Section 17
Industrial Engineering

BY

ERWIN M. SANIGA *Dana Johnson Professor of Information Technology and Professor of Operations Management, University of Delaware.*
SCOTT JONES *Professor, Department of Accounting & MIS, Alfred Lerner College of Business and Economics, University of Delaware.*
ASHLEY C. COCKERILL *Vice President and Event Coordinator, nanoTech Business, Inc.*
VINCENT M. ALTAMURO *President, VMA, Inc., Toms River, NJ.*
ANDREW M. DONALDSON *Project Director, Parsons E&C, Reading, PA.*
ROBERT F. GAMBON *Power Plant Design and Development Consultant.*
EZRA S. KRENDEL *Emeritus Professor of Operations Research and Statistics, Wharton School, University of Pennsylvania.*

17.1 OPERATIONS MANAGEMENT
by Erwin M. Saniga

17.2 COST ACCOUNTING
by Scott Jones

17.3 ENGINEERING STATISTICS AND QUALITY CONTROL
by Ashley C. Cockerill

17.4 METHODS ENGINEERING
by Vincent M. Altamuro

17.5 COST OF ELECTRIC POWER
by Andrew M. Donaldson and Robert F. Gambon

17.6 HUMAN FACTORS AND ERGONOMICS
by Ezra S. Krendel

17.7 AUTOMATIC MANUFACTURING
by Vincent M. Altamuro

17.1 OPERATIONS MANAGEMENT

by Erwin M. Saniga

REFERENCES: Meredith and Shafer, "Operations management for MBAs," 2d ed., John Wiley & Sons. Chase, Aquilano, and Jacobs, "Operations Management for Competitive Advantage," 9th ed., McGraw-Hill Irwin. Stevenson, "Production/Operations Management," 5th ed., Irwin. Montgomery, "Introduction to Statistical Quality Control," 3d ed., John Wiley and Sons. Hoerl, "Six Sigma and the Future of the Quality Profession," *Quality Progress*, June 1998, pp. 35–42. Mentch, "Manufacturing Process Quality Optimization Studies," *Journal of Quality Technology*, vol. 12, no. 3, July 1980, pp. 119–129. Dilworth, "Operations Management: Providing Value in Goods and Services," 3d ed., Dryden Press. Box, Jenkins, and Reinsel, "Time Series Analysis," 3d ed., Prentice Hall. Deming, "Out of the Crisis," MIT Center for Advanced Engineering Study.

Operations management is concerned with the production of the goods and services of an organization. In particular, operations management problems focus on determining the way an organization will produce high-quality goods or services in an efficient manner. We define efficiency as producing goods and services in a timely and cost-efficient manner. In the past, the field was limited to the study of manufacturing processes, but in the last several decades the emphasis has included the study of service processes, which now number more organizations than manufacturing in the United States. Many of the same principles of effective operations management that have been found to work in manufacturing organizations will optimize the efficiency and quality of service organizations.

Operations management has generated increased interest among organizations today because of the competitive advantage that accrues to the organization that has a high-quality product or service and manages its efficiency of production in an optimal way. The terminology that is common in business today is due to an emphasis on operations management strategies. Some examples include six sigma, lean manufacturing, supply chain management, and just-in-time inventory methods.

FORECASTING

There are three general types of forecasting models used in organizations today. These are time series models, regression models, and subjective models.

Time series models are models that are based upon the assumptions that the past behavior of the variable to be forecasted, say sales, is the best indicator of the future behavior of sales. Some of the common models that have been used in many firms include moving-average methods and exponential smoothing methods. More complex models due to Box and Jenkins are used when simpler models do not satisfy.

Moving-average methods are based upon simply averaging the past n observations of a stream of data to obtain a forecast for the next period. For example, a three-period moving average of sales would be obtained by averaging the past three periods of sales. This would be the forecast for the next period (or future periods) sales. The formula for a moving average forecast of sales is simply

$$S_{t+1} = \sum_{i=t-N+1} X_i / N$$

where S_t is the forecast in period t, X_i is sales in period i, and N is the number of periods in the moving average.

An obvious weakness of moving-average forecasts is that equal weight is placed on all observations. It seems more plausible to weigh the more recent observations higher in determining a forecast. **Exponential smoothing** is a common method that allows more recent observations to have more weight in the averaging scheme. The model for exponential smoothing is:

$$S_{t+1} = \alpha X_t + (1 - \alpha) S_t$$

where α is the smoothing coefficient.

Note that the model is dependent upon the choice of α. If α is near 1, there is very little smoothing taking place and the forecast follows the rapid changes in the data. Note that if α equals 1, the forecast for the next period is simply the actual observation for the previous period. If α is near zero, the forecast approaches the average of the past data. In practice, common values of α are between 0.1 and 0.3.

Trend and seasonalities are not accounted for in the exponential smoothing model presented above. One can account for trend in a model by using a model called **double exponential** smoothing, which is a simple extension of the model presented above. **Seasonalities** are common in many industries. Consider the demand for the sale of turkeys. One would expect there would be peaks around the Thanksgiving and Christmas holidays in the United States. The usual exponential smoothing or double exponential smoothing models would yield forecasts that lag the actual demand. One simple and effective method for handling these seasonalities is to first **deseasonalize** the data by calculating the seasonal component for the data. An example of this would be to calculate the percentage of sales occurring in each month by averaging the last 3 years of monthly data (assuming there is a monthly seasonality) and then dividing the data by this monthly index. One can then develop a forecast on the deseasonalized data and, when the forecast is obtained, multiply the forecast by the seasonal index for the month in which a forecast is desired.

More complicated methods for forecasting time series are due to George Box and Gwilym Jenkins and are based upon a set of models called ARIMA, or mixed autoregressive and moving-average models. These are more general and powerful models of the class mentioned above and have been shown to be extremely effective in forecasting in practice. Consult their book for a detailed explanation of this class of models.

A second class of forecasting models is based upon using explanatory variables in a **regression** model to forecast sales. The regression model may be represented by the function F, where

$$Y = F(X_1, X_2, \ldots, X_k) + \varepsilon$$

The variables labeled X_j, $j = 1, 2, \ldots, k$ are the explanatory variables, and the variable Y is called the dependent variable. The term ε represents the error. For example, we may hypothesize that sales (Y) can be explained by price (X_1) and average delivery time (X_2). We can build the model by entering data from the past where we have recorded average delivery time in a month, the price in a month, and sales in a month.

Human judgment is the basis for a third class of models. These are commonly used in practice and, at times, yield accurate forecasts. Perhaps the most common types of judgmental forecasts are those based upon management or sales force opinions obtained from discussions with customers. Surveys of customers are also used to build these types of forecasts.

INVENTORY MANAGEMENT

Inventory decisions have a major impact on the profitability of an organization. Organizations that carry large inventories incur large holding costs, which impact the bottom line. On the other hand, maintaining low inventories may prevent an organization from meeting the variability in demand. Additionally, low inventories may not allow flexibility in scheduling, may not allow the organization to take account of economic order quantities, or may not allow for variation that occurs in vendor-supplied inputs. Thus, the inventory decision is very important in organizations. Essentially, the inventory decision can be broken down into two major components. These are (*a*) determining when items should be ordered and (*b*) determining how much to order.

The standard criterion on which to base inventory decision is cost. Four costs are generally considered. These are holding or carrying costs, ordering costs, setup costs, and shortage costs. **Holding costs** include as a major component the cost of the capital tied up in the inventory and additional costs include the cost of storage, insurance, deterioration, obsolescence, etc. **Ordering costs** refer to the costs incurred when a purchase order is placed or when a production order is given. **Setup costs** include the clerical costs in changing production. **Shortage costs** are difficult to measure since they include a component for lost sales, lost customers, or penalties for late delivery.

When analyzing inventory decisions, analysts place inventory in two categories: independent demand inventory and dependent demand inventory. Independent demand items are items whose demand does not depend on the demand for another item. Dependent demand items are ones for which the demand depends on the demand for another item. For example, the demand for bicycles in an organization may be an independent demand item. Wheels for the bicycle, on the other hand, are dependent demand items, since their demand depends on the demand for bicycles. That is, if the demand estimate for bicycles is 500 units, the demand for wheels is 1,000 units.

The most common inventory model is labeled the **economic order quantity (EOQ)** model. It is used when one is interested in determining the optimal order quantity where the criterion is cost minimization and where the inventory demand item is independent. The assumptions of the EOQ model are that demand is satisfied on time, price is fixed (there are no price breaks), stock is depleted linearly, and there are no stockouts. Under these assumptions one can show that the EOQ is

$$\mathrm{EOQ} = (2\,R\,c_p/c_h)^{1/2}$$

where c_p is the ordering costs, R is the yearly demand, and c_h is the holding costs. In addition, it can be shown that the reorder point is when the inventory reaches a level of P, where

$$P = L \times R/52$$

where L is the lead time in weeks. Other similar models are developed for the situations where price breaks are available or where backorders are allowable.

An equivalent model to the EOQ is available for the situation where the analyst wishes to determine the **economic production quantity** or lot size (**EPQ**). Here, one can find the EPQ as

$$\mathrm{EPQ} = \{2c_p R/[c_h(1 - r/p)]\}^{1/2}$$

where r = daily usage and p = production rate per day.

When demand is not constant, organizations must provide a level of safety stock if they wish to maintain a certain service level. Safety stock, then, is the excess inventory carried over expected demand. In the standard EOQ model, which is also called the *fixed order quantity* model, the EOQ is calculated in the same way if the demand is constant or varies. But if the demand varies, the **reorder point** differs. If the demand varies, the reorder point must be calculated by using some knowledge of the distribution of demand. For example, if demand is normally distributed with standard deviation, the reorder point becomes

$$P = d\,L + z$$

where d is the average daily demand, L is the lead time in days, and z is the number of standard deviations associated with a particular service level probability.

Another way to address this problem is to use a fixed time period model. Here, order quantities vary from time period to time period but the reorder point remains the same. An advantage of this model is that one does not have to constantly account for the inventory; all one must keep track of is the time period. A disadvantage of this model is that it requires the maintenance of a higher level of inventory. Most textbooks present the models and discuss the implications of the fixed time period model.

There are other simpler systems that enjoy popularity in applications because they do not rely on any assumptions such as a normal distribution of demand or constant usage. Some common systems include the ABC system of classifying inventory and one- and two-bin systems.

The **ABC** system for inventory management essentially classifies inventory into one of three levels: A, B, and C. The classification is based upon the dollar value of inventory, which is determined by multiplying annual usage times the cost. Then, the A items are defined as the ones with the top 75 to 80 percent of dollar value. B items are the items with the next 15 percent of value and C items are those with the next 5 to 10 percent of value. Once the items are classified, management uses this information to determine where to apply the greatest control.

The **one-bin system** is very simple to manage but has a disadvantage in its cost that may be far from minimal compared to more complex systems. To use the one-bin system the analyst simply replenishes inventory up to a maximum level on a regular basis, say weekly for A items. The **two-bin system** is one in which there are two bins of the same inventory. Inventory is used from the first bin until it is used up; subsequently an order is placed for replenishment and inventory is used from the second bin. The second bin should contain enough inventory to cover needs during the order lead time. The amount in the second bin thus should equal P, where P is the order point defined above.

The methods discussed above constitute some of the more valuable methods of inventory management available to managers but, at times, various industries employ other methods that are more usable because of the difficulty in managing a large number of items. For example, shoes are distributed according to batches organized by sizes. Department stores use a method called “open to buy,” which refers to the budgeted value of inventory minus the amount that was currently spent for various departments. That is, instead of managing individual inventory items, the manager focuses on managing the cost of inventory for a department as a whole.

As mentioned earlier, the demand for inventory can be independent or dependent. When demand for an inventory item is dependent upon the demand for another item, such as the case for bicycle wheels examined before, the methods of managing inventory differ from the case where demand is independent. A common technique for managing dependent demand inventories is the method of **materials requirements planning (MRP)**, which is an accounting information system technique used for keeping track of inventory needs for these items. MRP has been popular for about 30 years, and computer programs and systems to perform the cumbersome calculations associated with MRP are readily available.

MRP is based upon knowledge of the bill of materials for an item, the master production schedule, and a record of the inventories currently in stock. The bill of materials is simply a list of the components of a particular product. For our bicycle example, the bill of materials would contain two wheels, two tires, a set of handlebars, a set of rear brakes, front brakes, and various other items. The master production schedule provides information on when to produce these bicycles and what quantity to produce. The MRP then uses these inputs along with lead times to provide the production manager with planned order releases along with various other kinds of information, depending upon the needs of the manager. In our simple example, suppose our forecasted demand is such that we need to produce 3,000 bicycles in March and in June; this would be the master production schedule. Suppose we have 3,000 wheels in inventory and the lead time for wheels is 1 month. The MRP would calculate that the organization needs 6,000 wheels in March and in June, but since the organization has 3,000 on hand it would show the necessary order releases are for 3,000 wheels on February 1 and for 6,000 wheels on May 1.

Of course, in practice MRP is much more complicated than the simple example presented here, for a number of reasons. For example, some firms produce many products and some of these have common parts, or dependent items. Order releases must account for these factors. Also, firms might want to take advantage of price breaks for larger lots or by ordering less often to minimize ordering costs. In MRP, this problem is called **lot sizing.** A variety of techniques for lot sizing are available and range from the simplest, called *lot-for-lot ordering*, to complex ones such as dynamic programming techniques—for example, the Wagner Whitin model. The lot-for-lot model, for example, indicates that orders are to be placed when they are needed for each period; here, holding costs are minimized.

Related to MRP is an approach to broaden the scope of MRP to the firm as a whole. That is, all of the factors going into production are considered together on a larger scale than just the production system. This includes the overall planning of the business. Marketing and finance functions, for example, might take part in the determination of the master production schedule, forecasting, the financing requirements of the process, etc. This approach is called **MRP II** and has been popular for about 20 years.

An approach to managing production systems that includes the management of its inventory is called a **just-in-time** approach. This approach was initially used at Toyota in Japan and is very much related to the teachings of W. Edwards Deming in that it requires that the production system be in a state of statistical control, that suppliers be kept to a minimum, that scrap and rework are to be avoided as a policy rather than as a cost consideration, that preventive maintenance be performed, that continuous improvement is in place, and that workers are team based. Because the system is in control and there is a minimum of scrap or rework, an innovative method of managing the inventory is possible. This method is to manage inventory, whether purchased from a vendor or work-in-process inventory, so that large stocks of inventory are not necessary to be held, thereby minimizing holding costs. Other common elements of just-in-time systems are that setups are quick, and therefore not costly, and that lot sizes are small. Additionally, just-in-time systems move work in a pull versus push way. Push systems work with less regard for the forecast or the demands of higher upstream stations. Pull systems, on the other hand, work in a just-in-time method such that the upstream demand is pulled from downstream. As a simple example, a firm might produce according to expected demand only rather than by efficient lot sizes.

A drawback encountered in just-in-time systems is that, since the supplier is penalized for late delivery, it often prepares and stores finished product which is then shipped to arrive at the customer just in time. The carrying cost for this storage, which the customer thinks has been avoided, is passed on to the customer in the form of an increased price.

One way in which to manage the information in a pull system is with a **kanban sheet,** which is simply a record of what work is taking place. For example, a kanban sheet might be attached to a lot of parts. When upstream demand requires that this lot is needed for production, the kanban sheet is placed in an area where the need for replenishment is indicated.

JIT systems are conceptually very attractive and have been used to advantage by a number of companies, especially those in Japan.

SCHEDULING

Scheduling the accomplishment of tasks within the operations of a manufacturing or service environment is dependent upon a number of factors. First, there are differences between services and manufacturing in that the concept of an inventory of service has little meaning. Second, within manufacturing there are different types of production systems used for different types of products. More specifically these can be broken down into systems in which the same product is made many times (flow shop), a system in which several products are made by the same manufacturer on the same production system, a system in which similar products are made to order (job shop), and finally a system in which a single product is made (project). Some examples of these are respectively automobile production, a private press, a machine shop producing milled products, and construction of a skyscraper.

In a service environment, operations scheduling is a matter of providing the service when it is demanded by the customer. Analytical approaches involve developing forecasts of when customers arrive and the probability distributions of these arrivals and using these to build service capacity so that desired service levels are achieved. These service levels are usually expressed in terms of the probability or percentage of the time customers are serviced within a time limit. For example, the credit card company MBNA answers more than 90 percent of its phone calls from customers within two rings.

Some organizations such as doctors' offices schedule by appointment. The demand is then known and the service supply can be provided to meet that customer demand. Airlines use reservations to accomplish much of the same objective. Of course the determination of feasible schedules involves forecasting as well.

Complex service operations such as hospitals have compound scheduling problems encompassing the issues mentioned above in addition to the consideration that multiple service providers of various types are needed. For example, scheduling surgery involves scheduling surgeons, nurses, anesthesiologists, operating rooms, etc.

Scheduling for the situation when a product is produced in a series of workstations, or a flowshop, is called assembly line balancing. The product spends a period of time in each workstation; this time is referred to as the *cycle time*. The objective of assembly line balancing is to allocate all of the tasks needed to complete a product into workstations. Criteria for evaluation of assembly line balances include the cycle time required, the number of workstations required, and the amount of slack in the assembly line, which is the sum of the idle time in the workstations.

To perform assembly line balancing for producing a particular product, one needs to establish the precedence relationships between the tasks and determine the task times for each of the tasks. In the situation where cycle time is specified, one can calculate the theoretical minimum number of workstations; it is the next highest integer obtained from the ratio S/C, where S is the sum of the task times and C is the cycle time. In assembly line balancing, one then assigns tasks to workstations such that precedence relationships are not violated and, in addition, the sum of the task times for each station is less than the cycle time.

When one assigns tasks to workstations, one finds that there usually will be alternative assignments. Researchers have investigated the performance of a number of heuristics to use in these situations. Some of the more common are to assign tasks to workstations such that the longest tasks are assigned first without, of course, violating precedence relationships, and to assign tasks according to positional weights, where positional weight is the sum of the task times and the times of all of the tasks that follow.

A graph showing the precedence relationships for an 8-task assembly line is shown in Fig. 17.1.1. The times for the various tasks are:

Task	1	2	3	4	5	6	7	8	Total
Time	2	4	3	4	5	4	6	4	32

Thus, if the cycle time is 7, the theoretical minimum number of workstations is $32/7 = 4.55 = 5$.

If we allocate tasks to stations such that the longer tasks are assigned first, we have the following balance:

Station	1	2	3	4	5	6
Tasks	1,2	5	3,6	4	7	8
Time	6	5	7	4	6	4
Slack	1	2	0	3	1	2

In practice the problem is usually more complex because task times are stochastic variables and actual results may differ from the expected. Additionally, some tasks may require special training, which may change the way in which tasks are assigned to workstations. Many other

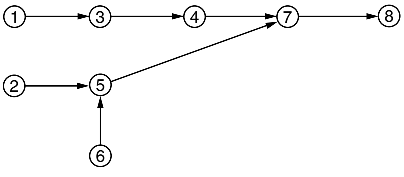


Fig. 17.1.1 Network showing precedence relationships for an 8-task assembly line. Code numbers within nodes signify events. Connecting lines with directional arrows indicate operations that are dependent on prerequisite operations.

situations occur that make the procedures described above a starting point in the assembly line balancing decision.

When a firm produces a number of products using similar equipment and locations, the scheduling problem becomes more complex because the production manager needs to make decisions on the size of the lot produced and the sequencing of the lots. In some cases the economic lot size can be determined by using methods discussed in the section on inventory management; methods are also available for multiple products. Sequencing becomes an issue of satisfying the needs of the customers, whether they be internal or external, and minimizing the cost of holding inventory.

A common tool used for evaluating various types of schedules is called the **Gantt chart.** This is a simple chart that provides a visual record of the order of processing tasks over time (Fig. 17.1.2). Here, M_1 and M_2 represent two machines in a job shop. The numbers within the blocks show the processing order, or schedule, to complete three jobs, each of which requires processing time on each of the two machines. Note that all the jobs can be done first on machine 1 or on machine 2 but not on both simultaneously.

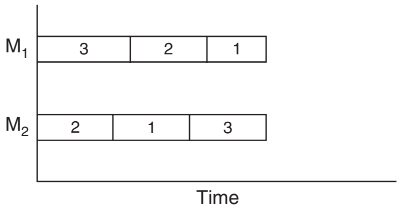


Fig. 17.1.2 Sample Gantt chart.

In job shop scheduling the object is to determine the sequence in which machines will process jobs. The evaluation criteria are varied. Some common criteria include **makespan,** which is the total time to process all jobs; **mean flowtime,** which is the average time a job is in the shop; and **average lateness** per job, which is the average time a job takes longer than the promised time.

If arrivals are static and there is one machine to process jobs, an optimal schedule in terms of minimizing mean flow time is to schedule jobs by processing time, with the shortest jobs being processed first. Note that makespan for this situation is the same for all processing orders.

When there are two machines and n jobs to process on these machines and the order is the same for each job, a method called Johnson's rule can be used to minimize makespan. A variant, Jackson's rule, can be used if jobs can be performed on machine 1 only, machine 2 only, first machine 1 then machine 2, or first machine 2 then machine 1. This rule minimizes the makespan of the sequence.

With three machines and n jobs, Johnson's rule can be used by employing a trick in the solution procedure. If there are m machines and n jobs, Campbell's heuristic can be used to minimize makespan. This is not an optimal procedure since there is not an efficient method available to find an exact solution.

If, on the other hand, jobs arrive intermittently throughout a time period, which is the usual case in a job shop, one must rely on schedules based on priority dispatch rules. In these cases, jobs with the smallest priority are scheduled first. One common rule is the shortest processing time rule. This rule can be dominant in terms of minimizing lateness, makespan, and flowtime. First come, first serve is another common priority dispatch rule. Some others include: random priority, least work remaining, total work fewest operations remaining, due date, slack, and critical ratio, which is the ratio of the time remaining before delivery divided by the remaining processing time. One can evaluate the effectiveness of all of these rules for a particular situation by constructing Gantt charts of each of the resulting schedules and comparing these on the basis of makespan, mean flow time, and average lateness. Implementation involves choosing that one that best fits the needs of the organization. Subsequently, one should monitor the results of the solution over time to see, if, in fact, it is providing the best solution in terms of optimizing the solution criteria.

Scheduling projects is somewhat different from scheduling for assembly lines for a single or multiple product or for job shops. Projects are different in that the firm is producing a single unit (or small number) of the product. Some examples include implementing an MRP or MRP II system in an organization, building a skyscraper, or doing a consulting project for an organization. For many projects the Gantt chart is a useful tool in that it lists tasks to be accomplished, the time it should take to accomplish them, and the sequence and procedure of the tasks.

Researchers have made improvements to this method with the use of PERT and CPM charts to help them schedule and better manage projects. **PERT** is an acronym for program evaluation and review technique and **CPM** is an acronym for critical path method. These methods were initially different in that PERT used probabilistic times for tasks and CPM used deterministic times, but after the more than 50 years in which these methods have been applied there is little to distinguish the two methods.

PERT and CPM rely on a network diagram, which is a graph which contains a number of nodes and arrows. Usually the arrows designate the tasks to be performed and the nodes represent the beginning and end of these tasks. See Fig. 17.1.1. The value of the network diagram is that it clearly displays all of the tasks that must be completed to complete the project, the precedence relationships between these tasks, and, in some graphs, the time to complete each task. Of special interest among the various paths from the beginning to the ending node is the critical path. The **critical path** is the path that is the longest in terms of the time it takes to go from the beginning to the ending node. It is of interest because any delay in accomplishing tasks on the critical path will result in a delay of the completion of the project. To shorten the length of the project, tasks on the critical path will have to be shortened.

One advantage of PERT and CPM are that the graphical display clearly indicates tasks and precedences. Once the critical path has been determined managers can closely monitor tasks on the critical path and, further, monitor tasks not on the critical path that have **slack**, which is the difference between the time it takes to complete a task and the time allowable to complete the task. Another advantage of the use of these techniques is that they help the manager to organize all information and data required to complete the project.

AGGREGATE PLANNING

Aggregate planning is the task of determining how much to produce of a product, what workforce level to use, and what inventory to carry for a period of time. One use of the aggregate plan is as an input into the master production schedule. It is dependent, of course, on the forecast for the product for each of the periods for which the plan is to be made and the amount on inventory on hand along with the costs of the various components.

Factors that can be used to match production to demand include the ability to hire employees and lay off employees; the ability to vary the number of shifts, the length of shifts and the amount of overtime; the ability to subcontract; and the ability to vary the amount of inventory held or backordered.

Aggregate plans are usually evaluated on the basis of objective factors such as cost and meeting demand on time, as well as subjective factors such as lost sales potential. Costs to be considered in most approaches include the basic cost of producing the product (fixed and variable costs); costs associated with hiring, layoffs, and overtime; costs of inventory (holding and backorder costs); and the cost of subcontracting.

There are a number of approaches used to develop aggregate plans in industry, and these range from simple trial-and-error evaluations of plans on a spreadsheet to sophisticated mathematical programming or artificial intelligence models to develop plans that are optimal in terms of particular criteria.

The most simple planning method, trial and error using a spreadsheet, consists of trying various plans and seeing what they yield in terms of meeting production schedules and, in addition, what costs are incurred in the various categories mentioned above. One of the important reasons for using this type of planning method is that a number of subjective considerations can be incorporated into the solution. These include constraints such as, say, maintaining a stable workforce, eliminating shortages or backorders, eliminating subcontracting, etc. Some of these may be hard to model when one is using sophisticated modeling methods but are relatively easy to incorporate into a trial and error method.

Sophisticated methods of aggregate planning include methods such as mathematical programming, of which linear programming is a subset (covered later in this section), and expert systems approaches to the problem. Mathematical programming methods involve deciding upon what criterion to optimize and what constraints are to be placed upon the system. The objective function may be one where the costs of producing the product (which include costs of hiring, layoffs, subcontracting, inventory holding and shortage costs, etc.) are minimized. Constraints are then written that ensure that, say, the amount of inventory available for a particular period is such that demand is satisfied in that period. Note that the inventory can be obtained from inventory on hand, from production in a variety of ways, and from subcontracting. Linear constraints such as these become part of the mathematical model. The mathematical program built in the way above then is optimized using a computer program and the outputs of this program are the optimal plan and the costs of this plan.

One of the advantages of the mathematical programming approach is that cost is minimized. A disadvantage is that some constraints are not easily modeled in some cases. For example, subcontractors may offer price breaks for certain quantities of products or inventory produced. In practice, most firms use the trial-and-error approach.

QUALITY MANAGEMENT

In the past, much of the emphasis on quality management was placed on inspection. An organization produced a product and inspected samples of that product, usually at the end of production. Technology consisted of implementing inspection sampling schemes.

In the early 1980s, many organizations began to modify their methods to adhere to the teachings of proponents such as W. Edwards Deming. Deming emphasized strategic changes in the management of quality. He argued that quality was not a cost to be minimized but rather a constraint that had to be met in order for a firm to remain in business. Some of his more revolutionary contributions were that rework should be eliminated, that high quality is associated with high efficiency, and that an organization should always try to get better. Deming labeled this **continuous improvement.** Deming had a list of 14 points that he called the basis of transformation of American industry. These are (Deming, 1993, p. 23):

1. Create constancy of purpose toward improvement of product and service, with the aim to become competitive and to stay in business, and to provide jobs.
2. Adopt the new philosophy. We are in a new economic age. Western management must awaken to the challenge, must learn its responsibilities, and take on leadership for change.
3. Cease dependence on inspection to achieve quality. Eliminate the need for inspection on a mass basis by building quality into the product in the first place.
4. End the practice of awarding business on the basis of price tag. Instead, minimize total cost. Move toward a single supplier for any one item, on a long-term relationship of loyalty and trust.
5. Improve constantly and forever the system of production and service, to improve quality and productivity, and thus constantly decrease costs.
6. Institute training on the job.
7. Institute leadership. The aim of supervision should be to help people and machines and gadgets to do a better job. Supervision of management is in need of overhaul, as is supervision of production workers.
8. Drive out fear, so that everyone may work effectively for the company.
9. Break down barriers between departments. People in research, design, sales, and production must work as a team, to foresee problems of production and in use that may be encountered with the product or service.
10. Eliminate slogans, exhortations, and targets for the workforce asking for zero defects and new levels of productivity. Such exhortations only create adversarial relationships, as the bulk of the causes of low quality and low productivity belong to the system and thus lie beyond the power of the workforce.
11. a. Eliminate work standards (quotas) on the factory floor. Substitute leadership.
    b. Eliminate management by objective. Eliminate management by numbers and numerical goals. Substitute leadership.
12. a. Remove barriers that rob the hourly worker of the right to pride of workmanship. The responsibility of supervisors must be changed from sheer numbers to quality.
    b. Remove barriers that rob people in management and in engineering of their right to pride of workmanship. This means, *inter alia*, abolishment of the annual or merit rating and of management by objective.
13. Institute a vigorous program of education and self-improvement.
14. Put everybody in the company to work to accomplish the transformation. The transformation is everybody's job.

One of the essential requirements of the Deming idea is that if a problem in the production of a product occurs, it should be fixed before any more of the product is produced. This concept, of course, can be applied to any organization, as all organizations produce either a product or a service. In service organizations, if the service that is provided is poor, one risks losing the customer forever. Moreover, some marketing researchers claim that a dissatisfied customer tells several other customers about the bad experience and this may magnify the cost of a poor service encounter.

STATISTICAL PROCESS CONTROL

The primary set of tools in managing the production of a good or service so that quality is optimized is called **statistical process control** or **SPC**. SPC has as its purpose the identification of the quality of a process, the identification of the factors that influence the quality of a process, the elimination of factors that cause excess variability in the process, the monitoring of a process so that problems are quickly detected and eliminated, and the continuous improvement of a process. Most of the tools in SPC are graphical; these have the added advantage of ease of use, understanding, and communication. Perhaps the most important of these tools is the **control chart,** which is a graph of a quality measurement over time. This quality measurement has been calculated from a sample taken from the process because sampling is more efficient than 100 percent inspection in most cases, when one considers cost. The control chart is used to determine the current performance of a process, to determine the capability of a process, and to monitor a process.

