. The traditional (non-automated) methods and equipment types are discussed in this section. Automated storage systems are discussed in the following section. Application characteristics for the different equipment types are summarized in Table 11.2.

TABLE 11.2 Application Characteristics of the Types of Storage Equipment and Methods

Storage Equipment	Advantages and Disadvantages	Typical Applications
Bulk storage	Highest density is possible Low accessibility Low cost per square foot	Storage of low turnover, large stock, or large unit loads
Rack systems	Low cost Good storage density Good accessibility	Palletized loads in warehouses
Shelves and bins	Some stock items not clearly visible	Storage of individual items on shelves and commodity items in bins
Drawer storage	Contents of drawer easily visible Good accessibility Relatively high cost	Small tools Small stock items Repair parts
Automated storage systems	High throughput rates Facilitates use of computerized inventory control system Highest cost equipment Facilitates integration with automated material handling systems	Work-in-process storage Final product warehousing and distribution center Order picking Kitting of parts for electronic assembly

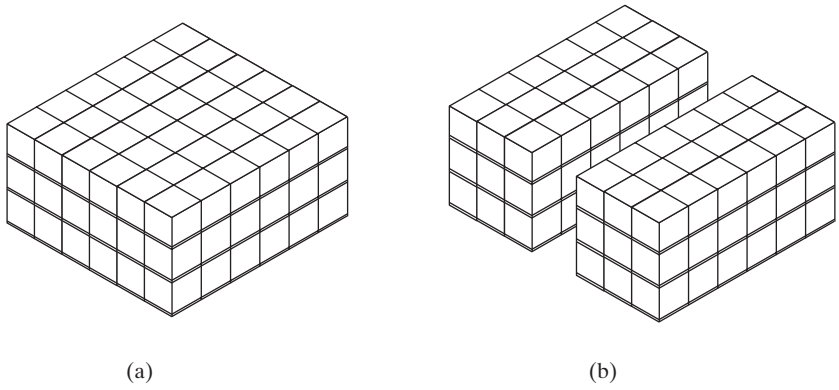


Figure 11.2 Two bulk storage arrangements: (a) high-density bulk storage provides low accessibility; (b) bulk storage with loads arranged to form rows and blocks for improved accessibility.

Bulk Storage. Bulk storage is the storage of stock in an open floor area. The stock is generally contained in unit loads on pallets or similar containers, and unit loads are stacked on top of each other to increase storage density. The highest density is achieved when unit loads are placed next to each other in both floor directions, as in Figure 11.2(a). However, this provides very poor access to internal loads. To increase accessibility, bulk storage loads can be organized into rows and blocks, so that natural aisles are created between pallet loads, as in Figure 11.2(b). The block widths can be designed to provide an appropriate balance between density and accessibility. Depending on the shape and physical support provided by the items stored, there may be a restriction on how high the unit loads can be stacked. In some cases, loads cannot be stacked on top of each other, either because of the physical shape or limited compressive strength of the individual loads. The inability to stack loads in bulk storage reduces storage density, removing one of its principal benefits.

Although bulk storage is characterized by the absence of specific storage equipment, material handling equipment must be used to put materials into storage and to retrieve them. Industrial trucks such as pallet trucks and powered forklifts (Section 10.2.1) are typically used for this purpose.

Rack Systems. Rack systems provide a method of stacking unit loads vertically without the need for the loads themselves to provide support. One of the most common rack systems is the pallet rack, consisting of a frame that includes horizontal load-supporting beams, as illustrated in Figure 11.3. Pallet loads are stored on these horizontal beams. Alternative storage rack systems include

- *Cantilever racks*, similar to pallet racks except the supporting horizontal beams are cantilevered from the vertical central frame. Elimination of the vertical beams at the front of the frame provides unobstructed spans, which facilitates storage of long materials such as rods, bars, and pipes.
- *Flow-through racks*. In place of the horizontal load-supporting beams in a conventional rack system, the flow-through rack uses long conveyor tracks capable of

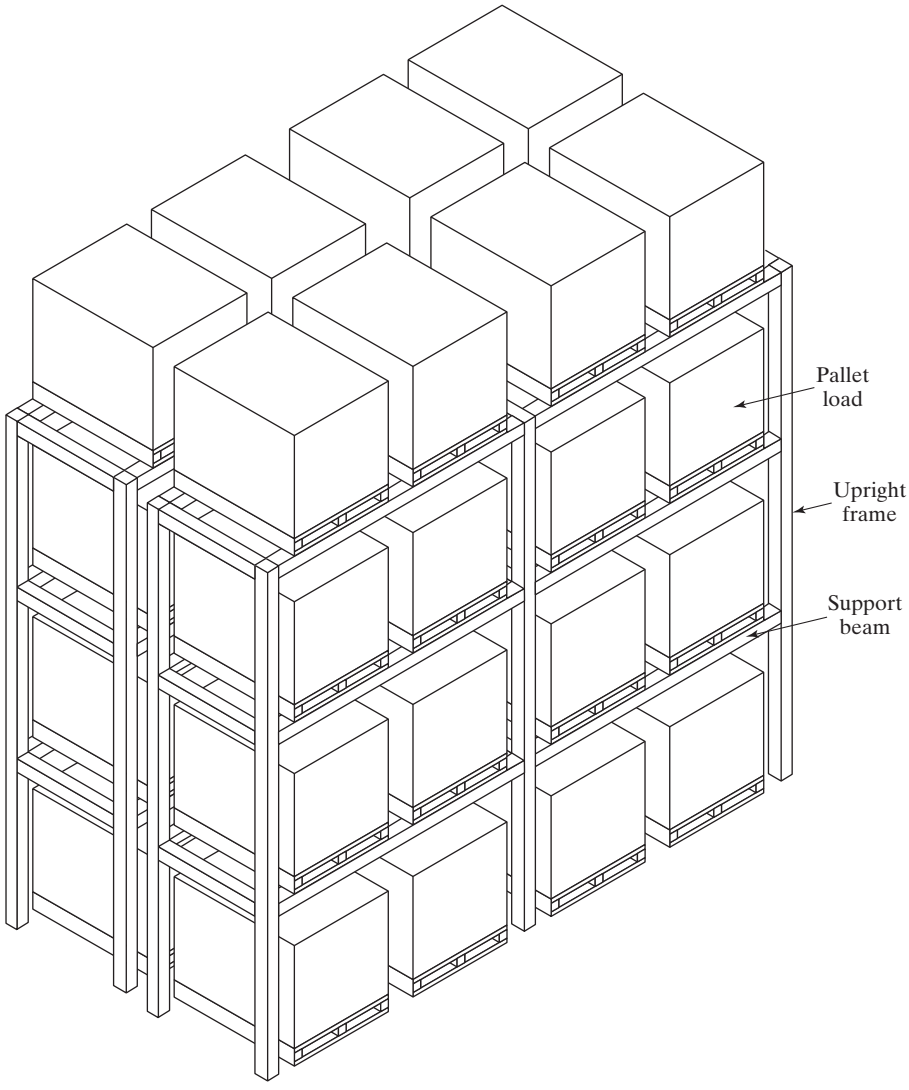


Figure 11.3 Pallet rack system for storage of unit loads on pallets.

supporting a row of unit loads. The unit loads are loaded from one side of the rack and unloaded from the other side, thus providing first-in-first-out stock rotation. The conveyor tracks are inclined at a slight angle to allow gravity to move the loads toward the output side of the rack system.

Shelving and Bins. Shelves represent one of the most common storage equipment types. A shelf is a horizontal platform, supported by a wall or frame, on which materials are stored. Steel shelving sections are manufactured in standard sizes, typically ranging from about 0.9 to 1.2 m (3 to 4 ft) long (in the aisle direction), from 0.3 to 0.6 m

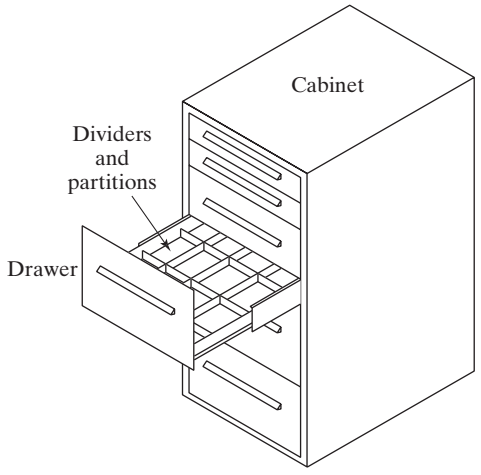


Figure 11.4 Drawer storage.

(12 to 24 in) wide, and up to 3.0 m (10 ft) tall. Shelving often includes bins, which are containers or boxes that hold loose items.

Drawer Storage. Finding items in shelving can sometimes be difficult, especially if the shelf is either far above or far below eye level for the storage attendant. Storage drawers, Figure 11.4, can alleviate this problem because each drawer pulls out to allow its entire contents to be readily seen. Modular drawer storage cabinets are available with a variety of drawer depths for different item sizes and are widely used for storage of tools and maintenance items.

11.3 AUTOMATED STORAGE SYSTEMS

The storage equipment described in the preceding section requires a human worker to access the items in storage. The storage system itself is static. Mechanized and automated storage systems are available that reduce or eliminate the amount of human intervention required to operate the system. The level of automation varies. In less automated systems, a human operator is required to handle each storage/retrieval transaction. In highly automated systems, loads are entered or retrieved under computer control, with no human participation except to input data to the computer. Table 11.2 lists the advantages and disadvantages as well as typical applications of automated storage systems.

An automated storage system represents a significant investment, and it often requires a new and different way of doing business. Companies have a variety of reasons for automating the storage function. Table 11.3 provides a list of possible objectives and reasons behind company decisions to automate their storage operations.

Automated storage systems divide into two general types: (1) fixed-aisle automated storage/retrieval systems and (2) carousel storage systems. A fixed-aisle AS/RS consists of a rack structure for storing loads and a storage/retrieval machine whose motions are linear (x–y–z motions), as pictured in Figure 11.5. By contrast, a carousel system uses

TABLE 11.3 Possible Objectives and Reasons for Automating a Company's Storage Operations

- To increase storage capacity
- To increase storage density
- To recover factory floor space presently used for storing work-in-process
- To improve security and reduce pilferage
- To improve safety in the storage function
- To reduce labor cost and/or increase labor productivity in storage operations
- To improve control over inventories
- To improve stock rotation
- To improve customer service
- To increase throughput

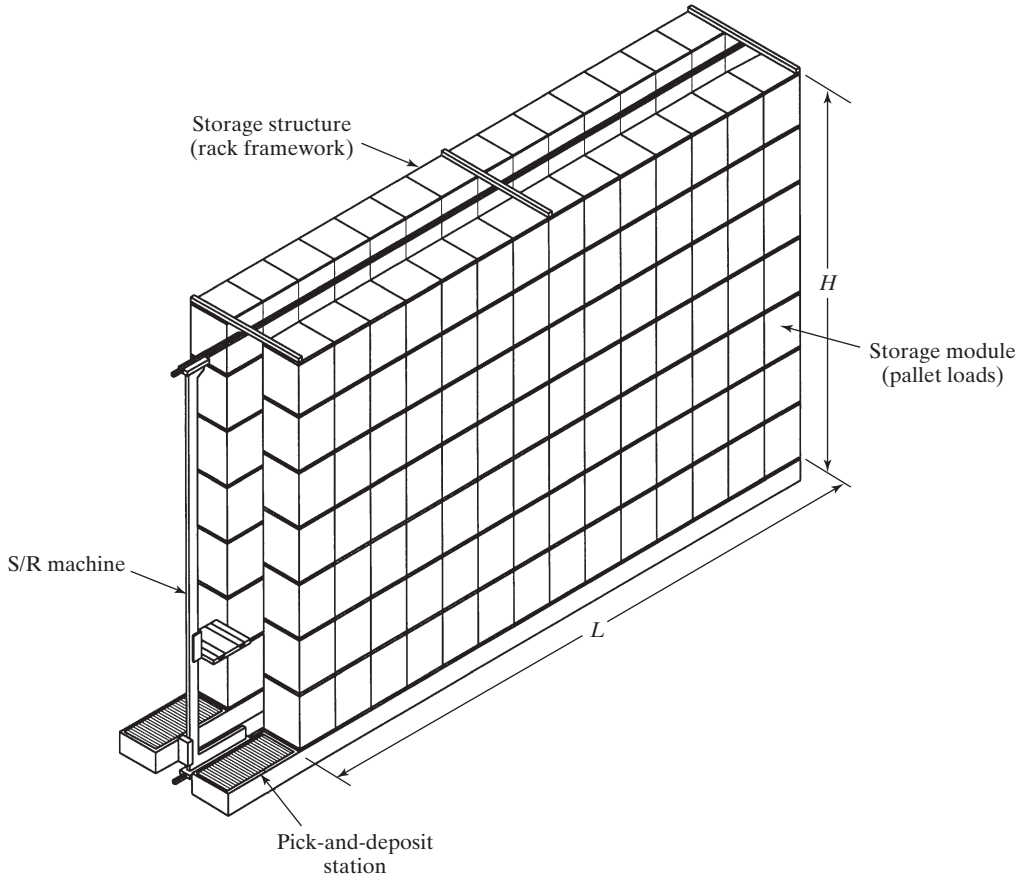


Figure 11.5 One aisle of a unit load automated storage/retrieval system (AS/RS).

storage baskets attached to a chain-driven conveyor that revolves around an oval track loop to deliver the baskets to a load/unload station, as in Figure 11.6. The differences between an AS/RS and a carousel storage system are summarized in Table 11.4. Both types include horizontal and vertical structures, with the horizontal configuration being much more common in both cases.

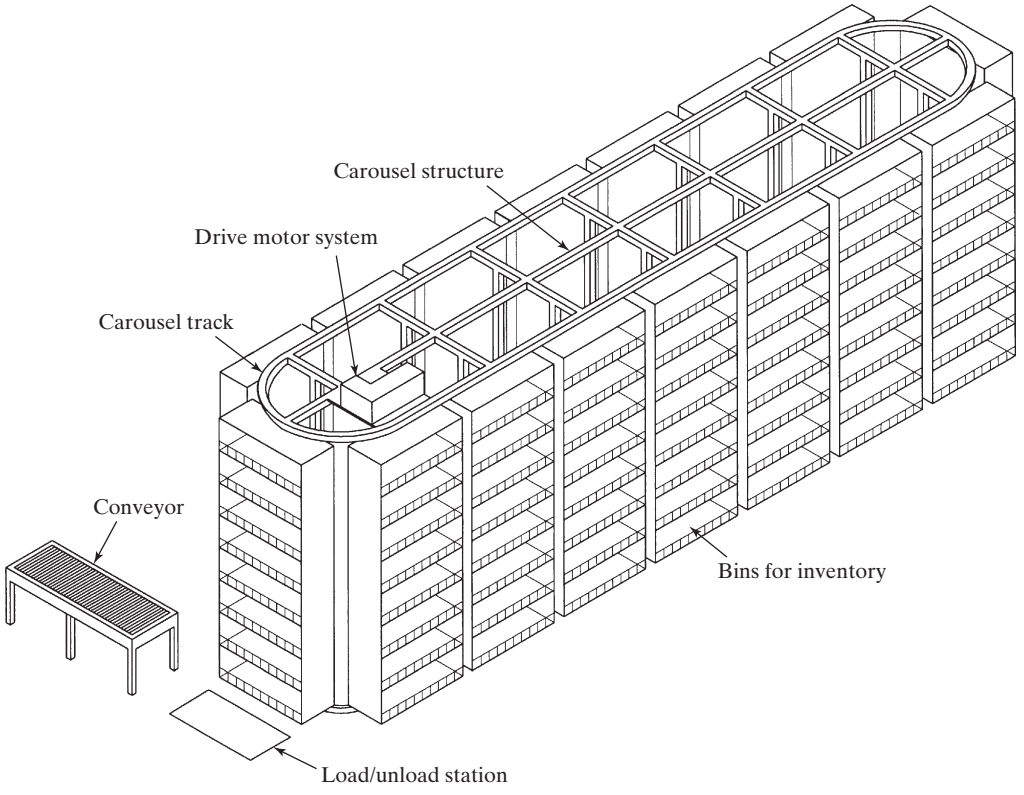


Figure 11.6 A horizontal storage carousel.

TABLE 11.4 Differences Between a Fixed-Aisle AS/RS and a Carousel Storage System

Feature	Fixed-Aisle AS/RS	Carousel Storage System
Storage structure	Rack system to support pallets or shelf system to support tote bins	Baskets suspended from overhead conveyor trolleys
Motions	Linear motions of S/R machine	Revolution of conveyor trolleys around oval track
Storage/retrieval operation	S/R machine travels to compartments in rack structure	Conveyor revolves to bring baskets to load/unload station
Replication of storage capacity	Multiple aisles, each consisting of rack structure and S/R machine	Multiple carousels, each consisting of oval track and storage bins

11.3.1 Fixed-Aisle Automated Storage/Retrieval Systems

A fixed-aisle automated storage/retrieval system (AS/RS) is a storage system consisting of one or more aisles of storage racks attended by storage/retrieval machines, usually one S/R machine per aisle. The system performs storage and retrieval operations with speed and accuracy under a defined degree of automation. The S/R machines (sometimes referred to as *cranes*) are used to deliver materials to the storage racks and to retrieve

materials from the racks. Each AS/RS aisle has one or more input/output stations where materials are delivered into the storage system and withdrawn from it. The input/output stations are called *pickup-and-deposit stations* (P&D stations) in AS/RS terminology. P&D stations can be manually operated or interfaced to some form of automated transport system such as a conveyor or an AGVS (automated guided vehicle system).

Figure 11.5 shows one aisle of an AS/RS that handles and stores unit loads on pallets. A wide range of automation is found in commercially available AS/RSs. At the most sophisticated level, the operations are computer controlled and fully integrated with factory and/or warehouse operations. At the other extreme, human workers control the equipment and perform the storage/retrieval transactions. Fixed-aisle automated storage/retrieval systems are custom-designed for each application, although the designs are based on standard modular components available from each respective AS/RS supplier.

AS/RS Types. Several important categories of fixed-aisle automated storage/retrieval system can be distinguished. The following are the principal types:

- *Unit load AS/RS.* The unit load AS/RS is typically a large automated system designed to handle unit loads stored on pallets or in other standard containers. The system is computer controlled, and the S/R machines are automated and designed to handle the unit load containers. The AS/RS pictured in Figure 11.5 is a unit load system. Other systems described below represent variations of the unit load AS/RS.
- *Deep-lane AS/RS.* The deep-lane AS/RS is a high-density unit load storage system that is appropriate when large quantities of stock are stored, but the number of separate stock types (SKUs) is relatively small. Instead of storing each unit load so that it can be accessed directly from the aisle (as in a conventional unit load system), the deep-lane system stores ten or more loads in a single rack, one load behind the next. Each rack is designed for "flow-through," with input on one side and output on the other side. Loads are picked up from one side of the rack by an S/R-type machine designed for retrieval, and another machine inputs loads on the entry side of the rack.
- *Miniload AS/RS.* This storage system is used to handle small loads (individual parts or supplies) that are contained in bins or drawers in the storage system. The S/R machine is designed to retrieve the bin and deliver it to a P&D station at the end of the aisle so that individual items can be withdrawn from the bins. The P&D station is usually operated by a human worker. The bin or drawer must then be returned to its location in the system. A miniload AS/RS is generally smaller than a unit load AS/RS and is often enclosed for security of the items stored.
- *Man-on-board S/RS.* A man-on-board (also called *man-aboard*) storage/retrieval system represents an alternative approach to the problem of retrieving individual items from storage. In this system, a human operator rides on the carriage of the S/R machine. Whereas the miniload system delivers an entire bin to the end-of-aisle pick station and must return it subsequently to its proper storage compartment, with the man-on-board system the worker picks individual items directly at their storage locations. This offers an opportunity to increase system throughput.
- *Automated item retrieval system.* These storage systems are also designed for retrieval of individual items or small product cartons; however, the items are stored

in lanes rather than bins or drawers. When an item is retrieved, it is pushed from its lane and drops onto a conveyor for delivery to the pickup station. The operation is somewhat similar to a candy vending machine, except that an item retrieval system has more storage lanes and a conveyor to transport items to a central location. The supply of items in each lane is periodically replenished, usually from the rear of the system so that there is flow-through of items, thus permitting first-in/first-out inventory rotation.

- *Vertical lift modules* (VLM). All of the preceding AS/RS types are designed around a fixed horizontal aisle. The same principle of using a center aisle to access loads is used in a VLM except that the aisle is vertical. The structure consists of two columns of trays that are accessed by an S/R machine (also called an *extractor*) that delivers the trays one-by-one to a load/unload station at floor level. Vertical lift modules, some with heights of 10 m (30 ft) or more, are capable of holding large inventories while saving valuable floor space in the facility.

AS/RS Applications. Most applications of fixed-aisle automated storage/retrieval systems have been associated with warehousing and distribution operations, but they can also be used to store raw materials and work-in-process in manufacturing. Three application areas can be distinguished: (1) unit load storage and handling, (2) order picking, and (3) work-in-process storage. Unit load storage and retrieval applications are represented by the unit load AS/RS and deep-lane storage systems. These kinds of applications are commonly found in warehousing for finished goods in a distribution center, but rarely in manufacturing. Deep-lane systems are used in the food industry. Order picking involves retrieving materials in less than full unit load quantities. Miniload, man-on-board, and item retrieval systems are used for this application area.

Work-in-process (WIP) storage is a more recent application of automated storage technology. While it is desirable to minimize the amount of work-in-process, WIP is unavoidable and must be effectively managed. Automated storage systems, either fixed-aisle storage/retrieval systems or carousel systems, represent an efficient way to store materials between processing steps, particularly in batch and job shop production.

The merits of an automated WIP storage system for batch and job shop production can best be seen by comparing it with the traditional way of dealing with work-in-process. The typical factory contains multiple work cells, each performing its own processing operations on different parts. At each cell, orders consisting of one or more parts are waiting on the plant floor to be processed, while other completed orders are waiting to be moved to the next cell in the sequence. It is not unusual for a plant engaged in batch production to have hundreds of orders in progress simultaneously, all of which represent work-in-process. The disadvantages of keeping all of this inventory in the plant include (1) time spent searching for orders, (2) parts or even entire orders becoming temporarily or permanently lost, sometimes resulting in repeat orders to reproduce the lost parts, (3) orders not being processed according to their relative priorities at each cell, and (4) orders spending too much time in the factory, causing customer deliveries to be late. These problems indicate poor control of work-in-process.

Automated storage/retrieval systems are also used in high-production operations. In the automobile industry, some final assembly plants use large capacity AS/RSs to temporarily store car and small truck bodies between major assembly steps. The AS/RS can be used for staging and sequencing the work units according to the most efficient production schedule [1].