The concept behind the control chart is that it is a tool for separating the variation of the process quality into two parts. Deming calls these two parts **special causes of variability** and **common causes of variability**. Some other common usage is assignable causes (special) and chance or natural (common) variability. Special-cause variability is due to problems with machines, operators, or raw materials and is usually large compared to common-cause variability, which is the variability in the process due to all of the other factors of production. When a process has had all of the special causes of variability removed, it is said to be stable or in a state of statistical control. Further reduction in the remaining common-cause variability is called process improvement but should not be attempted until the process is stable.

Other common tools for SPC include graphs such as histograms, scatterplots, stem and leaf diagrams, Pareto charts, and boxplots. Most textbooks in statistical quality control give in-depth coverage of these topics. Section 17.3, Engineering Statistics and Quality Control, covers some of the tools mentioned here as well as others in more detail.

There are several steps in SPC. The first step is to define quality. In the past product quality was measured by specifications determined by engineers. A more contemporary view of the measurement of quality is to use customer input into the design of the product or service and subsequently to use these specifications of the dimensions of quality as targets.

The second step in SPC is to measure the current quality of a process; this step is called process performance evaluation. One accurate method of measuring quality is to take random samples from the process over time. In SPC, though, one is interested in eliminating special causes of variability after one does a process performance evaluation so the most common form of sampling is by rational subgroups. These are samples chosen so that any differences due to machines, operators, or raw materials will be apparent in the data. While not as accurate in terms of providing an estimate of quality as a random sample, rational subgrouping provides a very accurate measure of quality and, more important, allows an analyst to have some insight into what special causes of variability are present. For example, a process performance evaluation might show that quality depends upon the supplier of raw material, as in a case where the product quality obtained from shifts in which a particular supplier's inputs are used is much worse than the quality when other suppliers' raw materials are used. The value of rational subgrouping is that these types of differences are apparent if the rational subgroups are cleverly chosen.

The process of removing special causes of variability is called *process capability analysis*. In the example above, the analyst might work with the particular supplier to find and eliminate the causes of poor-quality input.

After a process has undergone a process capability analysis and special causes of variability are removed, the process should be monitored. Monitoring involves taking samples from the process at particular times and calculating a statistic that measures quality. Some common statistics are the sample mean, the sample range, the sample proportion defective, and the sample number of defects per unit. These statistics are plotted on a control chart; for these statistics the control charts are respectively the X bar chart, the R chart, the p chart and the c chart. A control chart is used to visually determine if a special cause of poor quality has occurred; if it has, ideally the process is shut down and the special cause removed.

The last step in the SPC is process improvement, which ideally happens continuously. Thus, it has the name *continuous improvement*. In this step, the analyst tries to continuously reduce the common-cause variability through observational studies with rational subgrouping or experimental designs.

In practice there have been many general attempts to push the idea of quality in the workplace. Some advocates such as Deming, Juran, and Crosby came up with platforms for convincing managers that their methods would ensure successful quality efforts within an organization. Other general schemes to manage quality have appeared over the past several decades and include total quality management, ISO certification, the Malcom Baldridge national quality award, and reengineering. Perhaps the most successful is the current platform called six sigma which originated at Motorola under the leadership of Bob Galvin and at GE under Jack Welch.

Six sigma is a method that fully uses the methods discussed above in SPC. It gets its name from the fact that six sigma means there are not more than 3.4 defects per million, an admirable goal. Some of the major ideas behind six sigma include teaching everyone involved the standard statistical tools of SPC such as those discussed above, teaching a structured process such as SPC as a general problem-solving tool, making decisions analytically by using data rather than intuition or experience, and striving for a continuous need to reduce variation in a process. One of the advantages of six sigma is that its use is suggested in all areas of an organization, not just manufacturing.

Six sigma has generated admirable results because it stands on a strong theoretical foundation and is plausible. It is also not beyond the scope of anyone in its complexity. And, empirically, it has had a major financial impact on many firms.

LINEAR PROGRAMMING

At the heart of management's responsibility is the best or optimum use of limited resources including money, personnel, materials, facilities, and time. Linear programming, a mathematical technique, permits determination of the best use which can be made of available resources. It provides a systematic and efficient procedure which can be used as a guide in decision making.

As an example, imagine the simple problem of a small machine shop that manufactures two models, standard and deluxe. Each standard model requires 2 h of grinding and 4 h of polishing. Each deluxe model requires 5 h of grinding and 2 h of polishing. The manufacturer has three grinders and two polishers; therefore, in a 40-h week there are 120 h of grinding capacity and 80 h of polishing capacity. There is a profit of \$3 on each standard model and \$4 on each deluxe model and a ready market for both models. The management must decide on: (1) the allocation of the available production capacity to standard and deluxe models and (2) the number of units of each model in order to maximize profit.

To solve this linear programming problem, the symbol X is assigned to the number of standard models and Y to the number of deluxe models. The profit from making X standard models and Y deluxe models is $3X + 4Y$ dollars. The term *profit* refers to the **profit contribution,** also referred to as **contribution margin** or **marginal income.** The profit contribution per unit is the selling price per unit less the unit variable cost. Total contribution is the per-unit contribution multiplied by the number of units.

The restrictions on machine capacity are expressed in this manner: To manufacture one standard unit requires 2 h of grinding time, so that making X standard models uses $2X$ h. Similarly, the production of Y deluxe models uses $5Y$ h of grinding time. With 120 h of grinding time available, the grinding capacity is written as follows: $2X + 5Y \leq 120$ h of grinding capacity per week. The limitation on polishing capacity is expressed as follows: $4X + 2Y \leq 80$ h per week. In summary, the basic information is:

	Grinding time	Polishing time	Profit contribution
Standard model	2 h	4 h	\$3
Deluxe model	5 h	2 h	4
Plant capacity	120 h	80 h	

Two basic linear programming techniques, the graphic method and the simplex method, are described and illustrated using the above capacity-allocation-profit-contribution maximization data.

Graphic Method

Operations	Hours available	Hours required per model		Maximum number of models	
		Standard	Deluxe	Standard	Deluxe
Grinding	120	2	5	$\frac{120}{2} = 60$	$\frac{120}{5} = 24$
Polishing	80	4	2	$\frac{80}{4} = 20$	$\frac{80}{2} = 40$

The lowest number in each of the two columns at the extreme right measures the impact of the hours limitations. The company can produce 20 standard models with a profit contribution of \$60 (20 × \$3) or 24 deluxe models at a profit contribution of \$96 (24 × \$4). Is there a better solution?

To determine production levels in order to maximize the profit contribution of $\$3X + \$4Y$ when:

$$\begin{array}{ll} 2X + 5Y \le 120 \text{ h} & \text{grinding constraint} \\ 4X + 2Y \le \ \ 80 \text{ h} & \text{polishing constrain} \end{array}$$

a graph (Fig. 17.1.3) is drawn with the constraints shown. The two-dimensional graphic technique is limited to problems having only two variables—in this example, standard and deluxe models. However, more than two constraints can be considered, although this case uses only two, grinding and polishing.

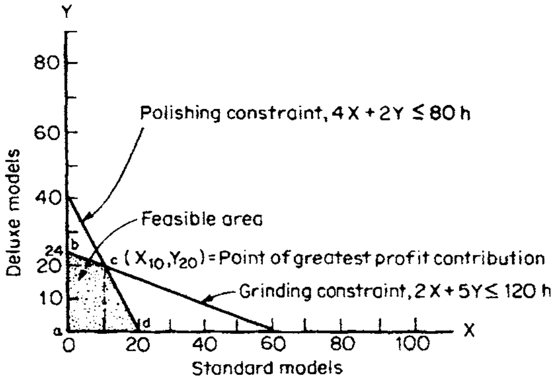


Fig. 17.1.3 Graph depicting feasible solution.

The **constraints** define the solution space when they are sketched on the graph. The solution space, representing the area of feasible solutions, is bounded by the corner points *a, b, c,* and *d* on the graph. Any combination of standard and deluxe units that falls within the solution space is a feasible solution. However, the *best* feasible solution, according to mathematical laws, is in this case found at one of the corner points. Consequently, all corner-point variables must be tried to find the combination which maximizes the profit contribution: $\$3X + \$4Y$.

Trying values at each of the corner points:

$a = (X = 0, Y = 0)$; \$3 (0) + \$4 (0) = \$ 0 profit
$b = (X = 0, Y = 24)$; \$3 (0) + \$4 (24) = \$ 96 profit
$c = (X = 10, Y = 20)$; \$3 (10) + \$4 (20) = \$110 profit
$d = (X = 20, Y = 0)$; \$3 (20) + \$4 (0) = \$ 60 profit

Therefore, in order to maximize profit the plant should schedule 10 standard models and 20 deluxe models.

Simplex Method The simplex method is considered one of the basic techniques from which many linear programming techniques are directly or indirectly derived. The method uses an iterative, stepwise process which approaches an optimum solution in order to reach an objective function of maximization (for profit) or minimization (for cost). The pertinent data are recorded in a tabular form known as the **simplex tableau.** The components of the tableau are as follows (see Table 17.1.1):

The **objective row** of the matrix consists of the coefficients of the objective function, which is the profit contribution per unit of each of the products.

The **variable row** has the names of the variables of the problem including slack variables. **Slack variables** S_1 and S_2 are introduced in order to transform the set of inequalities into a set of equations. The use of slack variables involves simply the addition of an arbitrary variable to one side of the inequality, transforming it into an equality. This arbitrary variable is called **slack variable,** since it takes up the slack in the inequality. The simplex method requires the use of equations, in contrast to the inequalities used by the graphic method.

The **problem rows** contain the coefficients of the equations which represent constraints upon the satisfaction of the objective function. Each constraint equation adds an additional problem row.

The **objective column** receives different entries at each iteration, representing the profit per unit of the variables. In this first tableau (the only one illustrated due to space limitations) zeros are listed because they are the coefficients of the slack variables of the objective function. This column indicates that at the very beginning every S_n has a net worth of zero profit.

The **variable column** receives different notations at each iteration by replacement. These notations are the variables used to find the profit contribution of the particular iteration. In this first matrix a situation of no (zero) production is considered. For this reason, zeros are marked in the objective column and the slacks are recorded in the variable column. As the iterations proceed, by replacements, appropriate values and notations will be entered in these two columns, objective and variable.

The **quantity column** shows the constant values of the constraint equations.

Based on the data used in the graphic method and with a knowledge of the basic components of the simplex tableau, the first matrix can now be set up.

Letting X and Y be respectively the number of items of the standard model and the deluxe model that are to be manufactured, the system of inequalities or the set of constraint equations is

$$\begin{aligned} 2X + 5Y &\le 120 \\ 4X + 2Y &\le \ \ 80 \end{aligned}$$

in which both X and Y must be positive values or zero ($X \ge 0$; $Y \ge 0$) for this problem.

The objective function is $3X + 4Y = P$; these two steps were the same for the graphic method.

The set of inequalities used by the graphic method must next be transformed into a set of equations by the use of slack variables. The inequalities rewritten as equalities are

$$\begin{aligned} 2X + 5Y + S_1 &= 120 \\ 4X + 2Y + S_2 &= \ \ 80 \end{aligned}$$

and the objective function becomes

$$3X + 4Y + 0S_1 + 0S_2 = P \qquad \text{to be maximized}$$

The first tableau with the first solution would then appear as shown in Table 17.1.1.

The tableau carries also the first solution which is shown in the **index row.** The index row carries values computed by the following steps:

1. Multiply the values of the quantity column and those columns to the right of the quantity column by the corresponding value, by rows, of the objective column.
2. Add the results of the products by column of the matrix.
3. Subtract the values in the objective row from the results in step 2. For this operation the objective row is assumed to have a zero value in the quantity column. By convention the profit contribution entered in the cell lying in the quantity column and in the index row is zero, a condition valid only for the first tableau; in the subsequent matrices it will be a positive value.

Index row:

Steps 1 and 2:	Step 3:
120(0) + 80(0) = 0	0 − 0 = 0
2(0) + 4(0) = 0	0 − 3 = −3
5(0) + 2(0) = 0	0 − 4 = −4
1(0) + 0(0) = 0	0 − 0 = 0
0(0) + 1(0) = 0	0 − 0 = 0

Table 17.1.1 First Simplex Tableau and First Solution

		0	3	4	0	0	Objective row
	Mix	Quantity	X	Y	S_1	S_2	Variable row
0	S_1	120	2	5	1	0	Problem rows
0	S_2	80	4	2	0	1	
		0	–3	–4	0	0	Index row
Objective column	Variable column	Quantity column					

In this first tableau the slack variables were introduced into the product mix, variable column, to find *a feasible* solution to the problem. It can be proven mathematically that beginning with slack variables assures a feasible solution. One possible solution might have S_1 take a value of 120 and S_2 a value of 80. This approach satisfies the constraint equation but is undesirable since the resulting profit is zero.

It is a rule of the simplex method that the optimum solution has not been reached if the index row carries any negative values at the completion of an iteration in a maximization problem. Consequently, this first tableau does not carry the optimum solution since negative values appear in its index row. A second tableau or matrix must now be prepared, step by step, according to the rules of the simplex method.

Duality of Linear Programming Problems and the Problem of Shadow Prices Every linear programming problem has associated with it another linear programming problem called its **dual.** This duality relationship states that for every maximization (or minimization) problem in linear programming, there is a unique, similar problem of minimization (or maximization) involving the same data which describe the original problem. The possibility of solving any linear programming problem by starting from two different points of view offers considerable advantage. The two paired problems are defined as the dual problems because both are formed by the same set of data, although differently arranged in their mathematical presentation. Either can be considered to be the **primal;** consequently the other becomes its dual.

Shadow prices are the values assigned to one unit of capacity and represent economic values per unit of scarce resources involved in the restrictions of a linear programming problem. To maximize or minimize the total value of the total output it is necessary to assign a quantity of unit values to each input. These quantities, as cost coefficients in the dual, take the name of "shadow prices," "accounting prices," "fictitious prices," or "imputed prices" (values). They indicate the amount by which total profits would be increased if the producing division could increase its productive capacity by a unit. The shadow prices, expressed by monetary units (dollars) per unit of each element, represent the least cost of any extra unit of the element under consideration, in other words, a kind of marginal cost. The real use of shadow prices (or values) is for management's evaluation of the manufacturing process.

QUEUING THEORY

Queuing theory or **waiting-line** theory problems involve the matching of servers, who provide, to randomly arriving customers, services which take random amounts of time. Typical questions addressed by queuing theory studies are: how long does the average customer wait before being waited on and how many servers are needed to assure that only a given fraction of customers waits longer than a given amount of time.

In the typical problem applicable to queuing theory solution, people (or customers or parts) arrive at a server (or machine) and wait in line (in a queue) until service is rendered. There may be one or more servers. On completion of the service, the person leaves the system. The rate at which people arrive to be serviced is often considered to be a random variable with a Poisson distribution having a parameter λ. The average rate at which services can be provided is also generally a Poisson distribution with a parameter μ. The symbol k is often used to indicate the number of servers.

MONTE CARLO

Monte Carlo simulation can be a helpful method in gaining insight to problems where the system under study is too complex to describe or the model which has been developed to represent the system does not lend itself to an analytical solution by other mathematical techniques. Briefly, the method involves building a mathematical model of the system to be studied which calculates results based on the input variables, or **parameters.** In general the variables are of two kinds: decision parameters, which represent variables which the analyst can choose, and stochastic, or random, variables, which may take on a range of values and which the analyst cannot control.

The random variables are selected from specially prepared probability tables which give the probability that the variable will have a particular value. All the random variables must be independent. That means that the probability distribution of each variable is **independent** of the values chosen for the others. If there is any correlation between the random variables, that correlation will have to be built into the system model.

For example, in a model of a business situation where market share is to be calculated, a decision-type variable representing selling price can be selected by the analyst. A variable for the price of the competitive product can be randomly selected. Another random variable for the rate of change of market share can also be randomly selected. The purpose of the model is to use these variables to calculate a market share suitable to those market conditions. The algebra of the model will take the effects of all the variables into account. Since the rate-of-change variable can take many values which cannot be accurately predicted, as can the competitive price variable, many runs will be made with different randomly selected values for the random variables. Consequently a range of probable answers will be obtained. This is usually in the form of a **histogram.** A histogram is a graph, or table, showing values of the output and the probability that those values will occur. The results, when translated into words, are expressed in the typical Monte Carlo form of: if such a price is chosen, the following probability distribution of market shares is to be expected.

17.2 COST ACCOUNTING

by Scott Jones

REFERENCES: Horngren and Foster, "Cost Accounting—A Managerial Emphasis," Prentice-Hall. Anthony and Govindarajan, "Management Control Systems," Irwin. Cooper and Kaplan, "The Design of Cost Management Systems—Text, Cases, and Readings," Prentice-Hall.

ROLE AND PURPOSE OF COST ACCOUNTING

Cost measurements and reporting procedures are integral components of management information systems, providing financial measurements of economic value and reports useful to many and varied objectives. In a functional organization, where lines of authority are drawn between engineering, production, marketing, finance, and so on, cost accounting information is primarily used by managers for control in guiding departmental units toward the attainment of specific organizational goals. In team-oriented organizations, authority is distributed to multifunctional teams empowered as a group to make decisions. The focus of cost accounting in this organizational structure is not so much on control, but on supporting decisions through the collection of relevant information for decision making. Cost accounting also provides insight for the attainment of competitive advantage by providing an information set and analytical framework useful for the analysis of product or process design, or service delivery alternatives.

The basic purpose of cost measurements and reporting procedures can be organized into a few fundamental areas. These are: (1) identifying and measuring the economic value to be placed on goods and services for reporting periodic results to external information users (creditors, stockholders, and regulators); (2) providing control frameworks for the implementation of specific organizational objectives; (3) supporting operational and strategic decision making aimed at achieving and sustaining competitive advantage.

MEASURING AND REPORTING COSTS TO STOCKHOLDERS

Accounting in general and cost accounting in particular are most visible to the general public in the role of **external reporting.** In this role, cost accounting is geared toward measuring and reporting periodic results, typically annual, to users outside of the company such as stockholders, creditors, and regulators. The annual report presents to these users management forecasts for the coming period and results of past years operations as reflected in **general-purpose financial statements.** (See also Sec. 17.1, "Operations Management.") Performance is usually captured through the presentation of three reports: the income statement, the balance sheet, and the statement of cash flows (see Fig. 17.2.1). These statements are audited by independent certified public accountants who attest that the results reported present a fair picture of the financial position of the company and that prescribed rules have been followed in preparing the reports. The accounting principles and procedures that guide the preparation of those reports are governed by **generally accepted accounting principles** (GAAP).

The underlying principles that guide GAAP financial statements encourage comparability among companies and across time. Therefore, these rules are usually too constraining for reports destined for internal use. External reports such as the balance sheet and income statement are prepared on the **accrual basis.** Under this basis, revenues are recognized (reported to stockholders) when earned, likewise costs are expensed against (matched with) revenue when incurred. This is to be contrasted with the **cash basis,** which recognizes revenues and expenses when cash is collected or paid. For example, if a drill press is purchased for use in the factory, a cash outflow occurs when the item is paid for. The cash basis would recognize the purchase price of the drill press as an expense when the payment is made. On the other hand, the accrual basis of accounting would capitalize the price paid, and this amount would be the cost (or basis) of the asset called *drill press.* In accrual accounting, an asset is something having future economic benefit, and therefore the cost of this asset must be distributed among the periods of time when it is used to generate revenue. The cost of the drill press would be expensed periodically by deducting a small amount of that cost from revenue as the drill press is used over its economic life, which may be several years. This periodic charge is called *depreciation.* To capture the effects that revenue-generating activities have on cash, GAAP financial statements also include the statement of cash flows. The statement of cash flows is not prepared on an accrual basis; rather, it reflects the amount of cash flowing into a company during a period, as well as the cash outflow. The first section of that statement, "Cash flows from operating activities," is essentially an income statement prepared on the cash basis.

Another application of cost accounting measurements for external users involves the preparation of reports such as income tax returns for governmental agencies. Federal, state, and local tax authorities prescribe specific accounting procedures to be applied in determining taxable income. These rules are conceptually similar to general-purpose financial reporting but differ mainly in technical aspects of the computations, which are modified to support whatever public finance goals may exist for a particular period. Whereas GAAP financial statements allow for the analysis of credit and investment opportunities, Internal Revenue Service regulations are designed to raise revenue, stimulate the economy, or both. Regulations may be primarily aimed at reducing the federal deficit and hence assign rather long "useful lives" to depreciable assets; at other times in history useful lives were shortened to stimulate investment and economic growth. For GAAP, management usually estimates the useful life of an asset for purposes of depreciation. For IRS purposes, the amount of depreciation is based on the **class life** of the asset, and the depreciation system elected by the taxpayer. Class life is based on whether the asset is specialized or general purpose, and in what industry the asset is employed.

CLASSIFICATIONS OF COSTS

The purposes of cost accounting require classifications of costs so that they are recognized (1) by the nature of the item (a natural classification), (2) in their relation to the product, (3) with respect to the accounting period to which they apply, (4) in their tendency to vary with volume or activity, (5) in their relation to departments, (6) for control and analysis, and (7) for planning and decision making.

Direct material and direct labor may be listed among the items which have a **variable** nature. Factory overhead, however, must be carefully examined with regard to items of a variable and a fixed nature. It is impossible to budget and control factory-overhead items successfully without regard to their tendency to be fixed or variable; the division is a necessary prerequisite to successful budgeting and intelligent cost planning and analysis.

In general, **variable expenses** show the following characteristics: (1) variability of total amount in direct proportion to volume, (2) comparatively constant cost per unit or product in the face of changing volume, (3) easy and reasonably accurate assignments to operating departments, and (4) incurrence controllable by the responsible department head.

The characteristics of **fixed expenses** are (1) fixed amount within a relative output range, (2) decrease of fixed cost per unit with increased output, (3) assignment to departments often made by managerial

Balance Sheet (Illustrative)

Black Carbon, Inc.

December 31, 20-

Assets

Current Assets			
Cash		$5,050,000	
Accounts Receivable (net)		6,990,000	
Inventories:			
Raw Materials and Supplies	$1,000,000		
Work in Process	1,800,000		
Finished Goods	2,900,000	5,700,000	
Investments		1,000,000	
Deferred Charges		340,000	
Total Current Assets			$19,080,000
Property, Plant and Equipment:			
Land		$4,000,000	
Buildings and Equipment	$75,500,000		
Less: Allowance for Depreciation	47,300,000	28,200,000	
Total Fixed Assets			32,200,000
Total Assets			$51,280,000

Liabilities

Current Liabilities:			
Accounts Payable and Accruals		$3,580,000	
Provision for Income Taxes:			
Federal	$2,250,000		
State	65,000	2,315,000	
Total Current Liabilities			$5,895,000
Long-term Debt			5,300,000
Total Liabilities			$11,195,000
Stockholders Equity:			
Common Stock—no par value			
Authorized—2,000,000 shares			
Outstanding—1,190,000 shares		$11,900,000	
Earnings retained in the business		28,185,000	
Total Stockholders' Equity			$40,085,000
Total Liabilities and Stockholders' Equity			$51,280,000

Income Statement (Illustrative)

Black Carbon Inc.

for the year 20-

Net Sales		$50,087,000
Cost of products:		
Material, Labor, and Overhead (excluding depreciation)	$32,150,000	
Depreciation	5,420,000	37,570,000
Gross Profit		$12,517,000
Less: Selling and Administrative Expenses		3,220,000
Profit from Operations		9,297,000
Other Deductions		305,000
		8,992,000
Other Income		219,000
Income before Federal and State Income Taxes		9,211,000
Less: Provision for Federal and State Income Taxes		4,055,000
Total Net Income		5,156,000
Dividends paid to shareholders		2,200,000
Income retained in the business		$ 2,956,000

Statement of Cash Flows (Illustrative)

Black Carbon Inc.

December 31, 20-

Cash flows from operating activities		
Cash received from customers	$49,525,000	
Cash paid to suppliers and employees	(40,890,000)	
Cash paid for interest	(1,050,000)	
Cash paid for income taxes	(2,860,000)	
Net cash provided by operating activities		$ 4,725,000
Cash flows from investing activities		
Purchase of equipment	(1,970,000)	
Net cash from investing activities		(1,970,000)
Cash flows from financing activities		
Principal payment on loans	(2,780,000)	
Dividend payments	(2,200,000)	
Net cash used by financing activities		(4,980,000)
Net increase (decrease) in cash		(2,225,000)
Cash at beginning of year		7,275,000
Cash at end of year		$5,050,000

Fig. 17.2.1 Examples of the balance sheet, income statement, and statement of cash flows based on the published annual report.

decisions or cost-allocation methods, and (4) control for incurrence resting with top management rather than departmental supervisors. Whether an expense is classified as fixed or variable may well be the result of managerial decisions.

Some **factory overhead** items are semivariable in nature; i.e., they vary with production but not in direct proportion to the volume. For practical purposes, it is desirable to resolve each semivariable expense item into its variable and fixed components.

A factory is generally organized along departmental lines for production purposes. This factory departmentalization is the basis for the important classification and subsequent accumulation of costs by departments to achieve (1) cost control and (2) accurate costing. The departments of a company generally fall into two categories: (1) producing, or productive, departments, and (2) nonproducing, or service, departments. A producing department is one in which manual and machine operations are performed directly upon any part of the product manufactured. A service department is one that is not directly engaged in production but renders a particular type of service for the benefit of other departments. The expense incurred in the operation of service departments represents a part of the total factory overhead that must be absorbed in the cost of the product.

For **product costing,** the factory may be divided into departments, and departments may also be subdivided into cost centers. As a product passes through a cost center or department, it is charged with a share of the indirect expenses on the basis of a departmental factory-overhead rate. For cost-control purposes, budgets are established for departments and cost centers. Actual expenses are compared with budget allowances in order to determine the efficiency of a department and to measure the manager's success in controlling expenses.

Factory overhead, which is charged to a product or a job on the basis of a predetermined overhead rate, is considered indirect with regard to the product or the job to which the expense is charged. Service-department expenses are prorated to other service departments and/or to the producing departments. The proration is accomplished by using some rational basis such as area occupied or number of workers. The prorated costs are termed **indirect departmental charges.** When all service-department expenses have been prorated to the producing departments, each producing department's total factory overhead will consist of its own direct departmental expense and the indirect (or prorated, or apportioned) charges. This total cost is charged to the product or the job on the basis of the predetermined factory-overhead rate.

A company's cost system provides the data required for establishing **standard costs** and for the preparation and operation of a **budget.**

The **budget** program enlists all members of management in the task of creating a workable and acceptable plan of action, welds the plan into a homogeneous unit, communicates to the managerial levels differences between planned activity and actual performance, and points out unfavorable conditions which need corrective action. The budget not only will help promote coordination of people, clarification of policy, and crystallization of plans, but with successful use will create greater internal harmony and unanimity of purpose among managers and workers.

The established **standard-cost** values for material, labor, and factory overhead form the foundation for the budget. Since standard costs are an invaluable aid in the process of setting prices, it is essential to set these standard costs at realistic levels. The measurement of deviations from established standards or norms is accomplished through the use of variance accounts.

Costs as a basis for planning are estimated costs which may be incurred if any one of several alternative courses of action is adopted. Different types of costs involve varying kinds of consideration in managerial planning and decision making.

METHODS OF ACCUMULATING COSTS IN RECORDS OF ACCOUNT

The balance sheet lists the components of inventory as raw materials, work in process, and finished goods. These accounts reflect the cost of unsold production at various stages of completion. The costs in work in process and finished goods are accumulated or tabulated in the record of accounts according to one of two methods:

1. The Job-Order Cost Method When orders are placed in the factory for specific jobs or lots of product, which can be identified through all manufacturing processes, a job cost system is appropriate. This method has certain characteristics. A manufacturing order often corresponds to a customer's order, though sometimes a manufacturing order may be for stock. The customer's order may be obtained on the basis of a bid price computed from an estimated cost for the job. The goods in each order are kept physically separate from those of other jobs. The costs of a manufacturing order are entered on a job cost sheet which shows the total cost of the job upon completion of the order. This cost is compared with the estimated cost and with the price which the customer agreed to pay.

2. The Process Cost Method When production proceeds in a continuous flow, when units of product are not separately identifiable, and when there are no specific jobs or lots of product, a process cost system is appropriate, for it has certain characteristics: work is ordered through the plant for a specific time period until the raw materials on hand have all been processed or until a specified quantity has been produced; goods are sold from the stock of finished goods on hand since a customer's order is not separately processed in the factory; the cost-of-production sheet is a record of the costs incurred in operating the process—or a series of processes—for a period of time. It shows the quantity produced in pounds, tons, gallons, or other units, and the cost per unit is obtained by dividing the total costs of the period by the total units produced. Performance is indicated by comparing the quantity produced and the cost per unit of the current period with similar figures of other periods or with standard cost figures.

ELEMENTS OF COSTS

The main items of costs shown on the income statement are factory costs which include direct materials, direct labor and factory overhead; and selling and administrative expenses. A breakdown of costs is shown in Figure 17.2.2.

Materials The cost of materials purchased is recorded from purchase invoices. When the materials are used in the factory, an assumption must be made as to **cost flow,** that is, whether to charge them to operations at average prices, at costs based on the first-in, first-out method of costing, or at costs based on the last-in, first-out method of costing. Each method will lead to a different cost figure, depending on how prices change. Each situation must be studied individually to determine which practice will give a maximum of accuracy in cost figures with a minimum of accounting and clerical effort. Once the choice has been made, records must be set up to charge materials to operations based on requisitions. Indirect material is necessary to the completion of the product, but its consumption with regard to the final product is either so small or so complex that it would be futile to treat it as a direct-material item.

Labor Labor also consists of two categories: direct and indirect. Direct labor, also called **productive labor,** is expended immediately on the materials comprising the finished product. Indirect labor, in contrast to direct labor, cannot be traced specifically to the construction

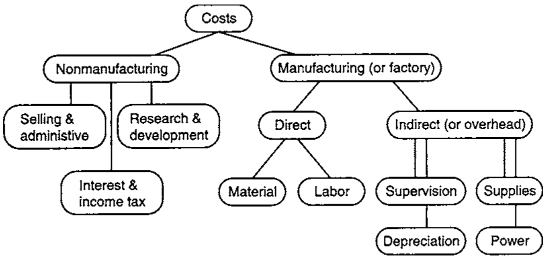


Figure 17.2.2 Summary diagram of cost relationships.

or composition of the finished product. The term includes the labor of supervisors, shop clerks, general helpers, cleaners, and those employees engaged in maintenance work.

Factory Overhead Indirect materials or factory supplies and indirect labor constitute an important segment of factory overhead. In addition, costs of fuel, power, small tools, depreciation, taxes on real estate, patent amortization, rent, inspection, supervision, social security taxes, health and accident insurance, workers' compensation insurance, and many others fall into this large category. These expenses must be collected and allocated to all jobs or units produced. Many expenses are definitely applicable to a specific department and are easily assigned thereto. Other expenses relate to the entire plant and must be prorated to departments on some suitable basis. For instance, heat might be prorated to departments on the basis of volume of space occupied. The expenses of the service departments are prorated to the producing departments on some basis such as service rendered in the case of a maintenance department or per dollar of payroll processed in the case of a cost department.

The charging of factory overhead to jobs or products is accomplished by means of an overhead or burden rate. This rate is essentially a ratio computed to show the relationship of the total burden of a department to some other easily measurable total figure for the department. For example, the total burden cost of a department may be divided by its direct-labor cost to give a percentage-of-direct-labor rate. This percentage applied to the direct-labor cost of a job or a product gives the amount of overhead chargeable thereto. Other common types of burden rates are the labor-hour rate (departmental expenses ÷ total direct-labor hours) and the machine-hour rate (departmental expenses ÷ total machine hours available). Labor rates are most commonly used. When, however, machines perform the greater amount of the work, machine-hour rates give better results. It must be clearly understood that these rates are computed in advance of production, generally at the beginning of the year. They are used throughout the fiscal period unless seasonal fluctuations or unusual changes in expense amounts necessitate the creation of a new rate. The determination of the overhead rate is closely tied up with overhead budgets.

Departmental Classification As mentioned above, the establishment of departmental lines is important not only for costing purposes but also for budgetary control purposes. Departmental lines are set up in order to (1) segregate basically different processes of production, (2) secure the smoothest possible flow of production, and (3) establish lines of responsibility for control over production and costs. When the costing methods are designed to fit in with the departmentalization of factory and office, costs can be accumulated within a department with production being on either the job-order or process cost method.

ACTIVITY-BASED COSTING

The method of assigning overhead to products based on labor hours or machine hours is referred to as **volume-based** overhead absorption because the amount assigned will vary strictly with the volume of either labor or machine time consumed. In applications where production costs may not be strictly driven by volume of labor hours, **activity-based** or **transaction** costing is appropriate. This situation typically occurs when there are many options or alternatives available to the customer. Typically, these products are produced in low volumes and have a high degree of complexity. An example would be the option of a premium radio in an automobile. Though the actual purchase cost of that radio would not be overlooked in pricing the automobile, the indirect costs would be overlooked in a volume-based system because the indirect costs associated with the premium radio, such as holding an additional item of inventory, documenting and producing separate receiving orders, added clerical and assembly coordination effort, increased engineering complexity, and so on would be grouped under the general category of overhead. On the other hand, in an activity-based system, indirect costs would be assigned to a unique cost pool, such as shown in Fig. 17.2.3 and 17.2.4, which compare the procedures of volume and activity-based costing. The striking difference is that the overhead cost pool used in volume systems is not present. In activity-based systems, many more cost pools are used and are closely related to some **causal aspect** of the process such as machine setup, receiving orders, or material movement. The costs are assigned to products based on the relative amounts of each cost driver consumed by that product. Therefore, low-volume options such as the premium radio receive a larger share of indirect costs relative to the actual volume of labor used to insert the component. The amount of each cost pool attributed to a product can then be spread over the number of units produced and a more accurate assignment of costs obtained.

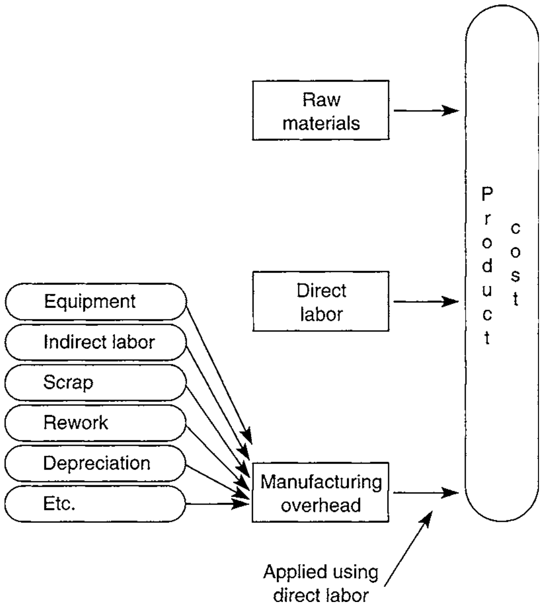


Fig. 17.2.3 Volume-based costing.

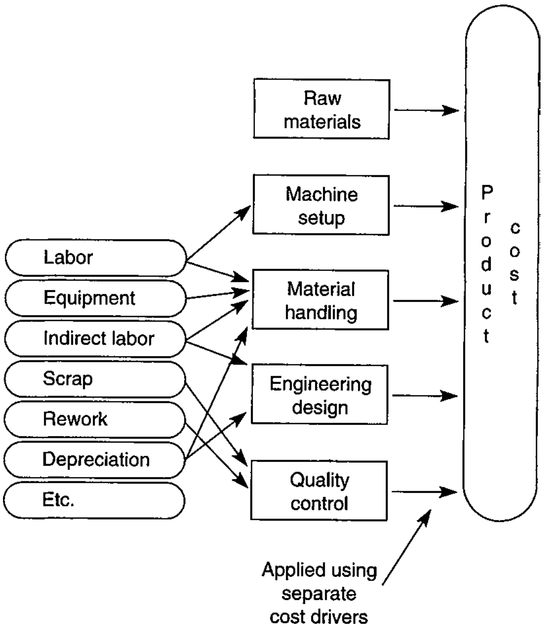


Fig. 17.2.4 Activity-based costing.

MANAGEMENT AND THE CONTROL FUNCTION

To be successful, management must integrate its own knowledge, skills, and practices with the know-how and experience of those who are entrusted with the task of carrying out company objectives. Management, together with its employees and workers, can achieve its objectives through performance of the three managerial functions: (1) planning and setting objectives, (2) organizing, and (3) controlling.

Planning is a basic function of the management process. Without planning there is no need to organize or control. However, planning must precede doing, and the budget is the most important planning tool of an enterprise.

Organizing is essentially the establishment of the framework within which the required activities are to be performed, together with a list of who should perform them. Creation of an organization requires the establishment of organizational or functional units generally known as departments, divisions, sections, floors, branches, etc.

Controlling is the process or procedure by which management ensures operative performance which corresponds with plans. The control process is pictured diagrammatically in Fig. 17.2.5. Recognition of accounting as an important tool in the controlling phase is evidenced through the role of performance reports in pointing out areas and jobs or tasks which require corrective action. These reports should make possible "management by exception."

The effectiveness of the control of costs depends upon proper communication through control and action reports from the accountant to the various levels of operating management. An organization chart is essential to the development of a cost system and cost reports which parallel the responsibilities of individuals for implementing management plans. The coordinated development of a company's organization with the cost and budgetary system will lead to "responsibility accounting." Responsibility accounting plays a key role in determining the type of cost system used.

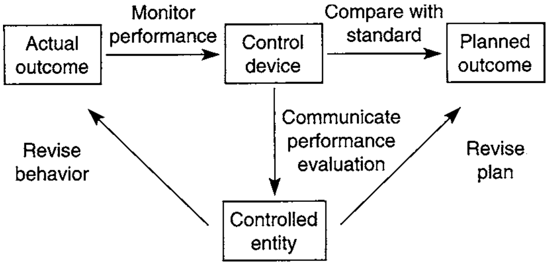


Fig. 17.2.5 The control process.

TYPES OF COST SYSTEMS

The construction of a cost system requires a thorough understanding of (1) the organizational structure of the company, (2) the manufacturing procedure, and (3) the type of information which management requires of the cost system.

1. The organization chart gives a graphic picture of the ranking authority of superintendents, department heads, and managers who are responsible for (*a*) providing the detailed information needed by the accounting division in order to install a successful system; (*b*) incurring expenditures in personnel, materials, and other cost elements, which the cost accountant must segregate and report to those in charge. The cost system with its operating accounts must correspond to organizational divisions of authority so that the individual supervisor, department head, or executive can be held "accountable" for the costs incurred in the department.

2. The manufacturing procedure and shop methods lead to a consideration of the type of pay (piece rate, incentive, day rate, etc.); the method of collecting hours worked; the control of inventories; the problem of costing tools, dies, jigs, and machinery; and many other problems connected with the factory.

3. The organizational setup on the one hand and the manufacturing procedure on the other form the background for the design of a cost system that is based on (*a*) recognition of the various cost elements, (*b*) departmentalization of factory and office, and (*c*) the chart of accounts.

Any cost system should be perfected so that it will (1) aid in the control and management of the company; (2) measure the efficiency of personnel, materials, and machines; (3) help in eliminating waste; (4) provide comparison within individual industries; (5) provide a means of valuing inventories; and (6) aid in establishing selling prices. In an organization departmentalized or segmented along product lines, it is often arbitrary to allocate certain indirect costs especially when common facilities or personnel are shared. This is because there is no objective basis to compute a division of common costs. Control methods for evaluating performance often rely on the **segment margin** statement.

Example of a Segment Margin Statement

	Product A	Product B	Total
Sales	$9,000	$11,000	$20,000
Direct material & labor	(4,000)	(5,000)	(9,000)
Contribution margin	5,000	6,000	11,000
Product specific overhead	(1,000)	(2,500)	(3,500)
Segment margin	4,000	3,500	7,500
Common costs			(2,000)
Operating income			$5,500

The cost system's value is greatly enhanced when it is interlocked with a **budgetary control system.** When budget figures are based upon standard costs, the greatest benefit will be derived from such a combination.

Basically, two types of cost systems exist: (1) the **actual** (or **historical**) and (2) the **standard** (or **predetermined**). The actual cost system accumulates and summarizes costs as they occur and determines a final product cost after all manufacturing operations have been completed. The job is charged with actual quantities and costs of materials used and labor expended; the overhead or burden is allocated on the basis of some predetermined overhead rate. This predetermined overhead rate shows that even the so-called **actual system** does not entirely live up to its name. Under a standard cost system all costs are predetermined in advance of production. Both the actual (historical) and the standard cost system may be used in connection with either (1) the job-order cost method or (2) the process cost method.

BUDGETS AND STANDARD COSTS

A budget provides management with the information necessary to attain the following major objectives of budgetary control: (1) an organized procedure for planning; (2) a means for coordinating the activities of the various divisions of a business; (3) a basis for cost control. The planning phase provides the means for formalizing and coordinating the plans of the many individuals whose decisions influence the conduct of a business. Sales, production, and expense budgets must be established. Their establishment leads necessarily to the second phase of coordination. Production must be planned in relation to expected sales, materials and labor must be acquired or hired in line with expected production requirements, facilities must be expanded only as foreseeable future needs justify, and finances must be planned in relation to volume of sales and production. The third phase of cost control is predicated on the idea that actual costs will be compared with budgeted costs, thus relating what actually happened with what should have happened. To accomplish this purpose, a good measure of what costs should be under any given set of conditions must be provided. The most important condition affecting costs is volume or rate of activity. By predetermining, through the use of the flexible budget, the expenses allowed for any given rate of activity and comparing it with the actual expense, a better measurement of the performance of an individual department is achieved and the control of costs is more readily accomplished.

In the construction of overhead budgets the volume or activity of the entire organization as well as of the individual department is of considerable importance in their relationship to existing capacity. Capacity must be looked upon as that fixed amount of plant, machinery, and personnel to which management has committed itself and with which it expects to conduct the business. Volume or activity is the variable factor in business related to capacity by the fact that volume attempts to make the best use of the existing capacity. To find a profitable solution to this relationship is one of the most difficult problems faced by business management and the accountant who tries to help with appropriate cost data. Volume, particularly of a department, is often expressed in terms of direct-labor hours. With different rates of capacity, a different cost per hour of labor will be computed. This relationship can be demonstrated in the following manner:

Percentage of productive capacity	60%	80%	100%
Direct-labor hours	600	800	1,000
Factory overhead			
Fixed overhead	$1,200	$1,200	$1,200
Variable overhead	600	800	1,000
Total	$1,800	$2,000	$2,200
Overhead rate per direct-labor hour	$3.00	$2.50	$2.20

The existence of fixed overhead causes a higher rate at lower capacity utilization. It is desirable to select that overhead rate which permits a full recovery of production costs by the end of the business cycle. The above tabulation reveals another important axiom with respect to fixed and variable overhead. Fixed overhead remains constant in total but varies in respect to cost per unit or hour. Variable overhead varies in total but remains fixed in relationship to the unit or hour.

Standard Costs The budget, as a statement of expected costs, acts as a guidepost which keeps business on a charted course. Standards, however, do not tell what the costs are expected to be but rather what they will be if certain performances are attained. In a well-managed business, costs never exceed the budget. They should constantly approach predetermined standards. The uses of standard costs are of prime importance for (1) controlling and reducing costs, (2) promoting and measuring efficiencies, (3) simplifying the costing procedures, (4) evaluating inventories, (5) calculating and setting selling prices. The success of a standard cost system depends upon the reliability and accuracy of the standards. To be effective, standards should be established for a definite period of time so that control can be exercised and variances from standards computed. Standards are set for materials, labor, and factory overhead. When actual costs differ from standard costs with respect to material and labor, two causes can generally be detected. (1) The price may be higher or lower or the rate paid a worker may be different; the difference is called a material price or a labor rate variance. (2) The quantity of the material used may be more or less than the standard quantity or the hours used by the worker may be more or less. The difference is called material-quantity variance or labor-efficiency variance, respectively. For factory overhead, the computation is somewhat more elaborate. Actual expenses are compared not only with standard expenses but also with budget figures. Various methods are in vogue, resulting in different kinds of overhead variances. Most accountants compute a controllable and a volume variance. The controllable variance deals chiefly with variable expenses and measures the efficiency of the manager's ability to hold costs within the budget allowance. The volume variance portrays fixed overhead with respect to the use or nonuse of existing capacity. It measures the success of management in its ability to fill capacity with sales or production volume. These two variances can be analyzed further into an expenditure and efficiency variance for the controllable variances and into an effectiveness and capacity variance for the volume variance. Such detailed analyses might bring forth additional information which would help management in making decisions. Of absolute importance for any cost system is the fact that the information must reach management promptly, with regularity, and in a report that is analytical, permitting quick comparison with targets and goals. Only in this manner can management, which includes all echelons from the foreman to the president, exercise control over costs and therewith over profits.

TRANSFER PRICING

A **transfer price** refers to the selling price of a good or service when both the buyer and seller are within the same organization. For example, one division of a company may produce a component, such as an engine, and transfer this component to an assembly division. For purposes of control, these organizational units may be treated as **profit centers** (responsible for earning a specified profit or return on investment). Accordingly, the transfer price is a revenue for the seller and a cost to the buyer. Because organizational control is at issue whenever interdivisional transfers are made, companies must often specify a policy to dictate the basis for determining a transfer price. Transfer prices should be based on market prices when available. Most taxing authorities require intercompany transfers to be made at market price as well. To solve situations of suboptimal resource usage (e.g., idle capacity) it is often possible to construct transfer prices based on manufacturing cost plus some allowance for profit. If the producing division is a **cost center** (responsible for controlling costs to achieve a certain budgeted level), in order to promote efficiency transfers are usually made on the basis of **standard cost.**

SUPPORTING DECISION MAKING

The analytical phase of cost accounting has become more important and influential in the last few years. Management must make many decisions, some of a short-range, others of a long-range nature. To base judgment upon good, reliable data and analyses is a major task for controllers and their staffs. Cost analysis comprises such matters as analysis of distribution costs, gross-profit analysis, break-even analysis, profit-volume analysis, differential-cost analysis, direct costing, capital-expenditure analysis, return on capital employed, and price analysis. A detailed discussion of each phase mentioned lies beyond the scope of this section, but a short description is appropriate.

Distribution-cost analysis deals with allocation of selling expenses to territories, customers, channels of distribution, products, and sales representatives. Once so allocated it might be possible to determine the most profitable and the least profitable commodity, product, territory, or customer. Segment margin statements are useful for this analysis. Standards have been introduced recently in these analyses. The Robinson-Patman Act, an amendment to Section 2 of the Clayton Act, gave additional impetus to the analytical phase of distribution costs. This act prohibits pricing the same product at different amounts when the amounts do not reflect actual cost differences (such as distribution or warranty).

Gross-profit analysis attempts to determine the causes for an increase or decrease in the gross profit. Any change in the gross profit is due to one or a combination of the following: (1) changes in the selling price of the products; (2) changes in the volume sold; (3) changes in the types of products sold, called the sales-mix; (4) changes in the cost elements. Cost elements are analyzed through budgetary control methods. Sales figures must be scrutinized to unearth the changes from the contemplated course and therewith from the final profit.

Break-even analysis, generally presented in the form of a break-even chart, constitutes one of the briefest and most easily understood devices for data presentation for policy-making decisions. The name "break-even" implies that point at which the company neither makes a profit nor suffers a loss from the operations of the business. A break-even chart can be defined as a portrayal in graphic form of the relation of production and sales to profit or, more briefly, a graphic variable income statement. The computation of the break-even point can be made by the following formula.

$$\text{Break-even sales volume} = \frac{\text{total fixed expenses}}{1 - \dfrac{\text{total variable expenses}}{\text{total sales volume}}}$$

EXAMPLE. Assume fixed expenses, $13,800,000; variable expenses, $27,000,000; total sales volume, $50,000,000. Computation: Break-even sales volume = $13,800,000/[1 − ($27,000,000/$50,000,000)] = $30,000,000.

Results can be obtained in chart form (Fig. 17.2.6).

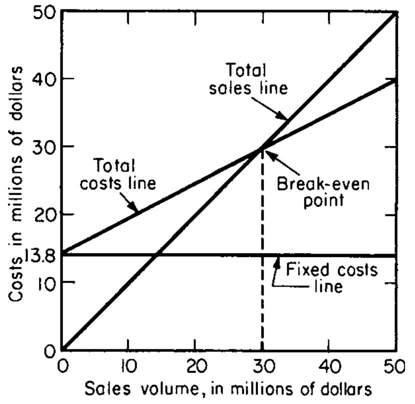


Fig. 17.2.6 Illustrative break-even chart.

Cost-volume-profit analysis deals with the effect that a change of volume, cost, price, and product-mix will have on profits. Managements of many enterprises attempt to stimulate the public to purchase their products by conducting intensive promotion campaigns in radio, press, mail, and television. The customer, however, makes the final decision. What management wants to know is which product or model will yield the most profitable margin; which is the least profitable; what effect a reduction in sales price will have on final profit; what effect a shift in volume or product-mix will have on product costs and profits; what the new break-even point will be under such changing conditions; what the effect of expected increases in wages or other operating costs on profit will be; what the effect will be on costs, profit, and sales volume should there be an expansion of the plant. Cost-volume-profit analysis can also be presented graphically in a so-called *volume-profit-analysis graph.* Using the same data as in the break-even chart, a volume-profit analysis graph takes the form shown in Fig. 17.2.7.

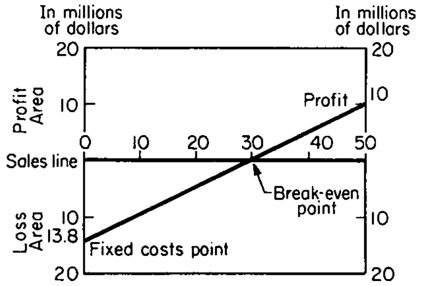


Fig. 17.2.7 Illustrative cost-volume-profit analysis graph.

Differential-cost analysis treats differences, as the title suggests. These differences, also called alternative courses, arise when management wants to know whether or not to take business at a special price, to risk a decline in price of total sales, to sacrifice volume for price, to shut down part of the plant, or to enlarge plant capacity. While accountants generally use the term **"differential,"** economists speak of **"marginal"** and engineers of **"incremental"** costs in connection with such a study. As in any of the previously discussed analyses, the classification of costs into their fixed and variable components is absolutely essential. However, while in break-even analysis the emphasis rests upon the fixed expenses, differential-cost studies stress the variable costs. The differential-cost statement presents only the differences in the following manner:

	Present business	Additional business	Total
Sales	$100,000	$10,000	$110,000
Variable costs	60,000	6,000	66,000
Marginal income	40,000	4,000	44,000
Fixed expenses	30,000	none	30,000
Profit	$ 10,000	$ 4,000	$ 14,000

This statement shows that additional business is charged with the variable expenses only because present business is absorbing all fixed expenses.

Direct costing is a costing method which charges the products with only those costs that vary directly with volume. Variable or direct costs such as direct materials, direct labor, and variable manufacturing expenses are examples of costs chargeable to the product. Costs that are a function of time rather than of production are excluded from the cost of the product. The only costs assignable to inventories are variable costs, and because they should vary in proportion to increases or decreases in production, the unit cost assigned to inventories should be uniform.

CAPITAL-EXPENDITURE DECISIONS

The preparation of a capital-expenditure budget must be preceded by an analytical and decision-making process by management. This area of managerial decisions not only is important to the success of the company but also is crucial in case of errors. Financial requirements, present and anticipated costs, profits, tax considerations, and legal, personnel, and market problems must be studied and reviewed before making the final decision.

Five evaluation techniques are generally accepted as representative tools for decision making: (1) **payback-** or **payout-period method;** (2) **average-return-on-investment method;** (3) **present-value method;** (4) **discounted-cash-flow method;** and (5) **economic value added (EVA).** None of these methods serves every purpose or every firm. The methods should, however, aid management in exercising judgment and making decisions. Of significance in the evaluation of a capital expenditure is the time value of money which is employed in the present value and the discounted-cash-flow methods. The **present value** means that a dollar received a year hence is not the equivalent of a dollar received today, because the use of money has a value. For this reason, the estimated results of an investment proposal can be stated as a cash equivalent at the present time, i.e., its present value. Present-value tables have been devised to facilitate application of present-value theory.

In the present-value method the discount rate is known or at least predetermined. In the **discounted-cash-flow** (DCF) method the rate is to be calculated and is defined as the rate of discount at which the sum of positive present values equals the sum of negative present values. The DCF method permits management to amortize corporate profits by selecting proposals with the highest rates of return as long as the rates are higher than the company's own **cost of capital** plus management's allowance for risk and uncertainty. Cost of capital represents the expected return for a given level of risk that investors demand for investing their money in a given firm or venture. However, when related to capital-expenditure planning, the cost of capital refers to a specific cost of capital from a particular financing effort to provide funds for a specific project or numerous projects. Such use of the concept connotes the marginal cost of capital point of view and implies linkage of the financing and investment decisions. It is, therefore, not surprising that the cost of capital differs, depending upon the sources. A company could obtain funds from (1) bonds, (2) preferred and common stock, (3) use of retained earnings, and (4) loans from banks. If a company obtains funds by some combination of these sources to achieve or maintain a particular capital structure, then the cost of capital (money) is the weighted average cost of each money source.

Return-on-Capital Concept This aids management in making decisions with respect to proposed capital expenditures. This concept can also be used for (1) measuring operating performance, (2) profit planning and decision making, and (3) product pricing. The return on capital may be expressed as the product of two factors: the percentage of profit to sales and the rate of capital turnover. In the form of an equation, the method appears as

$$\begin{array}{lcc} \dfrac{\text{Profit}}{\text{Sales}} = \text{profit margin} & \downarrow & \\ & \times \rightarrow = & \begin{array}{c}\text{return on}\\ \text{capital}\\ \text{investment}\end{array} \\ \dfrac{\text{Sales}}{\text{Capital}} = \begin{array}{c}\text{Investment}\\ \text{turnover}\end{array} & \uparrow & \end{array}$$

Whether for top executive, plant or product manager, plant engineer, sales representative, or accountant, the concept of return on capital employed tends to mesh the interest of the entire organization. An understanding and appreciation of the return-on-capital concept by all employees help in building an organization interested in achieving fair profits and an adequate rate of return.

COST MANAGEMENT

Often, programs of **continuous improvement** require that costs be computed according to the activity-based method. That method facilitates identifying and setting priorities for the elimination of non-value-adding activities. **Non-value-added** activities decrease cycle time efficiency, where cycle time efficiency is the sum of all value-added activity times divided by total cycle time. The engineer may redesign the product using common parts or through process redesign so as to eliminate those activities or cost pools that add to product cost without adding to value. Some examples of these activities are material movement, run setup, and queue time.

Efforts to manage product costs by eliminating non-value-adding activities are frequently the result of a need to attain a specific target cost. Traditionally selling price was determined by adding a required markup to total cost. Global competition has forced producers to accept a market price determined by competitive forces:

$$\text{Target cost} = \text{market selling price} - \text{required return on investment}$$

Accordingly, the company that stays in business is the one that can accept this price and still earn a return on investment. **Target costing** is a concerted effort to design, produce, and deliver the product at a cost that will assure long-term survival.

Engineers may also focus on process throughput as a means of cost management. The **theory of constraints** proposes that profit is maximized when process throughput is maximized. The underlying assumption is that costs are fixed in the short run and cost per unit is lowest when fixed costs are distributed over the largest possible volume. This goal is accomplished by reducing cycle time through the elimination of process bottlenecks.

17.3 ENGINEERING STATISTICS AND QUALITY CONTROL

by Ashley C. Cockerill

REFERENCES: Brownlee, "Statistical Theory and Methodology in Science and Engineering," Wiley. Conover, Two *k*-sample slippage tests, *Journal of the American Statistical Association,* **63:** 614–626. Conover, "Practical Nonparametric Statistics," Wiley. Duncan, "Quality Control and Industrial Statistics," Richard D. Irwin. Gibbons, "Nonparametric Statistical Inference," McGraw-Hill. Olmstead, A corner test for association, *Annals of Mathematical Statistics,* **18:** 495–513. Owen, "Handbook of Statistical Tables," Addison-Wesley. Pearson, '"Biometrika Tables for Statisticians," vol. 1, 2d ed., Cambridge University Press. Tukey, "A quick compact two sample test to Duckworth's specifications," *Technometrics,* **1:** 31–48. Wilks, "Mathematical Statistics," Wiley.

ENGINEERING STATISTICS AND QUALITY CONTROL

Statistical models and statistical methods play an important role in modern engineering. Phenomena such as turbulence, vibration, and the strength of fiber bundles have statistical models for some of their underlying theories. Engineers now have available to them batteries of computer programs to assist in the analysis of masses of complex data. Many textbooks are needed to cover fully all these models and methods; many are areas of specialization in themselves. On the other hand, every engineer has a need for easy-to-use, self-contained statistical methods to assist in the analysis of data and the planning of tests and experiments. The sections to follow give methods that can be used in many everyday situations, yet they require a minimum of background to use and need little, if any, calculation to obtain good results.

STATISTICS AND VARIABILITY

One of the primary problems in the interpretation of engineering and scientific data is coping with **variability.** Variability is inherent in every physical process; no two individuals in any population are the same. For example, it makes no real sense to speak of the tensile strength of a synthetic fiber manufactured under certain conditions; some of the fibers will be stronger than others. Many factors, including variations of raw materials, operation of equipment, and test conditions themselves, may account for differences. Some factors may be carefully controlled, but variability will be observed in any data taken from the process. Even tightly designed and controlled laboratory experiments will exhibit variability.

Variability or variation is one of the basic concepts of statistics. Statistical methods are aimed at giving objective, quantitative, and reproducible ways of assessing the effects of variability. In particular, they aim to provide measures of the uncertainty in conclusions drawn from observational data that are inherently variable.

A second important concept is that of a **random sample.** To make valid inferences or conclusions from a set of observational data, the data should be able to be considered a random sample. What does this mean? In an operational sense it means that everything we are interested in seeing should have an equal chance of being represented in the observations we obtain. Some examples of what not to do may help. If machine setup is an important contributor to differences, then all observations should not be taken from one setup. If instrumental variation can be important, then measurements on the same item should not be taken successively—time to "forget" the last reading should pass. A random sample of *n* items in a warehouse is not the first *n* that you can find. It is the *n* that is selected by a procedure guaranteed to give each item of interest an equal chance of selection. One should be guided by generalizations of the fact that the apples on top of a basket may not be representative of all apples in the basket.

CHARACTERIZING OBSERVATIONAL DATA: THE AVERAGE AND STANDARD DEVIATION

The two statistics most commonly used to characterize observational data are the **average** and the **standard deviation.** Denote by $x_1, x_2, \ldots, x_n$

the n individual observations in a random sample from some process. Then the average and standard deviation are defined as follows:
Average:

$$\bar{x} = \sum_{i=1}^{n} x_i/n$$

Standard deviation:

$$s = \left[\sum_{i=1}^{n} (x_i - \bar{x})^2/(n - 1)\right]^{1/2}$$

Clearly, the average gives one number around which the n observations tend to cluster. The standard deviation gives a measure of how the n observations vary or spread about this average. The square of the standard deviation is called the **variance.** If we consider a unit mass at each point x_i, then the variance is equivalent to a moment of inertia about an axis through $\bar{x}$. It is readily seen that for a fixed value of $\bar{x}$, greater spreads from the average will produce larger values of the standard deviation s. The average and the standard deviation can be used jointly to summarize where the observations are concentrated. Tchebysheff's theorem states: A fraction of at least $1 - (1/k^2)$ of the observations lie within k standard deviations of the average. The theorem guarantees lower bounds on the percentage of observations within k (also known as z in some textbooks) standard deviations of the average.