Automated storage systems help to regain control over WIP. Reasons that justify the installation of automated storage systems for work-in-process include:

- *Buffer storage in production*. A storage system can be used as a buffer storage zone between two processes whose production rates are significantly different. A simple example is a two-process sequence in which the first processing operation feeds a second process, which operates at a slower production rate. The first operation requires only one shift to meet production requirements, while the second step requires two shifts to produce the same number of units. An in-process buffer is needed between these operations to temporarily store the output of the first process.
- *Support of just-in-time delivery*. Just-in-time (JIT) is a manufacturing strategy in which parts required in production and/or assembly are received immediately before they are needed in the plant (Section 26.2). This results in a significant dependency of the factory on its suppliers to deliver the parts on time for use in production. To reduce the chance of stock-outs due to late supplier deliveries, some plants have installed automated storage systems as storage buffers for incoming materials. Although this approach subverts the objectives of JIT, it also reduces some of its risks.
- *Kitting of parts for assembly*. The storage system is used to store components for assembly of products or subassemblies. When an order is received, the required components are retrieved, collected into kits (tote pans), and delivered to the production floor for assembly.
- *Compatible with automatic identification systems*. Automated storage systems can be readily interfaced with automatic identification devices such as bar code readers. This allows loads to be stored and retrieved without needing human operators to identify the loads.
- *Computer control and tracking of materials*. Combined with automatic identification, an automated WIP storage system permits the location and status of work-in-process to be known.
- *Support of factory-wide automation*. Given the need for storage of work-in-process in batch production, an appropriately sized automated storage system can be an important subsystem in a fully automated factory.

Components and Operating Features of an AS/RS. Virtually all of the fixed-aisle automated storage/retrieval systems described earlier consist of the following components, shown in Figure 11.5: (1) storage structure, (2) S/R machine, (3) storage modules (e.g., pallets for unit loads), and (4) one or more pickup-and-deposit stations. In addition, a control system is required to operate the AS/RS.

The storage structure is the rack framework, made of fabricated steel, which supports the loads contained in the AS/RS. The individual storage compartments in the structure must be designed to hold the storage modules used to contain the stored materials. The rack structure may also be used to support the roof and siding of the building in which the AS/RS resides. Another function of the storage structure is to support the aisle hardware required to align the S/R machines with respect to the storage compartments of the AS/RS. This hardware includes guide rails at the top and bottom of the structure as well as end stops and other features required for safe operation.

The S/R machine is used to accomplish storage transactions, delivering loads from the input station into storage, and retrieving loads from storage and delivering them to the output station. To perform these transactions, the storage/retrieval machine must be

capable of horizontal and vertical travel to align its carriage (which carries the load) with the storage compartment in the rack structure. The S/R machine consists of a rigid mast on which is mounted an elevator system for vertical motion of the carriage. Wheels are attached at the base of the mast to permit horizontal travel along a rail system that runs the length of the aisle. A parallel rail at the top of the storage structure is used to maintain alignment of the mast and carriage with respect to the rack structure.

The carriage includes a shuttle mechanism to move loads into and from their storage compartments. The design of the shuttle must also permit loads to be transferred from the S/R machine to the P&D station or other material handling interface with the AS/RS. The carriage and shuttle are positioned and actuated automatically in the usual AS/RS. Man-on-board S/R machines are equipped for a human operator to ride on the carriage.

To accomplish the desired motions of the S/R machine, three drive systems are required: horizontal movement of the mast, vertical movement of the carriage, and shuttle transfer between the carriage and a storage compartment. Modern S/R machines are available with horizontal speeds up to 200 m/min (600 ft/min) along the aisle and vertical or lift speeds up to around 50 m/min (150 ft/min). These speeds determine the time required for the carriage to travel from the P&D station to a particular location in the storage aisle. Acceleration and deceleration have a more significant effect on travel time over short distances. The shuttle transfer is accomplished by any of several mechanisms, including forks (for pallet loads) and friction devices for flat-bottom tote pans.

The storage modules are the unit load containers of the stored material. These include pallets, steel wire baskets and containers, plastic tote pans, and special drawers (used in miniload systems). The storage modules are a standard size that is designed to fit in the storage compartments of the rack structure and can be handled automatically by the carriage shuttle of the S/R machine.

The pickup-and-deposit station is where loads are transferred into and out of the AS/RS. It is generally located at the end of the aisle for access by the external handling system that brings loads to the AS/RS and takes loads away. Pickup stations and deposit stations may be located at opposite ends of the storage aisle or combined at the same location. This depends on the origin of incoming loads and the destination of output loads. A P&D station must be compatible with both the S/R machine shuttle and the external handling system. Common methods to handle loads at the P&D station include manual load/unload, forklift truck, conveyor (e.g., roller), and AGVS.

The principal AS/RS controls problem is positioning the S/R machine within an acceptable tolerance at a storage compartment in the rack structure to deposit or retrieve a load. The locations of materials stored in the system must be determined to direct the S/R machine to a particular storage compartment. Within a given aisle in the AS/RS, each compartment is identified by its horizontal and vertical positions and whether it is on the right side or left side of the aisle. A scheme based on alphanumeric codes can be used for this purpose. Using this location identification scheme, each unit of material stored in the system can be referenced to a particular location in the aisle. The record of these locations is called the "item location file." Each time a storage transaction is completed, the transaction must be recorded in the item location file.

Computer controls and programmable logic controllers are used to determine the required location and guide the S/R machine to its destination. Computer control permits the physical operation of the AS/RS to be integrated with the supporting information and record-keeping system. It allows storage transactions to be entered in real time, inventory records to be accurately maintained, system performance to be monitored, and communications to be facilitated with other factory computer systems. These automatic controls

can be superseded or supplemented by manual controls when required under emergency conditions or for man-on-board operation of the machine.

11.3.2 Carousel Storage Systems

A carousel storage system consists of a series of bins or baskets attached to a chain-driven conveyor that revolves around an oval track loop to deliver the baskets to a load/unload station, as depicted in Figure 11.6. The purpose of the chain conveyor is to position bins at a load/unload station at the end of the oval. The operation is similar to the powered overhead rack system used by dry cleaners to deliver finished garments to the front of the store. Most carousels are operated by a human worker at the load/unload station. The worker activates the powered carousel to deliver a desired bin to the station. One or more parts are removed from or added to the bin, and then the cycle is repeated. Some carousels are automated by using transfer mechanisms at the load/unload station to move loads into and from the carousel.

Carousel Technology. Carousels can be classified as horizontal or vertical. The more common horizontal configuration shown in Figure 11.6 comes in a variety of sizes, ranging between 3 m (10 ft) and 30 m (100 ft) in length. Carousels at the upper end of the range have higher storage density, but the average access cycle time is greater. Accordingly, most carousels are 10–16 m (30–50 ft) long to achieve a proper balance between these competing factors.

A horizontal carousel storage system consists of welded steel framework that supports the oval rail system. The carousel can be either an overhead system (called a *top-driven* unit) or a floor-mounted system (called a *bottom-driven* unit). In the top-driven unit, a motorized pulley system is mounted at the top of the framework and drives an overhead trolley system. The bins are suspended from the trolleys. In the bottom-driven unit, the pulley drive system is mounted at the base of the frame, and the trolley system rides on a rail in the base. This provides more load-carrying capacity for the carousel storage system. It also eliminates the problem of dirt and oil dripping from the overhead trolley system onto the storage contents in top-driven systems.

The design of the individual bins and baskets of the carousel must be consistent with the loads to be stored. Bin widths range from about 50 to 75 cm (20 to 30 in), and depths are up to about 55 cm (22 in). Heights of horizontal carousels are typically 1.8–2.4 m (6–8 ft). Standard bins are made of steel wire to increase operator visibility.

Vertical carousels are constructed to operate around a vertical conveyor loop. They occupy much less floor space than the horizontal configuration, but require sufficient overhead space. The ceiling of the building limits the height of vertical carousels, and therefore their storage capacity is typically lower than for the average horizontal carousel.

Controls for carousel storage systems range from manual call controls to computer control. Manual controls include foot pedals, hand switches, and specialized keyboards. Foot pedal control allows the operator at the pick station to rotate the carousel in either direction to the desired bin position. Hand control involves use of a hand-operated switch that is mounted on an arm projecting from the carousel frame within easy reach of the operator. Again, bidirectional control is the usual mode of operation. Keyboard control permits a greater variety of control features than the previous control types. When the operator enters the desired bin position, the carousel is programmed to deliver the bin to the pick station by the shortest route (i.e., clockwise or counterclockwise motion of the carousel).

Computer control increases opportunities for automation of the mechanical carousel and for management of the inventory records. On the mechanical side, automatic loading and unloading is available on modern carousel storage systems. This allows the carousel to be interfaced with automated handling systems without the need for human participation in the load/unload operations. Data management features provided by computer control include the capability to maintain data on bin locations, items in each bin, and other inventory control records.

Carousel Applications. Carousel storage systems provide a relatively high throughput and are often an attractive alternative to a miniload AS/RS in manufacturing operations where their relatively low cost, versatility, and high reliability are recognized. Typical applications of carousel storage systems include (1) storage and retrieval operations, (2) transport and accumulation, (3) work-in-process, and (4) specialized uses.

Storage and retrieval operations can be efficiently accomplished using carousels when individual items must be selected from groups of items in storage. Sometimes called "pick and load" operations, these procedures are common in order-picking of tools in a toolroom, raw materials in a stockroom, service parts or other items in a wholesale firm, and work-in-process in a factory. In small electronics assembly, carousels are used for kitting of parts to be transported to assembly workstations.

In transport and accumulation applications, the carousel is used to transport and/or sort materials as they are stored. One example of this is in progressive assembly operations where the workstations are located around the periphery of a continuously moving carousel, and the workers have access to the individual storage bins of the carousel. They remove work from the bins to complete their own respective assembly tasks, then place their work into another bin for the next operation at some other workstation. Another example of transport and accumulation applications is sorting and consolidation of items. Each bin is defined for collecting the items of a particular type or customer. When the bin is full, the collected load is removed for shipment or other disposition.

Carousel storage systems often compete with automated storage and retrieval systems for applications where work-in-process is to be temporarily stored. Applications of carousel systems in the electronics industry are common.

One example of specialized use of carousel systems is electrical testing of products or components, where the carousel is used to store the item during testing for a specified period of time. The carousel is programmed to deliver the items to the load/unload station at the conclusion of the test period.

11.4 ANALYSIS OF STORAGE SYSTEMS

Several aspects of the design and operation of a storage system are susceptible to quantitative engineering analysis. This section examines capacity sizing and throughput performance for the two types of automated storage systems.

11.4.1 Fixed-Aisle Automated Storage/Retrieval Systems

While the methods developed here are specifically for fixed-aisle automated storage/retrieval systems, similar approaches can be used for analyzing traditional storage facilities, such as warehouses consisting of pallet racks and bulk storage.

Sizing the AS/RS Rack Structure. The total storage capacity of one storage aisle depends on how many storage compartments are arranged horizontally and vertically in the aisle, as indicated in Figure 11.7. This can be expressed as

$$\text{Capacity per aisle} = 2n_y n_z \tag{11.1}$$

where n_y = number of load compartments along the length of the aisle, and n_z = number of load compartments that make up the height of the aisle. The constant, 2, accounts for the fact that loads are contained on both sides of the aisle.

If the AS/RS is designed for a standard size unit load, then the compartment size must be standardized and its dimensions must be larger than the unit load dimensions. Let x and y = the depth and width dimensions of a unit load (e.g., a standard pallet size as given in Table 10.3), and z = the height of the unit load. The width, length, and height of the rack structure of the AS/RS aisle are related to the unit load dimensions and number of compartments as follows [6]:

$$W = 3(x + a) \tag{11.2a}$$

$$L = n_y(y + b) \tag{11.2b}$$

$$H = n_z(z + c) \tag{11.2c}$$

where W, L, and H are the width, length, and height of one aisle of the AS/RS rack structure, mm (in); x, y, and z are the dimensions of the unit load, mm (in); and a, b, and c are

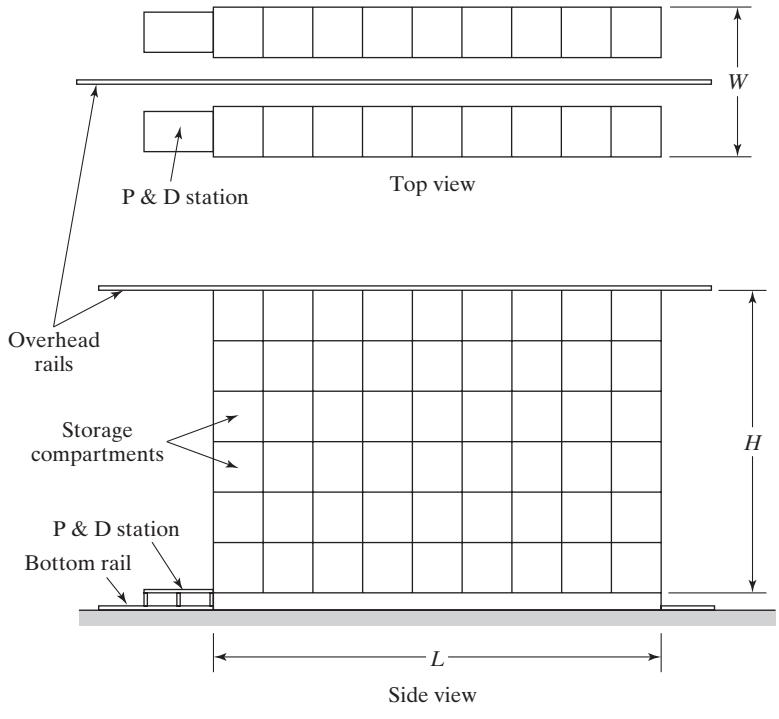


Figure 11.7 Top and side views of a unit load AS/RS, with nine storage compartments horizontally ($n_y = 9$) and six compartments vertically ($n_z = 6$).

allowances designed into each storage compartment to provide clearance for the unit load and to account for the size of the supporting beams in the rack structure, mm (in). For the case of unit loads contained on standard pallets, recommended values for the allowances [6] are: $a = 150$ mm (6 in), $b = 200$ mm (8 in), and $c = 250$ mm (10 in). For an AS/RS with multiple aisles, W is simply multiplied by the number of aisles to obtain the overall width of the storage system. The rack structure is built above floor level by 300–600 mm (12–24 in), and the length of the AS/RS extends beyond the rack structure to provide space for the P&D station.

EXAMPLE 11.2 Sizing an AS/RS System

Each aisle of a four-aisle AS/RS contains 60 storage compartments in the length direction and 12 compartments vertically. All storage compartments are the same size to accommodate standard-size pallets of dimensions: $x = 42$ in and $y = 48$ in. The height of a unit load $z = 36$ in. Using the allowances $a = 6$ in, $b = 8$ in, and $c = 10$ in, determine (a) how many unit loads can be stored in the AS/RS and (b) the width, length, and height of the AS/RS.

Solution: (a) The storage capacity is given by Equation (11.1): Capacity per aisle $= 2(60)(12) = 1{,}440$ unit loads. With four aisles, the total capacity is

$$\text{AS/RS capacity} = 4(1440) = \textbf{5,760 unit loads}$$

(b) From Equations (11.2), the dimensions of the storage rack structure can be computed as:

$$W = 3(42 + 6) = 144 \text{ in} = 12 \text{ ft/aisle}$$

$$\text{Overall width of the AS/RS} = 4(12) = \textbf{48 ft}$$

$$L = 60(48 + 8) = 3{,}360 \text{ in} = \textbf{280 ft}$$

$$H = 12(36 + 10) = 552 \text{ in} = \textbf{46 ft}$$

AS/RS Throughput. System throughput is defined as the hourly rate of S/R transactions that the automated storage system can perform (Section 11.1.1). A transaction involves depositing a load into storage or retrieving a load from storage. Either of these transactions alone is accomplished in a single-command cycle. A dual-command cycle accomplishes both transaction types in one cycle; because this reduces travel time per transaction, throughput is increased by using dual-command cycles.

Several methods are available to compute AS/RS cycle times to estimate throughput performance. The method presented here is recommended by the Material Handling Institute (MHI) [2]. It assumes (1) randomized storage of loads in the AS/RS (i.e., any compartment in the storage aisle is equally likely to be selected for a transaction), (2) storage compartments of equal size, (3) the P&D station located at the base and end of the aisle, (4) constant horizontal and vertical speeds of the S/R machine, and (5) simultaneous horizontal and vertical travel. For a single-command cycle, the load to be

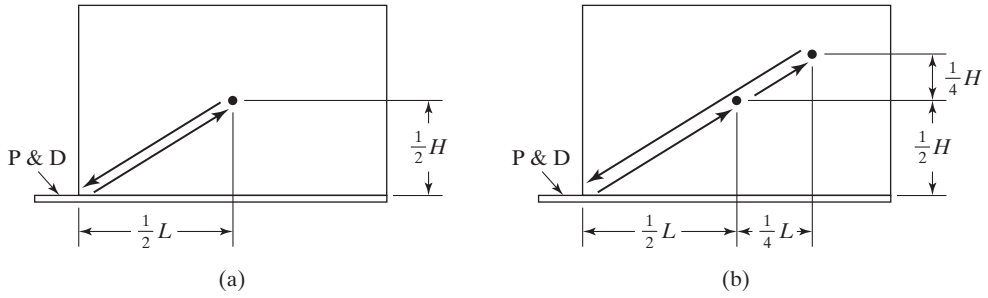


Figure 11.8 Assumed travel trajectory of the S/R machine for (a) single-command cycle and (b) dual-command cycle.

entered or retrieved is assumed to be located at the center of the rack structure, as in Figure 11.8(a). Thus, the S/R machine must travel half the length and half the height of the AS/RS, and it must return the same distance. The single-command cycle time can therefore be expressed by

$$T_{cs} = 2\,\text{Max}\left\{\frac{0.5L}{v_y}, \frac{0.5H}{v_z}\right\} + 2T_{pd} = \text{Max}\left\{\frac{L}{v_y}, \frac{H}{v_z}\right\} + 2T_{pd} \tag{11.3a}$$

where T_{cs} = cycle time of a single-command cycle, min/cycle; L = length of the AS/RS rack structure, m (ft); v_y = velocity of the S/R machine along the length of the AS/RS, m/min (ft/min); H = height of the rack structure, m (ft); v_z = velocity of the S/R machine in the vertical direction of the AS/RS, m/min (ft/min); and T_{pd} = pickup-and-deposit time, min. Two P&D times are required per cycle, representing load transfers to and from the S/R machine.

For a dual-command cycle, the S/R machine is assumed to travel to the center of the rack structure to deposit a load, and then it travels to 3/4 the length and height of the AS/RS to retrieve a load, as in Figure 11.8(b). Thus, the total distance traveled by the S/R machine is 3/4 the length and 3/4 the height of the rack structure, and back. In this case, cycle time is given by

$$T_{cd} = 2\,\text{Max}\left\{\frac{0.75L}{v_y}, \frac{0.75H}{v_z}\right\} + 4T_{pd} = \text{Max}\left\{\frac{1.5L}{v_y}, \frac{1.5H}{v_z}\right\} + 4T_{pd} \tag{11.3b}$$

where T_{cd} = cycle time for a dual-command cycle, min/cycle; and the other terms are defined earlier.

System throughput depends on the relative numbers of single- and dual-command cycles performed by the system. Let R_{cs} = number of single-command cycles performed per hour, and R_{cd} = number of dual-command cycles per hour at an assumed utilization level. An equation for the amounts of time spent in performing single-command and dual-command cycles each hour can be formulated as follows:

$$R_{cs}T_{cs} + R_{cd}T_{cd} = 60U \tag{11.4}$$

where U = system utilization during the hour. The right-hand side of the equation gives the total number of minutes of operation per hour. To solve Equation (11.4), the relative

proportions of R_{cs} and R_{cd} must be determined, or assumptions about these proportions must be made. When solved, the total hourly cycle rate is given by

$$R_c = R_{cs} + R_{cd} \tag{11.5}$$

where R_c = total S/R cycle rate, cycles/hr. Note that the total number of storage and retrieval transactions per hour will be greater than this value unless $R_{cd} = 0$, since there are two transactions accomplished in each dual-command cycle. Let R_t = the total number of transactions performed per hour; then

$$R_t = R_{cs} + 2R_{cd} \tag{11.6}$$

EXAMPLE 11.3 AS/RS Throughput Analysis

Consider the AS/RS from Example 11.2, in which an S/R machine is used for each aisle. The length of the storage aisle = 280 ft and its height = 46 ft. Suppose horizontal and vertical speeds of the S/R machine are 200 ft/min and 75 ft/min, respectively. The S/R machine requires 20 sec to accomplish a P&D operation. Determine (a) the single-command and dual-command cycle times per aisle and (b) throughput per aisle under the assumptions that storage system utilization = 90% and the number of single-command and dual-command cycles are equal.

Solution: (a) The single- and dual-command cycle times are calculated by Equations (11.3):

$$T_{cs} = \text{Max}\{280/200, 46/75\} + 2(20/60) = \textbf{2.066 min/cycle}$$

$$T_{cd} = \text{Max}\{1.5 \times 280/200, 1.5 \times 46/75\} + 4(20/60) = \textbf{3.432 min/cycle}$$

(b) From Equation (11.4), the single-command and dual-command activity levels each hour can be established as follows:

$$2.066R_{cs} + 3.432R_{cd} = 60(0.90) = 54.0 \text{ min}$$

According to the problem statement, the number of single-command cycles is equal to the number of dual-command cycles. Thus, $R_{cs} = R_{cd}$. Substituting this relation into the above equation,

$$2.066R_{cs} + 3.432R_{cs} = 54$$

$$5.498R_{cs} = 54$$

$$R_{cs} = 9.822 \text{ single command cycles/hr}$$

$$R_{cd} = R_{cs} = 9.822 \text{ dual command cycles/hr}$$

System throughput is equal to the total number of S/R transactions per hour from Equation (11.6):

$$R_t = R_{cs} + 2R_{cd} = 29.46 \text{ transactions/hr}$$

With four aisle, R_t for the AS/RS = $4(29.46)$ = **117.84 transactions/hr**

11.4.2 Carousel Storage Systems

In this section, the corresponding capacity and throughput relationships for a carousel storage system are developed. Because of their construction, carousel systems do not possess nearly the volumetric capacity of an AS/RS. However, according to sample calculations, a typical carousel system is likely to have a higher throughput rate than an AS/RS.

Storage Capacity. The size and capacity of a carousel can be determined with reference to Figure 11.9. Individual bins or baskets are suspended from carriers that revolve around an oval rail with circumference given by

$$C = 2(L - W) + \pi W \tag{11.7}$$

where C = circumference of the oval conveyor track, m (ft); and L and W are the length and width of the track oval, m (ft).

The capacity of the carousel system depends on the number and size of the bins (or baskets) in the system. Assuming standard-size bins are used, each of a certain volumetric capacity, the number of bins can be used as the measure of capacity. As illustrated in Figure 11.9, the number of bins hanging vertically from each carrier is n_b and n_c = the number of carriers around the periphery of the rail. Thus,

$$\text{Total number of bins} = n_c n_b \tag{11.8}$$

The carriers are separated by a certain distance so that they do not interfere with each other while traveling around the ends of the carousel. Let s_c = the center-to-center spacing of carriers along the oval track. Then the following relationship must be satisfied by the values of s_c and n_c:

$$s_c n_c = C \tag{11.9}$$

where C = circumference, m(ft); s_c = carrier spacing, m/carrier (ft/carrier); and n_c = number of carriers, which must be an integer value.

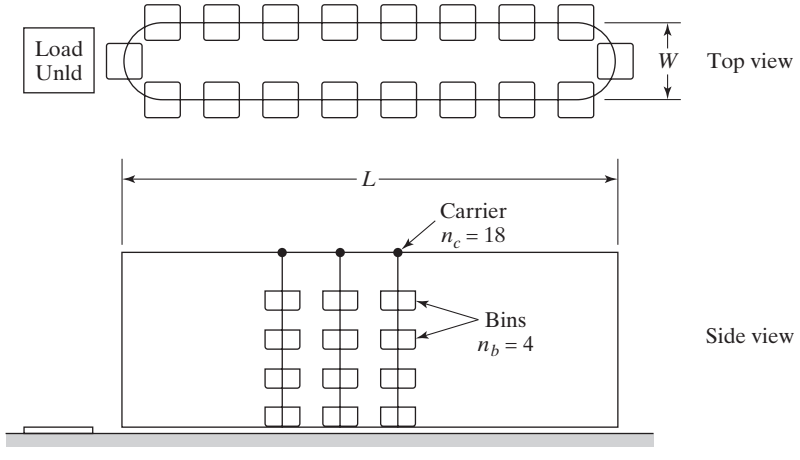


Figure 11.9 Top and side views of horizontal storage carousel with 18 carriers ($n_c = 18$) and four bins/carrier ($n_b = 4$). Key: Unld = unload.

Throughput Analysis. The storage/retrieval cycle time can be derived based on the following assumptions. First, only single-command cycles are performed; a bin is accessed in the carousel either to put items into storage or to retrieve one or more items from storage. Second, the carousel operates with a constant speed v_c; acceleration and deceleration effects are ignored. Third, random storage is assumed; that is, any location around the carousel is equally likely to be selected for an S/R transaction. And fourth, the carousel can move in either direction. Under this last assumption of bidirectional travel, it can be shown that the mean travel distance between the load/unload station and a bin randomly located in the carousel is $C/4$. Thus, the S/R cycle time is given by

$$T_c = \frac{C}{4v_c} + T_{pd} \tag{11.10}$$

where T_c = S/R cycle time, min; C = carousel circumference as given by Equation (11.7), m (ft); v_c = carousel velocity, m/min (ft/min); and T_{pd} = the average time required to pick or deposit items each cycle by the operator at the load/unload station, min. The number of transactions accomplished per hour is the same as the number of cycles and is given by the following:

$$R_t = R_c = \frac{60}{T_c} \tag{11.11}$$

EXAMPLE 11.4 Carousel Operation

The oval rail of a carousel storage system has length = 12 m and width = 1 m. There are 75 carriers equally spaced around the oval. Suspended from each carrier are six bins. Each bin has volumetric capacity = 0.026 m^3. Carousel speed = 20 m/min. Average P&D time for a retrieval = 20 sec. Determine (a) volumetric capacity of the storage system and (b) hourly retrieval rate of the storage system.

Solution: (a) Total number of bins in the carousel is

$$n_c n_b = 75 \times 6 = 450 \text{ bins}$$

Total volumetric capacity = $450(0.026)$ = **11.7 m^3**

(b) The circumference of the carousel rail is determined by Equation (11.7):

$$C = 2(12 - 1) + 1\pi = 25.14 \text{ m}$$

Cycle time per retrieval is given by Equation (11.10):

$$T_c = \frac{25.14}{4(20)} + 20/60 = 0.647 \text{ min}$$

Expressing throughput as an hourly rate, $R_t = 60/0.647$ = **92.7 retrieval transactions/hr**

Chapter 18

Group Technology and Cellular Manufacturing

CHAPTER CONTENTS

18.1 Part Families and Machine Groups
- 18.1.1 What is a Part Family?
- 18.1.2 Intuitive Grouping
- 18.1.3 Parts Classification and Coding
- 18.1.4 Production Flow Analysis

18.2 Cellular Manufacturing
- 18.2.1 Composite Part Concept
- 18.2.2 Machine Cell Design

18.3 Applications of Group Technology

18.4 Analysis of Cellular Manufacturing
- 18.4.1 Rank-order Clustering
- 18.4.2 Arranging Machines in a GT Cell
- 18.4.3 Performance Metrics in Cell Operations

Appendix 18A: Opitz Parts Classification and Coding System

Batch manufacturing is estimated to be the most common form of production in the United States, constituting more than 50% of total manufacturing activity. It is important to make mid-volume manufacturing, which is traditionally accomplished in batches, as efficient and productive as possible. In addition, there has been a trend to integrate the design and manufacturing functions in a firm. An approach directed at both of these objectives is group technology (GT).

Group technology is a manufacturing philosophy in which similar parts are identified and grouped together to take advantage of their similarities in design and production.

Similar parts are arranged into part families, where each part family possesses similar design and/or manufacturing characteristics. For example, a plant producing 10,000 different part numbers may be able to group the vast majority of these parts into 30 or 40 distinct families. It is reasonable to believe that the processing of each member of a given family is similar, and this should result in manufacturing efficiencies. The efficiencies are generally achieved by arranging the production equipment into cells (machine groups) to facilitate work flow. Organizing the production equipment into machine cells, where each cell specializes in the production of a part family, is called *cellular manufacturing*. The origins of group technology and cellular manufacturing can be traced to around 1925 (Historical Note 18.1).

Group technology and cellular manufacturing are applicable to a wide variety of production situations. The following conditions are when GT is most appropriate:

- The plant currently uses traditional batch production and a process-type layout, which results in much material handling, high in-process inventory, and long manufacturing lead times.
- It is possible to group the parts into part families. This is a necessary condition. Each GT machine cell is designed to produce a given part family, or a limited collection of part families, so it must be possible to identify part families made in the plant. Fortunately, in the typical mid-volume production plant, most of the parts can be grouped into part families.

Historical Note 18.1 Group Technology (GT)

In 1925, R. Flanders of the United States presented a paper before the American Society of Mechanical Engineers that described a way of organizing manufacturing at Jones and Lamson Machine Company that would today be called group technology. In 1937, A. Sokolovskiy of the former Soviet Union described the essential features of group technology by proposing that parts of similar configuration be produced by a standard process sequence, thus permitting flow-line techniques to be used for work normally accomplished by batch production. In 1949, A. Korling of Sweden presented a paper in Paris on "group production," whose principles are an adaptation of production line techniques to batch manufacturing. In the paper, he described how to decentralize work into independent groups, each containing the machines and tooling to produce "a special category of parts."

In 1959, researcher S. Mitrofanov of the Soviet Union published a book titled *Scientific Principles of Group Technology*. The book was widely read and is considered responsible for over 800 plants in the Soviet Union using group technology by 1965. Another researcher, H. Opitz in Germany, studied parts manufactured by the German machine tool industry and developed the well-known parts classification and coding system for machined parts that bears his name (Appendix 18A).

In the United States, the first application of group technology was at the Langston Division of Harris-Intertype in Camden, New Jersey, in the late 1960s. Traditionally a machine shop arranged as a process-type layout, the company reorganized into "family of parts" lines, each of which specialized in producing a given part configuration. Part families were identified by taking photos of about 15% of the parts made in the plant and grouping them into families. When the changes were implemented, productivity was improved by 50% and lead times were reduced from weeks to days.

There are two major tasks that a company must undertake when it implements group technology. These tasks represent significant obstacles to the application of GT.

1. *Identifying the part families.* If the plant makes 10,000 different parts, reviewing all of the part drawings and grouping the parts into families is a substantial and time-consuming task.
2. *Rearranging production machines into machine cells.* It is time-consuming and costly to plan and accomplish this rearrangement, and the machines are not producing during the changeover.

Group technology and cellular manufacturing offer substantial benefits to companies that have the perseverance to implement them:

- GT promotes standardization of tooling, fixturing, and setups.
- Material handling is reduced because the distances within a machine cell are much shorter than within the entire factory.
- Process planning and production scheduling are simplified.
- Setup times are reduced, resulting in lower manufacturing lead times.
- Work-in-process is reduced.
- Worker satisfaction usually improves when workers collaborate in a GT cell.
- Higher quality work is accomplished.

18.1 PART FAMILIES AND MACHINE GROUPS

The logical starting point in this chapter's coverage of group technology, cellular manufacturing, and related topics is the underlying concept of part families. This section describes part families and the grouping of machines into cells that specialize in producing those families.

18.1.1 What is a Part Family?

A part family is a collection of parts that are similar either in geometric shape and size or in the processing steps required in their manufacture. The parts within a family are different, but their similarities are close enough to merit their inclusion as members of the part family. Figures 18.1 and 18.2 show two different part families. The two

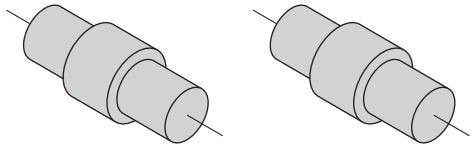


Figure 18.1 Two parts of identical shape and size but different manufacturing requirements: (a) 1,000,000 pc/yr, tolerance = ± 0.010 in., material = 1015 CR steel, nickel plate; and (b) 100 pc/yr, tolerance = ± 0.001 in., material = 18−8 stainless steel.

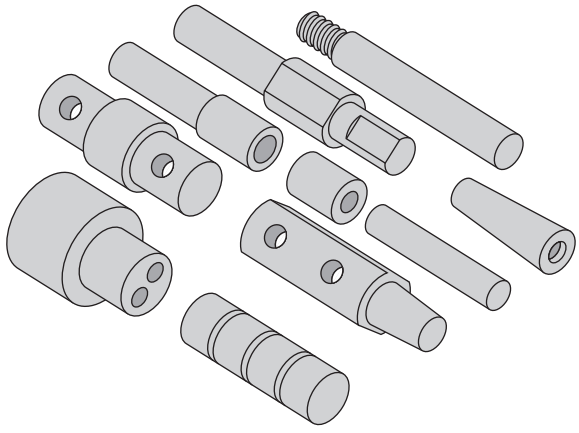


Figure 18.2 A family of parts with similar manufacturing process requirements but different design attributes. All parts are machined from cylindrical stock by turning; some parts require drilling and/or milling.

parts in Figure 18.1 are very similar in terms of geometric design, but quite different in terms of manufacturing because of differences in tolerances, production quantities, and materials. The parts shown in Figure 18.2 constitute a part family in manufacturing, but their different geometries make them appear quite different from a design viewpoint.

One of the important manufacturing advantages of grouping work parts into families can be explained with reference to Figures 18.3 and 18.4. Figure 18.3 shows a process-type plant layout for batch production in a machine shop. The various machine tools are arranged by function. There is a lathe department, milling machine department, drill press department, and so on. To machine a given part, the workpiece must be transported between departments, perhaps visiting the same department several times. This results in much material handling, large in-process inventories, many machine setups, long manufacturing lead times, and high cost. Figure 18.4 shows a production shop of equivalent capacity that has its machines arranged into cells. Each cell is organized to specialize in the production of a particular part family. Advantages are gained in the form of reduced workpiece handling, lower setup times, fewer setups (in some cases, no setup changeovers are necessary), less in-process inventory, and shorter lead times.

The biggest single obstacle in changing over to group technology from a conventional job shop is the problem of grouping the parts into families. There are three general methods for solving this problem. All three are time consuming and involve the analysis of much data by properly trained personnel. The three methods are (1) intuitive grouping, (2) parts classification and coding, and (3) production flow analysis.

18.1.2 Intuitive Grouping

This method, also known as the *visual inspection* method, is the least sophisticated and least expensive method. It is claimed to be the most common method that companies use to identify part families [35]. Intuitive grouping involves the classification of parts

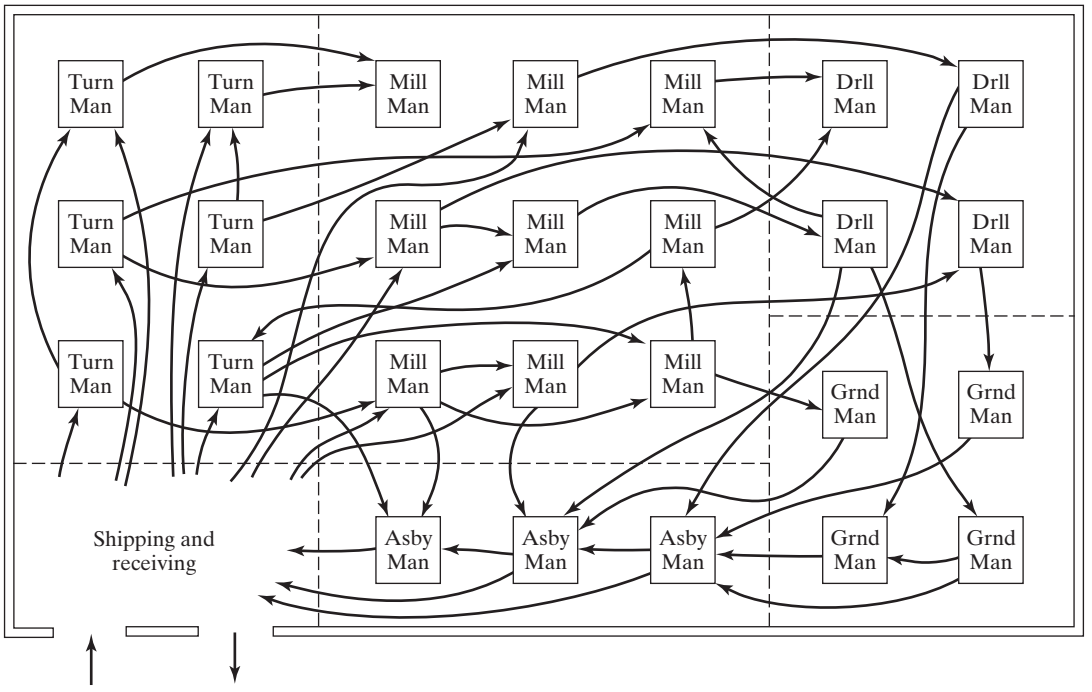


Figure 18.3 Process-type plant layout. (Key: Turn = turning, Mill = milling, Drll = drilling, Grnd = grinding, Asby = assembly, Man = manual operation; arrows indicate work flow through plant, and dashed lines indicate separation of machines into departments.)

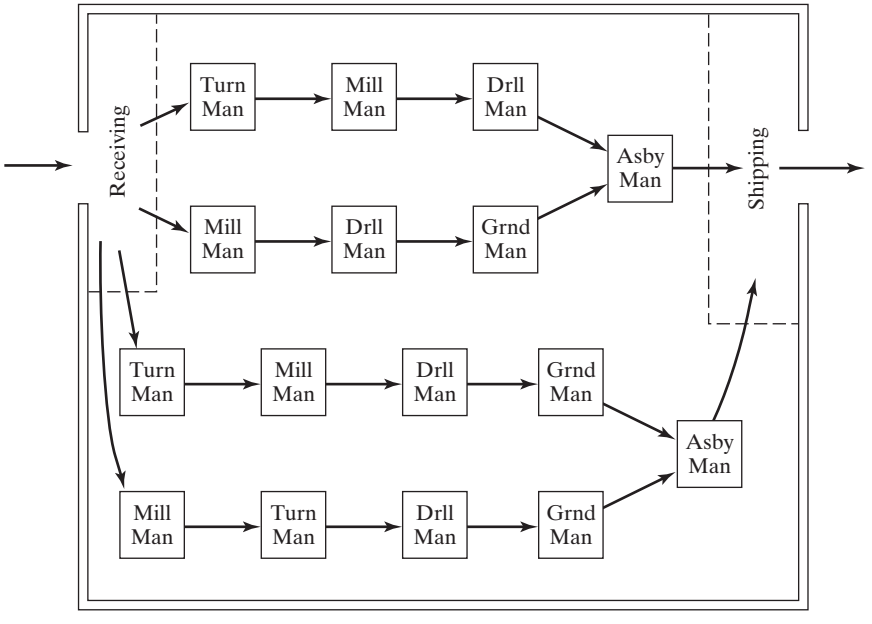


Figure 18.4 Group-technology layout. (Key: Turn = turning, Mill = milling, Drll = drilling, Grnd = grinding, Asby = assembly, Man = manual operation; arrows indicate work flow in machine cells.)

TABLE 18.1 Design and Manufacturing Attributes Typically Included in a Group Technology Classification and Coding System

Part Design Attributes	Part Manufacturing Attributes
Basic external shape	Major processes
Basic internal shape	Minor operations
Rotational or rectangular shape	Operation sequence
Length-to-diameter ratio (rotational parts)	Major dimension
Aspect ratio (rectangular parts)	Surface finish
Material types	Machine tool
Part function	Production cycle time
Major dimensions	Batch size
Minor dimensions	Annual production
Tolerances	Fixtures required
Surface finish	Cutting tools used in manufacture

into families by experienced technical staff in the plant who examine either the physical parts or their photographs and arrange them into groups having similar features. Two categories of part similarities can be distinguished: (1) design attributes, which are concerned with part characteristics such as geometry, size, and material, and (2) manufacturing attributes, which consider the processing steps required to make a part. Table 18.1 presents a list of common design and manufacturing attributes typically included in a part classification scheme. A certain amount of overlap exists between design and manufacturing attributes, because a part's geometry is largely determined by the manufacturing processes performed on it. Accordingly, by classifying parts into families, potential machine groups are also identified.

Although intuitive grouping is generally considered the least accurate of the three, it was the method used in one of the first major success stories of group technology in the United States, the Langston Division of Harris-Intertype in New Jersey (Historical Note 18.1).

18.1.3 Parts Classification and Coding

This method is the most time consuming of the three. In parts classification and coding, similarities among parts are identified and these similarities are related in a coding system that usually includes both a part's design and manufacturing attributes. Reasons for using a coding scheme include:

- *Design retrieval.* A designer faced with the task of developing a new part can use a design retrieval system to determine if a similar part already exists. Simply changing an existing part would take much less time than designing a whole new part from scratch.
- *Automated process planning.* The part code for a new part can be used to search for process plans for existing parts with identical or similar codes.
- *Machine cell design.* The part codes can be used to design machine cells capable of producing all members of a particular part family, using the composite part concept (Section 18.2.1).

To accomplish parts classification and coding, an analyst must examine the design and/or manufacturing features of each part. This is sometimes done by looking in tables to match the subject part against the features described and diagrammed in the tables. The Opitz classification and coding system uses tables of this kind (Appendix 18A). An alternative and more productive approach involves using a computerized classification and coding system, in which the user responds to questions asked by the computer. On the basis of the responses, the computer assigns a code number to the part. Whichever method is used, the classification results in a code number that uniquely identifies the part's attributes.

The principal functional areas that would use a parts classification and coding system are design and manufacturing. Accordingly, parts classification and coding systems fall into one of three categories: (1) systems based on part design attributes, (2) systems based on part manufacturing attributes, and (3) systems based on both design and manufacturing features. The typical design and manufacturing part attributes have previously been noted in Table 18.1.

In terms of the meaning of the symbols in the code, there are three structures used in classification and coding schemes:

1. *Hierarchical structure*, also known as a *monocode*, in which the interpretation of each successive symbol depends on the values of the preceding symbols
2. *Chain-type structure*, also known as a *polycode*, in which the interpretation of each symbol in the sequence is always the same; it does not depend on the values of preceding symbols
3. *Mixed-mode structure*, a hybrid of the two previous coding schemes.

To distinguish the hierarchical and chain-type structures, consider a two-digit code number for a part, such as 15 or 25. Suppose the first digit stands for the general shape of the part: 1 means the part is cylindrical (rotational), and 2 means the geometry is block-like. In a hierarchical structure, the interpretation of the second digit depends on the value of the first digit. If preceded by 1, the 5 might indicate a length-to-diameter ratio; and if preceded by 2, the 5 might indicate an aspect ratio between the length and width dimensions of the part. In the chain-type structure, the symbol 5 would have the same meaning whether preceded by 1 or 2. For example, it might indicate the overall length of the part. The advantage of the hierarchical structure is that in general more information can be included in a code of a given number of digits. The mixed-mode structure uses a combination of hierarchical and chain-type structures.

The number of digits in the code can range between 6 and 30. Coding schemes that contain only design data require fewer digits, perhaps 12 or fewer. Most classification and coding systems include both design and manufacturing data, and this usually requires 20–30 digits. This might seem like too many digits for a human reader to easily comprehend, but most of the data processing of the codes is accomplished by computer, for which a large number of digits is of minor concern.

A number of parts classification and coding systems are described in the literature (e.g., [15], [18], and [29]), including a number of commercial packages that were developed. However, none of the systems has been universally adopted. One reason is that a classification and coding system must be customized for each company, because each company's products are unique. A system that works for one company may not work

for another company. Another reason is the significant expense for the user company to implement a coding system.[1]

18.1.4 Production Flow Analysis

Production flow analysis (PFA) is an approach to part family identification and machine cell formation that was pioneered by J. Burbidge [7], [8], [9]. It is a method for identifying part families and associated machine groupings that uses the information contained on production route sheets rather than part drawings. Work parts with identical or similar routings are classified into part families. These families can then be used to form logical machine cells in a group-technology layout. Since PFA uses manufacturing data rather than design data to identify part families, it can overcome two possible anomalies that can occur in parts classification and coding. First, parts whose basic geometries are quite different may nevertheless require similar or even identical process routings. Second, parts whose geometries are quite similar may nevertheless require process routings that are quite different. Recall Figures 18.1 and 18.2.

The procedure in production flow analysis must begin by defining the scope of the study, which means deciding on the population of parts to be analyzed. Should all of the parts in the plant be included in the study, or should a representative sample be selected for analysis? Once this decision is made, then the procedure in PFA consists of the following steps:

1. *Data collection.* The minimum data needed in the analysis are the part number and operation sequence, which is contained in shop documents called route sheets or operation sheets. Each operation is usually associated with a particular machine, and so determining the operation sequence also determines the machine sequence.
2. *Sortation of process routings.* In this step, the parts are arranged into groups according to the similarity of their process routings. To facilitate this step, all operations or machines included in the shop are reduced to code numbers, such as those shown in Table 18.2. For each part, the operation codes are listed in the order in which they

TABLE 18.2 Possible Code Numbers Indicating Operations and/or Machines for Sortation in Production Flow Analysis (Highly Simplified)

Operation or Machine	Code
Cutoff	01
Lathe	02
Turret lathe	03
Mill	04
Drill—manual	05
NC drill	06
Grind	07

[1]Parts classification and coding was a popular topic in the trade literature during the 1980s, and, as mentioned, a number of commercial classification and coding systems were available. The companies included Brisch-Birn Inc. (Brisch System), Lovelace, Lawrence & Co. (Part Analog System), Manufacturing Data Systems, Inc. (CODE), Metcut Research Associates (CUTPLAN), and Organization for Industrial Research (Multi-Class). An Internet search revealed that most of these companies are no longer in business or their business no longer includes parts classification and coding. This is why coverage of this topic has been reduced in this edition of the book relative to earlier editions.

are performed. A sortation procedure is then used to arrange parts into "packs," which are groups of parts with identical routings. Some packs may contain only one part number, indicating the uniqueness of the processing of that part. Other packs will contain many parts, and these will constitute a part family.