Interval	Lower bound on % of measurements
$\bar{x} - 2s$ to $\bar{x} + 2s$	75%
$\bar{x} - 3s$ to $\bar{x} + 3s$	89%
$\bar{x} - 4s$ to $\bar{x} + 4s$	94%
$\bar{x} - 5s$ to $\bar{x} + 5s$	96%
$\bar{x} - 6s$ to $\bar{x} + 6s$	97%

Since the average and the standard deviation are computed from a sample, they are themselves subject to fluctuation. However, if μ, is the long-term average of the process and σ is the long-term standard deviation, then:

$$\text{Average } (\bar{x}) = \mu, \text{ process average}$$
$$\text{Average } (s) = \sigma, \text{ process standard deviation}$$

Furthermore, the intervals $\mu \pm k\sigma$ contain the same percentage of all values, as do the intervals $\bar{x} \pm ks$ for the sample; that is, at least 89 percent of all the long-term values will be contained in the interval $\mu - 3\sigma$ to $\mu + 3\sigma$, etc.

Range Estimate of the Standard Deviation

For $n \leq 20$ it is more convenient to compute the range r to estimate the standard deviation σ. The range is $x_{(n)} - x_{(1)}$, where $x_{(n)}$ is the largest value in a random sample of size n and $x_{(1)}$ is the smallest value. For example, if $n = 10$ and the observations are 310, 309, 312, 316, 314, 303, 306, 308, 302, 305, the range is $r = 316 - 302 = 14$. An estimate of the standard deviation σ is obtained by multiplying r by the factor f_n in Table 17.3.1. The average value of $r \cdot f_n$ is σ. Thus, in the example above, an estimate of σ and a value that can be used for s is $0.3249r = 0.3249(14) = 4.5486$.

Table 17.3.1 Average of Range $f_n = s$

Sample size	f_n	Sample size	f_n
2	0.8862	11	0.3152
3	0.5908	12	0.3069
4	0.4857	13	0.2998
5	0.4299	14	0.2935
6	0.3946	15	0.2880
7	0.3698	16	0.2831
8	0.3512	17	0.2787
9	0.3367	18	0.2747
10	0.3249	19	0.2711
		20	0.2677

PROCESS VARIABILITY—HOW MUCH DATA?

Since the output of all processes is variable, one can make reasonable decisions about the output only if one can obtain a measure of how much variability or spread one can expect to see under normal conditions. Variability cannot be measured accurately with a small amount of data. Methods for assessing how much data are needed are given for two general situations.

Specified Tolerances

A convenient statement about the variability or spread of a process can be based on the smallest and largest values in a random sample of the output. There are no practical limitations on its use. Suppose that we have a random sample of n values from our process. Denote the values by $X_1, X_2, \ldots, X_n$. After obtaining the values we find the smallest, $X_{(1)}$, and the largest, $X_{(n)}$. Now we want to assess what percent of all future values that this process might generate will be covered by $X_{(1)}$ and $X_{(n)}$. In statistics, $X_{(1)}$ and $X_{(n)}$ are called **tolerance limits.** If the process generates bolts and X is the diameter, then the engineering concept of tolerance and the statistical concept of tolerance are seen to be quite similar.

Let p be the percentage of all the process values that on a long-term basis will be between $X_{(1)}$ and $X_{(n)}$. Let P be a lower bound for this percentage p. Now consider the probability statement: Probability $(p \geq P) = C$. The quantity C we call **confidence.** Since it is a probability its value is between 0 and 1. As C approaches 1 our confidence in the percentage P increases. The interpretation of P and C can be explained in terms of Table 17.3.2.

Suppose that we take a random sample of size $n = 269$ values from our process output; the smallest value is 10 and the largest is 54. In Table 17.3.2 we see that 269 is the entry for $P = 99$ and $C = 0.75$. This tells us that at least $P = 99$ percent of all future values that this process will generate will be between 10 and 54, the smallest and largest values seen in the sample of 269. The confidence $C = 0.75 = {}^3/_4$ tells us that the chances are 3 out of 4 that our statement is correct. As we increase the sample size n, we increase the chances that our statement is correct. For example, if our sample size had been $n = 473$, then $C = 0.95$ and the chances are 95 in 100 that we are correct in making the statement that at least 99 percent of all process values will be between the 10 and 54 seen in the sample. Similarly, if the sample size had been 740, then the chances of being correct increase to 995 in 1000. If sample size n is decreased sufficiently, the confidence $C = 0.50 = {}^1/_2$ indicates that the chances are one in two of being right, *and* one in two of being *wrong*. Therefore, it is important to select n so as to keep C as large as possible. The cost of acquiring the data will determine the upper limit for n.

Further information on tolerance limits can be found in Wilks (1962) and Duncan (1986).

Wear-Out and Life Tests

A special case of coverage occurs if our interest is in a wear-out phenomenon or a life test. For example, suppose we put a number of incandescent light bulbs on test; our interest is in the length of time to failure. Clearly we do not want to wait until all specimens fail to draw a conclusion; it

Table 17.3.2

	Confidence, C			
P, %	0.995	0.99	0.95	0.75
99.9	7427	6636	4742	2692
99.5	1483	1325	947	538
99	740	661	473	269
98	368	330	235	134
97	245	219	156	89
96	183	163	117	67
95	146	130	93	53
90	71	64	46	26
80	34	31	22	13
75	27	24	18	10

NOTE: Sample size r required to have a confidence C that at least P percent of all future values will be included between the smallest and largest values in a random sample.

Table 17.3.3

	Confidence, *C*			
Q, %	0.995	0.99	0.95	0.75
0.1	5296	4603	2995	1386
1	528	459	299	138
2	263	228	149	69
3	174	152	99	46
4	130	113	74	34
5	104	90	59	28
10	51	44	29	14
15	33	29	19	9
20	24	21	14	7
25	19	17	11	5

NOTE: Sample size *r* required to have a confidence *C* that fewer than *Q* percent of future units will fail in a time shorter than the shortest life in the sample. For a more extensive table of values, see Owen (1962).

might take an inordinate length of time for the last one to fail. From a practical point of view we would probably be interested in those that fail first anyway. If the sample size is properly chosen, there will be important information as soon as we obtain the first failure.

In a random sample of size *n* let the failure times be $T_1 \geq T_2 \geq \ldots \geq T_n$. The value T_1 is thus the smallest value in a random sample of size *n*. Now let *q* be the percentage of future units that can be expected to fail in a time less than T_1, the smallest value in the sample. As before we can make a probability statement about *q*. Let *Q* be an upper bound to *q*. Then we can compute: Probability $(q \leq Q) = C$.

For example, suppose that we put a random sample of 299 items on test and the first one fails in time $T_1 = 151$ h. From Table 17.3.3 we see that 299 is an entry for $Q = 1$ and $C = 0.95$. Thus we can conclude that not more than 1 percent of future units should fail in a time less than 151 h. The confidence in the statement is 95 chances in 100 of being correct. Again referring to Table 17.3.3, we see that if $T_1 = 151$ were based on a sample of 528, then the confidence would be increased to 995 chances in 1000. Most importantly, Table 17.3.3 tells us that we need to test a very large sample if we want to have high confidence that only a small percentage of future units will fail in a time less than the smallest observed. The theory behind Table 17.3.3 can be found in Wilks (1962). For a more extensive table of values see Owen (1962). If $Q' = Q/100$, use

$$r = [\log(1 - C)]/[\log(1 - Q')]$$

CORRELATION AND ASSOCIATION

One of the most common problems in the analysis of engineering data is to determine if a correlation or an association exists between two variables *X* and *Y*, where the data occur in pairs (X_i, Y_i). The "corner test of association" developed by Olmstead and Tukey (1947) is a quick and simple test to assist in making this determination.

Corner Test

Conditions for Use Each pair (X_i, Y_i) should have been obtained independently; there are no other practical assumptions for its use. Of course, the user should consider the physical process generating the data when interpreting any correlation or association that is determined to exist.

Procedure

1. Make a scatter plot on graph paper of the data pairs (X_i, Y_i), with the usual convention that *X* is the horizontal scale and *Y* the vertical.

2. Determine the median X_m of the X_i values. Determine the median Y_m of the Y_i values.

The median splits the data into two parts so that there is an equal number of values above and below the median. Let *N* denote the total number of points. If *N* is odd, then *N* can be written as $2k + 1$ and the median is the $(k + 1)$st value as the values are ordered from the smallest to the largest. If *N* is even, then *N* can be written as $2k$. Then the median is taken to be midway between the *k*th and $(k + 1)$st values.

Table 17.3.4

i	*X*	*Y*	*i*	*X*	*Y*
1	10.45	4.1	18	9.65	3.8
2	13.81	2.7	19	7.44	5.4
3	12.22	1.6	20	10.70	7.6
4	9.05	4.3	21	13.38	6.0
5	17.86	2.6	22	13.00	10.4
6	14.54	0.1	23	13.90	10.7
7	19.99	3.7	24	11.94	9.4
8	8.73	3.5	25	14.11	10.7
9	4.66	5.3	26	0.93	12.9
10	13.88	3.9	27	3.18	12.5
11	5.10	4.4	28	13.13	6.5
12	3.98	4.1	29	13.45	11.7
13	8.12	6.3	30	12.70	9.6
14	12.26	6.6	31	15.95	8.5
15	10.30	6.5	32	7.30	16.6
16	5.40	11.9	33	7.78	8.8
17	10.39	5.8			

NOTE: Data are on paper samples. *X* is proportional to reciprocal of light transmission. *Y* is proportional to tensile strength.

3. Draw a vertical line through X_m.

4. Draw a horizontal line through Y_m.

5. The lines in (3) and (4) divide the graph into four quadrants. Label the upper right and lower left as plus. Label the upper left and lower right as minus.

6. Begin at the right side of the plot. Count the values, in order of decreasing *X*, until forced to cross the horizontal median Y_m. Give the count a plus sign if the values are in a plus quadrant, a minus sign if they are in a minus quadrant.

7. Repeat the procedure in (6), moving down from above until you have to cross the vertical median, moving from left to right until you have to cross the horizontal median, and moving up from below until you have to cross the vertical median.

8. Compute the algebraic sum of the four counts obtained in (6) and (7). Denote the sum by *T*.

Test If $|T| \geq 11$, then there is evidence of correlation between *X* and *Y*; $|T|$ is the value of *T* ignoring the sign.

EXAMPLE. Table 17.3.4 gives 33 pairs of values obtained from samples of a paper product. The *X* coordinate is proportional to the reciprocal of light transmitted by the sample. The *Y* coordinate is proportional to tensile strength.

1. The 33 pairs of values are plotted in Fig. 17.3.1.

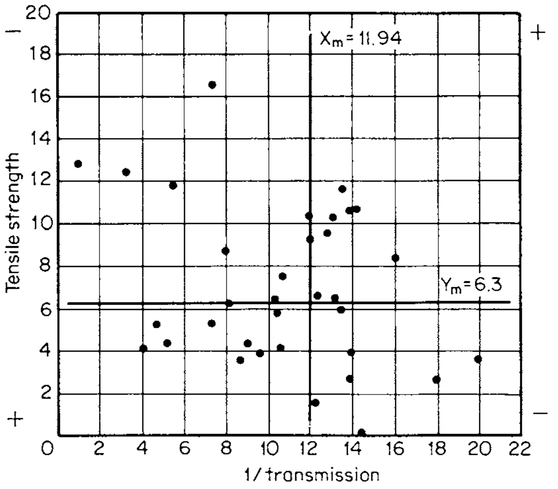


Fig. 17.3.1 Plot of example data used in conjunction with the Corner Test.

2. The median of X values is $X_m = 11.94$. The median of Y values is $Y_m = 6.3$.
3 to 5. The medians are shown in Fig. 17.3.1, and the quadrants are labeled.
6 to 7. The counts are as follows:
Right to left: -2 (points at 19.99 and 17.86 on X)
Top to bottom: -4 (points at 16.6, 12.9, 12.5, 11.7 on Y)
Left to right: -2 (points at 0.93 and 3.18 on X)
Bottom to top: -4 (points at 0.1, 1.6, 2.6, 2.7 on Y)
8. The algebraic sum of the counts is -12. Hence $T = -12$. And since $|T| \geq 11$, one can conclude that there is evidence of correlation or association between the variables X and Y.

COMPARISON OF METHODS OR PROCESSES

A common problem in engineering investigations is that of using experimental or observational data to assess the performance of two processes, two treatment methods, or the like. Often one process or treatment is a standard or the one in current use. The other is an alternative that is a candidate to replace the standard. Sometimes it is cheaper, and one hopes to see no performance difference. Sometimes it is supposed to offer superior performance, and one hopes to see a measurable difference in the variable of interest. In either case we know that the variable of interest will have a distribution of values; and if the two processes are to be measurably different the distribution of values should not overlap too much. For an objective assessment we need to have some way to calibrate the overlap. A quick and easy-to-use test for this purpose is the **outside count test** developed by Tukey (1959).

Two Methods—Outside Count Test

Conditions for Use Given two groups of measurements taken under conditions 1 and 2 (treatments, methods, etc.), we identify the direction of difference by insisting that the two groups have minimum overlap. Use 1 to denote the group with the smaller number of measurements and let N_1 be the number of measurements for that group. Let N_2 be the number of measurements for the other group. The number of observations for each group should be about the same.

The conditions to be satisfied are:

$$4 \le N_1 \le N_2 \le 30$$
$$N_2 \le (4N_1/3) + 3$$

Procedure

1. Count the number of values in the one group *exceeding* all values in the other.
2. Count the number of values in the other group *falling below* all those in the other.
3. Sum the two counts in (1) and (2). (It is required that neither count be zero. One group must have the largest value and the other the smallest.)

Test If the sum of the two counts in (3) is 7 or larger, there is sufficient evidence to conclude that the two methods are measurably different.

EXAMPLE. The following data represent the results of a trial of two methods for increasing the wear resistance of a grinding wheel. The data are proportional to wear:

Method 1: 13.06**, 9.52, 9.98, 8.83, 12.78, 9.00, 11.56, 8.10*, 12.21.
Method 2: 8.44, 9.64, 9.94, 7.30, 8.74, 6.30*, 10.78**, 7.24, 9.30, 6.66.

The smallest value for each method is marked with an asterisk; the largest value is marked with two asterisks.

Count 1: The largest value is 13.06 for method 1. The values 13.06, 12.78, 12.21, and 11.56 for method 1 exceed the largest value for method 2, viz., 10.78. Hence the count is 4.

Count 2: The smallest value for method 1 is 8.10. For method 2 the values 7.30, 6.30, 7.24, and 6.66 are less than 8.10. Hence the count is 4.

Count 3: The total count is $4 + 4 = 8 > 7$.

The data support the conclusion that method 2 produces measurably less wear than method 1.

Several Methods

The problem outlined in the preceding section can be generalized so that one can make a comparison of several processes, treatments, methods, or the like. Again, if there are differences among the methods, the values that we see should not overlap too much. We give you two easy-to-use tests. The first is for the situation where there is an equal amount of data for each method. For the second, the amount of data may differ. Each method will be demonstrated using the data in Table 17.3.5.

Table 17.3.5

	Supplier			
Sample no.	1	2	3	4
1	45.37	30.05	41.30	46.21
2	21.68	36.04	31.09	36.01
3	43.91	18.04	24.31	46.28
4	47.76*	32.91	15.64	21.80
5	23.81	41.67	54.85*	28.57
6	19.90	37.40	32.96	48.45
7	44.68	46.67*	45.48	33.49
8	11.81	27.93	45.14	53.07*
9	35.42	45.20	45.49	35.65
10	39.85	29.54	52.82	14.95

NOTE: The data are proportional to the time to failure of a standard cutting tool. Asterisks denote largest value in each group.

Several Methods—Overlap Test

Conditions for Use Independent data should be obtained for each of the k methods. *The number of values n should be the same* for each method.

Procedure

1. For each of the k methods, determine the *largest* value. Label it with an asterisk.
2. Scan the largest values. Label the group with the *largest largest* value as BIG. Label the group with the *smallest* largest value as SMALL, and its largest value as S.
3. In the group labeled BIG *count* the number of values that are larger than S, the largest value in SMALL. Denote this count by C.
4. Enter Table 17.3.6 for n values of k groups. If C exceeds the tabled value, then the data support a conclusion that the methods are different. Otherwise, they do not.

EXAMPLE. 1. In Table 17.3.5 the largest value for each of the four groups is marked with an asterisk.
2. Group 3 is BIG. Group 2 is SMALL; the largest value in Group 2 is $S = 46.67$.
3. The number of values in Group 3 larger than 46.67 is 2 (52.82, 54.85).
4. Enter Table 17.3.6 with $n = 10$ and $k = 4$. The entry is 5 which is greater than 2. Hence, the data do not support a conclusion that the time to failure for the cutting tools of the four suppliers is measurably different.

Several Methods—Rank Test

Conditions for Use Independent data should be obtained for each of the methods. The amount of data for each method may be different.

Procedure

1. Let n_i be the number of values in Group i.
2. Let $N = \sum_{1}^{k} n_i$ be the total number of values.
3. Rank each value from 1 to N beginning with the smallest. (If there are ties among t values, divide the successive ranks equally among them.)
4. Compute r_i, the sum of the ranks for the ith group. [*Note:* For a check $\sum_{1}^{k} r_i = N(N + 1)/2$.]

Table 17.3.6 95% Point for *k*-Sample Problems

	k							
n	3	4	5	6	7	8	9	10
5	4	4	4	4	4	4	4	4
6	4	4	4	5	5	5	5	5
7	4	5	5	5	5	5	5	5
8	4	5	5	5	5	5	5	5
9	5	5	5	5	5	5	6	6
10	5	5	5	5	6	6	6	6
12	5	5	5	6	6	6	6	6
14	5	5	6	6	6	6	6	6
16	5	5	6	6	6	6	6	6
18	5	6	6	6	6	6	6	7
20	5	6	6	6	6	6	7	7
25	5	6	6	6	6	7	7	7
30	5	6	6	6	7	7	7	7
40	5	6	6	7	7	7	7	7
>40	6	6	7	7	7	7	8	8

NOTE: *k* is the number of groups; *n* is the number of values per group. For other *n, k,* and percentage points see Conover (1968).

Test

1. Compute

$$T = [12/N(N + 1)]\left[\sum_{1}^{k} (r_i^2/n_i)\right] - 3(N + 1)$$

2. Go to Table 17.3.7; find the entry under $k - 1$.

If T exceeds the entry, then the data support the conclusion that the groups are different. Otherwise, they do not.

Table 17.3.7 Chi-Square Distribution

k	*w*	*k*	*w*
1	3.841	16	26.30
2	5.991	17	27.59
3	7.815	18	28.87
4	9.488	19	30.14
5	11.07	20	31.41
6	12.59	22	33.92
7	14.07	24	36.42
8	15.51	26	38.89
9	16.92	28	41.34
10	18.31	30	43.77
11	19.68	40	55.76
12	21.03	50	67.50
13	22.36	60	79.08
14	23.68	70	90.53
15	25.00	80	101.9

NOTE: Entries are $P(W > w) = p = 0.05$. For other values of *k* and *p,* see Pearson and Hartley (1962).

EXAMPLE. We again use the data shown in Table 17.3.5. In Table 17.3.8 the numerical values representing times to failure have been replaced by their ranks. To facilitate such ranking it is convenient to order the values in each group from smallest to largest. Then all values are ranked from smallest to largest. In Table 17.3.8 the values have been reordered this way. The ranks are in parentheses.

1. The number of values in each group is 10. Hence $n_i = 10$ for each value of i.
2. The total number of values $N = 40$.

3 and 4. The sum of the ranks r_i is shown for each group. [Note that $\Sigma r_i = 820 = (40)(41)/2$.]

$$\begin{aligned} T &= [12/N(N + 1)][\Sigma(r_i^2/10)] - 3(N + 1) \\ &= [12/(40)(41)][170182/10] - 3(41) \\ &= 124.523 - 123. = 1.523 \end{aligned}$$

Now go to Table 17.3.7 and obtain the entry under $k = 4 - 1 = 3$. The entry is 7.815, which is larger than 1.523. Hence, the data do not support a conclusion that the time to failure for the cutting tools for the four suppliers is measurably different.

Table 17.3.8

Supplier*				
1	2	3	4	
11.81 (1)	18.04 (4)	15.64 (3)	14.95 (2)	
19.90 (5)	27.93 (10)	24.31 (9)	21.80 (7)	
21.68 (6)	29.54 (12)	21.09 (14)	28.57 (11)	
23.81 (8)	30.05 (13)	32.96 (16)	33.49 (17)	
35.42 (18)	32.91 (15)	41.30 (24)	35.65 (19)	
39.85 (23)	36.04 (21)	45.14 (28)	36.01 (20)	
43.91 (26)	37.40 (22)	45.48 (31)	46.21 (33)	
44.68 (27)	41.67 (25)	45.49 (32)	46.28 (34)	
45.37 (30)	45.20 (29)	52.82 (38)	48.45 (37)	
47.76 (36)	46.67 (35)	54.85 (40)	53.07 (39)	
180	186	235	219	r_i
32400	34596	55225	47961	r_i^2

* These are the data of Table 17.3.5 with the values for each supplier listed from smallest to largest. The values in parentheses are the ranks of the time to failure values from smallest to largest.

GO/NO-GO DATA

Quite often the data that we encounter will be **attribute** or **go/no-go data;** that is, we will not have quantitative measurements but only a characterization as to whether an item does or does not have some attribute. For example, if a manufactured part has a specification that it should not be shorter than 2 in, we might construct a template; and if a part is to meet the specification, it should not fit into the template. After inspecting a series of units with the template our data would consist of a tabulation of "gos" and "no-gos"—those that did not meet the specification and those that did.

If the items that are checked for an attribute are obtained by random sampling, the resulting data will follow what is known as the **binomial distribution.** Its standard form is as follows:

p is the long-term fraction of failures
$q = 1 - p$ is the long-term fraction of successes
n is the size of the random sample.

Then the probability that our sample gives x failures and $n - x$ successes is

$$f(x;n,p) = \binom{n}{x} p^x q^{n-x}; x = 0,1,2, \ldots, n$$

where $$\binom{n}{x} = n!/x!(n - x)!$$

From $f(x; n, p)$, for a given n and p, we can calculate the probability of x failures in a sample size n. Similarly, by summing terms for different values of x, we can calculate the probability of having more than w failures or fewer than r failures, etc. Here we are not going to try to be so precise; rather we are going to try to show the general picture of the relationship between x, p, and n by the use of examples and the graph in Fig. 17.3.2.

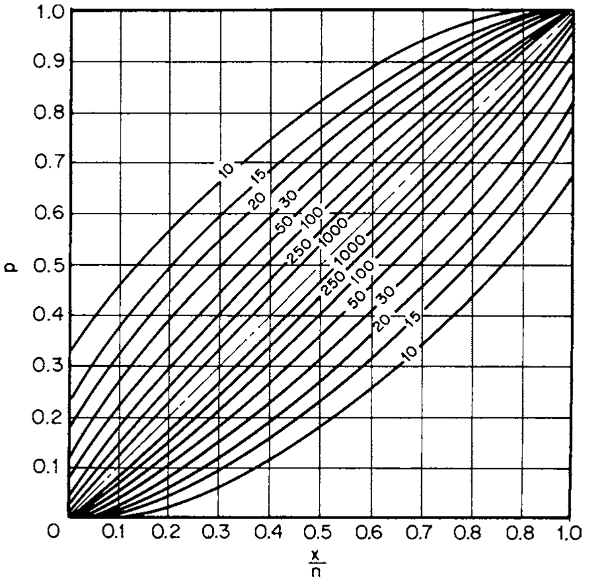


Fig. 17.3.2 Binomial distribution, 95 percent confidence bands. *(Reproduced with the permission of the Biometrika Trust from C. J. Clopper and E. S. Pearson, "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial,"* Biometrika, 26 (1934), p. 410.)

Estimating the Failure Rate

In a manufacturing process a general index of quality is the fraction of items which fail to pass a certain test. Suppose that we take a random sample of size $n = 100$ from the process and observe $x_0 = 10$ failures. Clearly we have met the conditions for a binomial distribution and an estimate of p, the long-term failure rate is $\hat{p} = x_0/n = 10/100 = 0.1$. However, we would also like to know the accuracy of the estimate. In other words, if we operate the process for a long time under these conditions and obtain a large sample, what might be the value of p? One simple way to assess the estimate of $\hat{p}$ is to find values p_1 and p_2 ($p_1 < p_2$) such that the following probabilities are satisfied for a fixed value of α:

$$\Pr[x \geq x_0 | p_1] = \alpha/2$$
$$\Pr[x \leq x_0 | p_2] = \alpha/2$$

These values are the solutions for p of the two equations.

$$\sum_{x=x_0}^{n} \binom{n}{x} p_1^x (1 - p_1)^{n-x} = \alpha/2$$
$$\sum_{x=0}^{x_0} \binom{n}{x} p_2^x (1 - p_2)^{n-x} = \alpha/2$$

General solutions for these equations for $\alpha = 0.05$ are shown in Fig. 17.3.2. If we go to Fig. 17.3.2 with $x/n = 0.1$ and read where the lines for $n = 100$ intersect the ordinate or p scale, we see that $p_1 = 0.07$ and $p_2 = 0.18$. We can then state that we have reasonable confidence (the probability is 0.95) that the long-term failure rate for the process is between 0.07 and 0.18.

Estimating the Sample Size

It should be evident that Fig. 17.3.2 can also be used to determine how large a sample is needed to estimate a proportion or a percentage with a specific accuracy or tolerance. Suppose that the proportion of interest is assumed to be around $p = 0.20$. Now enter Fig. 17.3.2 with $x/n = 0.20$. From the figure we see that if we take a sample of size 50, our estimate will have a range of about ± 0.10 (actually -0.10, $+0.13$). On the other hand, if the sample size is 250, the estimate will have a range of about ± 0.06.

Often one wants to compare the performance of two processes. As above, suppose that the rate p for our process is 0.20. We have a modification that we want to test; however, to be economical the modification has to bring the rate p down to 0.15 or less. If the modification is going to be assessed on the p's for the standard and the modification, then we do not want the uncertainty in their estimates to overlap and the uncertainty should be less than half of 0.05 where $0.05 = (0.20 - 0.15)$. Figure 17.3.2 shows that we would have to use a sample size of at least 1000. This demonstrates that attribute sampling is effective only when the items and their characterization are not expensive. Otherwise, it is best to go to measurements where smaller sample sizes can be used to assess differences.

A more detailed exposition of the binomial distribution and its uses can be found in Brownlee (1970).

Statistical Software Packages

There are numerous statistical software packages available that may be used to determine sample sizes, design experiments, fit curves to data, test for goodness of fit, and perform basic statistical calculations. The packages include both standardized and customized menus, features for importing and exporting data to and from external files, and well-presented analysis summaries and reports. For example, a widely used statistical package called JMP was developed and is periodically upgraded by the SAS Institute, Inc.

CONTROL CHARTS

When an industrial process is under control it is in a state of "statistical equilibrium." By equilibrium we mean that we can characterize its output by a fixed average μ and a fixed standard deviation σ. The variation in output is what one would expect to see from that μ and σ, as bounded by the values given in Tchebysheff's theorem, let us say. However, if control is lost, we tend to get a greater spread in values. In effect, the average μ, or the standard deviation σ is changing because of some cause. The causes of lack of control are manifold—it can be a change in raw materials, tool wear, instrumentaton failure, operator error, etc. The important thing is that one wants to be able to detect when this lack of control occurs and take the appropriate steps to make corrections.

One of the most frequently used statistical tools for analyzing the state of an industrial process is the control chart. The two most commonly used charts are those for the **average** and the **range.** The control chart procedure consists of these steps:

1. Choose a characteristic X which will be used to describe the product coming from the process.

2. At time t_i, take a small number of observations n on the process. The number n should be small enough so that it is reasonable to assume that conditions will not change during the course of obtaining the observstions.

3. For each set of n observations, compute the average $\bar{x}_i$ and the range r_i as defined in the section "Characterizing Observational Data."

There are two different control situations of interest.

4a. *Control standards given.* Suppose that from past operation of the process or from the need to meet certain specifications, a goal average μ^* and a goal standard deviation σ^* are specified. Then we set up two charts as follows:

Average chart:	Upper limit line:	$\mu^* + A\sigma^*$
	Central line:	μ^*
	Lower limit line:	$\mu^* - A\sigma^*$
Range chart:	Upper limit line:	$D_2\sigma^*$
	Central line:	$d\sigma^*$
	Lower limit line:	$D_1\sigma^*$

Table 17.3.9 Factors for Control-Chart Limits

Sample size n	For averages		For ranges				
	A	A_2	d	D_1	D_2	D_3	D_4
2	2.12	1.88	1.128	0	3.69	0	3.27
3	1.73	1.02	1.693	0	4.36	0	2.57
4	1.50	0.73	2.059	0	4.70	0	2.28
5	1.34	0.58	2.326	0	4.92	0	2.11
6	1.22	0.48	2.534	0	5.08	0	2.00
7	1.13	0.42	2.704	0.21	5.20	0.08	1.92
8	1.06	0.37	2.847	0.39	5.31	0.14	1.86
9	1.00	0.34	2.970	0.55	5.39	0.18	1.82
10	0.95	0.31	3.078	0.69	5.47	0.22	1.78

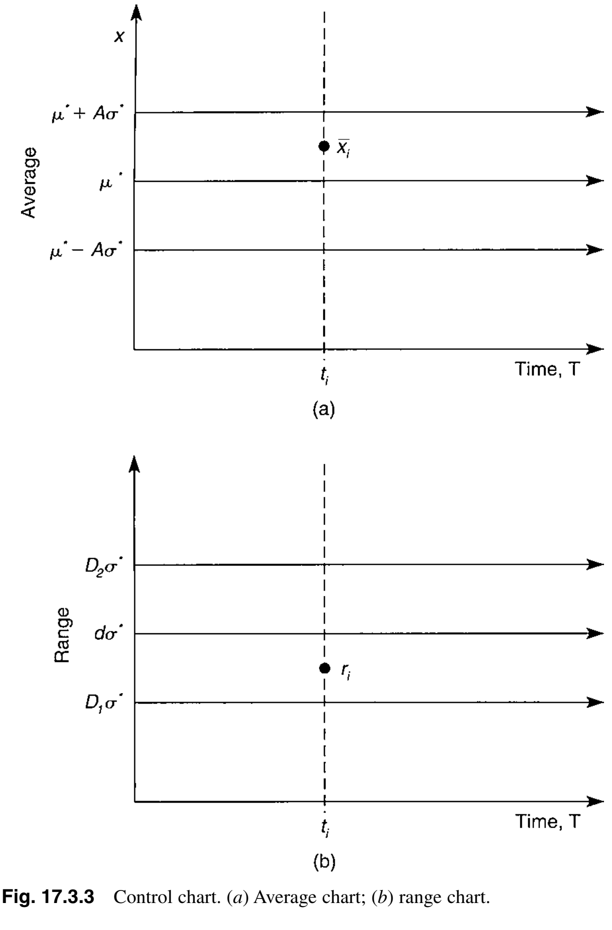
The values of A, d, D_1, and D_2 depend upon n and can be found in Table 17.3.9.

Plot the values of $\bar{x}_i$ and r_i obtained in (3) on the two charts as shown in Fig. 17.3.3. Whenever a value falls outside the limit lines, there is an indication of lack of control, and one is justified in seeking the causes for a change.

4b. *Control, no standards given.* Often one has no prior information about the process μ and σ, and one wants to determine if the process behaves as though it is in statistical equilibrium, and if not, take actions to get it there. In this case one has to determine the central lines for the charts from process data. To do this one first accumulates the data for 25 to 50 time periods as indicated in (2). Then two charts are set up as follows: Let K be the 25 to 50 time periods observed. Compute an overall average $\overline{X} = \Sigma_{i=1}^{k} \bar{x}_i/K$ and the average range $\overline{R} = \Sigma_{i=1}^{k} r_i/K$. Set up charts with limits defined from:

Average chart: Upper limit line: $\overline{X} + A_2\overline{R}$
Central line: $\overline{X}$
Lower limit line: $\overline{X} - A_2\overline{R}$
Range chart: Upper limit line: $D_4\overline{R}$
Central line: $\overline{R}$
Lower limit line: $D_3\overline{R}$

The values of A_2, D_3, and D_4 depend upon n and can be found in Table 17.3.9. The individual $\bar{x}_i$ and r_i are plotted on the charts, and again a value outside the limits is an indication of lack of control and is justification for seeking the cause for a change.



(a)

(b)

Fig. 17.3.3 Control chart. (*a*) Average chart; (*b*) range chart.

Process Capability Indices

If the process is in statistical control, an estimate for the process standard deviation can be obtained by using

$$\hat{\sigma} = \overline{R}/d$$

In turn, $\hat{\sigma}$ is used to calculate the **process capability indices** C_p and C_{pk}. These two indices compare the actual spread of the data with the desired range (usually as specified by a customer). C_p is used when the actual process average is equal to the goal average. C_{pk} is used when the actual process average is not equal to the goal average. The desired range is called the **specification tolerance** (ST) and is equal to the upper specification limit minus the lower specification limit, namely, $2A\sigma^*$.

C_p is given by the following equation:

$$C_p = \mathrm{ST}/6\hat{\sigma} = 2A\sigma^*/6\hat{\sigma}$$

If $C_p \geq 1$, the process is considered to be capable, which means that most or all of the data stayed within the desired range. If $C_p < 1$, the process is considered to be not capable and requires adjustment. Ideally, one should control the process variability so that $C_p \geq 2$.

The other index, C_{pk}, is given by

$$C_{pk} = C_p(1 - k)$$

where $$k = 2(\mu^* - \overline{X})/\mathrm{ST} = (\mu^* - \overline{X})/\mathrm{A}\sigma^*$$

and $$\overline{X} = \Sigma\,\bar{x}_i/K$$

The $(1 - k)$ factor modifies C_p so as to allow the actual average $\overline{X}$ to be different from the goal average μ^*. Ideally, one should control the process so that $C_{pk} \geq 1.5$. In process capability analysis, the indices should be calculated on a frequent basis, but the trends should be examined only monthly or quarterly in order to be meaningful.

Charts for Go/No-Go Data

The control chart concept can also be used for attribute or go/no-go data. The procedures are, in general, the same as outlined for averages and ranges. Briefly, they are as follows:

1. Select a sample of size n from the process; for best results n should be in the range of 50 to 100.
2. Let x_i denote the number of defective units in the sample of size n at time t_i; then $\hat{p}_i = x_i/n$ is an estimate of the process fraction defective.
3. Set control limits for a standard fraction defective p^* at

$$p^* \pm 3[p^*(1 - p^*)/n]^{1/2}$$

If no standard is given, then take $K = 25$ to 50 samples of size n to get a good estimate of the fraction-defective p. Define $\bar{p} = \Sigma_{i=1}^{K} \hat{p}_i/K$. In this case set control limits at

$$\bar{p} \pm 3[\bar{p}(1 - \bar{p})/n]^{1/2}$$

4. Interpret a $\hat{p}_i$ outside the limits as an indication of a change worthy of investigation.

Further information on control charts can be found in Duncan (1986).

17.4 METHODS ENGINEERING

by Vincent M. Altamuro

REFERENCES: ASME Standard Industrial Engineering Terminology. Barnes, “Motion and Time Study,” Wiley. Krick, “Methods Engineering,” Wiley. Maynard, Stegemerten, and Schwab, “Methods-Time-Measurement,” McGraw-Hill.

SCOPE OF METHODS ENGINEERING

Methods engineering is concerned with the selection, development, and documentation of the methods by which work is to be done. It includes the analysis of input and output conditions, assisting in the choice of the processes to be used, operations and work flow analyses, workplace design, assisting in tool and equipment selection and specifications, ergonomic and human factors considerations, workplace layout, motion analysis and standardization, and the establishment of work time standards. A primary concern of methods engineering is the integration of humans and equipment in the work processes and facilities.

PROCESS ANALYSIS

Process analysis is that step in the conversion of raw materials to a finished product at which decisions are made regarding what methods, machines, tools, inspections and routings are best. In many cases, the product’s specifications can be altered slightly, without diminishing its function or quality level, so as to allow processing by a preferred method. For this reason, it is desirable to have the product’s designer and the process engineer work together before specifications are finalized.

WORKPLACE DESIGN

Material usually flows through a facility, stopping briefly at stations where additional work is done on it to bring it closer to a finished product. These workstations, or workplaces, must be designed to permit performance of the required operations, to contain all the tooling and equipment needed to fit the capabilities and limitations of the people working at them, to be safe and to interface smoothly with neighboring workplaces.

Human engineering and ergonomic factors must be considered so that all work, tools, and machine activation devices are not only within the comfortable reach of the operator but are designed for safe and efficient operation. A workplace chart (Fig. 17.4.1) which analyzes the required actions of both hands is an aid in workplace design.

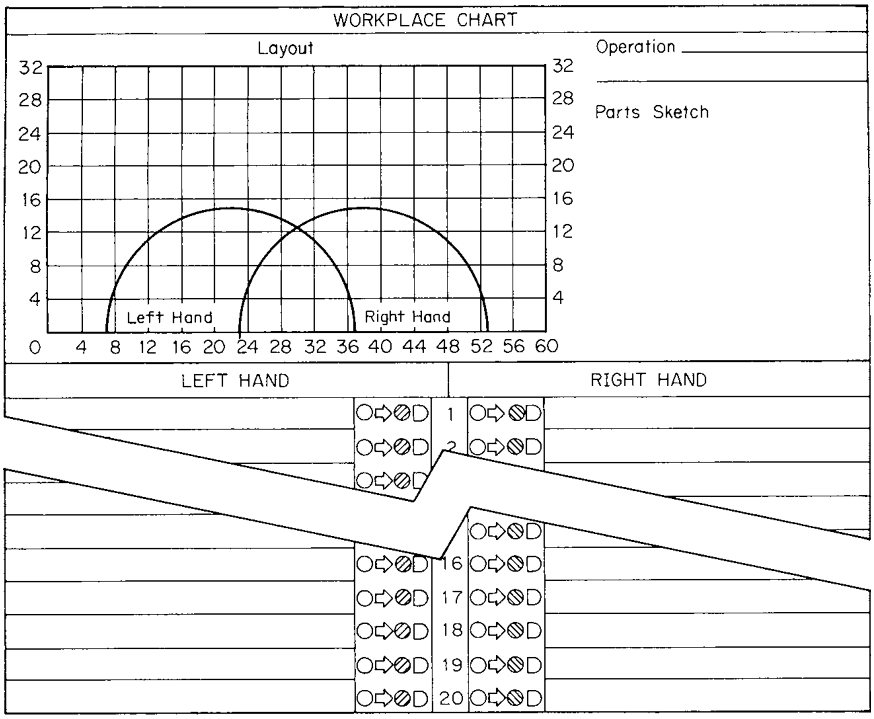


Fig. 17.4.1 Workplace layout chart.

METHODS DESIGN

Methods design is the analysis of the various ways a task can be done so as to establish the one best way. It includes *motion analysis*—the study of the actions the operator can use and the advantages and/or disadvantages of each variation—and *standardization of procedure*—the selection and recording of the selected and authorized work methods.

While "time and motion study" is the more commonly used term, it is more correct to use "motion and time study," as the motion study to establish the standard procedure must be done prior to the establishment of a standard time to perform that work.

According to ASME Standard Industrial Engineering Terminology, motion study is defined as

> . . . the analysis of the manual and the eye movements occurring in an operation or work cycle for the purpose of eliminating wasted movements and establishing a better sequence and coordination of movements.

In the same publication, time study is defined as

> . . . the procedure by which the actual elapsed time for performing an operation or subdivisions or elements thereof is determined by the use of a suitable timing device and recorded. The procedure usually but not always includes the adjustment of the actual time as the result of performance rating to derive the time which should be required to perform the task by a workman working at a standard pace and following a standard method under standard conditions.

Attempts have been made to separate the two functions and to assign each to a specialist. Although motion study deals with method and time study deals with time, the two are nearly inseparable in practical application work. The method determines the time required, and the time determines which of two or more methods is the best. It has, therefore, been found best to have both functions handled by the same individual.

ELEMENTS OF MOTION AND TIME STUDY

Figure 17.4.2 presents graphically the steps which should be taken to make a good motion and time study and shows their relation to each other and the order in which they must be performed.

METHOD DEVELOPMENT

The first steps are concerned with the development of the method. Starting with the drawing of the product, the operations which must be performed are determined and tools and equipment are specified. In large companies, this is usually done by a specialist called a process engineer. In smaller companies, processing is commonly done by the time study specialist.

Next, the detailed method by which each operation should be performed is developed. The procedures used for this are known as operation analysis and motion study.

OPERATION ANALYSIS

Operation analysis is the procedure employed to study all major factors which affect a given operation. It is used for the purpose of uncovering possibilities of improving the method. The study is made by reviewing the operation with an open mind and asking either of oneself or others questions which are likely to lead to methods-improving ideas. If this is done systematically, so that the possibility of overlooking factors which should be considered is minimized, worthwhile improvements are almost certain to result.

The 10 major factors explored during operation analysis, together with typical questions which should be asked about each factor, are as follows:

1. Purpose of operation
   - *a.* Is the result accomplished by the operation necessary?
   - *b.* Can the purpose of the operation be accomplished better in any other way?

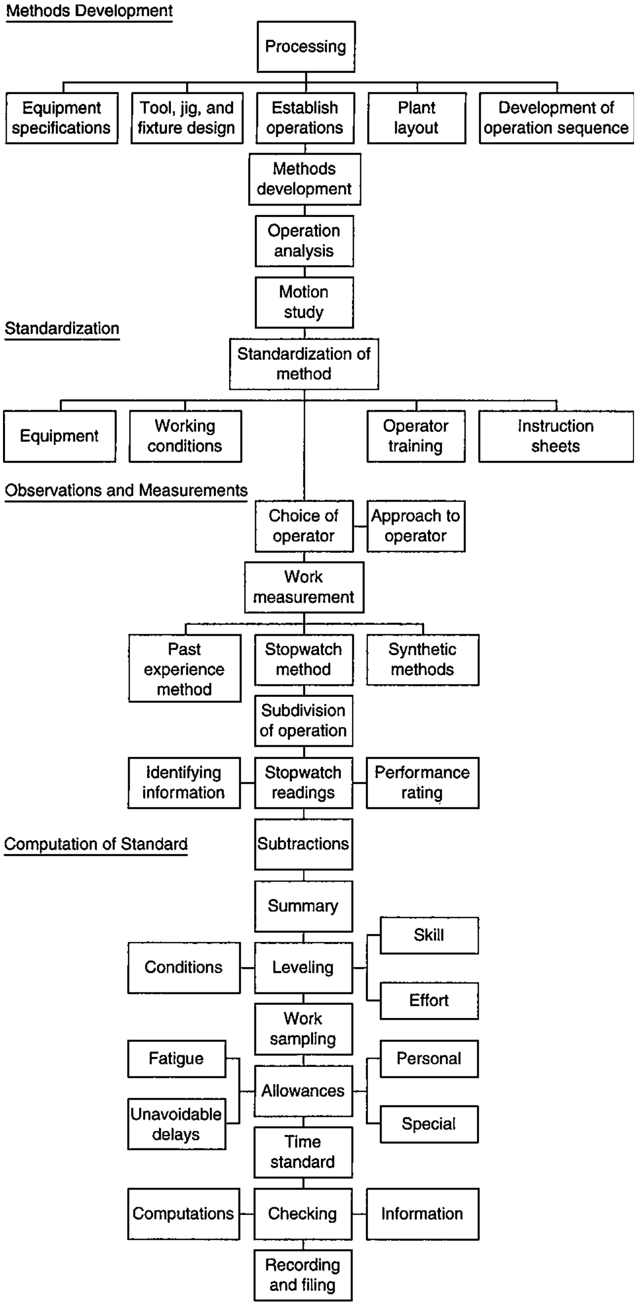


Fig. 17.4.2 Graphic analysis of the elements of motion and time study.

2. Design of part
   - *a.* Can motions be eliminated by design changes which will not affect the functioning and other desirable characteristics of the product?
   - *b.* Is the design satisfactory for automated assembly?
3. Complete survey of all operations performed on part
   - *a.* Can the operation being analyzed be eliminated by changing the procedure or the sequence of operations?
   - *b.* Can it be combined with another operation?
4. Inspection requirements
   - *a.* Are tolerance, allowance, finish, and other requirements necessary?

   - *b.* Will changing the requirements of a previous operation make this operation easier to perform?
5. Material
   - *a.* Is the material furnished in a suitable condition for use?
   - *b.* Is material utilized to best advantage during processing?
6. Material handling
   - *a.* Where should incoming and outgoing material be located with respect to the work station?
   - *b.* Can a progressive assembly line be set up?
7. Workplace layout, setup, and tool equipment
   - *a.* Does the workplace layout conform to the principles of motion economy?
   - *b.* Can the work be held in the machine by other means to better advantage?
8. Common possibilities for job improvement
   - *a.* Can "drop delivery" be used?
   - *b.* Can foot-operated mechanisms be used to free the hand for other work?
9. Working conditions
   - *a.* Has safety received due consideration?
   - *b.* Are new workers properly introduced to their surroundings, and are sufficient instructions given them?
10. Method
    - *a.* Is the repetitiveness of the job sufficient to justify more detailed motion study?
    - *b.* Should full automation be considered?

When the method has been developed, conditions are standardized, and the operators are trained to follow the approved method.

At this time, not before, the job is ready for time study. Suitable operators are selected, the purposes of the study are carefully explained to them, and the time study observations are made. During the study, time study specialists rate the performance being given by operators either by judging the skill and effort they are exhibiting or by assessing the speed with which motions are made as compared with what they consider to be a normal working pace. The final step is to compute the standard.

PRINCIPLES OF MOTION STUDY

Operation analysis is a primary analysis which eliminates inefficiencies. Motion study is a secondary analysis which refines the method still further. Motion study may and often does suggest further improvements in the factors considered during operation analyses, such as tools, material handling, design, and workplace layouts. In addition, it studies the human factors as well as the mechanical and sets up operations in conformance with the limitations, both physical and psychological, of those who must perform them.

The technique of motion study rests on the concept originally advanced by Frank B. and Lillian M. Gilbreth that all work is performed by using a relatively few basic operations in varying combinations and sequences. These Gilbreth Basic Elements have also been called "therbligs" and "basic divisions of accomplishment."

The basic elements together with their symbols (for definitions see ASME Industrial Engineering Terminology), grouped in accordance with their effect on accomplishment, are as follows:

Group 1 Accomplishes	
Reach	R
Move	M
Grasp	G
Position	P
Disengage	D
Release	RL
Examine	E
Do	DO

Group 2 Retards accomplishment	
Change direction	CD
Preposition	PP
Search	S
Select	SE
Plan	PL
Balancing delay	BD

Group 3 Does not accomplish	
Hold	H
Avoidable delay	AD
Unavoidable delay	UD
Rest to overcome fatigue	F

Group 1 is the useful group of basic elements or the ones that accomplish work. They do not necessarily accomplish it in the most effective way, however, and a study of these elements will often uncover possibilities for improvement.

Group 2 contains the basic elements that tend to retard accomplishment when present. In most cases, they do this by slowing down the group 1 basic elements. They should be eliminated wherever possible.

Group 3 is the nonaccomplishment group. The greatest improvements in method usually come from the elimination of the group 3 basic elements from the cycle. This is done by rearranging the motion sequence, by providing mechanical holding fixtures, and by improving the workplace layout.

An operation may be analyzed into its basic elements either by observation or by making a micromotion study of a motion picture of the operation.

Methods improvement may be made on any operation by eliminating insofar as possible the group 2 and group 3 basic elements and by arranging the workplace so that the group 1 basic elements are performed in the shortest reasonable time. In doing this, certain laws of motion economy are followed. The following, derived from the laws originally stated by the Gilbreths, are the most important.

1. When both hands begin and complete their motions simultaneously and are not idle during rest periods, maximum performance is approached.
2. When motions of the arms are made simultaneously in opposite directions over symmetrical paths, rhythm and automaticity develop most naturally.
3. The motion sequence which employs the fewest basic elements is the best for performing a given task.
4. When motions are confined to the lowest practical classification, maximum performance and minimum fatigue are approached. Motion classifications are: Class 1, finger motions; Class 2, finger and wrist motions; Class 3, finger, wrist, and forearm motions; Class 4, finger, wrist, forearm, and upper-arm motions; Class 5, finger, wrist, forearm, upper-arm, and body motions.

STANDARDIZING THE JOB

When an acceptable method has been devised, equipment, materials, and conditions must be standardized so that the method can always be followed. Information and records describing the standard method must be carefully made and preserved, for experience has shown that, unless this is done, minor variations creep in which may in time cause a major problem. In the case of repetitive work, a job is not standardized until each piece is delivered to operators in the same condition, and it is possible for them to perform their work on each piece by completing a set cycle of motions, doing a definite amount of work with the same equipment under uniform working conditions.

The operator or operators must then be taught to follow the approved method. Operator training is always important if reasonable production is to be obtained, but it is an absolute necessity where methods have been devised by motion study. It is quite apparent that the operators cannot be expected to discover for themselves the method which the time-study specialist developed as the result of hours of concentrated study. They must, therefore, be carefully trained if they are to be expected to reach standard production. In addition, an accurate time study cannot be made until the operator is following the approved method with reasonable proficiency.

WORK MEASUREMENT

Work measurement is the calculation of the amount of time it should take to do a standardized job. It utilizes the concept of a standard time. The standard time to perform a task is the agreed-upon and reproducible calculated time that a hypothetical typical person working at a normal rate of speed should take to do the job using the specified method with the proper tools and materials. It is the normal time determined to be required to complete one prescribed cycle of an operation, including noncyclic tasks, allowances, and unavoidable delays.

Time Study Methods There are several bases upon which time standards may be calculated. They include:

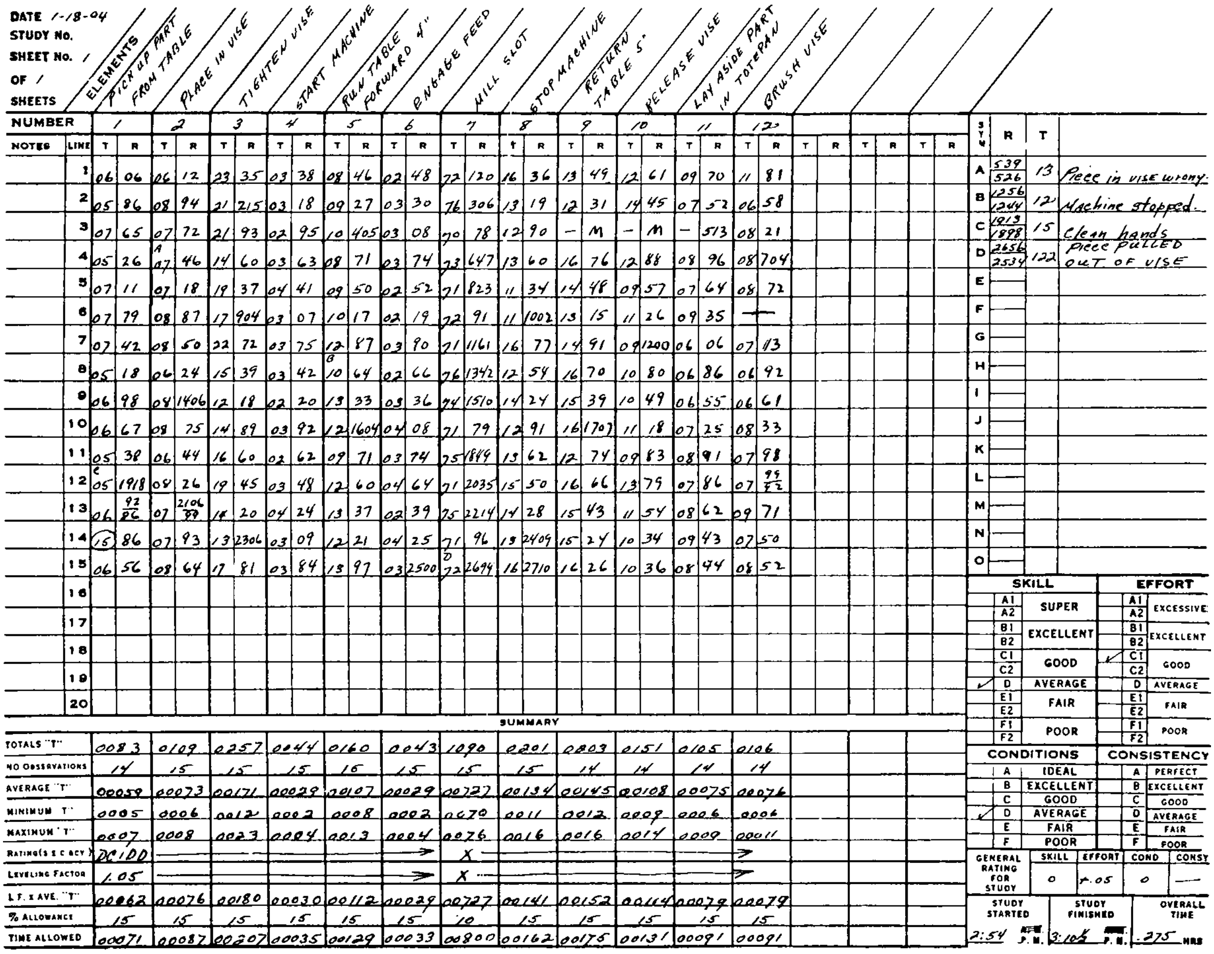
1. *Application of past experience.* The time required to do the operation in the past, either recorded or remembered, may be used as the present standard or as a basis for estimating a standard for a similar operation or the same operation being done under changed conditions.
2. *Direct observation and measurement.* The operation may be observed and its time recorded as it is actually performed and adjustments may be made to allow for the estimated pace rate of the operator and for special allowances. A stop watch or other recording instrument may be used or work sampling may be employed, which makes statistical inferences based upon random observations.
3. *Synthetic techniques.* A time standard for an actual or proposed operation may be constructed from the sum of the times to perform its several components. The times of the components are extracted from standard charts, tables, graphs, and formulas in manuals or in computer databases and totaled to arrive at the overall time for the entire operation. Standard data or predetermined motion times may be used. Methods-time measurement (MTM), basic-motion times (BMT), work factor (WF), and others are some of the systems available.

TIME STUDY OBSERVATIONS

The time study specialist can study any operator he or she wishes so long as that operator is using the accepted method. By applying what is known as the leveling procedure, the time study specialist should arrive at the same final time standard regardless of whether studying the fastest or the slowest worker.

The manner in which the operator is approached at the beginning of the study is important. This is particularly true if the operator is not accustomed to being studied. Time study specialists should be courteous and unassuming and should show a recognition of and a respect for the problems of the operator. They should be frank in their dealings with the operator and should be willing at any time to explain what they are doing and how they do it.

The first step in making time study observations is to subdivide the operation into a number of smaller operations which will be studied and timed separately. These subdivisions are known as elements, or elemental operations. Each element is exactly described in a few well-chosen words, which are recorded on the top of the time study form. Figure 17.4.3 shows how this is done. The beginning and ending points



DATE 1-18-04
STUDY No.
SHEET No. 1
OF 1
SHEETS

ELEMENTS		PICK UP PART FROM TABLE		PLACE IN VISE		TIGHTEN VISE		START MACHINE		RUN TABLE FORWARD 4"		ENGAGE FEED		MILL SLOT		STOP MACHINE		RETURN TABLE 5"		RELEASE VISE		LAY ASIDE PART IN TOTEPAN		BRUSH VISE							
NUMBER		1		2		3		4		5		6		7		8		9		10		11		12							
NOTES	LINE	T	R	T	R	T	R	T	R	T	R	T	R	T	R	T	R	T	R	T	R	T	R	T	R	T	R	T	R	T	R
	1	06	06	06	12	23	35	03	38	08	46	02	48	72	120	16	36	13	49	12	61	09	70	11	81						
	2	05	86	08	94	21	215	03	18	09	27	03	30	76	306	13	19	12	31	14	45	07	52	06	58						
	3	07	65	07	72	21	93	02	95	10	405	03	08	70	78	12	90	—	M	—	M	—	513	08	21						
	4	05	26	A 07	46	14	60	03	63	08	71	03	74	73	647	13	60	16	76	12	88	08	96	08	704						
	5	07	11	07	18	19	37	04	41	09	50	02	52	71	823	11	34	14	48	09	57	07	64	08	72						
	6	07	79	08	87	17	904	03	07	10	17	02	19	72	91	11	1002	13	15	11	26	09	35	—	—						
	7	07	42	08	50	22	72	03	75	12	87	03	90	71	1161	16	77	14	91	09	1200	06	06	07	13						
	8	05	18	06	24	15	39	03	42	B 10	64	02	66	76	1342	12	54	16	70	10	80	06	86	06	92						
	9	06	98	08	1406	12	18	02	20	13	33	03	36	74	1510	14	24	15	39	10	49	06	55	06	61						
	10	06	67	08	75	14	89	03	92	12	1604	04	08	71	79	12	91	16	1707	11	18	07	25	08	33						
	11	05	38	06	44	16	60	02	62	09	71	03	74	75	1849	13	62	12	74	09	83	08	91	07	98						
	12	C 05	1918	08	26	19	45	03	48	12	60	04	64	71	2035	15	50	16	66	13	79	07	86	07	99 92						
	13	06	92 86	07	2106 99	14	20	04	24	13	37	02	39	75	2214	14	28	15	43	11	54	08	62	09	71						
	14	⑮	86	07	93	13	2306	03	09	12	21	04	25	71	96	13	2409	15	24	10	34	09	43	07	50						
	15	06	56	08	64	17	81	03	84	13	97	03	2500	D 72	2694	16	2710	16	26	10	36	08	44	08	52						
	16																														
	17																														
	18																														
	19																														
	20																														

SUMMARY

TOTALS "T"	0083	0109	0257	0044	0160	0043	1090	0201	0803	0151	0105	0106		
NO OBSERVATIONS	14	15	15	15	16	15	15	15	14	14	14	14		
AVERAGE "T"	00059	00073	00171	00029	00107	00029	00727	00134	00145	00108	00075	00076		
MINIMUM "T"	0005	0006	0012	0002	0008	0002	0070	0011	0012	0009	0006	0006		
MAXIMUM "T"	0007	0008	0023	0004	0013	0004	0076	0016	0016	0014	0009	00011		
RATING(S E C ACY)	DC 100					→	X					→		
LEVELING FACTOR	1.05					→	X					→		
L.F. x AVE. "T"	00062	00076	00180	00030	00112	00029	00727	00141	00152	00114	00079	00079		
% ALLOWANCE	15	15	15	15	15	15	10	15	15	15	15	15		
TIME ALLOWED	00071	00087	00207	00035	00129	00033	00800	00162	00175	00131	00091	00091		

SYM	R	T	
A	539 526	13	Piece in vise wrong.
B	1256 1244	12	Machine stopped.
C	1913 1898	15	Clean hands
D	2656 2534	122	Piece pulled out of vise
E			
F			
G			
H			
I			
J			
K			
L			
M			
N			
O			

SKILL			EFFORT		
	A1 A2	SUPER		A1 A2	EXCESSIVE
	B1 B2	EXCELLENT		B1 B2	EXCELLENT
	C1 C2	GOOD	✓	C1 C2	GOOD
✓	D	AVERAGE		D	AVERAGE
	E1 E2	FAIR		E1 E2	FAIR
	F1 F2	POOR		F1 F2	POOR

CONDITIONS			CONSISTENCY		
	A	IDEAL		A	PERFECT
	B	EXCELLENT		B	EXCELLENT
	C	GOOD		C	GOOD
✓	D	AVERAGE		D	AVERAGE
	E	FAIR		E	FAIR
	F	POOR		F	POOR

GENERAL RATING FOR STUDY	SKILL	EFFORT	COND	CONSY
	0	+.05	0	—

STUDY STARTED	STUDY FINISHED	OVERALL TIME
2:54 P.M.	3:10½ P.M.	.275 HRS

Fig. 17.4.3 Face of a time study form.

of these elements must be clearly recognizable so that the chances of overlapping watch readings will be minimized.