3. *PFA chart.* The processes used for each pack are then displayed in a PFA chart, a simplified example of which is illustrated in Table 18.3.[2] The chart is a tabulation of the process or machine code numbers for all of the part packs. In some of the GT literature [27], the PFA chart is referred to by the term *part-machine incidence matrix*. In this matrix, the entries have a value $x_{ij} = 1$ or 0: a value of $x_{ij} = 1$ indicates that the corresponding part i requires processing on machine j, and $x_{ij} = 0$ indicates that no processing of component i is accomplished on machine j. For clarity in presenting the matrix, the 0s are often indicated as blank (empty) entries, as in Table 18.3.
4. *Cluster analysis.* From the pattern of data in the PFA chart, related groupings are identified and rearranged into a new pattern that brings together packs with similar machine sequences. One possible rearrangement of the original PFA chart is shown in Table 18.4, where different machine groupings are indicated within blocks. The blocks might be considered as possible machine cells. It is often the case (but not in Table 18.4) that some packs do not fit into logical groupings. These parts might be analyzed to see if a revised process sequence can be developed that fits into one of the groups. If not, these parts must continue to be fabricated through a conventional process layout. Section 18.4.1 examines a systematic technique called *rank-order clustering* that can be used to perform the cluster analysis.

The weakness of production flow analysis is that the data used in the technique are derived from existing production route sheets. In all likelihood, these route sheets have been prepared by different process planners over many years, during which new equipment may have been installed and old equipment retired. Consequently, the routings may contain operations and machine selections that are biased by the process planners' backgrounds, experiences, and expertise. Thus, the final machine groupings obtained in the analysis may be suboptimal. Notwithstanding this weakness, PFA has

TABLE 18.3 PFA Chart, Also Known as a Part-Machine Incidence Matrix

	Parts (i)								
Machines (j)	A	B	C	D	E	F	G	H	I
1	1			1				1	
2					1				1
3			1		1				1
4		1				1			
5	1							1	
6			1						1
7		1				1	1		

[2]For clarity in the part-machine incidence matrices and related discussion, parts are identified by alphabetic character and machines by number. In practice, numbers would be used for both.

TABLE 18.4 Rearranged PFA Chart, Indicating Possible Machine Groupings

	Parts (*i*)								
Machines (*j*)	C	E	I	A	D	H	F	G	B
3	1	1	1						
2		1	1						
6	1		1						
1				1	1	1			
5				1		1			
7							1	1	1
4							1		1

the virtue of requiring less time than a complete parts classification and coding procedure. This is attractive to many firms wishing to introduce group technology into their plant operations.

18.2 CELLULAR MANUFACTURING

Whether part families have been determined by intuitive grouping, parts classification and coding, or production flow analysis, there are advantages in producing those parts using GT machine cells rather than a traditional process-type machine layout. When the machines are grouped, the term *cellular manufacturing* is used to describe this work organization. ***Cellular manufacturing*** is an application of group technology in which dissimilar machines or processes have been aggregated into cells, each of which is dedicated to the production of a part or product family, or a limited group of families. Typical objectives in cellular manufacturing are similar to those of group technology:

- To shorten manufacturing lead times by reducing setup, work-part handling, waiting times, and batch sizes.
- To reduce work-in-process inventory. Smaller batch sizes and shorter lead times reduce work-in-process.
- To improve quality. This is accomplished by allowing each cell to specialize in producing a smaller number of different parts. This reduces process variability.
- To simplify production scheduling. The similarity among parts in the family reduces the complexity of production scheduling. Instead of scheduling parts through a sequence of machines in a process-type shop layout, the system simply schedules the parts through the cell.
- To reduce setup times. This is accomplished by using group tooling (cutting tools, jigs, and fixtures) that have been designed to process the part family, rather than part tooling, which is designed for an individual part. This reduces the number of individual tools required as well as the time to change tooling between parts.

Hyer and Wemmerlov [21] make an interesting comparison between manufacturing cells and job shops that use a conventional process layout. As described in Section 2.3.1, a

process layout consists of production departments in each of which the equipment is *similar* (e.g., lathe department, drill press department, milling department). This organization lends itself to processing of *dissimilar* parts. A manufacturing cell consists of *dissimilar* equipment that is organized to produce *similar* parts (part families).

Two aspects of cellular manufacturing are considered in this section: (1) the composite part concept and (2) machine cell design.

18.2.1 Composite Part Concept

Part families are defined by the fact that their members have similar design and/or manufacturing features. The composite part concept takes this part family definition to its logical conclusion. The ***composite part*** for a given family is a hypothetical part that includes all of the design and manufacturing attributes of the family. In general, an individual part in the family will have some of the features that characterize the family, but not all of them.

There is always a correlation between part design features and the production operations required to generate those features. Round holes are made by drilling, cylindrical shapes are made by turning, flat surfaces by milling, and so on. A production cell designed for the part family would include those machines required to make the composite part. Such a cell would be capable of producing any member of the family, simply by omitting those operations corresponding to features not possessed by the particular part. The cell would be designed to allow for size variations within the family as well as feature variations.

To illustrate, consider the composite part in Figure 18.5(a). It represents a family of rotational parts with features defined in part (b) of the figure. Associated with each feature is a certain machining operation, as summarized in Table 18.5. A machine cell to produce this part family would be designed with the capability to accomplish all seven operations required to produce the composite part (last column in the table). To produce a specific member of the family, operations would be included to fabricate the required features of the part. For parts without all seven features, unnecessary operations would

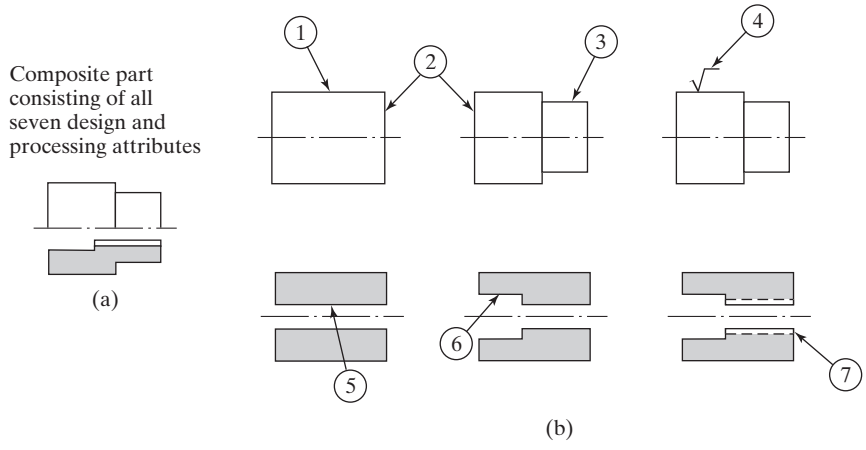


Figure 18.5 Composite part concept: (a) the composite part for a family of machined rotational parts, and (b) the individual features of the composite part. See Table 18.5 for key to individual features and corresponding manufacturing operations.

TABLE 18.5 Design Features of the Composite Part in Figure 18.5 and the Manufacturing Operations Required to Shape Those Features

Label	Design Feature	Corresponding Manufacturing Operation
1	External cylinder	Turning
2	Cylinder face	Facing
3	Cylindrical step	Turning
4	Smooth surface	External cylindrical grinding
5	Axial hole	Drilling
6	Counterbore	Counterboring
7	Internal threads	Tapping

simply be omitted. Machines, fixtures, and tools would be organized for efficient flow of work parts through the cell.

In practice, the number of design and manufacturing attributes is greater than seven, and allowances must be made for variations in overall size and shape of the parts in the family. Nevertheless, the composite part concept is useful for visualizing part families and the machine cell design problem.

18.2.2 Machine Cell Design

Design of the machine cell is critical in cellular manufacturing. The cell design determines to a great degree the performance of the cell. This section discusses types of cells, cell layouts, and the key machine concept.

Types of Machine Cells. GT cells can be distinguished as either (1) assembly cells, which produce families of subassemblies or products, or (2) part cells, which process families of parts. Assembly cells are discussed in Section 15.6.

Machine cells for part family production can be classified according to the number of machines and the degree to which the material flow is mechanized between machines. Four common GT cell configurations are (1) single-machine cell, (2) group-machine cell with manual handling, (3) group-machine cell with semi-integrated handling, and (4) flexible manufacturing cell or flexible manufacturing system.

As its name indicates, the ***single-machine cell*** consists of one machine plus supporting fixtures and tooling. This type of cell can be applied to work parts whose attributes allow them to be made on one basic type of process, such as turning or milling. For example, the composite part of Figure 18.5 could be produced on a conventional turret lathe with the possible exception of the cylindrical grinding operation (step 4).

The ***group-machine cell with manual handling*** is an arrangement of more than one machine used collectively to produce one or more part families, and there is no provision for mechanized parts movement between machines in the cell. Instead, the human operators who run the cell perform the material handling function. The cell is often organized into a U-shaped layout, as shown in Figure 18.6. This layout is considered appropriate when there is variation in the work flow among the parts made in the cell. It also allows the multifunctional workers in the cell to move easily between machines [26]. Other advantages of U-shaped cells in batch-model assembly applications, compared to a conventional paced assembly line, include (1) easier changeover from one model to the next, (2) improved quality, (3) visual control of work-in-process, (4) lower initial investment because the cells

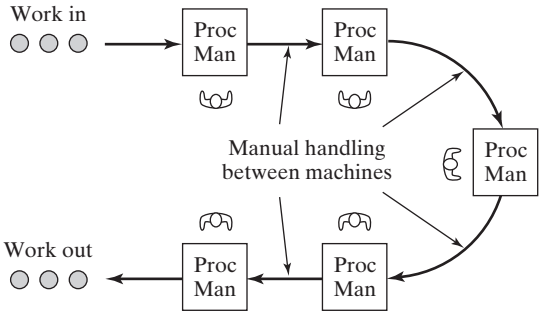


Figure 18.6 Machine cell with manual handling between machines. A U-shaped machine layout is shown. (Key: Proc = processing operation (mill, turn, etc.), Man = manual operation; arrows indicate work flow.)

are simpler and no powered conveyor is required, (5) greater worker satisfaction due to job enlargement and absence of pacing, and (6) more flexibility to adjust to increased demand simply by adding more cells [14].

The group-machine cell with manual handling is sometimes achieved in a conventional process layout without rearranging the equipment. This is done by simply assigning certain machines to be included in the machine group, and restricting their work to specified part families. This allows many of the benefits of cellular manufacturing to be achieved without the expense of rearranging equipment in the shop. Obviously, the material handling benefits of GT are minimized with this organization.

The ***group-machine cell with semi-integrated handling*** uses a mechanized handling system, such as a conveyor, to move parts between machines in the cell. The ***flexible manufacturing system*** (FMS) combines a fully integrated material handling system with automated processing stations. The FMS is the most highly automated of the group-technology machine cells. The following chapter is devoted to this form of automation, and discussion of it is deferred until then.

Machine Cell Layouts. Various layouts are used in GT cells. The U-shape in Figure 18.6 is a popular configuration in cellular manufacturing. Other GT layouts include in-line, loop, and rectangular, shown in Figure 18.7 for the case of semi-integrated handling.

Determining the most appropriate cell layout depends on the routings of parts produced in the cell. Four types of part movement can be distinguished in a mixed-model part production system. They are illustrated in Figure 18.8 and defined as follows, where the forward direction of work flow is from left to right in the figure: (1) ***repeat operation***, in which a consecutive operation is carried out on the same machine, so that the part does not actually move; (2) ***in-sequence move***, in which the part moves forward from the current machine to an immediate neighbor; (3) ***bypassing move***, in which the part moves forward from the current machine to another machine that is two or more machines ahead; and (4) ***backtracking move***, in which the part moves backward from the current machine to another machine.

When the application consists exclusively of in-sequence moves, an in-line layout is appropriate. A U-shaped layout also works well here and has the advantage of closer

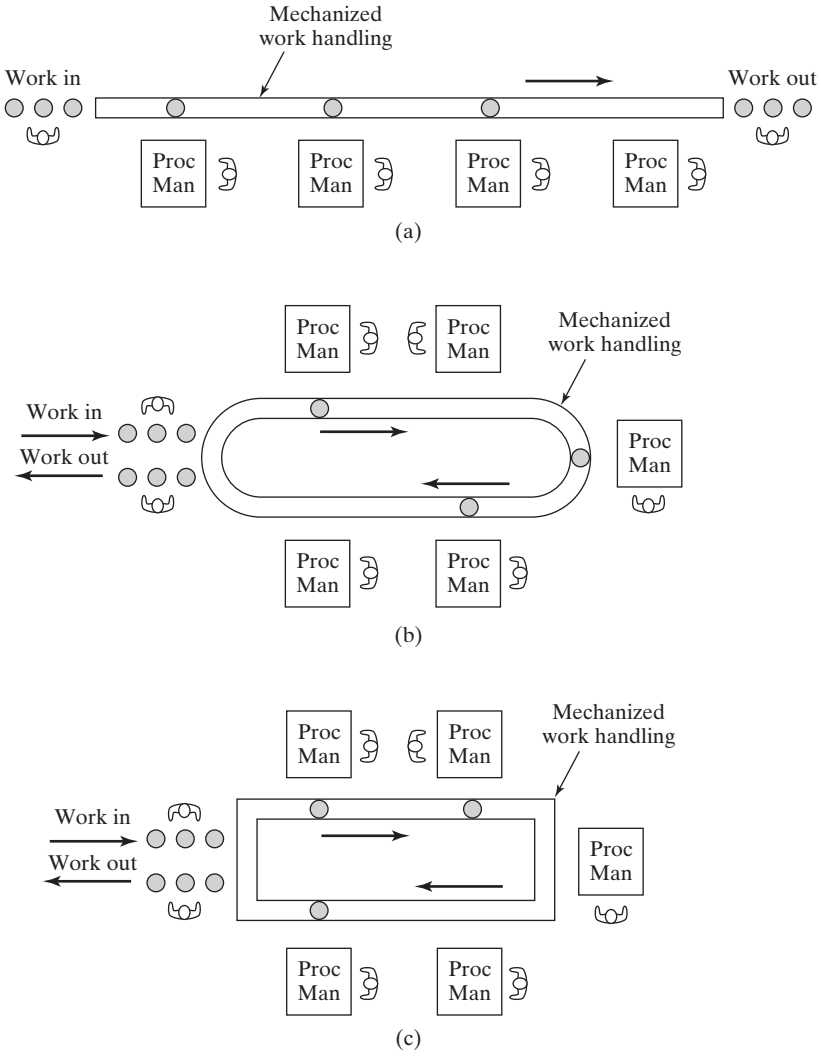


Figure 18.7 Machine cells with semi-integrated handling: (a) in-line layout, (b) loop layout, and (c) rectangular layout. (Key: "Proc" = processing operation (mill, turn, etc.), "Man" = manual operation; arrows indicate work flow.)

interaction among the workers in the cell. When the application includes repeated operations, multiple stations (machines) are often required. For cells requiring bypassing moves, the U-shape layout is appropriate. When backtracking moves are needed, a loop or rectangular layout allows recirculation of parts within the cell. Additional factors that must be accommodated by the cell design include:

- *Amount of work to be done by the cell.* This includes the quantity of parts per year and the processing (or assembly) time per part at each station. These factors determine the workload that must be accomplished by the cell and therefore the number of machines that must be included, as well as total operating cost of the cell and the investment that can be justified.

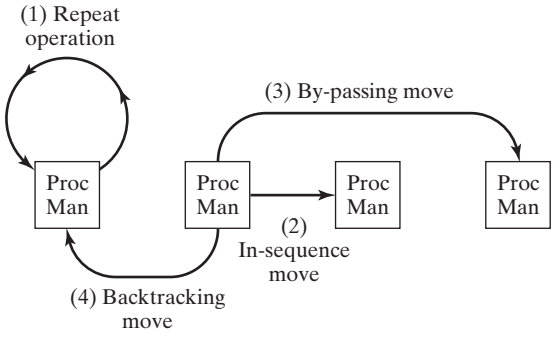


Figure 18.8 Four types of part moves in a mixed-model production system. The forward flow of work is from left to right.

- *Part size, shape, weight, and other physical attributes.* These factors determine the size and type of material handling and processing equipment that must be used.

Key Machine Concept. In some respects, a GT machine cell operates like a manual assembly line, and it is desirable to spread the workload as evenly as possible among the machines in the cell. On the other hand, there is typically a certain machine in a cell (or perhaps more than one machine in a large cell) that is more expensive to operate than the other machines or that performs certain critical operations in the plant. This machine is referred to as the *key machine*. It is important that the utilization of this key machine be high, even if it means that the other machines in the cell have relatively low utilizations. The other machines are referred to as *supporting machines*, and they should be organized in the cell to keep the key machine busy. In a sense, the cell is designed so that the key machine becomes the bottleneck in the system.

The key machine concept is sometimes used to plan the GT machine cell. The approach is to decide what parts should be processed through the key machine and then determine what supporting machines are required to complete the processing of those parts.

There are generally two measures of utilization that are of interest in a GT cell: the utilization of the key machine and the utilization of the overall cell. The utilization of the key machine can be measured using the usual definition (see Section 18.4.3). The utilization of each of the other machines can be evaluated similarly. The cell utilization is obtained by taking a simple arithmetic average of all the machines in the cell.

18.3 APPLICATIONS OF GROUP TECHNOLOGY

In the chapter introduction, group technology was defined as a "manufacturing philosophy." GT is not a particular technique, although various tools and techniques, such as parts classification and coding and production flow analysis, have been developed to implement it. The group-technology philosophy can be applied in a number of areas. The discussion here focuses on the two main areas of manufacturing and product design.

GT Manufacturing Applications. The most common applications of GT are in manufacturing, and the most common application in manufacturing involves the formation of cells of one kind or another. Not all companies rearrange machines to form cells.

There are three ways in which group-technology principles can be applied in manufacturing [20]:

1. *Informal scheduling and routing of similar parts through selected machines*. This approach achieves setup advantages, but no formal part families are defined and no physical rearrangement of equipment is undertaken.
2. *Virtual machine cells*. This approach involves the creation of part families and dedication of equipment to the manufacture of these part families, but without the physical rearrangement of machines into cells. The machines in the virtual cell remain in their original locations in the factory. Use of virtual cells seems to facilitate the sharing of machines with other virtual cells producing other part families [22].
3. *Formal machine cells*. This is the conventional GT approach in which a group of dissimilar machines are physically relocated into a cell that is dedicated to the production of one or a limited set of part families (Section 18.2.2). The machines in a formal machine cell are located in close proximity to one another in order to minimize part handling, throughput time, and work-in-process.

Other GT applications in manufacturing include process planning, family tooling, and numerical control (NC) part programs. Process planning of new parts can be facilitated by identifying part families. The new part is associated with an existing part family, and generation of the process plan for the new part follows the routing of the other members of the part family. This is done in a formalized way if parts classification and coding is used. The approach is discussed in the context of automated process planning (Section 24.2.1).

Ideally, all members of the same part family require similar setups, tooling, and fixturing. This generally results in a reduction in the amount of tooling and fixturing needed. Instead of using a special tool kit developed for each part, a GT system uses a tool kit developed for each part family. The concept of a ***modular fixture***, also known as a *flexible fixture*, can often be exploited, in which a common base fixture is used that can accommodate adaptations to rapidly switch between different parts in the family.

A similar approach can be applied in NC part programming. ***Parametric programming*** [25] involves the preparation of a common NC program that covers the entire part family, and the program is then adapted for individual members of the family by inserting dimensions and other parameters applicable to the particular part. Parametric programming reduces both part programming time and setup time.

GT Product Design Applications. The application of group technology in product design is principally for design retrieval systems that reduce part proliferation. It has been estimated that the cost of releasing a new part design ranges between $2,000 and $12,000 [32]. In a survey of industry reported in Wemmerlov and Myer [31], it was concluded that in about 20% of new part situations, an existing part design could have been used. In about 40% of the cases, an existing part design could have been used with modifications. The remaining cases required new part designs. If the cost savings for a company generating 1,000 new part designs per year were 75% when an existing part design could be used (assuming that there would still be some cost of time associated with the new part for engineering analysis and design retrieval) and 50% when an existing design could be modified, then the total annual savings to the company would be $700,000 to $4,200,000, or 35% of the company's total design expense due to part releases. The level of design savings described here requires an efficient design retrieval procedure. Most design retrieval procedures are based on parts classification and coding systems.

Other design applications of group technology involve simplification and standardization of design parameters such as tolerances, inside radii on corners, chamfer sizes on outside edges, hole sizes, and thread sizes. These measures simplify design procedures and reduce part proliferation. Design standardization also pays dividends in manufacturing by reducing the required number of distinct lathe tool nose radii, drill sizes, and fastener sizes. There is also a benefit in reducing the amount of data and information that the company must handle. Fewer part designs, design attributes, tools, fasteners, and so on mean fewer and simpler design documents, process plans, and other data records.

18.4 ANALYSIS OF CELLULAR MANUFACTURING

Many quantitative techniques have been developed to deal with problems in group technology and cellular manufacturing. Two problem areas are considered in this section: (1) grouping parts and machines into families, and (2) arranging machines in a GT cell. The first problem area has been the subject of academic research, and several publications are listed in references [2], [3], [12], [13], [23], and [24]. The technique described here for solving the part and machine grouping problem is rank-order clustering [23]. The second problem area has also been the subject of research, and several reports are listed in references [1], [7], [9], and [19]. In Section 18.4.2, a heuristic approach by Hollier is introduced [19].

18.4.1 Rank-order Clustering

The problem addressed here is determining how machines in an existing plant should be grouped into machine cells. The problem is the same whether the cells are virtual or formal (Section 18.3). It is basically the problem of identifying part families. After part families have been identified, the machines to produce a given part family can be selected and grouped together.