The timing is done with the aid of a stopwatch, or less frequently, with a special type of "time study machine." There are several types of stopwatches as well as several methods of recording watch readings in common use. The study illustrated by Fig. 17.4.3 was made using a decimal-hour stopwatch that reads directly in ten-thousandths of an hour. The readings were recorded using what is known as the continuous method of recording. In this method, the watch runs continuously from the beginning of the study to the end. Thus every moment of time is accounted for, something that may be important if the correctness of the study is ever questioned. The watch is read at the end of each elemental operation, and the reading is recorded in the "R" column under the proper element description. The elapsed time for each element is later secured by subtracting successive readings. This observation procedure gives results as accurate as any other and more accurate than some.

Occasionally variations from the regular sequence of elemental operations occur. The time study specialist must be prepared to handle such situations when they happen. These variations may be divided into four general classes as follows: (1) elements performed out of order, (2) elements missed by the time study specialist, (3) elements omitted by the operator, (4) foreign elements.

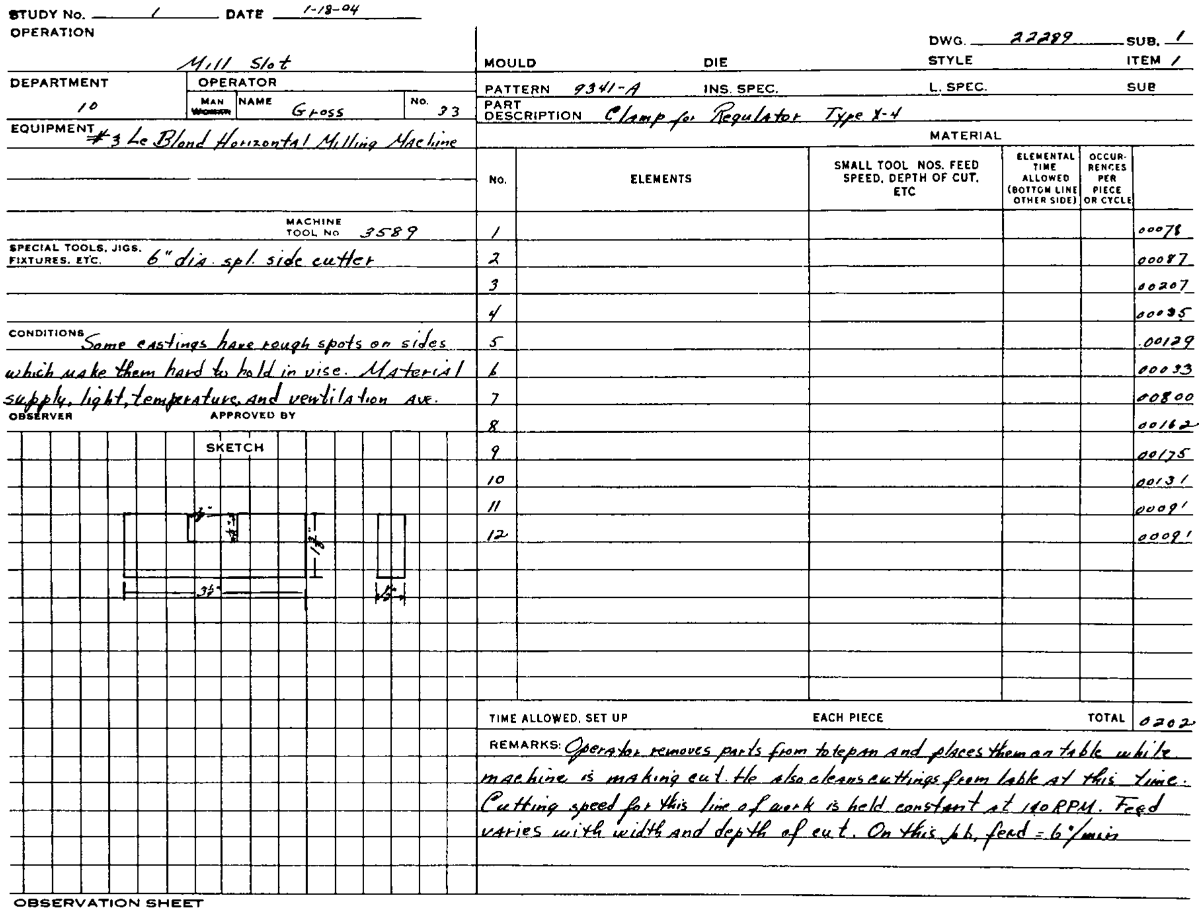
The time study illustrated by Fig. 17.4.3 contains examples of each of these kinds of irregularity. Elements 12 and 1 on lines 12 and 13 were performed out of order. On line 3, the time study specialist missed obtaining the watch readings for elements 9 and 10. On line 6, element 12 was omitted by the operator. Foreign elements A, B, C, and D occurred during regular elements 2, 5, 1, and 7, respectively. A study of these examples will show how the time study specialist handles variations from the regular sequence of elements which occur during the making of a time study.

A time study to be of value for future use must tell the whole story of a job in such a way that it will be understood by anyone familiar with the time study procedure. This will not be possible unless all identifying and other pertinent information is recorded at the time the study is made. Records should be made to show complete identification of the operator; the part or assembly; the machines, tools, and equipment used; the operation; the department in which the operation was performed; and the conditions existing at the time the study was made. Sketches are generally a desirable part of this description. Figure 17.4.4 shows the information which would be recorded on the reverse side of the time study form illustrated by Fig. 17.4.3.

PERFORMANCE RATING

The objective of a time study is to determine the time which a worker giving average performance will require to do the job under average or normal conditions. It is important to understand that when the time study specialist speaks of average performance, he or she is not referring to the mathematical average of all human beings, or even the average of all persons engaged in a given occupation. Average performance is established by definition and not statistically. It represents the time study specialist's conception of the normal, steady, but unhurried performance which may reasonably be expected from anyone qualified for the work. If sufficient inducement is offered by incentives or otherwise, this performance may be considerably surpassed.

If all operators available for study worked at the average performance level, the task of establishing a standard would be easy. It would



STUDY No. 1 DATE 1-18-04

OPERATION Mill Slot

DEPARTMENT 10

OPERATOR MAN ~~WOMAN~~ NAME Gross NO. 33

EQUIPMENT #3 Le Blond Horizontal Milling Machine

MACHINE TOOL No 3589

SPECIAL TOOLS, JIGS, FIXTURES, ETC. 6" dia. spl. side cutter

CONDITIONS Some castings have rough spots on sides which make them hard to hold in vise. Material supply, light, temperature, and ventilation ave.

OBSERVER

APPROVED BY

SKETCH

DWG. 22289 SUB. 1

MOULD DIE STYLE ITEM 1

PATTERN 9341-A INS. SPEC. L. SPEC. SUB

PART DESCRIPTION Clamp for Regulator Type X-4

MATERIAL

No.	ELEMENTS	SMALL TOOL NOS. FEED SPEED, DEPTH OF CUT, ETC	ELEMENTAL TIME ALLOWED (BOTTOM LINE OTHER SIDE)	OCCURRENCES PER PIECE OR CYCLE	
1					00078
2					00087
3					00207
4					00035
5					.00129
6					00033
7					00800
8					00162
9					00175
10					00131
11					00091
12					00091

TIME ALLOWED, SET UP EACH PIECE TOTAL 0202

REMARKS: Operator removes parts from totepan and places them on table while machine is making cut. He also cleans cuttings from table at this time. Cutting speed for this line of work is held constant at 140 RPM. Feed varies with width and depth of cut. On this job, feed = 6"/min

OBSERVATION SHEET

Fig. 17.4.4 Back of a time study form.

Skill			*Conditions*			*Effort*		
+0.15	A1	Superskill	+0.06	A	Ideal	+0.13	A1	Excessive
+0.13	A2		+0.04	B	Excellent	+0.12	A2	
+0.11	B1	Excellent	+0.02	C	Good	+0.10	B1	Excellent
+0.08	B2		0.00	D	Average	+0.08	B2	
+0.06	C1	Good	−0.03	E	Fair	+0.05	C1	Good
+0.03	C2		−0.07	F	Poor	+0.02	C2	
0.00	D	Average				0.00	D	Average
−0.05	E1	Fair				−0.04	E1	Fair
−0.10	E2					−0.08	E2	
−0.16	F1	Poor				−0.12	F1	Poor
−0.22	F2					−0.17	F2	

Fig. 17.4.5 Leveling factors for performance rating.

be necessary merely to average the elapsed elemental times determined from time study and add an allowance for fatigue and personal and unavoidable delays. It is seldom, however, that a performance is observed which is rated throughout as average. Therefore, to establish a standard which represents the time which would be taken had an average performance been observed, it is necessary to use some method of adjusting the recorded elemental times when other than average performance is timed.

One of the well-known methods of doing this is the **leveling procedure.** When properly applied it gives excellent results. It must be correctly understood, of course, and the time study specialist who uses it must be thoroughly trained to apply it correctly.

The procedure recognizes that when the correct method is being followed, skill, effort, and working conditions will affect the level at which the operator works. These factors are judged during the making of the time study. Skill is defined as *proficiency at following a given method.* This is not subject to variation at will by the operator but develops with practice over a period of time. Effort is defined as *the will to work.* It is controllable by the operator within the limits imposed by skill. Conditions are those conditions which affect the operator and not those which affect the method.

Definitions have been established for different degrees of skill, effort, and conditions. Numerical factors have been established by extensive research for each degree of skill, effort, and conditions. These are shown by Fig. 17.4.5. The algebraic sum of these numerical values added to 1.0 gives the leveling factor by which all actual elemental times are multiplied to bring them to the average or normal level. The leveling factor represents in effect the amount in percent which actual performance times are above and below the average performance level.

ALLOWANCES FOR FATIGUE AND PERSONAL AND UNAVOIDABLE DELAYS

The leveled elemental time values are net elapsed times adjusted to the average performance level. They do not provide for delays and other legitimate allowances. Something, therefore, must be added to take care of such things as fatigue, and special conditions of the work.

Fatigue allowances vary according to the nature of the work. Flat percentages are determined for each general class of work, such as bench work, machine-tool operation, hard physical labor, and so on. Personal allowances are the same for most classes of work. Unavoidable delay allowances vary with the nature of the work and the conditions under which it is performed. Peculiar conditions surrounding specific jobs sometimes require additional special allowances.

It is apparent, therefore, that the proper allowance factor to use can only be determined by a study of the class of work to which it is to be applied. Allowances are determined either by a series of all-day time studies or by a statistical method known as work sampling, or both. When an allowance factor has once been established, it is then applied to all time studies made on that class of work thereafter.

DEVELOPING THE TIME STANDARD

When time study observations have been completed, a series of calculations are made to develop the time standard. Elapsed times are determined by subtracting successive watch readings. Each subtraction is recorded between the two watch readings that determine its value. Elapsed time is noted in ink to ensure a permanent record and to distinguish it from the watch readings which are usually recorded in pencil. A study of Fig. 17.4.3 will show how subtractions are entered on the time study form and later summarized.

The several elapsed times for each element are next carefully compared and examined for abnormal values. If any are found, they are circled so that they can be distinguished and excluded from the summary.

The remaining elapsed times for each element are added and are averaged by dividing by the number of elapsed time readings. The results are average elapsed times which represent the time taken by the operator during that particular study. These times must be adjusted by multiplying them by a leveling factor to bring them to the average performance level. This factor is determined by the rating of skill, effort, and conditions made during the period of observation.

Each average elapsed time is multiplied by the leveling factor, except when the element is not controlled by the operator. An element that is outside the control of the operator, such as element 7 in Fig. 17.4.3 which is a cut with power feed, should not be leveled, because it is unaffected by the ability of the operator. As long as the proper feed and speed are used, the time for performing this element will be the same whether the worker is an expert or a learner.

If workers were able to work continuously, the leveled time would be the correct value to allow for doing the operation studied, but constant application to the job is neither possible nor desirable. In the course of a day, there are certain to be occasional interruptions and delays, for which due allowance must be made in establishing the final standard. Therefore, each elemental time is increased by an allowance which covers time that will be consumed by personal and unavoidable delays, fatigue, and any special factors that may affect the job.

The numbers and descriptions of the elemental operations together with their allowed time are transcribed on the back of the time study form as shown by Fig. 17.4.4. The number of times an elemental operations occurs on each piece or cycle of the operation is taken into account, and the total time allowed for each element is determined and recorded. The final standard for the operation is the sum of the amounts recorded in the “time-allowed” column. When all computations have been checked and all supporting records have been properly identified and filed, the task of developing the time standard is complete.

TIME FORMULAS AND STANDARD DATA

On repetitive work, time study is a satisfactory tool of work measurement. A single time study may be sufficient to establish a standard which will cover the work of one or more operators for a long period of time.

As quantities become smaller, however, the cost of establishing standards by individual time study increases until at length it becomes prohibitive. In the extreme case, where products are manufactured in quantities of one, it would require at least one time study specialist for each operator if standards were established by detailed time study, and the standards would not be available until after the jobs had been completed.

In order to simplify the task of setting standards on a given class of work and in order to improve the consistency of the standards, standard data are frequently used by time study specialists. A compilation of standard data in its simplest form is merely a list of all the different elements that have occurred during all the time studies made on a given class of work, with representative time values for each element. Every element that differs even slightly from any other element has its own time value.

When a job comes into the shop on which no standard has previously been established, time study specialists analyze the job either mentally or by direct observation and determine the elements required to perform it. They then select time values from the standard data for each element. Their sum gives the standard for the job.

This method, although a decided improvement from a time, cost, and consistency standpoint over individual time study, is capable of further refinement and improvement. On a given class of work, certain elements will be performed—for example, "pick up part"—on every piece produced, while others—such as "secure in steady rest"—will be performed only when a piece has certain characteristics. In some cases, the performance of a certain element will always require the performance of another element, e.g., "start machine" will always require the subsequent performance of the element "stop machine." Then again, the time for performing certain elements—for example, "engage feed"—will be the same regardless of the characteristics of the part being worked upon, while the time for performing certain other elements—like "lay part aside"—will be affected by the size and shape of the part.

Thus it is possible to make certain combinations and groupings which will simplify the task of applying standard data. Time study specialists construct various charts, and tables which they still call standard data, or, in the ultimate refinement, develop time formulas. A time formula is a convenient arrangement of standard data which simplifies their accurate application. Much of the analysis which is necessary when applying standard data is done once and for all at the time the formula is derived. The job characteristics which make the performance of certain elements or groups of elements necessary are determined, and the formula is expressed in terms of these characteristics.

Figure 17.4.3 illustrates a detailed time study made to establish a standard on a simple milling machine operation. The same standard can be derived much more quickly from the following time formula:

$$\text{Curve } A + \text{Table } T = \text{each piece time}$$

Curve *A* combines the times for the variable elements "pick up part from table," "place in vise," and "lay aside part in totepan" with the times for the constant elements "tighten vise," "start machine," "run table forward 4 in," "engage feed," "stop machine," "release vise," and "brush vise." Table *T* combines the times for "mill slot" and "return table." The standard time for milling a slot in a brass clamp of any size is computed by determining the variable characteristics of the job from the drawing—in this case, the volume of the clamp and the perimeter of the cut—and adding together the time read from curve *A* and the time read from Table *T.*

The amount of time which the use of time formulas will save the time study specialist is readily apparent. It takes a certain amount of time and no little know-how to develop a time formula, but once it is available, the job of establishing accurate standards becomes a simple, fairly routine task. The time required to make and work up a time study will be from 1 to 100 or more hours, depending upon the length of the operation cycle studied. The time required to establish a standard from a time formula will, in the majority of cases, range from 1 to 15 min, depending upon the complexity of the formula and the amount of time required to determine the characteristics of the job. Where all the necessary information may be obtained from the drawing of the part, the standard may generally be computed in less than 5 min.

USES OF TIME STANDARDS

Some of the more common uses of time standards are in connection with

1. Wage incentive plans
2. Plant layout
3. Plant capacity studies
4. Production planning and control
5. Standard costs
6. Budgetary control
7. Cost reduction activities
8. Product design
9. Tool design
10. Top-management controls
11. Equipment selection
12. Bidding for new business
13. Machine loading
14. Effective labor utilization
15. Material-handling studies

Time standards can be established not only for direct labor operations but also for indirect work, such as maintenance and repair, inspection, office and clerical operations, engineering, and management. They can also be set for machinery and equipment, including robots.

17.5 COST OF ELECTRIC POWER

by Andrew M. Donaldson and Robert F. Gambon

REFERENCES: Department of Energy (DOE), "Electric Plant Cost and Power Production Expenses," 1991. Electric Power Research Institute, "Technical Assessment Guide," 1993. Oak Ridge National Laboratory, "CONCEPT-V, A Computer Code for Conceptual Cost Estimates of Steam Electric Power Plants." "Handy-Whitman Index of Public Utility Construction Costs," Whitman, Requardt and Associates. Grant and Ireson, "Principles of Engineering Economy," Ronald Press.

In simplest terms, the cost of electric power is the cost of the initial energy source plus the cost of converting that energy into electricity (generation), plus the cost of delivering that electricity to the consumer (transmission and distribution).

That initial energy source is the potential and kinetic energy of hydroelectric power; the chemical energy of fossil, biomass, and waste power; or the atomic energy of nuclear power. In addition, there are other energy sources for the conversion to electricity. These other sources include energy derived from solar radiation, wind, and ocean waves, but these sources are limited, inefficient, and currently not cost-effective without subsidy. These alternative generation approaches are not addressed herein.

The cost of the initial energy and its conversion to electricity constitute generation and is the major factor in electric power cost and is the primary concern of this section. The largest contributor to electricity cost of fossil fuel plants is delivered fuel followed by plant capital cost amortization, and operations personnel and maintenance expenses. The higher the fuel cost and the lower the plant efficiency, the greater the effect of fuel cost on the final electric price. For natural gas, which is generally the most expensive fossil fuel, a combination of lower plant capital cost and/or higher plant efficiency is necessary for cost competitiveness. Electricity production with natural gas in a 33 percent efficient Rankine cycle steam boiler and turbine makes little sense today because the same fuel can be used in a large-frame combustion turbine combined-cycle plant and approaches 60 percent efficiency. The traditionally lower delivered cost of coal allows it to be used economically in the Rankine cycle. However, for many recent years, low gas prices and restrictive environmental regulations combined with the high efficiency and lower construction cost of combustion turbine combined-cycle plants have almost stopped coal plant construction. Only larger electricity producers conscious of the need for fuel diversity in their generation base have even considered coal in the last 5 to 10 years.

Environmental regulations have severely restricted coal use, but improving emission control equipment and increasing natural gas prices are likely to result in a resurgence of coal plants in the early twenty-first century.

For hydroelectric, geothermal, nuclear, and waste fuel plants, the highest cost contributor to electricity price is paying for the capital investment that is the power plant, followed by operations and maintenance.

Of course, economics have been overshadowed by public opinion and politics when it comes to nuclear and large hydroelectric plants, which have not been constructed in the United States in 20 years.

Generation costs consist of production expenses (fuel plus operating and maintenance expenses) and fixed charges on investment (cost of investment capital, depreciation or amortization, taxes, and insurance).

Interconnected power supply systems must price power to include transmission costs, distribution costs, and commercial expenses. Power price savings are realized by scheduling the installation and operation of a range of plant types to optimize overall generation costs. Generally, efficient plants burning lowest-cost fuels or operating on river water flows that are available the year round are assigned to continuous baseload operation at or near full capacity. Plants bearing low unit-kilowatt investment costs and lower efficiencies typically are installed to provide peaking capacity. Operation of these units during short daily periods of peak load limits their energy output and thus minimizes characteristic cost penalties associated with their poorer station efficiency or their requirement for higher-priced fuels. These higher unit cost per kilowatt but higher energy density **peaking plants** supply peak energy needs but are idle most of the time. This approach is preferable to building high capital cost but more efficient central stations that would sit idle most of the time until load growth catches up to capacity.

Power generating stations also participate in regional **transmission grid** systems supplementing capacity and reserve needs of neighboring systems, and by the daily and seasonal exchange of off-peak, low-incremental-cost energy. Historically, there have been significant regional differences in the price of power. These differences have reflected such factors as availability of hydroelectric energy, the cost of fossil fuels, labor costs, ownnership type and investment composition, local tax structure, and the opportunities pooling and coordination provide to exploit the economies of scale. This exchange of capacity and reserves to serve neighboring grid and spread the benefits of economies of scale, combined with regional power cost differences, led to the advent of state-by-state deregulation of the power industry, and proliferation of “nonutility” generators.

The early industrial areas of this country in the northern Midwest and New England areas were increasingly served by aging and decreasingly efficient power plants, often burning oil or requiring long-distance coal deliveries. These fuels have shown volatility on the world market and escalating delivery costs. The flight of heavy power-demanding industry from these areas allowed residential and light industrial growth to be picked up by the aging plants without new, more efficient capacity being added. The population density increased but the same old plants were kept in operation, often past their original life expectancy.

Since the early 1980s regulators and well-intentioned environmentalists have opposed nearly every new power plant. Existing plants, with their higher cost of replacement, were under no replacement pressure, since installing new, efficient, cost-effective power plants was delayed because of this environment. Regulators that refused requests from utilities for price increases, which would have allowed eventual new construction, further exacerbated the situation, preserving inefficiency.

Regional retail power price differences like 12 cents per kilowatt in areas of New York versus 6 cents per kilowatt in Georgia contributed to industry and job flight away from areas with high-priced electricity. As a result, business continues to look for lower electric rates, while politicians search for opportunities to keep jobs and voters in the area. Out of this dilemma, the idea of deregulation, in search of lower-cost electric supplies, was born. Opening some markets produced temporary downward pressure on prices even in regulated areas but also produced chaos and threatened the reliability of the national electric system that powers the U.S. economy.

The deregulation promises of lower prices were often temporary, as were the companies offering the lower prices. The California deregulation situation, with tremendous electricity price swings and state-assisted bankruptcy of one of California’s largest utilities, and the Enron scandal exposed in 2001, have severely reduced the number of nonutility generators and the likelihood of deregulation. However, some states, such as Pennsylvania, have had some success in the deregulation of electricity. The electrical supply, distribution, reliability, and price chaos of the early twentieth century electrical system, which led to utility regulation, was just repeated for entry into the twenty-first century.

This section, while aware of the short-term volatility resulting from regulation, deregulation, and politics will concentrate on the long-term aspects affecting the cost of power.

Cost or price is a basic characteristic of power supply. Along with relative abundance, reliability, and high quality of service, low prices have encouraged the widespread use of electric energy that has come to be associated with our national way of life. Industrial use of electric energy is particularly heavy in the electroprocess and metallurgical industries. The price of electricity has a significant effect on the end-product cost in these industries. In many manufacturing and process industries, quality of service in terms of voltage regulation, frequency control, and reliability is a major concern. Uninterrupted supply is of crucial importance in many process industries where a power failure causes material waste or damage to equipment, in addition to loss of production revenue. Because of the flexibility and convenience of electric power, future increases in industrial, residential, and commercial consumption are anticipated despite deregulation, fuel price increases, environmental considerations, and energy conservation efforts.

Various types of power-generating units are owned and operated by industry and private sector investors to provide power (and thermal energy) to industrial facilities and often sell excess power to the local utility at the avoided cost of incremental power for the utility. Inside-the-fence cogenerators at paper mills and other industrial facilities that are electrical and thermal consumers but provide only a fraction of their electric generation to the grid are not considered in this section. However, the calculations and information in this section can assist in deciding the advisability of purchasing versus generating electric power. Larger cogenerating facilities built under the “qualifying facility” (QF) sections of the **Public Utilities Regulatory Policies Act (PURPA)** that have electricity production for wholesale to the grid or local utility as their main objective are included in this section. With the PURPA regulations no longer restricting the development of **nonutility generators (NUGs),** a significant portion of new generation in the 1990s was provided by NUGs. Of course, many “nonutility generators” had regulated utility parents.

CONSTRUCTED PLANT COSTS

A **central station** serving a transmission system is designed to meet not only the existing and prospective loads of the system in which it is to function but also the pooling and integration obligations to adjacent systems. Service requirements which establish station size, type, location, and design characteristics ultimately affect cost of delivered power. Selection of plant type and the overall philosophy followed in design must accommodate a combination of objectives which may include high operating efficiency, minimum investment, high reliability and availability, maximum reserve capability margins, rapid load change capability, quick-start capability, or service adaptability as spinning reserve.

For a given plant, the design must account for siting factors such as environmental impacts, subsoil conditions, local meteorology and air quality, quality and quantity of available water supply, access for construction, transmission intertie, fuel delivery and storage, and maintainability. In addition, plant siting and design will be significantly affected by legal restrictions on effluents which may have adverse impacts on the environment. In the case of **nuclear plants,** siting must also consider the proximity of population centers and the size of exclusion areas. Reactor plant design must bear the investments required to control radioactive releases and to provide safeguard systems which protect against accidents. No new nuclear power plants are contemplated to begin the licensing process in the United States until 2010 or beyond. **Fossil-fueled power plants** may be sited adjacent to fuel supplies or in proximity to load centers, thereby increasing transmission costs on the one hand or fuel delivery charges on the other. Depending primarily on climate, plants may be enclosed, semienclosed, or of the outdoor type. Spare auxiliary components can be installed to improve reliability. Increased investments in sophisticated heat cycles and controls, for improved equipment performance, can achieve higher plant efficiencies. **Combustion-turbine, simple and combined-cycle plant** outputs are restricted by both increased elevation and high ambient air temperatures. Hydroelectric sites are frequently very distant from load centers and thus require added costs for extensive transmission facilities. Also, **hydroelectric facilities** may provide for flood control, navigation, or recreation as by-products of power production. In such instances, total cost should be properly allocated to the various product elements of the multipurpose project.

An **industrial power plant** provided to meet the requirements of an isolated load entails design considerations and exhibits cost characteristics which differ from those of a central station power plant assigned an integrated role within a connected generating system. Industrial power plants often produce both thermal and electric power. Industrial facilities must often accommodate both base- and peak-load requirements. They may be designed to provide for on-site reserve capacity or spinning reserve capacity. Frequency control and voltage regulation must be viewed as a special problem because of the limited capability of a single plant to meet load changes.

Table 17.5.1 provides typical installed cost data for central station generating plants. The figures represent costs of facilities in place, excluding interest during construction. The cost of land, waste-disposal facilities, and fuel is not included. Costs apply to plants completed in 2003. Interest during construction can be estimated by multiplying the simple interest rate per year by the construction period in years and dividing by 2 to reflect the carrying costs on the average commitment of capital toward equipment and labor during construction. Escalation effects for plants to be completed beyond this date may be extrapolated in accordance with anticipated cost trends for labor, material, and equipment. Historical cost trends, by region, as experienced in the power industry, which can be helpful in forecasting future costs, may be determined by use of the figures in Table 17.5.2. Escalation can significantly affect plant costs on future projects, especially in view of the 10- to 12-year engineering and construction periods historically experienced for nuclear facilities and the corresponding 5 to 6 years required for fossil-fueled power plants. In addition to rising equipment, construction labor, and material costs, major factors influencing the upward trend in plant costs include increased investment in environmental control systems, an emphasis on improved quality assurance and plant reliability, and a concern for safety, particularly in the nuclear field.

Table 17.5.1 Typical Investments Costs* (2003 Price Level)

Plant description			
Type	Net capacity, MW	Fuel	Total investment cost in $/net kW
Combustion turbine			
Simple cycle	200	Gas	250–400
Combined cycle	500	Gas	400–600
Conventional steam plant			
Fossil	500	Coal	1,200–1,500
	500	Oil	1,100–1,300
	500	Gas	900–1,000

* Capital investments exclude costs for the following: initial fuel supply, cost of decommissioning for nuclear plants, main transformers, switchyard, transmission facilities, waste disposal, land and land rights, and interest during construction.

Conventional Steam-Electric Plants Conventional fossil plant investment costs given in Table 17.5.1 are for 500-MW nominal units which are deemed to be representative of future central station fossil units. Costs will vary from those in the table due to equipment arrangement, pollution control systems, foundations, and cooling-water-system designs dictated by plant site conditions. The variety of cycle arrangements and steam conditions selected also affects plant capital cost. Plant designers try to economically balance investment and operating costs for each plant. Present-day parameters, in the face of current economics, call for a drum boiler, regenerative reheat cycles at initial steam pressures of 2,400 psig with superheat and reheat temperatures of 1,000°F (539°C). Similar temperature levels are employed for 3,500-psig initial steam pressure supercritical-reheat and double-reheat cycles which require higher investment outlays in exchange for efficiency or heat rate improvements of between 5 and 10 percent. Increased investment costs are also caused by more generous boiler-furnace sizing, and larger fuel and ash-handling, precipitator, and scrubber facilities required for burning poor quality coals. Investment increments are also required to provide partial enclosure of the turbine building and furnace structure and to fully enclose the boiler by providing extended housing to weatherprotect duct work and breeching.