The rank-order clustering technique, first proposed by King [23], is specifically applicable in production flow analysis. It is an efficient and easy-to-use algorithm for grouping machines into cells. In a starting part-machine incidence matrix that might be compiled to document the part routings in a machine shop (or other job shop), the occupied locations in the matrix are organized in a seemingly random fashion. Rank-order clustering works by reducing the part-machine incidence matrix to a set of diagonalized blocks that represent part families and associated machine groups. Starting with the initial part-machine incidence matrix, the algorithm consists of the following steps:

1. In each row of the matrix, read the series of 1s and 0s (blank entries = 0s) from left to right as a binary number. Rank the rows in order of decreasing value. In case of a tie, rank the rows in the same order as they appear in the current matrix.
2. Numbering from top to bottom, is the current order of rows the same as the rank order determined in the previous step? If yes, go to step 7. If no, go to the following step.
3. Reorder the rows in the part-machine incidence matrix by listing them in decreasing rank order, starting from the top.
4. In each column of the matrix, read the series of 1s and 0s (blank entries = 0s) from top to bottom as a binary number. Rank the columns in order of decreasing value. In case of a tie, rank the columns in the same order as they appear in the current matrix.

5. Numbering from left to right, is the current order of columns the same as the rank order determined in the previous step? If yes, go to step 7. If no, go to the following step.
6. Reorder the columns in the part-machine incidence matrix by listing them in decreasing rank order, starting with the left column. Go to step 1.
7. Stop.

For readers unaccustomed to evaluating binary numbers in steps 1 and 4, it might be helpful to convert each binary value into its decimal equivalent. For example, the entries in the first row of the matrix in Table 18.3 are read as 100100010. This converts to its decimal equivalent as follows: $(1 \times 2^8) + (0 \times 2^7) + (0 \times 2^6) + (1 \times 2^5) + (0 \times 2^4) + (0 \times 2^3) + (0 \times 2^2) + (1 \times 2^1) + (0 \times 2^0) = 256 + 32 + 2 = 290$. Decimal conversion becomes impractical for the large numbers of parts found in practice, so it is preferable to compare the binary numbers.

In Example 18.1, it is possible to divide the parts and machines into three mutually exclusive part-machine groups. This represents the ideal case because the part families and

EXAMPLE 18.1 Rank-order Clustering Technique

Apply the rank-order clustering technique to the part-machine incidence matrix in Table 18.3.

Solution: Step 1 consists of reading the series of 1s and 0s in each row as a binary number. This is done in Table 18.6(a), converting the binary value for each row to its decimal equivalent. The values are then rank-ordered in the far right-hand column. In step 2, it is seen that the row order is different from the starting matrix. Therefore, the rows are reordered in step 3. In step 4, the series of 1s and 0s in each column are read from top to bottom as a binary number (again this has been converted to the decimal equivalent) and rank the columns in order of decreasing value, as shown in Table 18.6(b). In step 5, it is observed that the column order is different from the preceding matrix. Proceeding from step 6 back to steps 1 and 2, and a reordering of the columns provides a row order that is in descending value and the algorithm is concluded (step 7). The final solution is shown in Table 18.6(c). A comparison of this solution with Table 18.4 reveals that they are the same part-machine groupings.

TABLE 18.6(a) First Iteration (Step 1) in the Rank-order Clustering Technique Applied to Example 18.1

Binary values	2^8	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0		
					Parts					Decimal Equivalent	
Machines	A	B	C	D	E	F	G	H	I		Rank
1	1			1				1		290	1
2					1				1	17	7
3			1		1				1	81	5
4		1				1				136	4
5	1							1		258	2
6			1						1	65	6
7		1				1	1			140	3

TABLE 18.6(b) Second Iteration (Steps 3 and 4) in the Rank-order Clustering Technique Applied to Example 18.1

	Parts									
Machines	A	B	C	D	E	F	G	H	I	Binary values
1	1			1				1		2^6
5	1							1		2^5
7		1				1	1			2^4
4		1				1				2^3
3			1		1				1	2^2
6			1						1	2^1
2					1				1	2^0
Decimal	96	24	6	64	5	24	16	96	7	
equivalent Rank	1	4	8	3	9	5	6	2	7	

TABLE 18.6(c) Solution of Example 18.1

	Parts								
Machines	A	H	D	B	F	G	I	C	E
1	1	1	1						
5	1	1							
7				1	1	1			
4				1	1				
3							1	1	1
6							1	1	
2							1		1

associated machine cells are completely segregated. However, it is not uncommon for an overlap in processing requirements to exist between machine groups. That is, a given part type needs to be processed by more than one machine group. One way of dealing with the overlap is simply to duplicate the machine that is used by more than one part family, placing the same machine type in both cells. Other approaches, attributed to Burbidge [23], include (1) change the routing so that all processing can be accomplished in the primary machine group, (2) redesign the part to eliminate the processing requirement outside the primary machine group, and (3) purchase the parts from an outside supplier.

18.4.2 Arranging Machines in a GT Cell

After part-machine groupings have been identified, the next problem is to organize the machines into the most logical sequence. A simple yet effective method is suggested by Hollier [19][3] that uses data contained in from–to charts (Section 10.3.1) and is intended to place the machines in an order that maximizes the proportion of in-sequence moves

[3]Hollier [19] presented six heuristic approaches to solving the machine arrangement problem, of which only one is described here. He presents a comparison of the six methods in his paper.

within the cell. The method is based on the use of from–to ratios determined by summing the total flow from and to each machine in the cell. The algorithm can be reduced to three steps:

1. *Develop the from–to chart*. The data contained in the chart indicate numbers of part moves between the machines (or workstations) in the cell. Moves into and out of the cell are not included in the chart.
2. *Determine the "from–to ratio" for each machine*. This is accomplished by summing all of the "From" trips and "To" trips for each machine (or operation). The "From" sum for a machine is determined by adding the entries in the corresponding row, and the "To" sum is determined by adding the entries in the corresponding column. For each machine, the "from–to ratio" is calculated by taking the "From" sum for each machine and dividing by the respective "To" sum.
3. *Arrange machines in order of decreasing from–to ratio*. Machines with a high from–to ratio distribute more work to other machines in the cell but receive less work from other machines. Conversely, machines with a low from–to ratio receive more work than they distribute. Therefore, machines are arranged in order of descending from–to ratio; that is, machines with high ratios are placed at the beginning of the work flow, and machines with low ratios are placed at the end of the work flow. In case of a tie, the machine with the higher "From" value is placed ahead of the machine with a lower value.

EXAMPLE 18.2 Group-technology Machine Sequence Using the Hollier Method

A GT cell has four machines: 1, 2, 3, and 4. An analysis of 50 parts processed on these machines has been summarized in the from–to chart in Table 18.7. Additional information: 50 parts enter the machine grouping at machine 3, 20 parts leave after processing at machine 1, and 30 parts leave machine 4 after processing. Determine the most logical machine sequence using the Hollier method.

TABLE 18.7 From–To Chart for Example 18.2

	To			
From	1	2	3	4
1	0	5	0	25
2	30	0	0	15
3	10	40	0	0
4	10	0	0	0

Solution: Summing the "From" trips and "To" trips for each machine yields the "From" and "To" sums in Table 18.8. The from–to ratios are listed in the last column on the right. Arranging the machines in order of descending from–to ratio, the machines in the cell should be sequenced as follows:

$$3 \rightarrow 2 \rightarrow 1 \rightarrow 4$$

TABLE 18.8 From–To Sums and From–To Ratios for Example 18.2

	To					
From	1	2	3	4	**"From" sums**	**From–to ratio**
1	0	5	0	25	**30**	**0.60**
2	30	0	0	15	**45**	**1.0**
3	10	40	0	0	**50**	**∞**
4	10	0	0	0	**10**	**0.25**
"To" sums	**50**	**45**	**0**	**40**	**135**	

It is helpful to use a graphical technique, such as the network diagram (Section 10.3.1), to conceptualize the work flow in the cell. The network diagram for the machine arrangement in Example 18.2 is presented in Figure 18.9. The flow is mostly in-line; however, there is some bypassing and backtracking of parts that must be considered in the design of any material handling system that might be used in the cell. A powered conveyor would be appropriate for the forward flow between machines, with manual handling for the back flow.

Three ratings can be defined to compare solutions to the machine sequencing problem: (1) percentage of in-sequence moves, (2) percentage of bypassing moves, and (3) percentage of backtracking moves. Each rating is computed by adding all of the values representing that type of move and dividing by the total number of moves. It is desirable for the percentage of in-sequence moves to be high, and for the percentage of backtracking moves to be low. The Hollier method is designed to achieve these goals. Bypassing moves are less desirable than in-sequence moves, but certainly better than backtracking.

EXAMPLE 18.3 Rating Machine Sequences

Compute (a) the percentage of in-sequence moves, (b) the percentage of bypassing moves, and (c) the percentage of backtracking moves for the solution in Example 18.2.

Solution: From Figure 18.9, the number of in-sequence moves = 40 + 30 + 25 = 95, the number of bypassing moves = 10 + 15 = 25, and the number of backtracking

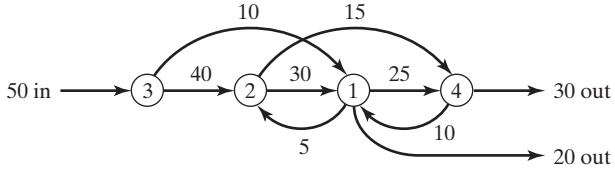


Figure 18.9 Network diagram for machine cell in Example 18.2. Flow of parts into and out of the cells is included.

moves = 5 + 10 = 15. The total number of moves = 135 (totaling either the "From" sums or the "To" sums). Thus,

(a) Percentage of in-sequence moves = 95/135 = 0.704 = 70.4%

(b) Percentage of bypassing moves = 25/135 = 0.185 = 18.5%

(c) Percentage of backtracking moves = 15/135 = 0.111 = 11.1%

18.4.3 Performance Metrics in Cell Operations

Some of the equations developed in Chapter 3 for factory operations can be adapted to the operations of group-technology cells. Suppose the cell consists of n machines (workstations) and produces a family of parts with n_f family members. Let i = a subscript to identify machines $(i = 1, 2, \ldots, n)$, and let j = a subscript to identify family members $(j = 1, 2, \ldots, n_f)$. The production time of family member j on machine i is given by T_{pij}, which is determined as follows:

$$T_{pij} = \frac{T_{suij} + Q_j T_{cij}}{Q_j} \tag{18.1}$$

where T_{suij} = the setup or changeover time to prepare for family member j on machine i, min; T_{cij} = operation cycle time for family member j on machine i, min/pc; and Q_j = batch quantity for family member j, pc. Unlike conventional batch production, one would expect the T_{suij} value to be minimal; in the ideal case, there would be no lost time for changeover in cellular manufacturing $(T_{suij} = 0)$. Similarly, batch quantity Q_j would be low, perhaps a batch size of one $(Q_j = 1)$. With these values, $T_{pij} = T_{cij}$. However, Equation (18.1) allows for changeover time between different family members, and for the family members to be run in batches if there is a changeover time.

The production rate R_{pij} for family member j on machine i is the reciprocal of production time, multiplied by 60 to express it as an hourly rate:

$$R_{pij} = \frac{60}{T_{pij}} \tag{18.2}$$

Let f_{ij} = the fraction of time during steady-state operation that machine i is processing family member j. Under normal conditions, it follows that for each machine i,

$$0 \le \sum_j f_{ij} \le 1 \text{ where } 0 \le f_{ij} \le 1 \text{ for all } i. \tag{18.3}$$

The value of Σf_{ij} for each machine is the utilization of that machine within the cell. That is,

$$U_i = \sum_j f_{ij} \tag{18.4}$$

where U_i = utilization of machine i. If $\Sigma f_{ij} = 1$, the machine is fully utilized. More likely, Σf_{ij} will be less than 1 for at least some of the machines in the cell. The average utilization of the cell is the average of the machine utilizations:

$$U = \frac{\sum_{i=1}^{n} \sum_{j} f_{ij}}{n} = \frac{\sum_{j} U_i}{n} \tag{18.5}$$

Each family member is processed through n_{oj} operations (machines) in the cell. The production rate of the cell is given by

$$R_p = \sum_{i=1}^{n} \sum_{j} f_{ij} R_{pij} \Big/ n_{oj} \tag{18.6}$$

where R_p = average hourly production rate (pc/hr) of the cell; n_{oj} = the number of operations required to produce family member j, and the other terms are defined earlier.

One of the advantages of cellular manufacturing is reduced lead time to get parts through the cell compared to a job shop. The manufacturing lead time is the sum of setup time, run time, and nonoperation time. The nonoperation time consists of waiting time and move time within the cell. For any family member, this can be expressed as follows:

$$MLT_j = \sum_{i=1}^{n_{oj}} (T_{suij} + Q_j T_{cij} + T_{noij}) \tag{18.7}$$

where MLT_j = manufacturing lead time for part family member j, min; T_{suij} = setup (changeover) time for operation i on family member j, min; Q_j = batch quantity of family member j being processed in the cell, pc; T_{cij} = cycle time for operation i on family member j, min/pc; T_{noij} = nonoperation time associated with operation i, min; and i indicates the operation sequence in the processing: $i = 1, 2, \ldots, n_{oj}$. One would expect the setup time to be minimal in a group-technology cell, depending on how similar the family members are. The nonoperation time in a GT cell would also be expected to be significantly less than in a conventional job shop or batch production situation. The average manufacturing lead time for the part family is given by the following:

$$MLT = \frac{\sum_{j=1}^{n_f} MLT_j}{n_f} \tag{18.8}$$

where MLT = average manufacturing lead time for the n_f family members, min; and MLT_j = lead time for family member j from Equation (18.7).

The work-in-process within the cell can be determined from the production rate and manufacturing lead time. As in Chapter 3, the determination is based on Little's formula (Section 3.1.3, footnote 2):

$$WIP = R_p(MLT) \tag{18.9}$$

where WIP = work-in-process in the plant, pc; R_p = average hourly production rate from Equation (18.6), pc/hr; and MLT = average manufacturing lead time from Equation (18.8), hr.

EXAMPLE 18.4 Performance Metrics for a GT Cell

A group-technology cell has three machines and is used to process a family of four similar parts. The table below lists production quantities (Q_j), production times (T_{pij}), and machine fractions for each family member (f_{ij}). Assume the nonoperation times (T_{no}) are all the same at 30 min per machine. Determine (a) average hourly production rate for the cell, (b) utilization of each machine and average utilization of the cell, (c) manufacturing lead time, and (d) work-in-process.

		Machine 1		Machine 2		Machine 3	
Part	Q_j	T_{p1} (min)	f_{1j}	T_{p2} (min)	f_{2j}	T_{p3} (min)	f_{3j}
A	1	3.0	0.2	4.5	0.3	2.25	0.15
B	1	2.0	0.2	4.0	0.4	3.0	0.3
C	1	5.0	0.25	4.0	0.2	3.0	0.15
D	1	4.0	0.3	1.333	0.1	2.667	0.2

Solution: A spreadsheet calculator was used to perform the calculations. Hourly production rates for each machine and family member were computed using Equation (18.2). The quantities of each family member produced in 1 hr were $Q_{ij} = f_{ij}R_{pij}$. The total for each column represents the hourly output of all parts from each machine (17.5 pc/hr). Finally, the MLT values were obtained from $MLT_j = T_{p1j} + T_{p2j} + T_{p3j} + 3T_{no}$. These were averaged to obtain 99.7 min = 1.661 hr.

R_{p1}	R_{p2}	R_{p3}	Q_1	Q_2	Q_3	MLT
20.00	13.33	26.67	4.00	4.00	4.00	99.75
30.00	15.00	20.00	6.00	6.00	6.00	99.00
12.00	15.00	20.00	3.00	3.00	3.00	102.00
15.00	45.00	22.50	4.50	4.50	4.50	68.00
			17.50	17.50	17.50	398.75
					Average MLT =	99.7

Summary: (a) Hourly production rate for each machine and for the cell R_p = **17.5 parts/hr**

Note that all three machines have the same production rate, which means that the workload in the cell is balanced.

(b) Utilization for each machine is given by Σf_{ij} for each machine i. Thus, U_1 = **0.95**, U_2 = **1.0**, and U_3 = **0.80**. Average cell utilization U = **0.917**.

(c) Average manufacturing lead time MLT = 99.7 min = **1.661 hr**

(d) Average work-in-process WIP = (17.5 parts/hr)(1.661 hr) = **29.1 parts**

Chapter 19

Flexible Manufacturing Cells and Systems

CHAPTER CONTENTS

19.1 What is a Flexible Manufacturing System?
  19.1.1 Flexibility
  19.1.2 Types of FMS
19.2 FMC/FMS Components
  19.2.1 Workstations
  19.2.2 Material Handling and Storage System
  19.2.3 Computer Control System
19.3 FMS Application Considerations
  19.3.1 FMS Applications
  19.3.2 FMS Planning and Implementation Issues
  19.3.3 FMS Benefits
19.4 Analysis of Flexible Manufacturing Systems
  19.4.1 Bottleneck Model
  19.4.2 Extended Bottleneck Model
  19.4.3 Sizing the FMS
  19.4.4 What the Equations Tell Us
19.5 Alternative Approaches to Flexible Manufacturing
  19.5.1 Mass Customization
  19.5.2 Reconfigurable Manufacturing Systems
  19.5.3 Agile Manufacturing

The flexible manufacturing system (FMS) is a type of machine cell to implement cellular manufacturing. It is the most automated and technologically sophisticated of the group-technology (GT) cells. An FMS typically possesses multiple automated stations

and is capable of variable routings among stations. Its flexibility allows it to cope with soft product variety (Section 2.3). An FMS integrates into one highly automated manufacturing system many of the concepts and technologies discussed in previous chapters, including flexible automation (Section 1.2.1), CNC machines (Chapter 7), distributed computer control (Section 5.3.3), automated material handling and storage (Chapters 10 and 11), and group technology (Chapter 18). The concept for flexible manufacturing systems originated in Britain in the early 1960s (Historical Note 19.1). The first FMS installations in the United States occurred around 1967. These initial systems performed machining operations on families of parts using NC machine tools.

FMS technology can be applied in production situations similar to those identified for cellular manufacturing:

- Presently the plant either produces parts in batches or uses manned GT cells, and management wants to automate.
- It is possible to group a portion of the parts made in the plant into part families, whose similarities permit them to be processed on the machines in the flexible manufacturing system. Part similarities can be interpreted to mean that (1) the parts belong to a common product and/or (2) the parts possess similar geometries. In either case, the processing requirements of the parts must be sufficiently similar to allow them to be made on the FMS.

Historical Note 19.1 Flexible Manufacturing Systems [21], [23], [24]

The flexible manufacturing system was first conceptualized for machining, and it required the prior development of numerical control. The concept is credited to David Williamson, a British engineer employed by Molins during the mid 1960s. Molins applied for a patent for the invention that was granted in 1965. The concept was called *System 24* because it was believed that the group of machine tools comprising the system could operate 24 hr a day, 16 hr of which would be unattended by human workers. The original concept included computer control of the NC machines, production of a variety of parts, and use of tool magazines that could hold various tools for different machining operations.

One of the first flexible manufacturing systems to be installed in the United States was a machining system at Ingersoll-Rand Company in Roanoke, Virginia, in the late 1960s by Sundstrand, a machine tool builder. Other systems introduced soon after included a Kearney & Trecker FMS at Caterpillar Tractor and Cincinnati Milacron's "Variable Mission System." Most of the early FMS installations in the United States were in large companies, such as Ingersoll-Rand, Caterpillar, John Deere, and General Electric Company. These large companies had the financial resources to make the major investments necessary, and they also possessed the prerequisite experience in NC machine tools, computer systems, and manufacturing systems to pioneer the new FMS technology. Flexible manufacturing systems were also installed in other countries around the world. In the Federal Republic of Germany (West Germany, now Germany), a manufacturing system was developed in 1969 by Heidleberger Druckmaschinen in cooperation with the University of Stuttgart. In the USSR (now Russia), a flexible manufacturing system was demonstrated at the 1972 Stanki Exhibition in Moscow. The first Japanese FMS was installed around the same time by Fuji Xerox. By around 1985, the number of FMS installations throughout the world had increased to about 300. About 20–25% of these were located in the United States. In recent years, there has been an emphasis on smaller, less expensive flexible manufacturing cells.

- The parts or products made by the facility are in the mid-volume, mid-variety production range. The appropriate production volume range is 5,000–75,000 parts per year [14]. If annual production is below this range, an FMS is likely to be an expensive alternative. If production volume is above this range, then a more specialized production system should probably be considered.

The differences between installing a flexible manufacturing system and implementing a manually operated machine cell are the following: (1) the FMS requires a significantly greater capital investment because new equipment is being installed, whereas the manually operated machine cell might only require existing equipment to be relocated, and (2) the FMS is technologically more sophisticated for the human resources who must make it work. However, the potential benefits are substantial. They include increased machine utilization, reduced factory floor space, greater responsiveness to change, lower inventory and manufacturing lead times, and higher labor productivity. Section 19.3.3 elaborates on these benefits.

This chapter addresses the following questions: What makes an FMS flexible? What are their components and applications? And how is FMS technology implemented? In Section 19.4, a mathematical model is presented for assessing the performance of an FMS.

19.1 WHAT IS A FLEXIBLE MANUFACTURING SYSTEM?

A flexible manufacturing system (FMS) is a highly automated GT machine cell, consisting of one or more processing stations (usually CNC machine tools), interconnected by an automated material handling and storage system and controlled by a distributed computer system. The reason the FMS is called flexible is that it is capable of processing a variety of different part styles simultaneously at the various workstations, and the mix of part styles and quantities of production can be adjusted in response to changing demand patterns.

An FMS relies on the principles of group technology. No manufacturing system can be completely flexible. There are limits to the range of parts or products that can be made in an FMS. Accordingly, a flexible manufacturing system is designed to produce parts (or products) within a defined range of styles, sizes, and processes. In other words, an FMS is capable of producing a single part family or a limited range of part families.

A more appropriate term for FMS would be *flexible automated manufacturing system.* The use of the word "automated" would distinguish this type of production technology from other manufacturing systems that are flexible but not automated, such as a manned GT machine cell. The word "flexible" would distinguish it from other manufacturing systems that are highly automated but not flexible, such as a conventional transfer line.[1]

19.1.1 Flexibility

The issue of manufacturing system flexibility was discussed in Section 13.2.4, and the three capabilities that a manufacturing system must possess in order to be flexible were identified as (1) the ability to identify the different incoming part or product styles processed by the system, (2) quick changeover of operating instructions, and (3) quick changeover of physical

[1]Notwithstanding the appropriateness of the term *flexible automated manufacturing system*, the current terminology (*flexible manufacturing system*) is well established in the commercial and research literature.

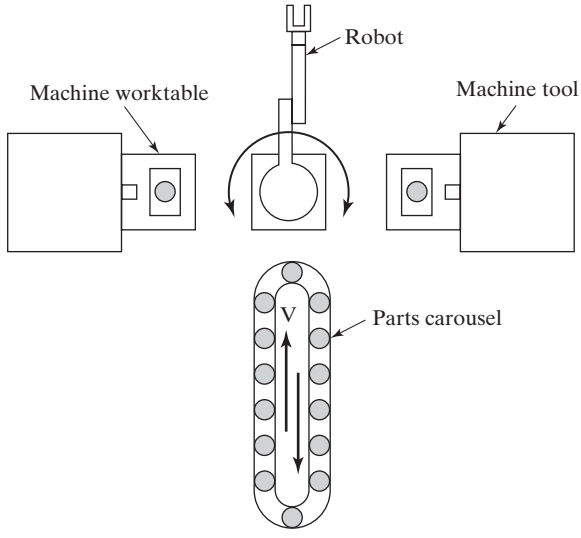


Figure 19.1 Automated manufacturing cell with two machine tools and robot. Is it a flexible cell?

setup. Flexibility is an attribute that applies to both manual and automated systems. In manual systems, the human workers are often the enablers of the system's flexibility.

To develop the concept of flexibility in an automated manufacturing system, consider a machine cell consisting of two CNC machine tools that are loaded and unloaded by an industrial robot from a parts storage system, perhaps in the arrangement depicted in Figure 19.1. The cell operates unattended for extended periods of time. Periodically, a worker must unload completed parts from the storage system and replace them with new work parts. By any definition, this is an automated manufacturing cell, but is it a flexible manufacturing cell? One might argue yes, it is flexible because the cell consists of CNC machine tools, and CNC machines are flexible because they can be programmed to machine different part configurations. However, if the cell only operates in a batch mode, in which the same part style is produced by both machines in lots of several hundred units, then this does not qualify as flexible manufacturing.

To qualify as being flexible, an automated manufacturing system should satisfy the following four tests of flexibility:[2]

1. *Part-variety test.* Can the system process different part or product styles in a mixed-model (non-batch) mode?
2. *Schedule-change test.* Can the system readily accept changes in production schedule, that is, changes in part mix and/or production quantities?
3. *Error-recovery test.* Can the system recover gracefully from equipment malfunctions and breakdowns, so that production is not completely disrupted?

[2]These four tests, as they are called here, are sometimes referred to as types or dimensions of flexibility [3], [7], [21], and [25]. The part-variety test is called *machine flexibility* or *production flexibility.* The schedule-change test is called *mix flexibility* or *volume flexibility.* The error-recovery test is called *routing flexibility*, and the new-part test is called *product flexibility.* Other names for flexibility have been developed by other authors and researchers.

4. *New-part test.* Can new part designs be introduced into the existing part mix with relative ease if their features qualify them as being members of the part family for which the system was designed? Also, can design changes be made in existing parts without undue challenge to the system?

If the answer to all of these questions is "yes" for a given manufacturing system, then the system is flexible. The most important tests are (1) and (2). Test (3) is applicable to multi-machine systems but in single-machine systems when the one machine breaks down it is difficult to avoid a halt in production. Test (4) would seem to not apply to systems designed for a part family whose members are all known in advance. However, such a system may have to deal with design changes to members of that existing part family.

Getting back to the robotic work cell, the four tests of flexibility are satisfied if the cell (1) can machine different part configurations in a mix rather than in batches; (2) permits changes in production schedule (changes in part mix); (3) is capable of continuing to operate even though one machine experiences a breakdown (e.g., while repairs are being made on the broken machine, its work is temporarily reassigned to the other machine), and (4) can accommodate new part designs if the NC part programs are written off-line and then downloaded to the system for execution. The fourth capability requires the new part to be within the part family intended for the FMS, so that the tooling used by the CNC machines as well as the end effector of the robot are compatible with the new part design.

19.1.2 Types of FMS

Each FMS is designed for a specific application, that is, a specific family of parts and processes. Therefore, each FMS is custom-engineered and unique. Given these circumstances, one would expect to find a great variety of system designs to satisfy a wide variety of application requirements.

Flexible manufacturing systems can be distinguished according to the kinds of operations they perform: processing operations or assembly operations. An FMS is usually designed to perform one or the other but rarely both. A difference that is applicable to machining systems is whether the system will process rotational parts or nonrotational parts. Flexible machining systems with multiple stations that process rotational parts are less common than systems that process nonrotational parts. Two other ways to classify flexible manufacturing systems are by number of machines and level of flexibility.

Number of Machines. Flexible manufacturing systems have a certain number of processing machines. The following are typical categories: (1) single-machine cell, (2) flexible manufacturing cell, and (3) flexible manufacturing system.

A ***single-machine cell*** consists of one CNC machining center combined with a parts-storage system for unattended operation, as in Figure 19.2. Completed parts are periodically unloaded from the parts-storage unit, and raw work parts are loaded into it. The cell can be designed to operate in a batch mode, a flexible mode, or a combination of the two. When operated in a batch mode, the machine processes parts of a single style in specified lot sizes and is then changed over to process a batch of the next part style. When operated in a flexible mode, the system satisfies three of the four flexibility tests. It is capable of (1) processing different part styles, (2) responding to changes in production schedule, and (4) accepting new part introductions. Test (3), error recovery, cannot be satisfied because if the single machine breaks down, production stops.

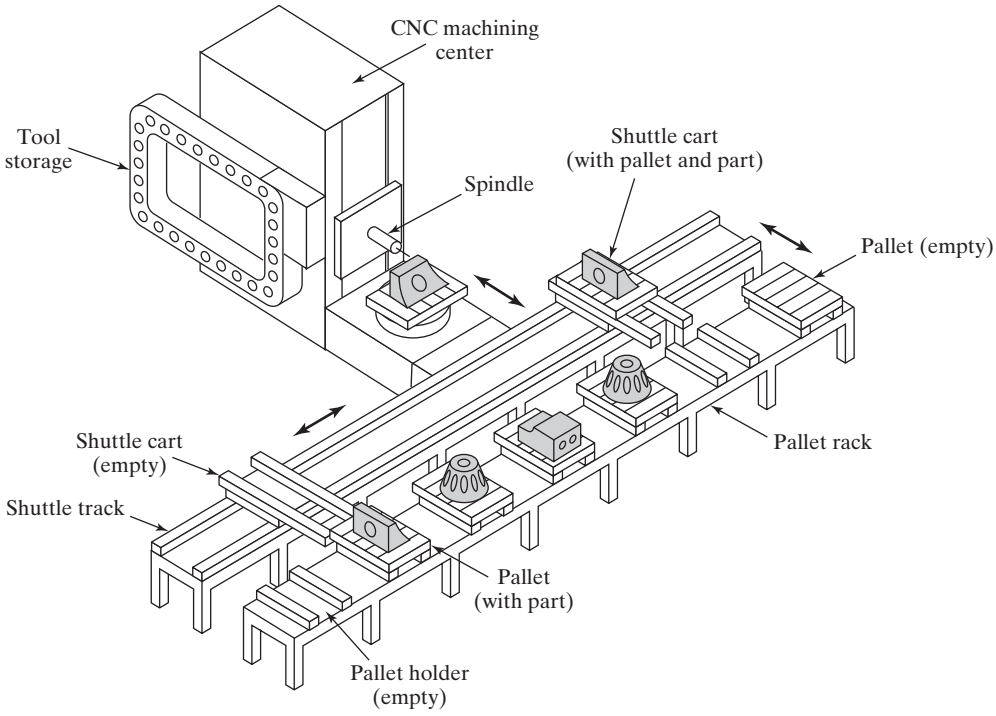


Figure 19.2 Single-machine cell consisting of one CNC machining center and parts-storage unit.

A ***flexible manufacturing cell*** (FMC) consists of two or three processing workstations (typically CNC machining centers or turning centers) plus a parts-handling system. The parts-handling system is connected to a load/unload station. The handling system usually includes a limited parts-storage capacity. One possible FMC is illustrated in Figure 19.3. A flexible manufacturing cell satisfies the four flexibility tests discussed previously.

A ***flexible manufacturing system*** (FMS) has four or more processing stations connected mechanically by a common parts-handling system and electronically by a distributed computer system. Thus, an important distinction between an FMS and an FMC is the number of machines: an FMC has two or three machines, while an FMS has four or more.[3] There are usually other differences as well. One is that the FMS generally includes nonprocessing workstations that support production but do not directly participate in it. These other stations include part/pallet washing stations, inspection stations, and so on. Another difference is that the computer control system of an FMS is generally more sophisticated, often including functions not always found in a cell, such as diagnostics and tool monitoring. These additional functions are needed more in an FMS than in an FMC because the FMS is more complex.

[3]The dividing line that separates an FMS from an FMC is defined here to be four machines. It should be noted that not all practitioners would agree with that dividing line; some might prefer a higher value, while others would prefer a lower value. Also, the distinction between cell and system seems to apply only to flexible manufacturing systems that are automated. The manually operated counterparts of these systems, discussed in Chapter 18, seem to always be referred to as cells, no matter how many workstations are included.

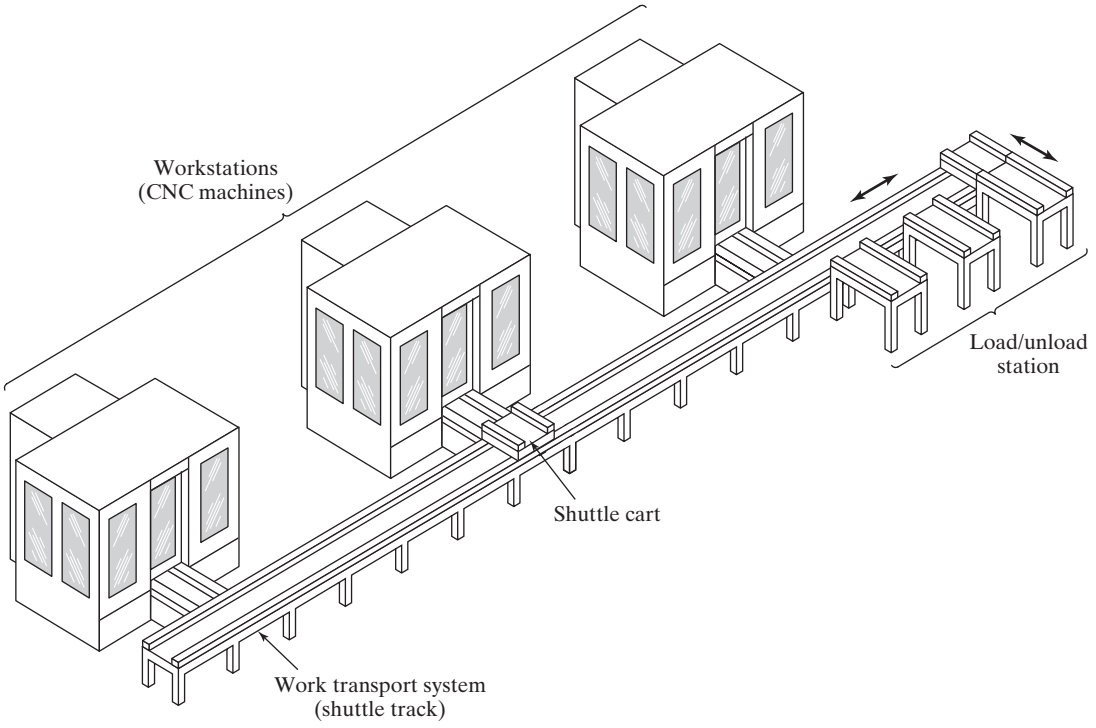


Figure 19.3 A flexible manufacturing cell consisting of three identical processing stations (CNC machining centers), a load/unload station, and a parts-handling system.

Table 19.1 compares the three systems in terms of the four flexibility tests.

Level of Flexibility. Another way to classify flexible manufacturing systems is according to the level of flexibility designed into the system. Two categories of flexibility are discussed here: (1) dedicated and (2) random-order.

TABLE 19.1 Four Tests of Flexibility Applied to the Three Types of Manufacturing Cells and Systems

	Four Tests of Flexibility			
System Type	1. Part Variety	2. Schedule Change	3. Error Recovery	4. New Part
Single-machine cell	Yes, but processing is sequential, not simultaneous.	Yes	Limited recovery due to only one machine.	Yes
Flexible manufacturing cell (FMC)	Yes, simultaneous production of different parts.	Yes	Error recovery limited by fewer machines than FMS.	Yes
Flexible manufacturing system (FMS)	Yes, simultaneous production of different parts.	Yes	Machine redundancy minimizes effect of machine breakdowns.	Yes

A ***dedicated FMS*** is designed to produce a limited variety of part styles, and the complete population of parts is known in advance. The part family may be based on product commonality rather than geometric similarity. The product design is considered stable, so the system can be designed with a certain amount of process specialization to make the operations more efficient. Instead of being general purpose, the machines can be designed for the specific processes required to make the limited part family, thus increasing the production rate of the system. In some instances, the machine sequence may be identical or nearly identical for all parts processed, so a transfer line may be appropriate, in which the workstations possess the necessary flexibility to process the different parts in the mix. Indeed, the term *flexible transfer line* is sometimes used for this case (Section 16.2.1).

A ***random-order FMS*** is more appropriate when the following circumstances apply: (1) the part family is large, (2) there are substantial variations in part configurations, (3) new part designs will be introduced into the system and engineering changes will be made to parts currently produced, and (4) the production schedule is subject to change from day-to-day. To accommodate these variations, the random-order FMS must be more flexible than the dedicated FMS. It is equipped with general-purpose machines to deal with the product variations and is capable of processing parts in various sequences (random order). A more sophisticated computer control system is required for this FMS type.

The trade-off between flexibility and productivity can be seen in these two system types. The dedicated FMS is less flexible but capable of higher production rates. The random-order FMS is more flexible but at the cost of lower production rates. Table 19.2 presents a comparison of the dedicated FMS and random-order FMS in terms of the four flexibility tests.

19.2 FMC/FMS COMPONENTS

The three basic components of a flexible manufacturing system are (1) workstations, (2) material handling and storage system, and (3) computer control system. In addition, even though an FMS is highly automated, people are required to manage and operate the system. Functions typically performed by humans include (1) loading raw work parts into the system, (2) unloading finished parts (or assemblies) from the system, (3) changing and setting tools, (4) performing equipment maintenance and repair, (5) performing NC part programming, (6) programming and operating the computer system, and (7) managing the system.

TABLE 19.2 Four Tests of Flexibility Applied to Dedicated and Random-Order Systems

	Four Tests of Flexibility			
System Type	1. Part Variety	2. Schedule Change	3. Error Recovery	4. New Part
Dedicated FMS	Limited. All parts are known in advance.	Limited changes can be tolerated.	Usually limited by sequential processes.	No. New part introductions are difficult.
Random-order FMS	Yes. Substantial part variations are possible.	Frequent and significant changes are possible.	Machine redundancy minimizes effect of machine breakdowns.	Yes. System is designed for new part designs.

19.2.1 Workstations

The processing or assembly equipment used in an FMC or FMS depends on the type of work accomplished by the system. In one designed for machining operations, the principal types of processing station are CNC machine tools. However, the FMS concept is applicable to other processes as well. Following are the types of workstations typically found in an FMS.

Load/Unload Stations. The load/unload station is the physical interface between the FMS and the rest of the factory. It is where raw work parts enter the system and finished parts exit the system. Loading and unloading can be accomplished either manually (the most common method) or by automated handling systems. If manually performed, the load/unload station should be ergonomically designed to permit convenient and safe movement of work parts. Mechanized cranes and other handling devices are installed to assist the operator with parts that are too heavy to lift by hand. A certain level of cleanliness must be maintained at the workplace, and air hoses or other washing facilities are used to flush away chips and ensure clean mounting and locating points. The station is often raised slightly above floor level using an open-grid platform to permit chips and cutting fluid to drop through the openings for subsequent recycling or disposal.

The load/unload station includes a data entry unit and monitor for communication between the operator and the computer system. Through this system, the operator receives instructions regarding which part to load onto the next pallet to adhere to the production schedule. When different pallets are required for different parts, the correct pallet must be supplied to the station. When modular fixturing is used, the correct fixture must be specified and the required components and tools must be available at the workstation to build it. When the part loading procedure has been completed, the handling system must launch the pallet into the system. These conditions require communication between the computer system and the operator(s) at the load/unload station.

Machining Stations. The most common applications of flexible manufacturing systems are machining operations. The workstations used in these systems are therefore predominantly CNC machine tools. Most common are CNC machining centers, which possess features that make them compatible with the FMS, including automatic tool changing and tool storage, use of palletized work parts, CNC, and capacity for distributed numerical control (Section 7.2.3). Machining centers are available with automatic pallet changers that can be readily interfaced with the FMS part-handling system. Machining centers are generally used for nonrotational parts. For rotational parts, turning centers are used; and for parts that are mostly rotational but require multi-tooth rotational cutters (milling and drilling), mill-turn centers and multitasking machines can be used. These equipment types are described in Section 14.2.3.

Assembly. Some flexible manufacturing systems are designed to perform assembly operations. Flexible automated assembly systems are gradually replacing manual labor in the assembly of products typically made in batches. Industrial robots are often used as the automated workstations in these flexible assembly systems. They can be programmed to perform tasks with variations in sequence and motion pattern to accommodate the different product styles assembled in the system. Other examples of flexible assembly workstations are the programmable component placement machines widely used in electronics assembly.

Other Stations and Equipment. Inspection can be incorporated into a flexible manufacturing system, either by including an inspection operation at a processing workstation or by including a station specifically designed for inspection. Coordinate measuring machines (Section 22.3), special inspection probes that can be used in a machine tool spindle (Section 22.3.4), and machine vision (Section 22.5) are three possible technologies for performing inspection on an FMS. Inspection is particularly important in flexible assembly systems to ensure that components have been properly added at the workstations. The topic of automated inspection is examined in more detail in Chapter 21.

In addition to the above, other operations and functions are often accomplished on a flexible manufacturing system. These include cleaning parts and/or pallet fixtures, central coolant delivery systems for the entire FMS, and centralized chip-removal systems often installed below floor level.

19.2.2 Material Handling and Storage System

The second major component of an FMS is its material handling and storage system. This section covers the functions of the handling system, types of handling equipment used in an FMS, and types of FMS layout.

Functions of the Handling System. The material handling and storage system in a flexible manufacturing system performs the following functions:

- *Random independent movement of work parts between stations.* Parts must be moved from any machine in the system to any other machine to provide various routing alternatives for different parts and to make machine substitutions when certain stations are busy or broken down.
- *Handling a variety of work part configurations.* For nonrotational parts, this is usually accomplished by using modular pallet fixtures in the handling system. The fixture is located on the top face of the pallet and is designed to accommodate a variety of part styles by means of common components, quick-change features, and other devices that permit a rapid changeover for a given part. The base of the pallet is designed for the material handling system. For rotational parts, industrial robots are often used to load and unload turning machines and to move parts between stations.
- *Temporary storage.* The number of parts in the FMS will typically exceed the number of parts being processed at any moment. Thus, each station has a small queue of parts, perhaps only one part, waiting to be processed; this helps to maintain high machine utilization.
- *Convenient access for loading and unloading work parts.* The handling system must include locations for load/unload stations.
- *Compatibility with computer control.* The handling system must be under the direct control of the computer system which directs it to the various workstations, load/unload stations, and storage areas.

Material Handling Equipment. The types of material handling systems used to transfer parts between stations in an FMS include a variety of conventional material transport equipment (Chapter 10), in-line transfer mechanisms (Section 16.1.1), and industrial robots (Chapter 8). The material handling function in an FMS is often shared between two systems: (1) a primary handling system and (2) a secondary handling system.

The primary handling system establishes the basic layout of the FMS and is responsible for moving parts between stations.

The secondary handling system consists of transfer devices, automatic pallet changers, and similar mechanisms located at the FMS workstations. The function of the secondary handling system is to transfer work from the primary system to the machine tool or other processing station and to position the parts with sufficient accuracy to perform the processing or assembly operation. Other purposes served by the secondary handling system include (1) reorientation of the work part if necessary to present the surface that is to be processed, and (2) buffer storage of parts to minimize work-change time and maximize station utilization. In some FMS installations, the positioning and registration requirements at the individual workstations are satisfied by the primary work-handling system. In these cases, there is no secondary handling system.

FMS Layout Configurations. The material handling system establishes the FMS layout. Most layout configurations found in today's flexible manufacturing systems can be classified into one of four categories: (1) in-line layout, (2) loop layout, (3) open field layout, and (4) robot-centered cell. The types of material handling equipment utilized in these four layouts are summarized in Table 19.3.

In the in-line layout, the machines and handling system are arranged in a straight line. In its simplest form, the parts progress from one workstation to the next in a well-defined sequence with work always moving in one direction and no back-flow, as in Figure 19.4(a). The operation of this type of system is similar to a transfer line (Chapter 16), except that a variety of work parts are processed in the system. For in-line systems requiring greater routing flexibility, a linear transfer system that permits movement in two directions can be used. One possible arrangement is shown in Figure 19.4(b), in which a secondary work-handling system is located at each station to separate parts from the primary line. The secondary handling system provides temporary storage of parts at each station.

The in-line layout can be combined with an integrated parts-storage system, as in Figure 19.5. Depending on the capacity of the storage system, this arrangement can be used for "lights out" operation of the FMS, in which workers load parts into the system during the day shift, and the FMS operates unattended during the two overnight shifts. A single-machine manufacturing system with integrated storage is shown in Figure 14.3.

In the loop layout, the workstations are organized in a loop that is served by a parts-handling system in the same shape, as shown in Figure 19.6(a). Parts usually flow in one

TABLE 19.3 Material Handling Equipment Typically Used as the Primary Handling System for FMS Layouts

Layout Configuration	Typical Material Handling System
In-line layout	In-line transfer system (Section 16.1.1) Conveyor system (Section 10.2.4) Rail-guided vehicle system (Section 10.2.3) Overhead rail-guided vehicle system with robotic part handling
Loop layout	Conveyor system (Section 10.2.4) In-floor towline carts (Section 10.2.4)
Open field layout	Automated guided vehicle system (Section 10.2.2) In-floor towline carts (Section 10.2.4)
Robot-centered layout	Industrial robot (Chapter 8)

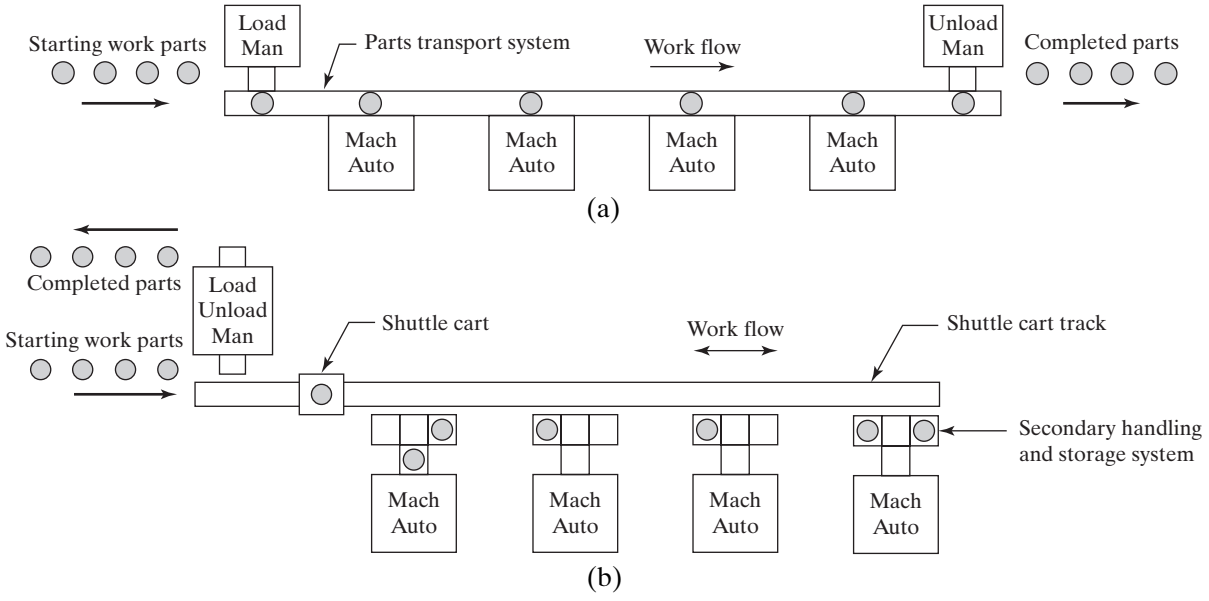


Figure 19.4 FMS in-line layouts: (a) one-direction flow similar to a transfer line, (b) linear transfer system with secondary parts-handling and storage system at each station to facilitate flow in two directions. Key: Load = parts loading station, Unload = parts unloading station, Mach = machining station, Man = manual station, Auto = automated station.

direction around the loop with the capability to stop and be transferred to any station. A secondary handling system is shown at each workstation to allow parts to move around the loop without obstruction. The load/unload station(s) are typically located at one end of the loop. An alternative form of loop layout is the rectangular layout. As shown in Figure 19.6(b), this arrangement might be used to return pallets to the starting position in a straight line machine arrangement.