Nuclear Plants The construction of nuclear power plant capacity in the United States has been suspended for over 20 years. A new generation of smaller light-water reactor nuclear power plants is currently in the conceptual design stage. Costs associated with these units, including siting, licensing, and fuel cycle, remain unclear, as they are precommercial.

In general, light-water reactor nuclear plants require considerably higher investments than do fossil-fueled plants, reflecting the need for leakproof reactor pressure containment structures, radiation shielding, and a host of reactor plant safety-related devices and redundant equipment. Also, light-water reactor plants operate at lower initial steam conditions than do fossil-fueled plants and, because of their poorer turbine

Table 17.5.2 U.S. Cost Trends of Electric Plant Construction by Region (1973 Index is 100)

	North Atlantic	South Atlantic	North Central	South Central	Plateau	Pacific
	Total: Steam generating plants*					
1993	329	308	318	304	317	333
1994	341	321	331	319	331	347
1995	354	332	345	329	340	357
1996	361	339	352	333	346	366
1997	373	350	362	347	357	376
1998	380	357	368	351	365	382
1999	385	361	374	355	369	388
2000	396	372	386	365	379	397
2001	420	391	404	391	394	419
2002	431	397	417	395	400	424
2003	445	412	438	412	423	445

*As of January 1 of each year.

SOURCE: “Handy-Whitman Index of Public Utility Construction Costs,” compiled and published by Whitman, Requardt & Associates, LLP, 801 Caroline St., Baltimore, MD 21231.

cycle efficiency, require larger steam flows and increased equipment sizes at added investment.

Combustion Turbine/Combined-Cycle Plants Simple and combined-cycle plants are offered by a number of vendors in the 100- to 1000-MW size range. These plants consist of multiple installations of combustion gas turbines arranged to exhaust to waste-heat steam generators which may be equipped for supplementary firing of fuel. Steam produced is supplied to a conventional steam turbine cycle. Advantages of combustion turbine–based plants are lower unit investment costs, efficient thermal performance, increased flexibility (which allows independent operation of the gas-turbine portion of the plant), shorter installation schedules, reduced cooling-water requirements, and the reduction in sulfur oxide and particulate emissions characteristic of gaseous fuels. The disadvantages of combustion turbines are the high costs of fuel and combustion turbine maintenance.

Hydroelectric Plants Hydroelectric generation offers unique advantages. Fuel, a heavy contributor to thermal plant operating costs, is eliminated. Also, hydro facilities last longer than do other plant types; thus they carry lower depreciation rates. They have lower maintenance and operating expenses, eliminate air and thermal discharges, and because of their relatively simple design, exhibit attractive availability and forced-outage rates. Quick-start capability and rapid response to load change ideally suit hydro turbines to spinning reserve and frequency-control assignments.

The constructed cost of a hydroelectric station is strongly site dependent. Overall costs fluctuate significantly with variations in dam costs, intake and discharge system requirements, pondage required to firm up capacity, and with the cost of relocating facilities within the areas inundated by the impoundments. For a given investment in structures, available head and flow quantity may vary considerably, resulting in a wide range of outputs and unit investment costs. Installed plant costs are competitive with other facilities. Cost prediction for future hydroelectric construction is difficult, particularly in view of the decreasing availability of economical sites and restrictions imposed by concern for the ecological and social consequences of disrupting the natural flow patterns of rivers and streams. As with nuclear, no new large hydroelectric plants have been constructed in over 20 years. A number of dams and associated power plants have been demolished or otherwise taken out of service to save fish or restore waterways to some past condition. Any new capacity in this area has come from upgrading existing hydroturbines for greater output and efficiency.

Pumped-Storage Plants Pumped-storage plants involve a special application of hydroelectric generation, allowing the use of off-peak energy supplied at incremental charges by low-operating-cost thermal stations to elevate and store water for the daily generation of energy during peak-load hours. Pumped hydro projects must justify the inefficiencies of storage pumping and hydroelectric reconversion of off-peak thermal plant energy by investment cost savings over competing peaking plants. Installation of a pumped hydro station calls for a suitable high head site which minimizes required water storage and upper and lower reservoir areas and an available makeup source to supply the evaporative losses of the closed hydraulic loop. Despite the added complications of installing both pumping and generating units, or of utilizing reversible motor-generator pump-turbines, costs for pumped hydro stations generally fall below those for conventional hydroelectric stations. Differences in gross head, impoundment, and siting make plant cost comparisons difficult.

Geothermal Plants Geothermal generation utilizes the earth's heat by extracting it from steam or hot water found within the earth's crust. Prevalent in geological formations underlying the western United States and the Gulf of Mexico, geothermal energy is predominantly unexploited, but it is receiving increased attention in view of escalating demands on limited worldwide fossil-fuel supplies. Because natural geothermal heat supplants fuel, the atmospheric release of combustion products is eliminated. Nevertheless, noxious gases and chemical residues, usually contained in geothermal steam and hot water, must be treated when geothemral resources are tapped. There is a current lack of significant cost data covering geothermal plants. Because boiler and associated fuel-handling facilities are eliminated, investment in these generating plants is considerably less than the cost of comparable fossil-fueled units. However, overall investment chargeable to geothermal facilities includes significant exploration and drilling costs which are site dependent and cannot be accurately predicted without extensive geophysical investigation.

Environmental Considerations Environmental protection has become a dominant factor in the siting and design of new power generating stations. Both stack emissions to the atmosphere and thermal discharges to natural water courses must be significantly reduced in order to meet increasingly stringent environmental criteria. In many cases older plants are being required to reduce emission levels to achieve legislated ambient air-quality standards and to control thermal discharges by the use of closed cooling systems to prevent aquatic thermal pollution.

Control of **air pollution** in fossil-fueled power plants includes the reduction of particulates, sulfur oxides, and nitrogen oxides in flue gas emissions. Particulate collection can be achieved by electrostatic precipitators, baghouse filters, or as part of stack gas scrubbing. Stack gas scrubbing is required for all new coal-fired power plants, regardless of sulfur content of the fuel. Fossil-fueled power plants are believed to be a major contributor to acid rain.

Scrubbers reduce sulfur oxide emissions by contacting flue gas with a sorbate composed of metal (usually sodium, magnesium, or calcium) hydroxides in solution which act as bases to produce sulfate and sulfite precipitates when they contact sulfur oxides in the flue gas. Wet scrabbers contact flue gas with a sorbate in solution. Dry scrubbers evaporate sorbate solution into the gas stream.

Fossil-fueled power plants are required to burn low-sulfur fuels. Restrictions on the use of oil in new power plants and its high cost have virtually eliminated new central station steam power plants designed to utilize oil.

Control of nitrogen oxides NO_x is attained primarily through modifications to flame propagation and the combustion processes.

NO_x control for combustion turbines is typically staged combustion. New burner designs and the use of water or steam injection into the combustion zone have effectively reduced emissions to acceptable rates. For steam boilers, changes to the burners and combustion zone were deemed insufficient and introduction of ammonia or urea spray has been incorporated with or without a catalyst [selective catalytic reduction (SCR) and selective noncatalytic reduction (SNCR), respectively]. The two systems, SCR and SNCR, operate at different temperatures. The SCR system includes ammonia injection across the flue gas in an area of the gas path at approximately 700°F. The flue gas then passes through a vanadium pentoxide (V_2O_5) catalyst. This combination removes NO_x with a removal efficiency of approximately 90 percent. SNCR operates at approximately 1600 to 2100°F but does not require a catalyst. Its removal efficiency is approximately 70 percent.

In order to avoid plant discharges of waste heat to the aquatic environment, evaporative-type closed-cooling cycles are employed in lieu of once-through cooling system designs. Closed-loop cooling systems in current use employ evaporative **cooling towers** or **cooling ponds.**

More advanced closed-cooling-loop designs use dry and wet-dry cooling towers. These represent alternates to conventional evaporative systems where makeup water is in short supply or where visible vapor plumes or ice formed by vapor discharge present hazards. The penalties for dry-tower cooling are significantly higher than those for conventional evaporative designs. Large, more costly water-to-air heat-transfer surfaces are required, and characteristically higher condensing temperatures result in higher turbine backpressure, restricting plant capability and reducing efficiency.

Evaporative cooling systems (wet towers) have a significant advantage over hybrid (wet and dry sections) and all-dry cooling systems in initial installed and operating costs. A dry cooling tower system can be \$200 to \$400 per kW more expensive. For a large facility the operating penalty can be \$1,000 per kW or more because of the additional fan power requirements. Hybrid systems, which are a combination of an evaporative section and a dry heat exchanger section, are often used in areas where water usage is a concern or fogging issues are present. The wet/dry hybrid arrangement utilizes the required amount of wet and dry surface to meet the site constraints. In the Northeast, many smaller facilities (under 100 MW) have used the wet/dry arrangement to minimize fogging or icing of nearby highways. In the western half of the United States, wet/dry towers have been used to conserve the precious and sometimes scarce water resources. Another consideration for the addition of any dry surface beyond the typical evaporative cooling system is the effect on efficiency. Use of a dry cooling system can easily increase the condensing temperature for the Rankine cycle by 20 to 30°F. The efficiency and output penalty of the increase in heat rejection temperature can approach 5 to 10 percent, which can reduce overall efficiency by several percent or more.

FIXED CHARGES

Costs that are established by the amount of capital investment in plant and which are fixed regardless of production level are termed **fixed charges.** Annual fixed charges are ordinarily expressed as a percentage of investment and include interest or the cost of money, funds applied to amortize investment or to allow for replacement of depreciated plant, and charges covering property taxes and insurance. Additionally, fixed charges may include an interim replacement allowance to cover the replacement cost of plant equipment not expected to last the full life of the plant.

For investor-owned utilities, the **cost of money** employed for plant expansion depends upon financial market conditions in general and upon the attitude of investors with regard to a particular utility enterprise or specific project. Funds for investor-owned utility expansion are derived from both the risk capital (equity) and debt capital (bond) markets. Prevailing return rates for investments in utility plant facilities are influenced by the rate-setting practices of public utility regulatory agencies and by supply-and-demand factors in the investment market.

Public utility facilities owned by state and municipal government organizations are generally financed by long-term revenue bonds, most of which qualify for tax-free-income status. Interest rates currently fall between 4 and 6 percent and reflect the tax relief on interest income enjoyed by the bondholders.

Generating facilities are often financed by industrial concerns whose primary business is the production of a manufactured product. In these instances, the annual cost of money invested in power facilities will be established by considering alternate investment of the required funds in manufacturing plant. Inasmuch as returns on equity capital invested in manufacturing industries are usually on the order of 10 to 20 percent, the rate of return for industrial power plants will tend to be set at these higher levels.

Nonutility Generators (NUGs) generally form a project-specific limited liability corporation, or LLC, and borrow through nonrecourse financing at or above market rates. They put up a minimum of capital. In the early days of NUGs, the project would get equity contributions from the equipment vendors and contractors to cover the equity required by the banks as "down payment." The plant and its revenue stream would collateralize the balance. Later revenue from earlier operating plants and real equity was required to satisfy the loan. As we enter the twenty-first century, project financing has seen numerous project defaults and the decline of many independent NUGs and utility-spawned NUGs. Equity requirements for new projects have increased dramatically, as well as the expected interest rates. Lenders expect multiple layers of guarantee from all participants. The short-lived electricity trading market has shrunk and the expectation of huge short-term profits that could justify and secure merchant generating has disappeared. New NUG generation will require guaranteed fuel costs, a reasonable expectation of electric sales through renewable sales contracts, and/or an electric price tied profitably to fuel and operating costs.

Corporate federal taxes are levied on equity capital income. Thus, corporate earnings on equity investment must exceed the return paid the investor by an amount sufficient to cover the tax increment.

The amortization of debt capital or the provision for **depreciation** over the physical life of utility plant facilities may be effected by several methods. Straight-line depreciation requiring uniform charges in each year over a predetermined period of useful service is commonly applied because of its simplicity. The percentage method of depreciation assumes a constant percentage decrease in the value of capital investment from its value the previous year, thereby resulting in annual depreciation charges which progressively diminish. The sinking-fund method of economic analysis assumes equal annual payments which, when invested at a given interest rate, will accumulate the capital value of facilities less their salvage value, over a predetermined useful service life. Table 17.5.3 illustrates the representative useful life for alternate utility facilities.

Table 17.5.3 Representative Useful Life of Alternate Utility Facilities

Facility	Representative useful service life, years
Steam-electric generating plant	30
Hydroelectric plant	50
Combustion turbine-combined cycle	30
Nuclear plant	30
Transmission and distribution plant	40

Use may be made of interest tables which show, for any rate of interest and any number of years, the equal annual payment (sinking-fund) rate which will amortize an investment and additionally will yield an annual return on investment equal to the interest rate.

$$\text{Equal annual payment} = \frac{i(1+i)^n}{(1+i)^n - 1}$$

where n = number of years of life, and i = interest rate or rate of return.

Property taxes and property insurance premiums are normally established as a function of plant investment and thus are properly included as fixed charges. Property taxes vary with the location of installed facilities and with the rates levied by the various governmental authorities having jurisdiction. In general, public power authorities will be free of taxes, although public enterprises often render payments to government in lieu of taxes. Annual property tax rates for private enterprise will amount to perhaps 2 to 4 percent of investment, while property insurance may account for annual costs of between 0.3 and 0.5 percent.

A representative makeup of fixed charges on investment in a conventional steam-electric station having a useful service life of 30 years is shown in various classes of ownership in Table 17.5.4.

Table 17.5.4 Typical of Fixed-Charge Rate for Conventional Fossil-Fueled Steam-Electric Plant with 30-Year Economic Life

	Investor-owned utility, %	Government-owned utility, %	Industrial-commercial ownership, %
Rate of return or interest	6.4	5.0	8.0
Amortization or depreciation	1.2	1.5	0.9
Federal income tax	1.5	0.0	4.3
Local taxes (or payment in lieu of taxes)	2.0	2.0	2.0
Insurance	0.3	0.3	0.3
Total	11.4	8.8	15.5

Table 17.5.5 Operating Expense for Fuel for Representative Heat Rates and Fuel Prices

	Nominal size, MW	Typical heat rate, Btu/kWh	Fuel price, ¢/MBtu	Fuel cost, mills/kWh
Conventional coal fired	500	9,800	200	19.6
Advanced light-water reactor	600	10,700	60	6.4
Combustion turbine/combined cycle	500	7,000	400	28
Combustion turbine/simple cycle	200	8,000	400	32

NOTE: Operating fuel cost, mills/kWh = (Btu/kWh) × (¢/Btu) × 10^{-5}

OPERATING EXPENSES

Fossil Fuels Currently, fossil fuels contribute approximately three-fourths of the primary energy consumed by the United States in the production of electric energy. Generating station demands for fossil fuels are continuing to increase as the electric utility and industrial power markets grow. Price comparisons of fossil fuel are generally made on the basis of delivered cost per million Btu. This cost includes mine-mouth or well-head price, plus the cost of delivery by pipeline or carrier. Price comparisons must recognize that solid and, to a lesser extent, liquid fuels require plant investments and operating expenditures for fuel receipt, storage, handling and processing facilities, and for ash collection and removal. High transportation cost contributions on a Btu basis will be incurred by high-moisture and ash-content coals with low heating values. It is, therefore, advantageous to fire lignite and subbituminous coals at mine-mouth generating plants.

Coal represents our most abundant indigenous energy resource, with enough economically recoverable supplies at current use rates to last well into the twenty-first century. About half of the recoverable coal reserves have a sulfur content above 1 percent and are considered high-sulfur coal. Low-sulfur coals are found chiefly in the low-load areas of the mountainous West, and delivered cost at the major markets east of the Mississippi include high transportation charges. Increasingly, coal production is bearing the cost of more rigid enforcement of stringent mine safety regulations, coal washing environmental compliance, and the charges associated with strip-mine land restoration. Delivered price depends upon transportation economies as may be affected by barging, unit train haulage, or pumping in slurry pipelines. Delivered price levels for coal fuel vary with plant location. Plants conveniently located with respect to eastern coal reserves report delivered-coal prices in the general range of $1.25 to $2.50 per million Btu, depending upon sulfur and ash content. Low-sulfur western subbituminous coals have delivered prices of between $1.00 to $2.00 per million Btu.

The domestic supply of **petroleum** is now outstripped by nationwide demand. The United States has become dependent on overseas sources to meet growing energy demands. The consequences of this trend are a continued unfavorable balance of payments and dependence on foreign oil supplies from politically unstable areas in the Middle East and North Africa. Residual and distillate oil prices have risen because of short supply and pressure by the major oil producers on worldwide market price levels. Blends of low-sulfur oils delivered during 2004 to generating plants along the eastern seaboard range from $3.50 per million Btu to as high as $4.50 per million Btu for the 0.3 percent sulfur fuel required for firing in some metropolitan areas. During this same period distillate oil commanded a nationwide price ranging from approximately $5.00 to $7.00 per million Btu.

Consumption of **natural gas** as a power fuel is increasing because it is clean burning, convenient to handle, and generally requires smaller and cheaper furnaces. Historically, its limited supply made it best-suited for consumption by residential and commercial users and meeting industrial process needs. Present-day power plant use of natural gas is on the rise. The proliferation of larger, more efficient combustion turbines like the 150+ MW "F series" machines designed for combined-cycle application has increased the consumption of natural gas for power generation. During the 1990s almost all new generation was natural gas-fired combustion turbine combined-cycle.

Fuel prices for 2004 ranged from $4.00 to $6.00 per million Btu. Price levels of natural gas, severely regulated in the past by government controls imposed at the well, have been falling sharply in response to current supply-and-demand factors and gas deregulation.

Nuclear Fuel Reactor plant fuel costs present a special case. Actually, the initial core loading which will support operation of a nuclear plant over its early years of life requires a single purchase prior to commissioning of the plant. For comparison with fossil-fuel prices, therefore, nuclear-fuel cycle costs, including first-core investment, periodic charges for reload fuel, and spent-fuel shipment and processing costs, are ordinarily extrapolated at assigned load factors over the life of the plant and are converted to an economic equivalent expressed in dollars per million Btu of released fission heat. Nuclear-fuel costs are not only influenced by ore prices and by fuel fabrication and processing costs, they are also sensitive to investment and uranium-enrichment costs. Future costs are subject to inflationary pressures and cost-saving technology changes such as extended burn-up cycles. Charges up to $1 per million Btu are representative of the levelized fuel prices for nuclear plants.

Operating Costs of Fuel Fuel price contributions to energy-generation costs will reflect start-up and will depend upon plant efficiencies which, at low loads, show considerable departure from the best-point performance achieved at or near full unit loadings. These factors significantly affect the operating expenses of load-following utility system units as well as plants assigned fluctuating demands in manufacturing or industrial service. As an example, calculated performance for a nominal 500-MW, 2,400-psig coal-fired regenerative reheat steam unit shows a best-point heat rate of 9,800 Btu/net kWh at rated output. Load reduction yields heat rates of 10,500 and 12,400 Btu/kWh at loadings of 250 MW and 125 MW, respectively. Typical operating-expense ranges for given fuel prices and estimated full-load heat rates may be determined directly where continuous operation at or near unit rating is assumed (Table 17.5.5).

Operating Labor and Maintenance In addition to fuel costs, operating expenses include labor costs for plant operation and maintenance, plus charges for operating supplies and maintenance materials, general administrative expenses, and other costs incidental to normal plant operation. Operating labor and maintenance costs vary considerably with unit size, operating regimen, plant-design conditions, type of facility, and the local labor market. Representative figures appear in Table 17.5.6.

Table 17.5.6 Representative Operating and Maintenance Costs (2004 Price Level)

Type of plant	Nominal size, MW	Fuel	Operating and maintenance costs, mills/kWh
Conventional fossil	500	Coal	5–9
Advanced lightwater reactor	600	Nuclear	5–12
Combustion turbine/ combined cycle	500	Gas	6–11
Combustion turbine/ simple cycle	200	Gas	7–12
Conventional hydro	300	—	6–9

NOTE: Unit kilowatt-hour costs include labor, maintenance materials, operating supplies, and incidental expenses. Costs shown assume base-load operation.

Environmental Controls Systems and equipment required for air and water pollution abatement generally carry increased fuel and maintenance labor and materials costs. Reductions in plant output resulting from the higher condensing pressures associated with cooling-tower operation or the added auxiliary power for stack gas clean-up systems lower plant efficiency and increase fuel consumption.

OVERALL GENERATION COSTS

The total cost of power generation may now be estimated by reference to the preceding material assuming type of ownership, capital structure, plant type, fuel, and loading regimen. Table 17.5.7 comprises an illustrative tabulation of the factors determining the overall generation cost of an investor-owned generating facility.

It should be noted that capacity factor, or the ratio of average-actual to peak-capable load carried by a given generating facility, will have a significant effect on generation expenses. In addition to the effects of part-load operation on fuel costs as previously discussed, capacity factor will determine the plant generation which will support fixed charges. High-capacity-factor operation will spread fixed charges over a large number of kilowatt-hours of output, thereby reducing unit generation costs. During its initial life, a thermal plant is usually operated at high-capacity factor. As inevitable obsolescence brings newer and more efficient equipment into service, a unit's baseload position on the utility system load duration curve is relinquished, and capacity factor tends to drop. This decline in capacity factor must be acknowledged in estimating output and generation costs over the life of a given facility.

Most modern electric utilities incorporate computerized systems designed to economically dispatch power generated at each production plant feeding the load. Individual generating-unit loads are assigned in a manner that can be demonstrated to result in minimum overall cost; i.e., at each system power level, load is shared between units so that all operate at the same incremental production cost. Telemetered data reflecting system load and generation is transmitted to central dispatch computers by multiplexing via power line carrier, microwave, or telephone lines. The communication schemes include channels for transmitting load adjustment commands developed by the computer to on-line generating units. Loading instructions account for the unit production efficiencies and transmission losses associated with each dispatch assignment.

TRANSMISSION COSTS

Because of a growing scarcity of urban sites, increasing emphasis on environmental protection, and public attitudes, it has become more and more difficult to site major power stations near centers of load. As a consequence, added cost of transmission along with attendant resistive power losses add significantly to the overall cost of service. Because of the distances involved, these costs are generally greatest for nuclear facilities, mine-mouth stations, and hydroelectric plants where remoteness or remote resources strongly govern siting. Transmission plant investment also reflects a trend toward interconnection of neighboring utilities. Designed to improve service reliability by the pooling of reserves and to effect savings by capacity and energy interchanges, such interties must carry substantial ratings so that emergency power transfers can be accommodated without exceeding system stability limits. Costs for major overhead transmission ties (1,000 MVA and up) are estimated to range between $400,000 and $900,000 per circuit mile depending on terrain and voltage level. Investments are also sensitive to factors of climate and proximity to urban areas that can increase right-of-way costs significantly. Where underground transmission is elected, installed investment costs can be as much as 10 times the cost of overhead lines.

Table 17.5.7 Total Cost of Generating Power
(Plant type: 500-MW conventional fossil; plant net heat rate: 9,800 Btu/kWh)

	Generating costs, mills/kWh
Coal fuel @ $2.00/mBtu	19.6
Operating and maintenance	7.0
Fixed charges @ 11.4% per annum	24.0
Total operating costs	50.6

NOTES: Fixed charges are based on assumed initial plant investment of 1,580/kW and 7,500 h/year of operation. Values shown are typical and could vary significantly for individual plants.

Transmission and distribution systems have seen a large-scale change in ownership in deregulated areas of the country as regional grids were sold, divested, or handed to new owner/operators. These grid systems that had been maintained by high-cash-flow integrated utilities were now the responsibility of lower-cash-flow entities receiving only a small percentage of the retail price of electricity. Extensive blackouts, like the August 2003 blackout that darkened Ohio to New York City and parts of Canada, can result from insufficient transmission line investment, publicly restricted transmission line additions, and capacity increases, as was demonstrated in this situation.

The choice of **transmission voltage** level and whether ac or dc is used for bulk power transfer depends on the amount of power transmitted, the transmission distance involved, and at each voltage level, the cost of line and substation equipment. Voltages generally employed are 230, 345, 500, and 765 kV ac and 5000 + kV dc. Where long distances on the order of 400 mi (650 km) or more are encountered, dc transmission becomes economically attractive. For shorter lines, however, savings in fewer conductors and lighter transmission towers are nullified by the high cost of dc-ac conversion equipment at both line terminals.

POWER PRICES

The price of power delivered must account for the production costs at each generating station in addition to transmission and distribution costs. As previously noted, these costs depend upon labor rates, fuel prices, and material charges. They reflect investment levels in generating plant and the fixed-charge rates established by funding patterns, type of ownership, and expected equipment service life. Overall system production costs are affected by the investment requirements of specific mixes of generating equipment types, and by the manner in which load is shared by units, i.e., how production is allocated between highly efficient base-load stations and the less-efficient peaking equipment which normally runs for only a few hours each day. Also, important cost reductions are achieved in hydro systems by controlling natural and stored water flows to allow optimized sizing and scheduling of hydroelectric output, thereby reducing needs for thermal peaking capacity and decreasing the generation requirements of high fuel cost fossil plants.

Power prices cover the investment charges, maintenance costs, and capacity and energy losses chargeable to the transmission and distribution plant. They also include the administrative costs incurred to maintain corporate enterprise and the commercial expense of metering and billing. The advertising and public relations cost of competition or the cost of preparing for competition in a deregulated market must be considered.

Generally, power prices must provide a return to cover the average cost of power production throughout a given system. For smaller utility systems and commercial and industrial companies, each plant must provide an adequate return based solely on the individual plant's costs rather than an average as discussed above. Rate schedules and supply contracts, however, are drawn to reflect the reduced cost of off-peak energy produced by available generating units during periods of low system load. Additionally, prices for high load factor service often recognize the cost reductions effected by spreading fixed charges over increased units of energy output. Large blocks of capacity and energy supplied for industrial use are often priced by establishing an annual charge for capacity which equals the fixed charges on investment in committed generation and transmission plant, plus charges for energy representing the sum of the variable kilowatt-hour production costs for fuel, maintenance, and operation.

Historically, utilities have applied rate schedules which promote consumption by applying progressively lower rates to blocks of increased energy usage. Rationale for such pricing is the savings that load growth can

Table 17.5.8 Average Cost, in Cents per Kilowatt-Hour, for Consumers by Sector, Census Division, and State, 2002

Census division, State	Residential	Commercial	Industrial	Other	All Sectors
New England	**11.18**	**9.91**	**8.52**	**10.50**	**10.16**
Connecticut	10.96	9.35	7.68	10.36	9.73
Maine	11.98	10.47	11.24	32.82	11.36
Massachusetts	10.97	10.14	8.77	9.79	10.18
New Hampshire	11.77	10.09	8.83	12.07	10.49
Rhode Island	10.21	8.84	8.04	8.09	9.19
Vermont	12.78	11.10	7.90	19.26	10.87
Middle Atlantic	**11.32**	**10.14**	**6.03**	**9.45**	**9.59**
New Jersey	10.38	8.87	7.83	14.04	9.31
New York	13.58	12.46	5.16	9.05	11.29
Pennsylvania	9.71	8.03	6.06	11.10	8.01
East North Central	**8.06**	**7.20**	**4.58**	**6.09**	**6.50**
Illinois	8.39	7.49	5.01	5.56	6.97
Indiana	6.91	5.98	3.95	9.75	5.34
Michigan	8.28	7.36	4.95	10.43	6.92
Ohio	8.29	7.68	4.68	5.70	6.66
Wisconsin	8.18	6.54	4.43	8.08	6.28
West North Central	**7.37**	**6.01**	**4.22**	**5.72**	**5.97**
Iowa	8.35	6.56	4.06	4.92	6.01
Kansas	7.67	6.28	4.53	9.30	6.31
Minnesota	7.49	5.88	4.19	7.36	5.84
Missouri	7.06	5.88	4.42	6.20	6.09
Nebraska	6.73	5.62	3.89	6.37	5.55
North Dakota	6.39	5.85	3.98	3.68	5.45
South Dakota	7.40	6.24	4.54	3.63	6.26
South Atlantic	**7.90**	**6.43**	**4.24**	**6.42**	**6.56**
Delaware	8.70	6.98	5.11	10.62	7.05
District of Columbia	7.82	7.38	4.95	6.60	7.37
Florida	8.16	6.64	5.23	7.43	7.31
Georgia	7.63	6.46	3.95	8.31	6.24
Maryland	7.71	6.09	3.88	10.18	6.21
North Carolina	8.19	6.51	4.70	6.70	6.74
South Carolina	7.72	6.48	3.85	6.44	5.83
Virginia	7.79	5.87	4.13	5.15	6.23
West Virginia	6.23	5.41	3.81	10.01	5.11
East South Central	**6.57**	**6.33**	**3.71**	**6.32**	**5.39**
Alabama	7.12	6.63	3.82	7.46	5.71
Kentucky	5.65	5.30	3.09	4.61	4.26
Mississippi	7.28	6.83	4.40	8.76	6.24
Tennessee	6.41	6.45	4.15	8.92	5.72
West South Central	**7.70**	**6.69**	**4.48**	**6.31**	**6.33**
Arkansas	7.25	5.68	4.01	6.52	5.61
Louisiana	7.10	6.64	4.42	7.05	5.99
Oklahoma	6.73	5.75	3.81	5.06	5.59
Texas	8.05	6.95	4.66	6.55	6.62
Mountain	**7.87**	**6.64**	**4.86**	**5.57**	**6.52**
Arizona	8.27	7.28	5.20	4.56	7.21
Colorado	7.37	5.67	4.52	6.64	6.00
Idaho	6.59	5.71	4.34	5.18	5.58
Montana	7.23	6.53	3.70	7.14	5.75
Nevada	9.43	9.06	7.25	6.54	8.42
New Mexico	8.50	7.22	4.48	6.23	6.73
Utah	6.79	5.60	3.84	4.69	5.39
Wyoming	6.97	5.71	3.55	5.93	4.68
Pacific Contiguous	**10.46**	**11.38**	**8.26**	**6.29**	**10.28**
California	12.90	13.22	10.83	6.68	12.50
Oregon	7.12	6.59	4.72	9.44	6.32
Washington	6.29	6.11	4.56	4.94	5.80
Pacific Noncontiguous	**14.20**	**12.46**	**10.26**	**14.63**	**12.35**
Alaska	12.05	10.13	7.65	14.04	10.46
Hawaii	15.63	14.11	11.02	16.85	13.39
U.S. Total	**8.46**	**7.86**	**4.88**	**6.73**	**7.21**

SOURCE: DOE website www.eia.doe.gov/cneaf/electricity/esr/table1abcd.xls#A238.

realize through economies of scale, as well as the improved utilization of existing utility plant. However, regulatory pressure, reflecting a policy of minimizing the cost to the consumer, impact on the environment, and the critical need to conserve high-cost imported fuel, has favored a marginal cost-pricing system more nearly reflecting the actual cost of production and transmission of a particular user's supply of power. Rate setting under this conservationist approach calls for flat, rather than reduced, rates as usage increases and for high unit energy charges during peak-load periods.