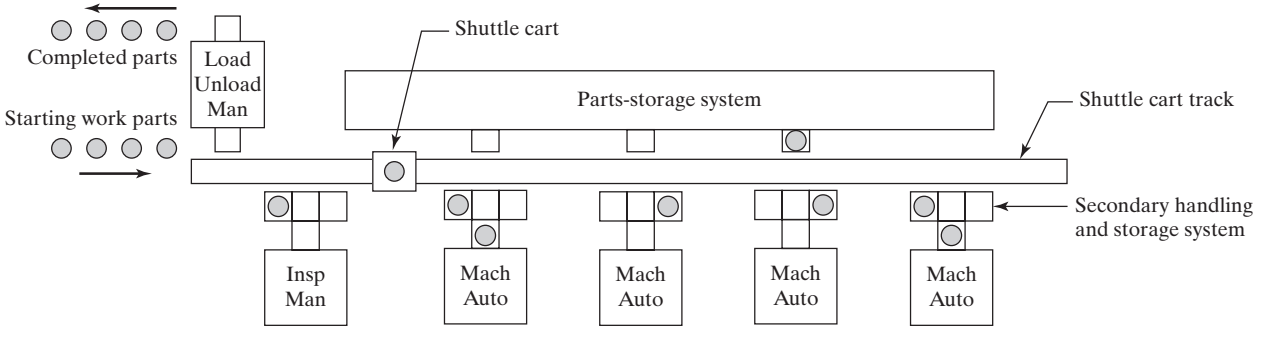


Figure 19.5 FMS in-line layout with integrated part-storage system. Key: Load = parts loading station, Unload = parts unloading station, Mach = machining station, Man = manual station, Auto = automated station.

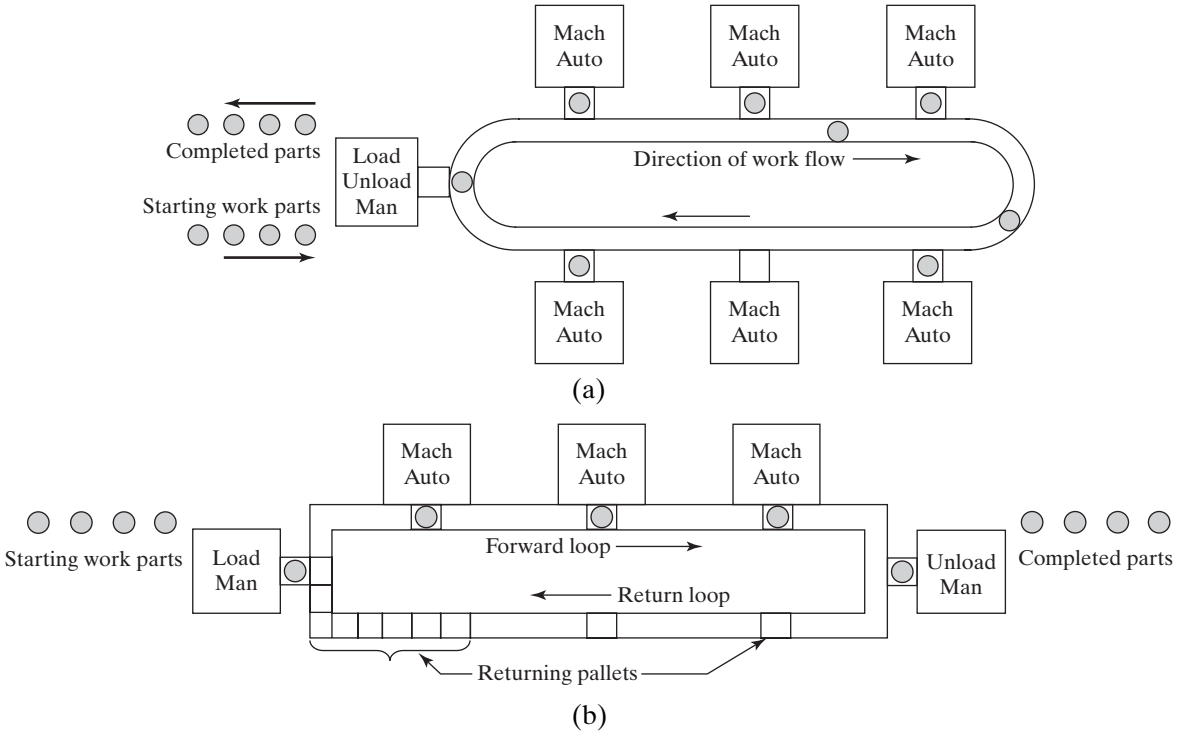


Figure 19.6 (a) FMS loop layout with secondary parts-handling system at each station to allow unobstructed flow on the loop, and (b) rectangular layout for recirculation of empty pallets to the parts loading station. Key: Load = parts loading station, Unload = parts unloading station, Mach = machining station, Man = manual station, Auto = automated station.

The open field layout consists of multiple loops and branches, and may include sidings as well, as illustrated in Figure 19.7. This layout type is generally appropriate for processing large families of parts. The number of different machine types may be limited, and parts are routed to different workstations depending on which one becomes available first.

The robot-centered layout (Figure 19.1) uses one or more robots as the material handling system. Industrial robots can be equipped with grippers that make them well suited for the handling of rotational parts, and robot-centered FMS layouts are often used to process cylindrical or disk-shaped parts. As an alternative to a robot-centered cell, a robot can be mounted on a floor-installed rail-guided vehicle or suspended from an overhead gantry crane to service multiple CNC turning centers in an in-line layout. The configuration would be similar to the layout shown in Figure 19.5, with a part-storage system on one side of the rail and CNC machines on the other side.

19.2.3 Computer Control System

The FMS includes a distributed computer control system (Section 5.3.3) that is interfaced to the workstations, material handling system, and other hardware components. A typical FMS computer system consists of a central computer and microcomputers controlling the individual machines and other components. The central computer coordinates the

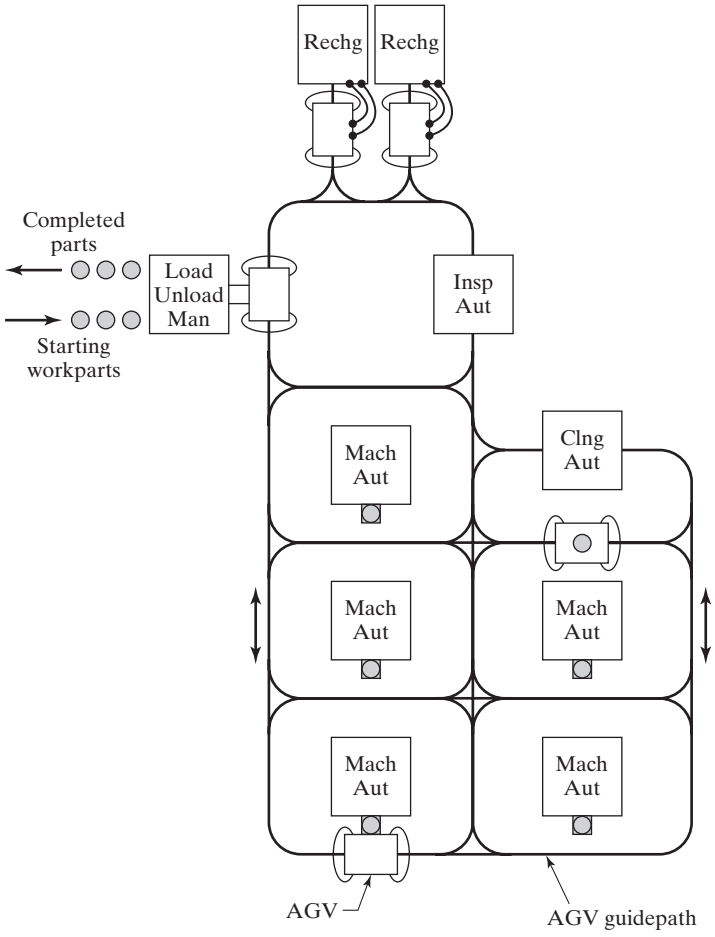


Figure 19.7 FMS open field layout. Key: Load = parts loading station, Unload = parts unloading station, Mach = machining station, Man = manual station, Aut = automated station, AGV = automated guided vehicle, Rechg = AGV battery recharging station, Clng = cleaning, Insp = inspection.

activities of the components to achieve smooth overall operation of the system. In addition, an uplink from the FMS to the corporate host computer is provided. Functions performed by the FMS computer control system can be divided into the following categories:

1. *Workstation control.* In a fully automated FMS, the individual processing or assembly stations generally operate under some form of computer control, such as CNC.
2. *Distribution of control instructions to workstations.* Part programs are stored in the central computer and downloaded to machines. Distributed numerical control (Section 7.2.3) is used for this purpose. The DNC system allows submission of new programs and editing of existing programs as needed.
3. *Production control.* The mix and rate at which the various parts are launched into the system must be managed, based on specified daily production rates for each part

type, numbers of raw work parts available, and number of applicable pallets.[4] This is accomplished by routing an applicable pallet to the load/unload area and providing instructions to the operator to load the desired work part.

4. *Traffic control.* This refers to the management of the primary material handling system that moves parts between stations. Traffic control is accomplished by actuating switches at branches and merging points, stopping parts at machine tool transfer locations, and moving pallets to load/unload stations.
5. *Shuttle control.* This function is concerned with the operation and control of the secondary handling system at each workstation. This must be coordinated with traffic control and synchronized with the operation of the machine tool it serves.
6. *Tool control.* This is concerned with managing two aspects of the cutting tools: (a) tool location, which involves keeping track of the cutting tools at each workstation and making sure that the correct tools are available at each station for the parts that are to be routed to that station; and (b) tool life monitoring, which involves comparing the expected tool life for each cutting tool with the cumulative machining time of the tool, and alerting a worker when a tool replacement is needed.
7. *Performance monitoring and reporting.* The computer control system collects data on the operation and performance of the flexible manufacturing system. The data are periodically summarized, and reports on system performance are prepared for management. The collected data include proportion uptime and utilization of each machine, daily and weekly production quantities, and cutting tool status (tool locations and tool life monitoring). In addition to reports on these data, management can request instantaneous status information on the current condition of the system.
8. *Diagnostics.* This function is used to indicate the probable source of the problem when a malfunction occurs. It can also be used to plan preventive maintenance and identify impending system failures. The purpose of the diagnostics function is to reduce breakdowns and downtime, and to increase availability of the system.

19.3 FMS APPLICATION CONSIDERATIONS

This section covers several topics related to the application and implementation of FMS technology as well as the benefits that are associated with FMS installations.

19.3.1 FMS Applications

Flexible automation is applicable to a variety of manufacturing operations. FMS technology is most widely applied in machining operations. Other applications include sheet metal pressworking and assembly.

Flexible Machining Systems. Historically, most of the applications of flexible machining systems have been in milling and drilling operations (nonrotational parts), using CNC machining centers. FMS applications for turning (rotational parts) were much less common until recently, and the systems that are installed tend to consist of fewer machines.

[4]The term *applicable pallet* refers to a pallet that is fixtured to accept a work part of a given style or geometry.

For example, single-machine cells consisting of parts-storage units, parts-loading robots, and CNC turning centers are widely used today, although not always in a flexible mode.

Unlike rotational parts, nonrotational parts are often too heavy for a human operator to easily and quickly load into the machine tool. Accordingly, pallet fixtures were developed so that these parts could be loaded onto the pallet off-line using hoists, and then the part-on-pallet could be moved into position in front of the machine tool spindle. Nonrotational parts also tend to be more expensive than rotational parts, and the manufacturing lead times tend to be longer. These factors provide a strong incentive to produce them as efficiently as possible, using FMS technology.

EXAMPLE 19.1 Vought Aerospace FMS

A flexible manufacturing system was installed at Vought Aerospace in Dallas, Texas, by Cincinnati Milacron. The system is used to machine approximately 600 different aircraft components. The FMS consists of eight CNC horizontal machining centers plus inspection modules. Part handling is accomplished by an automated guided vehicle system (AGVS) using four vehicles. Loading and unloading of the system is done at two stations. These load/unload stations consist of storage carousels that permit parts to be stored on pallets for subsequent transfer to the machining stations by the AGVS. The system is capable of processing a sequence of single, one-of-a-kind parts in a continuous mode, so a complete set of components for one aircraft can be made efficiently without batching.

Other FMS Applications. Additional manufacturing operations in which efforts have been made to develop flexible automated systems include sheet metal stamping [38] and assembly [36]. The following example illustrates the development efforts in the pressworking area.

EXAMPLE 19.2 Flexible Fabricating System

The term *flexible fabricating system* (FFS) is sometimes used in connection with systems that perform sheet metal pressworking operations. One FFS concept was developed by Wiedemann Division of Cross & Trecker Company. The system was designed to unload sheet metal stock from the automated storage/retrieval system (AS/RS), move the stock by rail-guided cart to the CNC punch press operations, and then move the finished parts back to the AS/RS, all under computer control.

19.3.2 FMS Planning and Implementation Issues

Implementation of a flexible manufacturing system represents a major investment and commitment by the user company. It is important that the installation of the system be preceded by a thorough planning and design process, and that its operation be characterized by good management of all resources: machines, tools, pallets, parts, and people. The

coverage in this section is organized around (1) planning and design issues and (2) operations management issues.

Planning and Design Issues. The initial phase of FMS planning must consider the parts that will be produced by the system. The issues are similar to those in cellular manufacturing. They include the following:

- *Part family considerations.* Any flexible manufacturing system must be designed to process a limited range of part or product styles. In effect, the part family to be processed on the FMS must be defined. Part families can be based on product commonality as well as part similarity. The term *product commonality* refers to different components used on the same product. Many successful FMS installations are designed to accommodate part families defined by this criterion. This allows all of the components required to assemble a given product unit to be completed just prior to assembly.
- *Processing requirements.* The types of parts and their processing requirements determine the types of processing equipment that will be used in the system. In machining applications, nonrotational parts are produced by machining centers, milling machines, and similar machine tools; rotational parts are machined by turning centers and similar equipment.
- *Physical characteristics of the work parts.* The size and weight of the parts determine the sizes of the machines and the size of the material handling system that must be used.
- *Production volume.* Quantities to be produced by the system determine how many machines of each type will be required. Production volume is also a factor in selecting the most appropriate type of material handling equipment for the system.

After the part family, production volumes, and similar part issues have been decided, the design of the system is initiated. Important factors that must be specified in FMS design include:

- *Types of workstations.* The types of machines are determined by part processing requirements. Consideration of workstations must also include the load/unload station(s).
- *Variations in process routings* and *FMS layout.* If variations in process sequence are minimal, then an in-line flow is appropriate. For a system with higher product variety, a loop might be more suitable. If there is significant variation in the processing, an open field layout is appropriate.
- *Material handling system.* Selection of the material handling equipment and layout are closely related, because the type of handling system determines the layout. The material handling system includes both primary and secondary handling systems (Section 19.2.2).
- *Work-in-process* and *storage capacity.* The level of work-in-process (WIP) allowed in the FMS is an important variable in determining its utilization and efficiency. If the WIP level is too low, then stations may become starved for work, causing reduced utilization. If the WIP level is too high, then congestion may result. The WIP level should be planned, not just allowed to happen. Storage capacity in the FMS must be compatible with WIP level.

- *Tooling.* Tooling decisions include types and numbers of tools at each station, and the degree of duplication of tooling at different stations. Tool duplication at stations increases the flexibility with which parts can be routed through the system.
- *Pallet fixtures.* In machining systems for nonrotational parts, it is necessary to select the number of pallet fixtures used in the system. Factors influencing the decision include allowed WIP levels and differences in part style and size. Parts that differ too much in configuration and size require different fixturing.

Operations Management Issues. Once the FMS is installed, its resources must be optimized to meet production requirements and achieve operational objectives related to profit, quality, and customer satisfaction. The operational problems that must be addressed include the following [23], [25], [33], [34]:

- *Scheduling and dispatching.* Scheduling of production in the FMS is dictated by the master production schedule (Section 25.1). Dispatching is concerned with launching of parts into the system at the appropriate times. Several of the following problem areas are related to scheduling.
- *Machine loading.* This problem is concerned with deciding which parts will be processed on which machines and then allocating tooling and other resources to those machines to accomplish the required production schedule.
- *Part routing.* Routing decisions involve selecting the routes that should be followed by each part in the production mix in order to maximize use of workstation resources.
- *Part grouping.* Part types must be grouped for simultaneous production, given limitations on available tooling and other resources at workstations.
- *Tool management.* Managing the available tools involves making decisions on when to change tools and how to allocate tools to workstations in the system.
- *Pallet and fixture allocation.* This problem is concerned with the allocation of pallets and fixtures to the parts being produced in the system. Different parts require different fixtures, and before a given part style can be launched into the system, a fixture for that part must be made available. Modular fixtures (Section 18.3) are used to increase pallet and fixture interchangeability.

19.3.3 FMS Benefits

A number of benefits can be expected in successful FMS applications. The principal benefits are the following:

- *Increased machine utilization.* Flexible manufacturing systems achieve a higher average utilization than machines in a conventional job or batch machine shop. Reasons for this include (1) 24 hr per day operation, (2) automatic tool changing of machine tools, (3) automatic pallet changing at workstations, (4) queues of parts at stations, and (5) dynamic scheduling of production that compensates for irregularities.
- *Fewer machines required.* Because of higher machine utilization, fewer machines are required compared to a batch production plant of equivalent capacity.
- *Reduction in factory floor space.* Compared to a batch production plant of equivalent capacity, an FMS generally requires less floor area.

- *Greater responsiveness to change.* A flexible manufacturing system improves response capability to part design changes, introduction of new parts, changes in production schedule and product mix, machine breakdowns, and cutting tool failures. Adjustments can be made in the production schedule from one day to the next to respond to rush orders and special customer requests.
- *Reduced inventory requirements.* Because different parts are processed together rather than separately in batches, work-in-process is less than in batch production. For the same reason, final parts inventories are also reduced compared to make-to-stock production systems.
- *Lower manufacturing lead times.* Closely correlated with reduced work-in-process is the time spent in process by the parts. This means faster customer deliveries.
- *Reduced direct labor requirements* and *higher labor productivity.* Higher production rates and lower reliance on direct labor mean greater productivity per labor hour with an FMS than with conventional production methods.
- *Opportunity for unattended production.* The high level of automation in a flexible manufacturing system allows it to operate for extended periods of time without human attention. In the most optimistic scenario, parts and tools are loaded into the system at the end of the day shift, the FMS continues to operate throughout the night, and the finished parts are unloaded the next morning.

19.4 ANALYSIS OF FLEXIBLE MANUFACTURING SYSTEMS

Many of the design and operational problems identified in Section 19.3.2 can be addressed using quantitative analysis techniques. Flexible manufacturing systems constitute an active area of interest in operations research, and many of the important contributions are included in the list of references at the end of this chapter. FMS analysis techniques can be classified into (1) deterministic models, (2) queueing models, (3) discrete event simulation, and (4) other approaches, including heuristics.

Deterministic models are useful in obtaining starting estimates of system performance. Later in this section, a deterministic model is presented that is useful in the beginning stages of FMS design to provide rough estimates of system parameters such as production rate, capacity, and utilization. Deterministic models do not permit evaluation of operating characteristics such as the buildup of queues and other dynamics that can impair system performance. Consequently, deterministic models tend to overestimate FMS performance. On the other hand, if actual system performance is much lower than the estimates provided by these models, it may be a sign of either poor system design or poor management of FMS operations.

Queueing models can be used to describe some of the dynamics not accounted for in deterministic approaches. These models are based on the mathematical theory of queues. They permit the inclusion of queues, but only in a general way and for relatively simple system configurations. The performance measures that are calculated are usually average values for steady-state operation of the system. Examples of queueing models to study flexible manufacturing systems are described in several of the references [4], [31], and [34]. Probably the most well known of the FMS queueing models is CAN-Q [29], [30].

In the later stages of design, discrete event simulation probably offers the most accurate method for modeling the specific aspects of a given flexible manufacturing system [28], [39]. The computer model can be constructed to closely resemble the details of a

complex FMS operation. Characteristics such as layout configuration, number of pallets in the system, and production scheduling rules can be incorporated into the simulation model. Indeed, the simulation can be helpful in optimizing these characteristics.

Other techniques that have been applied to analyze FMS design and operational problems include mathematical programming [32] and various heuristic approaches [1], [13].

19.4.1 Bottleneck Model

Important aspects of FMS performance can be mathematically described by a deterministic model called the bottleneck model, developed by Solberg [31].[5] Although it has the limitations of a deterministic approach, the bottleneck model is simple and intuitive. It can be used to provide starting estimates of FMS design parameters such as production rate, number of workstations, and similar measures. The term *bottleneck* refers to the fact that the output of the production system has an upper limit, given that the product mix flowing through the system is fixed. The model can be applied to any production system that possesses this bottleneck feature, for example, a manually operated group technology cell or a production job shop. It is not limited to flexible manufacturing systems.

Terminology and Symbols. The features, terms, and symbols for the bottleneck model, as they might be applied to a flexible manufacturing system, are defined as follows:

- *Part mix.* The mix of the various part or product styles produced by the system is defined by p_j, where p_j = the fraction of the total system output that is of style *j*. The subscript $j = 1, 2, \ldots n_f$, where n_f = the total number of different part styles (family members) made in the FMS during the time period of interest. The values of p_j must sum to unity, that is,

$$\sum_{j=1}^{n_f} p_j = 1.0 \tag{19.1}$$

- *Workstations and servers.* The flexible manufacturing system has a number of distinctly different workstation types *n.* In the terminology of the bottleneck model, each workstation type may have more than one server, which simply means that it is possible to have two or more machines of the same type and capable of performing the same operations. Using the terms *stations* and *servers* in the bottleneck model is a precise way of distinguishing between machines that accomplish identical operations and those that accomplish different operations. Let s_i = the number of servers at workstation *i*, where $i = 1, 2, \ldots, n$. The load/unload station is included as one of the stations in the FMS.
- *Process routing.* For each part or product, the process routing defines the sequence of operations, the workstations where operations are performed, and the associated processing times. The sequence includes the loading operation at the beginning of processing on the FMS and the unloading operation at the end of processing. Let T_{cijk} = processing cycle time, which is the total time that a production unit occupies a given workstation server, not counting any waiting time at the station. In the

[5]Solberg's model has been simplified somewhat in this coverage, and the notation and performance measures have been adapted to be consistent with the discussion in this and other chapters.

notation for T_{cijk}, the subscript i refers to the station, j refers to the part or product style, and k refers to the sequence of operations in the process routing. For example, the fourth operation in the process plan for part A is performed on machine 2 and takes 8.5 min; thus, $T_{c2A4} = 8.5$ min. Note that process plan j is unique to part j. The bottleneck model does not conveniently allow for alternative process plans for the same part.

- *Part-handling system.* The material handling system used to transport parts or products within the FMS can be considered to be a special case of a workstation. Let it be designated as station $n + 1$, and the number of carriers in the system (conveyor carts, AGVs, monorail vehicles, etc.) is analogous to the number of servers in a regular workstation. Let s_{n+1} = the number of carriers in the part-handling system.
- *Transport time.* Let T_r = the mean transport time (repositioning time) required to move a part from one workstation to the next station in the process routing. This value could be computed for each individual transport based on transport velocity and distances between stations in the FMS, but it is more convenient to simply use an average transport time for all moves in the FMS. The same kind of average value was used for repositioning time in Chapter 15 on manual assembly lines.
- *Operation frequency.* The operation frequency is defined as the expected number of times a given operation in the process routing is performed for each work unit. For example, an inspection might be performed on a sampling basis, once every four units; hence, the frequency for this operation would be 0.25. In other cases, the part may have an operation frequency greater than 1.0, for example, for a calibration procedure that may have to be performed more than once on average to be completely effective. Let f_{ijk} = operation frequency for operation k in process plan j at station i.

FMS Operational Parameters. Using the above terms, certain operational parameters of the system can be defined. The average workload for a given station is defined as the mean total time spent at the station per part. It is calculated as

$$WL_i = \sum_j \sum_k T_{cijk} f_{ijk} p_j \tag{19.2}$$

where WL_i = average workload for station i, min; T_{cijk} = processing cycle time for operation k in process plan j at station i, min; and f_{ijk} = operation frequency for operation k in part j at station i; and p_j = part-mix fraction for part j.

The part-handling system (station $n + 1$) is a special case, as noted previously. The workload of the handling system is the mean transport time multiplied by the average number of transports required to complete the processing of a work part. The average number of transports is equal to the mean number of operations in the process routing minus one. That is,

$$n_t = \sum_i \sum_j \sum_k f_{ijk} p_j - 1 \tag{19.3}$$

where n_t = mean number of transports, and the other terms are defined earlier.

EXAMPLE 19.3 Determining n_t

Consider a manufacturing system with two stations: (1) a load/unload station and (2) a machining station. The system processes just one part, part A, so the part-mix fraction $p_A = 1.0$. The frequency of all operations is $f_{iAk} = 1.0$. The parts are loaded at station 1, routed to station 2 for machining, and then sent back to station 1 for unloading (three operations in the routing). Using Equation (19.3),

$$n_t = 1(1.0) + 1(1.0) + 1(1.0) - 1 = 3 - 1 = 2$$

Looking at it another way, the process routing is (station 1) → (station 2) → (station 1). Counting the number of arrows gives the number of transports: $n_t = 2$.[6]

The workload of the handling system can now be computed:

$$WL_{n+1} = n_t T_r \tag{19.4}$$

where WL_{n+1} = workload of the handling system, min; n_t = mean number of transports by Equation (19.3); and T_r = mean transport time per move, min.

System Performance Metrics. Measures to assess performance of a flexible manufacturing system include production rate of all parts, production rate of each part style, utilization of the different workstations, and number of busy servers at each workstation. These measures can be calculated under the assumption that the FMS is producing at its maximum possible rate. This rate is constrained by the bottleneck station in the system, which is the station with the highest workload per server. The workload per server is simply the ratio WL_i/s_i for each station. Thus, the bottleneck is identified by finding the maximum value of the ratio among all stations. The comparison must include the handling system, because it might be the bottleneck.

Let WL^* and s^* equal the workload and number of servers, respectively, for the bottleneck station. The maximum production rate of all parts of the FMS can be determined as the ratio of s^* to WL^*. It is the maximum production rate, because it is limited by the capacity of the bottleneck station,

$$R_p^* = \frac{s^*}{WL^*} \tag{19.5}$$

where R_p^* = maximum production rate of all part styles produced by the system, which is determined by the capacity of the bottleneck station, pc/min; s^* = number of servers at the bottleneck station, and WL^* = workload at the bottleneck station, min/pc. It is not difficult to grasp the validity of this formula as long as all parts are processed through the bottleneck station. A little more thought is required to appreciate that Equation (19.5) is also valid even when not all the parts pass through the bottleneck station, as long as the product mix (p_j values) remains constant. In other words, if those parts not passing

[6]Counting the arrows works only when $f_{ijk} = 1$ for all i, j, and k. When one or more f_{ijk} = fractions, then this is a fractional move, and the counting of arrows gets complicated. The safest approach is to use Equation (19.3).

through the bottleneck are not allowed to increase their production rates to reach their respective bottleneck limits, then these parts' rates will be limited by the part-mix ratios.

The value of R_p^* includes parts of all styles produced in the system. Individual part production rates can be obtained by multiplying R_p^* by the respective part-mix ratios. That is,

$$R_{pj}^* = p_j(R_p^*) = p_j \frac{s^*}{WL^*} \tag{19.6}$$

where R_{pj}^* = maximum production rate of part style j, pc/min; and p_j = part-mix fraction for part style j.

The mean utilization of each workstation is the proportion of time that the servers at the station are working and not idle. This can be computed as:

$$U_i = \frac{WL_i}{s_i}(R_p^*) = \frac{WL_i}{s_i}\frac{s^*}{WL^*} \tag{19.7}$$

where U_i = utilization of station i; WL_i = workload of station i, min/pc; s_i = number of servers at station i; and R_p^* = overall production rate, pc/min. The utilization of the bottleneck station is 100% at R_p^*.

To obtain the average station utilization, simply compute the average value for all stations, including the transport system:

$$\overline{U} = \frac{\sum_{i=1}^{n+1} U_i}{n+1} \tag{19.8}$$

where $\overline{U}$ is an unweighted average of the workstation utilizations.

An alternative and perhaps more meaningful measure of overall FMS utilization can be obtained using a weighted average, where the weighting is based on the number of servers at each station for the n regular stations in the system, including the load/unload station but excluding the transport system. The argument for omitting the transport system is that the utilization of the processing stations is the important measure of FMS utilization. The purpose of the transport system is to serve the processing stations, and therefore its utilization should not be included in the average. The overall FMS utilization is calculated as:

$$\overline{U}_s = \frac{\sum_{i=1}^{n} s_i U_i}{\sum_{i=1}^{n} s_i} \tag{19.9}$$

where $\overline{U}_s$ = overall FMS utilization, s_i = number of servers at station i, and U_i = utilization of station i.

Finally, the number of busy servers at each station is of interest. All of the servers at the bottleneck station are busy at the maximum production rate, but the servers at the other stations are idle some of the time. The values can be calculated as

$$BS_i = WL_i(R_p^*) = WL_i \frac{s^*}{WL^*} \tag{19.10}$$

where BS_i = number of busy servers on average at station i and WL_i = workload at station i.

EXAMPLE 19.4 Bottleneck Model

A flexible machining system consists of a load/unload station and two machining workstations. Station 1 is the load/unload station with one server (human worker). Station 2 performs milling and consists of three identical CNC milling machines. Station 3 performs drilling and consists of two identical CNC drill presses. The stations are connected by a part-handling system that has two carriers. The mean transport time is 2.5 min. The FMS produces three parts, A, B, and C. The part-mix fractions and process routings for the three parts are presented in the table below. The operation frequency $f_{ijk} = 1.0$ for all i, j, and k. Determine (a) maximum production rate of the FMS, (b) corresponding production rates of each product, (c) utilization of each station, (d) average utilization of the processing stations, and (e) number of busy servers at each station.

Part j	Part Mix p_j	Operation k	Description	Station i	Process Time T_{cijk} (min)
A	0.4	1	Load	1	4
		2	Mill	2	25
		3	Drill	3	10
		4	Unload	1	2
B	0.35	1	Load	1	4
		2	Mill	2	20
		3	Drill	3	15
		4	Unload	1	2
C	0.25	1	Load	1	4
		2	Mill	2	15
		3	Unload	1	2

Solution: The computations were performed using a spreadsheet calculator with the results shown below. Equations used to compute the entries are given in the top row of the table. Station 2 has the highest *WL/s* ratio (6.9167) so it is the bottleneck station.

Equation		(19.2)	(19.4)		(19.7)	(19.10)
Station	Servers	*WL*	*WL*	*WL/s*	*U*	*BS*
1 (Load/unload)	1	6		6	**0.867**	**0.867**
2 (Mill)	3	20.75		6.917*	**1**	**3**
3 (Drill)	2	9.25		4.625	**0.668**	**1.337**
4 (Part handling)	2		6.875	3.438	**0.497**	**0.994**

* Highest value of *WL/s* denotes bottleneck station.

(a) Overall production rate is given by Equation (19.5) using station 2 values:

$$R_p^* = s^*/WL^* = 3/20.75 = 0.1446 \text{ pc/min} = \mathbf{8.675\ pc/hr}$$

(b) To determine the production rate of each product, multiply R_p^* by its respective part-mix fraction.

$$R_{p\text{A}}^* = 8.675(0.4) = \mathbf{3.470\ pc/hr}$$

$$R_{p\mathrm{B}}^* = 8.675(0.35) = \mathbf{3.036\ pc/hr}$$
$$R_{p\mathrm{C}}^* = 8.675(0.25) = \mathbf{2.169\ pc/hr}$$

(c) The utilization of each station can be computed using Equation (19.7). The values are shown in the table.

(d) Average utilization of the processing stations is based on Equation (19.9):

$$\overline{U}_s = \frac{0.8675 + 3 + 1.3373}{6} = \mathbf{0.8675}$$

(e) Mean number of busy servers at each station is determined using Equation (19.10). The values are shown in the table.

19.4.2 Extended Bottleneck Model

The bottleneck model assumes that the bottleneck station is utilized 100% and that there are no delays due to queues in the system. This implies, on the one hand, that there are a sufficient number of parts in the system to avoid starving of workstations and, on the other hand, that there will be no delays due to queueing. Solberg [31] argued that the assumption of 100% utilization makes the bottleneck model overly optimistic and that a queueing model which accounts for process time variations and delays would more realistically describe the performance of a flexible manufacturing system.

An alternative approach, developed by Mejabi [24], addresses some of the weaknesses of the bottleneck model without resorting to queueing computations (which can be difficult). He called his approach the *extended bottleneck model.* This extended model assumes a closed queueing network in which there are always a certain number of work parts in the FMS. Let N = this number. When one part is completed and exits the FMS, a new raw work part immediately enters the system, so that N remains constant. The new part may or may not have the same process routing as the one that just departed.

N plays a critical role in the operation of the manufacturing system. If N is smaller than the number of workstations, then some of the stations will be idle due to starving, sometimes even the bottleneck station. In this case, the production rate of the FMS will be less than R_p* calculated in Equation (19.5). If N is larger than the number of workstations, then the system will be fully loaded, with parts waiting in front of stations. In this case, R_p* will provide a good estimate of the production capacity of the system. However, work-in-process (*WIP*) will be high, and manufacturing lead time (MLT) will be long.

In effect, *WIP* corresponds to N, and MLT is the sum of processing times at the stations, transport times between stations, and any waiting time experienced by the parts in the system:

$$MLT = \sum_{i=1}^{n} WL_i + WL_{n+1} + T_w \tag{19.11}$$

where $\sum_{i=1}^{n} WL_i$ = summation of average workloads over all stations in the FMS, min; WL_{n+1} = workload of the part-handling system, min; and T_w = mean waiting time experienced by a part due to queues at the stations, min.

WIP (i.e., N) and MLT are correlated. If N is small, then MLT will take on its smallest possible value because waiting time will be short or even zero. If N is large, then MLT will be long and there will be waiting time for parts in the system. Thus there are two alternative cases, and adjustments must be made in the bottleneck model to account for them. To do this, Mejabi used the well-known Little's formula[7] from queueing theory. Using the symbols developed here, Little's formula is expressed as:

$$N = R_p(MLT) \tag{19.12}$$

where N = number of parts in the system, pc; R_p = production rate of the system, pc/min; and MLT = manufacturing lead time (time spent by a part in the system), min. Now consider the two cases:

Case 1: When N is small, production rate is less than in the bottleneck case because the bottleneck station is not fully utilized. In this case, the waiting time T_w of a unit is theoretically zero, and Equation (19.11) reduces to

$$MLT_1 = \sum_{i=1}^{n} WL_i + WL_{n+1} \tag{19.13}$$

where the subscript in MLT_1 is used to identify case 1. Production rate can be estimated using Little's formula:

$$R_p = \frac{N}{MLT_1} \tag{19.14}$$

and production rates of the individual parts are given by

$$R_{pj} = p_j R_p \tag{19.15}$$

As indicated waiting time is assumed to be zero:

$$T_w = 0 \tag{19.16}$$

Case 2: When N is large, the estimate of maximum production rate provided by Equation (19.5) should be valid: $R_p^* = s^*/WL^*$, where the asterisk $(^*)$ denotes that production rate is constrained by the bottleneck station in the system. The production rates of the individual products are given by

$$R_{pj}^* = p_j R_p^* \tag{19.17}$$

In this case, average manufacturing lead time is evaluated using Little's formula:

$$MLT_2 = \frac{N}{R_p^*} \tag{19.18}$$

The mean waiting time a part spends in the system can be estimated by rearranging Equation (19.11) to solve for T_w:

$$T_w = MLT_2 - \left(\sum_{i=1}^{n} WL_i + WL_{n+1} \right) \tag{19.19}$$

[7]Little's formula is usually given as $L = \lambda W$, where L = expected number of units in the system, λ = processing rate of units in the system, and W = expected time spent by a unit in the system.

The decision on whether to use case 1 or case 2 depends on the value of N. The dividing line between cases 1 and 2 is determined by whether N is greater than or less than a critical value given by

$$N^* = R_p^*\left(\sum_{i=1}^{n} WL_i + WL_{n+1}\right) = R_p^*(MLT_1) \tag{19.20}$$

where N^* = critical value of N, the dividing line between the bottleneck and non-bottleneck cases. If $N < N^*$, then case 1 applies. If $N \geq N^*$, then case 2 applies.

EXAMPLE 19.5 Extended Bottleneck Model

Use the extended bottleneck model on the data in Example 19.4 to compute hourly production rate, manufacturing lead time, and waiting time for four values of N: (a) $N = 4$, (b) $N = 6$, (c) $N = 7$, and (d) $N = 10$.

Solution: First compute the critical value of N, using Equation (19.20). From Example 19.4 $R_p^* = 8.675/60 = 0.1446$ pc/min. Also needed is the value of MLT_1. Using previously calculated values from Example 19.4 in Equation (19.13),

$$MLT_1 = 6 + 20.75 + 9.25 + 6.875 = 42.875 \text{ min} = \sim 42.9 \text{ min}$$

The critical value of N is given by Equation (19.20):

$$N^* = 0.1446(42.875) = 6.2 \text{ pc}$$

(a) $N = 4$ is less than the critical value, so the equations for case 1 apply.

$$MLT_1 = \mathbf{42.9 \text{ min}}$$

$$R_p = \frac{N}{MLT_1} = \frac{4}{42.875} = 0.0933 \text{ pc/min} = \mathbf{5.6 \text{ pc/hr}}$$

$$T_w = \mathbf{0}$$

(b) $N = 6$ is again less than the critical value, so case 1 applies.

$$MLT_1 = \mathbf{42.9 \text{ min}}$$

$$R_p = \frac{6}{42.875} = 0.1399 \text{ pc/min} = \mathbf{8.4 \text{ pc/hr}}$$

$$T_w = \mathbf{0}$$

(c) $N = 7$ is greater than the critical value, so case 2 applies.

$$R_p^* = \frac{s^*}{WL^*} = \frac{3}{20.75} = 0.1446 \text{ pc/min} = \mathbf{8.7 \text{ pc/hr}}$$

$$MLT_2 = \frac{7}{0.1446} = \mathbf{\sim 48.4 \text{ min}}$$

$$T_w = 48.4 - 42.9 = \mathbf{\sim 5.5 \text{ min}}$$

(d) $N = 10$ is greater than the critical value, so case 2 applies.

$$R_p^* = 0.1446 \text{ pc/min} = \textbf{8.7 pc/hr} \text{ calculated in part (c)}$$

$$MLT_2 = \frac{10}{0.1446} = \textbf{~69.167 min}$$

$$T_w = 69.2 - 42.9 = \textbf{~26.3 min}$$

The results of this example typify the behavior of the extended bottleneck model, shown in Figure 19.8. Below N^* (Case 1), MLT has a constant value, and R_p decreases proportionally as N decreases. Manufacturing lead time cannot be less than the sum of the processing and transport times, and production rate is adversely affected by low values of N because stations become starved for work. Above N^* (Case 2), R_p has a constant value equal to R_p^* and MLT increases. No matter how large N is, the production rate cannot be greater than the output capacity of the bottleneck station. Manufacturing lead time increases because backlogs build up at the stations.

These observations might tempt the reader to conclude that the optimum N value occurs at N^*, because MLT is at its minimum possible value and R_p is at its maximum possible value. However, caution must be exercised in the use of the extended bottleneck model (and the same caution applies even more so to the conventional bottleneck model, which disregards the effect of N). It is intended to be a rough-cut method to estimate performance in the early phases of FMS design. More reliable estimates of performance can be obtained using computer simulations of detailed models of the FMS—models that include considerations of layout, material handling and storage system, and other system design factors.

19.4.3 Sizing the FMS

The bottleneck model can be used to calculate the number of servers required at each workstation to achieve a specified production rate. Such calculations would be useful during the initial stages of FMS design in determining the "size" (number of workstations and servers) of the system. To make the computation, the part mix, process routings, and processing times must be known so that workloads can be calculated for each station to

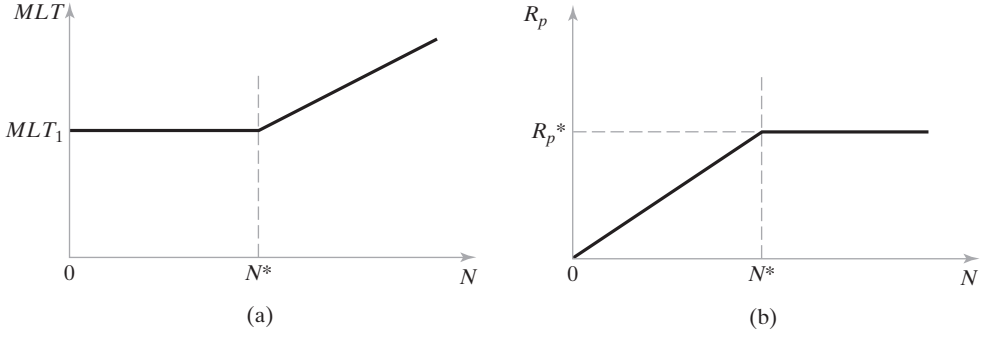


Figure 19.8 General behavior of the extended bottleneck model: (a) manufacturing lead time MLT as a function of N; and (b) production rate R_p as a function of N.

be included in the FMS. Given the workloads, the number of servers at each station i is determined as

$$s_i = \text{Minimum Integer} \geq R_p(WL_i) \tag{19.21}$$

where s_i = number of servers at station i; R_p = specified production rate of all parts to be produced by the system, pc/min; and WL_i = workload at station i, min. The following example illustrates the procedure.

EXAMPLE 19.6 Sizing the FMS

Suppose the part mix, process routings, and processing times for the family of parts to be machined on a proposed FMS are those given in Example 19.4. The FMS will operate 24 hr/day, 5 days/wk, 50 wk/yr. Determine (a) the number of servers that will be required at each station i to achieve an annual production rate of 60,000 parts/yr and (b) the utilization of each workstation.

Solution: (a) The number of hours of FMS operation per year will be $24 \times 5 \times 50 = 6{,}000$ hr/yr. The hourly production rate is given by:

$$R_p = \frac{60{,}000}{6{,}000} = 10.0 \text{ pc/hr} = 0.1667 \text{ pc/min}$$

The workloads at each station were previously calculated in Example 19.4: $WL_1 = 6.0$ min, $WL_2 = 20.75$ min, $WL_3 = 9.25$ min, $WL_4 = 6875.0$ min, and $WL_5 = 10.06$ min. Using Equation (19.21), the following number of servers are required at each station:

s_1 = minimum integer $\geq 0.1667(6.0) = 1.000 =$ **1 server**

s_2 = minimum integer $\geq 0.1667(20.75) = 3.458$ rounded up to **4 servers**

s_3 = minimum integer $\geq 0.1667(9.25) = 1.54$ rounded up to **2 servers**

s_4 = minimum integer $\geq 0.1667(6.875) = 1.146$ rounded up to **2 servers**

(b) The utilization at each workstation is determined as the calculated value of s_i divided by the resulting minimum integer value $\geq s_i$.

$U_1 = 1.0/1 = 1.0 =$ **100%**

$U_2 = 3.458/4 = 0.865 =$ **86.5%**

$U_3 = 1.54/2 = 0.77 =$ **77%**

$U_4 = 1.146/2 = 0.573 =$ **57.3%**

The maximum value is at station 1, the load/unload station. This is the bottleneck station.

Because the number of servers at each workstation must be an integer, station utilization may be less than 100% for most of the stations. In Example 19.6, the load/unload station has a utilization of 100%, but all of the other stations have utilizations less than 100%. It's a shame that the load/unload station is the bottleneck on the overall production rate of the FMS. It would be much more desirable for one of the production stations to be the bottleneck.