Cogeneration facilities designed for the **dual-purpose production of power and process steam** permit investment savings, principally in steam-generation plants, which result in combined production charges falling below the total cost of separate, single-purpose production of power and of steam. Such savings can permit proportionate decreases in the prices ordinarily charged for separate single-purpose production of each of the products, or they may be assigned in total to reduce the price of one or the other by-product. This latter option is often exercised in the case of dual-purpose water product plants arranged for seawater flash evaporation using power-turbine extraction as a process steam source. Where severe shortages of fresh water exist, social considerations favor the total assignment of dual-purpose savings to the water product. Thus, minimal prices for desalted product water are achieved, while dual-purpose power is marketed at prices competing with single-purpose power generation costs. Similarly, assignment of the total savings of dual-purpose power and process steam production to the power product may justify on-site industrial plant power generation in preference to outside purchases of higher-priced utility system power supplies.

Since the 1980s, some facilities qualifying under the PURPA regulations have provided very low-priced steam to a host industrial plant as a means to satisfy the PURPA regulations and sell electricity to the local utility, whether the utility wanted the facility or not. Changes to the pricing structures payable by the local utilities to the "qualifying facility" and the eventual end of the PURPA statute eliminated the "free steam" approach from cogeneration.

Where hydro facilities supply power in combination with irrigation, flood control, navigation, or recreational benefits, power costs are largely sensitive to the allocation of investment charges against each of the multipurpose project functions. Should generation be treated as a by-product, power prices can be reduced drastically to reflect equipment operating expenses and the limited fixed charges covering investment in only the generating plant itself.

Cheap fuel, advances in design, the economies of scale, and the economic application of alternate generating unit types produced downward trends in the price of electric service in the 1960s. In the 1970s these trends were reversed by inflationary effects on plant costs, high interest rates, and the increased fuel prices. A dwindling number of favorable plant sites, licensing delays, and the added costs of environmental impact controls combined to cause further upward pressure on power prices in the 1980s. However, falling interest rates and fuel prices began to slow the rate of increase in the cost of electric power. These stabilizing influences have continued into the 1990s, tempered by stricter environmental regulation costs. Increases in electric rates are expected into the foreseeable future.

The twenty-first century has started with good and bad deregulation cases, corporate financial problems (such as Enron), and increased terrorist attacks on the U.S. infrastructure. All of these factors, coupled with a significant rise in the price of natural gas in the first several years of the new millennium, have resulted in continual escalation of the price of electricity. Table 17.5.8 shows the electricity prices by sector for each region and state in the United States for the year 2002.

17.6 HUMAN FACTORS AND ERGONOMICS

by Ezra S. Krendel

References: "Aviation Safety and Pilot Control," NRC, National Academy Press, Washington, DC, 1997. Allen, McRuer, et al., "Computer-Aided Procedures for Analyzing Man-Machine System Dynamic Interactions," Vol. I, "Methodology and Application Examples," Vol. II, "Simplified Pilot-Modeling for Divided Attention Operations," Vol III, "Users Guide," WADC TR-89-3070, June 1989. Badler, Phillips, and Webber, "Simulating Humans: Computer Graphics, Animation and Control," Oxford University Press, 1993. Boff, Kaufman, and Thomas (eds.), "Handbook of Perception and Human Performance," Vols. I and II, John Wiley & Sons, 1986. Card, Moran, and Newell, "The Psychology of Human-Computer Interaction," Lawrence Erlbaum Associates, 1983. Kleinman, Baron, and Levison, "An Optimal Control Model of Human Response, Part I: Theory and Validation," *Automatica*, **6,** 1970. Konz and Johnson, "Work Design: Occupational Ergonomics," 6th ed., Holcomb Hathaway, 2004. Krendel and Wodinsky, "Search in an Unstructured Visual Field," *Jour. Optical Soc.*, **50,** 1960. McRuer, Pilot-Induced Oscillations and Human Dynamic Behavior, *NASA CR-4683* July 1995. McRuer, Clement, Thompson, and Magdaleno, "Minimum Flying Qualities," Vol. II, "Pilot Modeling for Flying Qualities Applications," Systems Technology, Inc. Hawthorne, CA, 1990. McRuer, "Human Dynamics in Man-Machine Systems," *Automatica,* **16,** 1980. McRuer and Krendel, "Mathematical Models of Human Pilot Behavior," AGARDograph No. 188, 1974. Salvendy (ed.), "Handbook of Human Factors," 2d ed., John Wiley & Sons, 1997. Sundin, Örtengren, and Sjöberg, "Proactive Human Factors Engineering Analysis in Space Station Design Using the Computer Manikin Jack," SAE Technical Paper 2000-01-2166. Thompson, "Program CC Version 5.0 for Windows," Systems Technology Inc., Hawthorne, CA, 2001.

SCOPE

The mission of human factors engineering/ergonomics (HFE/E) is to improve the performance, reliability, efficiency, and the risk management of systems in which humans work in concert with machines, man/machine/systems (MMS). This discipline developed many of its techniques during and after World War II. The life-or-death stakes and the advances in military technology made even minor improvements in the performance and reliability of manned military systems highly desirable and major improvements essential. The design of MMS interact with experimental psychology, physiology, and physical anthropology and with aeronautical, electrical, industrial, mechanical, systems, computer, and cognitive engineering.

MMS cover a wide range of enterprises, from piloting a jet aircraft to microsurgery. Achieving the goals of HFE/E in MMS requires a knowledge of humans as sensors, manipulators, responders, information processors, and decision makers. The expected performance of the "man" in MMS is affected by: training and motivation, variability of behavior among humans, personal stress, work load, age, fatigue, substance abuse, and by environmental conditions: vibration, noise, temperature, gravity, and distractions from the task. Some of the references for this section are sources for details and data on the impact of these many human-related variables. Other references provide engineering theory and practice for the design of MMS. What follows is an introduction to HFE/E findings that are likely to be helpful to mechanical engineers.

VISION

The visual pathway in psychomotor tasks, written or iconic messages, and searching for targets is primary in most MMS. Other senses, in declining order of importance to MMS design, are: hearing, which is superior to vision for alerting or warning signals; kinesthesis and vestibular senses, which facilitate positional awareness; and touch and pain.

High-resolution vision requires that the image of the object seen be projected by the eye's lens upon an approximately 2° central sector of the retina, known as the *fovea*, composed entirely of photoreceptors called *cones*. Cones generate color vision and respond to light intensity from the limit of brightness tolerance to that of candlelight reflected from a newspaper. As it scans its field of view, the eye continually moves in pulses of approximately 50 ms, pausing to fixate its fovea on objects in this field. Each fixation pause is about 250 to 300 ms, but can be as long as 1 minute when the viewer is uncertain about the image on his fovea.

The density of cones on the retina falls steeply beyond its area of concentration on the fovea and reaches a flat minimum at an eccentricity of about 20°. Beyond the fovea a second set of retinal photoreceptors, rods, appears among the cones. Rods provide vision from low brightness levels to the darkest conditions for which vision is still possible. Rods reach their maximum sensitivity to low light levels after about 30 min in the dark. In a twilight level of brightness comparable to snow under a full moon, rods and cones operate together. The number of rods on the retina increases in density until reaching a maximum at an eccentricity of about 20°. Although much greater than that of cones, rod density gradually declines as it approaches the limiting eccentricity of 80°. This peripheral vision supports orientation and motion detection.

Visual perception is the brain's interpretation of visual images. The perception of the size and distance of an object begins when the eyes' lenses, controlled by the ciliary muscles, accommodate so as to focus upon the object. Estimates of an object's apparent distance are influenced by: this muscle action, the visual cues leading to the object, and the anticipated size of the object. These cues may be inadequate or they may be inconsistent with one another, and as a consequence a perception of distance that differs from reality may emerge. The vision of an automobile driver may be momentarily focused on the dashboard displays at a distance of 0.5 m. Refocusing on an object appearing unexpectedly in the road ahead may take as long as 0.4 s. The object may be unfamiliar and helpful visual cues obscured by darkness or fog. The probability of an accident is increased by this briefly held misperception of object size or distance.

Some of the characteristics of vision can support a two-parameter model for the probability of success over time in searching with the naked eye for a target alone in a visual field. This simplified model must be elaborated when the proposed application differs greatly from the experimental conditions under which it was validated. This is a common condition when most HFE/E models or data are used.

Assuming the search is random and p_{sg} is the probability of finding the target during a single glimpse or one fixation, then the probability of finding the target after k glimpses is:

$$P_k = 1 - (1 - p_{\text{sg}})^k \tag{17.6.1}$$

Each glimpse plus movement time is T seconds, therefore the elapsed search time is $t = kT$; the probability of finding the target by this time is:

$$P(t) = 1 - e^{-mt} \qquad t_{\text{mean}} = 1/m \qquad m \approx p_{\text{sg}}/T \tag{17.6.2}$$

This model was confirmed for $0.006 \le p_{\text{sg}} \le 0.80$, different size sectors of the sky, different target sizes, and different target contrasts. In unpracticed random searches for a command in a menu on a monitor, the single-glimpse probability of detection, p_{sg}, was 0.08. After practice, p_{sg} increased to 0.60. Were the search systematic with no overlap of fixation areas the probability of detection, P_{sys}, for search for a limited time t would be:

$$P_{\text{sys}}(t) = p_{\text{sg}}\, t/T, \tag{17.6.3}$$

where $P_{\text{sys}} \le 1$.

This model for the probability of finding targets becomes more elaborate when the target must be found from among camouflaged decoys as well as from among decoys differing from it by contrast, shape, area, color, and velocity as well as by area of search, numbers and spatial density of decoys and of targets, and the use of auditory aids.

PSYCHOMOTOR PERFORMANCE

A program of skilled psychomotor behavior can be constructed in several ways: by incrementally aggregating a sequence of discrete reaction times and movements, by continuous closed-loop control known as manual control or tracking, by complex open-loop activities, and by various combinations.

Manual reaction time (RT), like the eyes' fixation time, increases with uncertainty. This increase in RT can be quantified when a subject must select the correct manual response out of n possible responses to a corresponding visual signal out of n probable signals. For situations in which the independent or the sequential probability, p_i, for the occurrence of each of these n signals can be estimated, uncertainty can be expressed as information theory entropy, H:

$$H = \sum\nolimits_i^n p_i \log_2 (1/p_i + 1) \tag{17.6.4}$$

For the range $0 \le H \le 3$, there is an empirical equation, known as the Hick-Hyman law for RT, for these manual responses:

$$\text{RT} = a + bH \tag{17.6.5}$$

The constants $a \approx 200$ ms and $b \approx 150$ ms are influenced by practice and by the compatibility between the geometry of the physical presentation of the stimulus and that of the mechanism for responding.

For rapid discrete or repetitive actions, usually by the hands, movement time (MT) is proportional to the difficulty of the task as defined by Fitts' law. Equation (17.6.6) defines the index of difficulty, I_D, in terms of target width W and movement amplitude A. MTs from a wide range of rapid movements, as in operating a key pad or sorting items into bins, can be described. In applications, d is about 100 ms and c, which depends on the movement geometry, is approximately 200 ms. Estimates of MT enable comparisons to be made among different operating procedures.

$$I_D = \log_2 (2A/W) \tag{17.6.6}$$

$$\text{MT} = c + d\, I_D \tag{17.6.7}$$

To be useful, measurements of RT and MT must come from a stable plateau of skilled performance that can be determined by units of error or of time to complete the given task. For many psychomotor tasks, skilled learning follows a power function

$$T_n = T_1 n^{-\beta} \tag{17.6.8}$$

where T_1 is the time to perform the task on the first trial. T_n the time to perform the task on the nth trial, and n is the number of trials. In the development of psychomotor skill, $0.2 \le \beta \le 0.6$.

SKILL AND ERRORS

The highest skill level in tracking is attained after extensive training when an operator fully familiar with the dynamics of the machine, the manipulator under her control, and the appropriate responses to the input signals can extract what coherence is present in these signals, develop pathways that reorganize the perceptual system, adapt her behavior to create a repertory of special responses, and select from this repertory the appropriate response for the best MMS performance. The procedure resulting in this performance is known as successive organization of perception (SOP). SOP results in a virtual display that provides the operator with the information that would have been present in the actual display had it been augmented. In closed-loop control situations, the precognitive mode response can be anticipatory and compensate for inherent human control lags. In an open-loop mode, the response may be programmed as in executing the sequential actions in preparing to stop a car on approaching a red light or otherwise responding to behavioral cues.

As psychomotor skills develop with training, errors diminish, and a plateau of stable, skilled performance is reached. Transient errors occur and are shed as training progresses. Posttraining errors are departures from the expected performance of a motivated, skilled operator. These errors may place the MMS in danger. Their number increases under the following pressures on human performance: divided attention, multitasking, and physiologically impairment such as fatigue, hypoxia, and alcohol or drug use. Errors can be divided into two classes: (1) mistakes, errors in the thought processes behind an action or decision, and (2) slips, errors in the detection or interpretation of sensory signals and unintended responses. The thinking processes which lead to mistakes have been reconstructed; for example, inappropriate expectations can result in misinterpreting a situation acting accordingly and making a malign error. The basic sources for such reconstructions are accident reports, anecdotes, interviews with participants, and introspection on the part of the analyst. If the mistake is made under time pressure, it is difficult for the operator to reexamine the situation, sort out his misconceptions, and examine his options in a deliberate manner. Mistakes can be decreased by training methods, personnel selection, and operating procedures that emphasize the examination of options and impose redundancy for critical actions to lessen the chance that operators will persevere in erroneous behavior.

By applying HFE/E findings in the selection and positioning of information displays, controls, monitors, and keyboards, and by accommodating the physical dimensions of the operator, the designer can make slips such as misreading a display or reaching for the wrong switch less likely. In the unlikely event that an error does occur, emergency corrective procedures and warnings must be part of the MMS design. Warnings can address different sensory modalities with audio signals, speech, annunciators, lights, and vibration, either separately or in various combinations determined by the task being performed and the environment of the person or persons to be warned.

The probabilities of slips occurring can be estimated from the existing databases in HFE/E and from manned and unmanned simulator studies. These estimates, together with the failure probabilities for the inanimate components of the MMS, can be incorporated into flowcharts for the MMS to determine those locations where further risk management procedures are necessary.

MANUAL CONTROL

The body of knowledge in manual control developed from the USAF's mid-twentieth-century interest in mathematically modeling the dynamic response of pilots in order to provide aeronautical engineers with data for designing aircraft of high performance, stability, and positive evaluations from their pilots. Classical control theory provided the structure for this model. The core was a quasi-linear transfer function or describing function for the pilot's dynamics represented as $Y_p(j\omega), Y_p(s)$, or Y_p, depending on the projected analysis. The quasi-linear human controller's response to an input comprises two parts: describing function components that correspond to the response of the equivalent linear elements driven by that input, and a "remnant" component that represents the difference between the response of the actual system and a system composed of equivalent linear elements. Test pilots operating fixed- and moving-base flight simulators as well as actual aircraft provided the empirical data for $Y_p(j\omega)$. Aircraft dynamics, the controlled element, $Y_c(j\omega)$, confine the dynamic response of motivated pilots to a limited range so that intersubject variability is suppressed. Consequently, adequate measurements for $Y_p(j\omega)$ can be obtained from a small number of skilled, motivated test subjects.

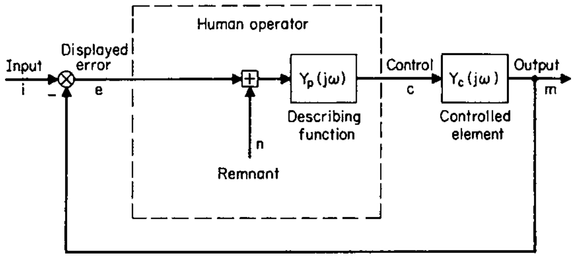


Fig. 17.6.1 Simplified block diagram for manual control.

The compensatory mode, Fig. 17.6.1, the most common and extensively studied mode of closed loop control, is present in vehicle control and many other configurations. In this mode, the human tracks a moving visual signal so as to keep its position as close to a reference marker as she can. This signal is the MMS's error $e(t)$, which is the difference between $i(t)$, the input to the MMS, and $m(t)$, the MMS output. Detecting cyclical and coherent structures in $e(t)$ and from these developing estimates of $m(t)$ and $i(t)$ are the first step in the SOP progression of evolving virtual displays. Maintaining system stability and minimizing the closed-loop system's error are basic design goals for closed-loop systems. If the open-loop gain $|Y_p(j\omega)Y_c(j\omega)|$ in the compensatory mode is too high, instability can occur in the closed-loop task. An expert pilot may unwittingly regress from a higher level in the SOP procedure to the compensatory mode. In this mode the pilot may inadvertently generate excessive open-loop gain and destabilize an otherwise stable system. Potentially destructive pilot-induced oscillations of the aircraft about its flight path are the result.

Compensatory pursuit and precognitive modes of the SOP procedures are presented in Figure 17.6.2. In the pursuit mode the operator

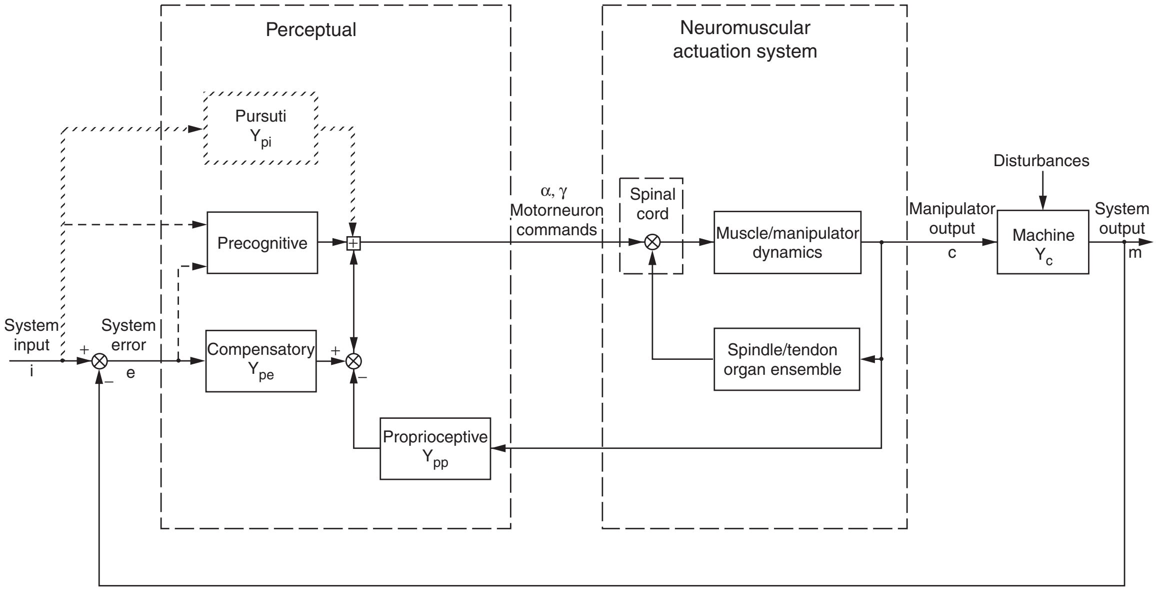


Fig. 17.6.2 Compensatory, pursuit, and precognitive pathways.

attempts to follow $i(t)$ with a cursor driven by the MMS dynamics. The perceptual pathways composing SOP for the human controller can be separated into sensory mechanisms that generate internal signals and to a central processor that integrates them, extracts coherence, and determines the human operator's dynamic responses. The describing functions and remnant depend explicitly on: the task variables, MMS inputs and disturbances, controlled element dynamics, and the operator-centered variables. The effects of these variables have been integrated into a set of rules for applying and adjusting human-operator describing functions for compensatory closed-loop control. The following conditions must apply before using these rules:

1. The forcing functions that act as inputs to the MMS are unpredictable, have bandwidth ω_i below 6.2 rad/s, and continuous waveforms.
2. The controlled element is a low-order system or can be so approximated and has no highly resonant modes over the input bandwidth.
3. The display and controls are reasonably well scaled and smooth so as to minimize the effects of thresholds, detents, and friction.
4. The other task demands are sufficiently light to permit the operator to devote the majority of his or her attention to minimizing the displayed error.

McRUER'S RULE

Under the foregoing conditions, which are commonly met in most operational control tasks, theory, experiments, and practice have shown that the operator compensates for the dynamics of the controlled element so that the open-loop characteristics of the MMS act like an integrator in series with an effective time delay in the frequency range most critical to system stability and performance. The frequency within this range where the open-loop gain of the MMS is unity is the crossover frequency ω_c. By compensating in this way, the operator creates a desirable, stable control system in accordance with good engineering design. This behavior by the human controller is known as **McRuer's rule,** and it is expressed by the crossover model**:**

$$Y_p(j\omega)Y_c(j\omega) \approx \frac{\omega_c e^{-j\omega\tau}}{j\omega}. \qquad (17.6.9)$$

Many of the basic qualities of feedback systems relate to ω_c, which is a close measure of the bandwidth of the closed-loop system. In the range of input bandwidth frequencies from very low up to and approaching ω_c, the output $m(t)$ of the closed loop system follows the system input $i(t)$, and the system error $e(t)$, is reduced. At increasingly higher input bandwidths beyond ω_c these properties are lost as $e(t)$ increases and $i(t)$ and $m(t)$ are no longer similar. For desirable closed-loop performance, ω_c must exceed the largest frequency in the input or in disturbances to the system that have appreciable power.

The effective time delay τ is the sum of perceptual and neuromuscular lags as affected by the demands of the tracking task.

As the controlled-element dynamics demand more lead from the operator to conform to the crossover model, the concomitant increased anticipation required of the operator imposes a cost. This is reflected in increases in τ, and affects the operator's evaluation of the MMS. Table 17.6.1 lists crossover model parameters.

As an illustration consider automobile heading control. At low to moderate speeds, this can be approximated by the controlled element $Y_c(j\omega) = K_c/j\omega$. Since the heading rate is proportional to the position of the steering wheel, the driver need only generate proportional control, i. e., zero lead. The effective time delay will then be 0.18 s and the maximum effective closed-loop system bandwidth will be the maximum crossover frequency or 4.7/rad/s.

Attitude control of a spacecraft with damper off can be approximated in the region of human control by $Y_c(j\omega) = K_c/(j\omega)^2$. Maintaining stability demands more of the operator than did automobile heading control, for the operator must generate a low-frequency lead. The cost incurred by control is an increase in the effective time delay τ to 0.32 s. A significant consequence is the reduction of the maximum system crossover frequency, and thus the maximum effective closed loop system bandwidth, to 3.3 rad/s.

For a single-loop control system of maximum crossover frequency ω_c, the attainable relative mean square error coherent with an input bandwidth, ω_i and variance, σ_i can be estimated:

$$\overline{e_i^2}/\sigma_i^2 \approx \frac{(\omega_i/\omega_c)^2}{3} \qquad (17.6.10)$$

Thus for similar MMS input characteristics, the estimated attainable mean square error for attitude control of a spacecraft is twice the estimated attainable mean square error for the less demanding heading control of an automobile.

MODELING THE HUMAN IN THE MMS

The crossover model is a first approximation in the process of elaborating models for human operators of increasingly more complex control systems in which there may be competing demands for the operator's attention. The process has been described in a reductionist fashion by analyzing the input-output behavior of the human into subsystem components. The sensory mechanism subsystems in Fig. 17.6.2 include optics, the retina, oculomotor muscles, eye dynamics, and the dynamics of linear acceleration and angular velocities acting on the vestibular and kinesthetic sensors. The central elements act to integrate and to equalize the outputs of these subsystems so as to generate commands to the neuromuscular actuation system, which is divisible into dynamic subsystems. These components are adjusted according to **McRuer's rule** so that the crossover model obtains and then is integrated to describe the operator's behavior in a structural-isomorphic model. In contrast, the algorithmic or optimal control model (OCM) proceeds in a holistic fashion and mimics the human's total response by the application of optimal control computational methods to those human properties subject to adaptation and hence optimization. The OCM proceeds by minimizing a quadratic performance index with an optimum linear predictor operating on estimated delayed state vectors to emulate the human's control actions. The computational procedure results in very high-order transfer functions for the human. PC-compatible programs, for example, Program CC Version 5, have been written for both the classical model and the OCM. Since **McRuer's rule** is the best understood, most widely applicable empirically tested description of human dynamics, the crossover model serves as the standard for comparing the classical model with the OCM model. This has been done for $Y_c = K_c/s$ and the comparison is good for frequencies less than 15 rad/s.

Manual control models present one perspective of the virtual human. Other perspectives are reach, strength, movements, and eye point of regard. Digital models exist which allow 3-D virtual humans to be placed in 3-D CAD-generated virtual environments. In the design of automobile, aircraft, spacecraft, or submarine interiors as well as factory work stations and large construction machinery, engineers are able to examine a human's activity in virtual environments and to sort through design options. One such digital human model, Jack/Jill, manufactured by EDS, has 69 segments, 68 interconnected joints, a 17-segment spine, 16-segment hands, coupled shoulder/clavicle joints, and 135 degrees of freedom to mimic the motions and positions of a wide range of humans.

Table 17.6.1 Crossover Model Parameters, ω_i = 2.5 rad/s

Controlled-element dynamics	K_c	$K_c/(j\omega)$	$K_c/(j\omega)^2$
Operator lead unit to establish control	−1 (integral)	0 (proportional)	+1 (low frequency lead)
Effective time delay τ, seconds	0.14	0.18	0.32
Maximum crossover frequency ω_c, rad/s	5.7	4.7	3.3

Joint limits as derived from NASA studies and sizes and shapes based on SAE measurements and ANSUR 88 data round out the engineer's ability to select digital manikins for specific problems and representative of 5th, 50th, or 95th percentile human actors. The virtual human operator can thus be used to simulate a population of potential users and determine the diversity permitted by a particular design.

Risk management of slips could be addressed by Jack/Jill simulating a control room operator confronted with an arrangement of manual controls such as knobs or levers. By selecting 5th and 95th percentile operators and adding statistical variability to their head position, eye point of regard, and reach, the probabilities for slip errors occurring for configurations of manual controls under different time pressures could be estimated.

17.7 AUTOMATIC MANUFACTURING

by Vincent M. Altamuro

The production of many products involves both **fabrication** and **assembly** activities. Fabrication is the making of the component piece parts that are later assembled into the final product. Fabrication methods include casting, molding, forging, forming, stamping, and machining. Assembly may be manual or automatic. Automatic manufacturing can range from **semiautomatic** to **fully automatic,** depending on the number of human operations required. Automatic manufacturing installations may also be called **fixed** and **flexible** according to how easily they can be altered to make product variations. These several available modes allow the creation of a mode that draws on all of them—the **autofacturing** mode, that combination of the manual and the specific types of automatic operations which best achieves the quality, operational, and economic objectives sought.

The selection of assembly mode must be consistent with the design of the product, the personnel skills, the equipment available, and other factors. Manual assembly is suitable for low volume output and relatively short production runs with high product-to-product variations. Fixed automatic assembly is suitable for high-speed, high-volume production of uniform products. Also called *hard, dedicated,* or *conventional* automation, it can produce low-unit-cost items of uniformly high quality once its machinery is perfected and until a product variation is wanted. If a product's assembly cost is the sum of its *setup cost* (the cost of getting everything ready to make it) and its *run cost* (the cost of making it), then manual assembly has a relatively low setup cost but a high run cost, while fixed automatic assembly has the reverse. Flexible automatic assembly is suitable for midrange-volume production with variations in product specifications. Also called *soft* or *programmable* automation, it uses robots and other computer-based equipment that are more easily movable and adjustable so as to reduce setup costs and still enjoy low run costs. It fits between fully manual and fully fixed automatic assembly, as shown in Fig. 17.7.1.

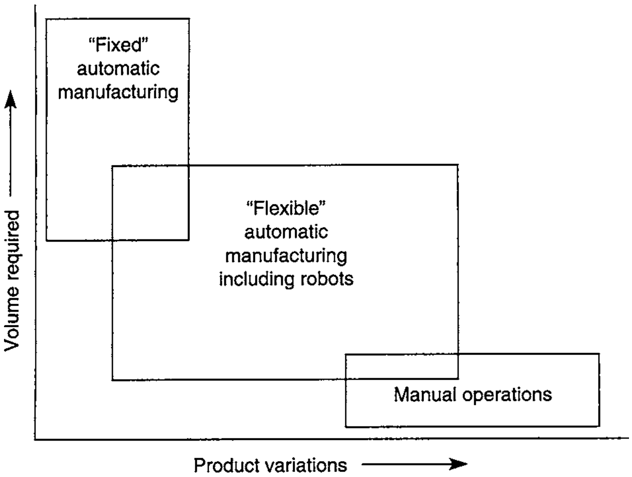


Fig. 17.7.1 The most appropriate assembly mode, as a function of output volume and variations.

Autofacturing is a production system that is comprised primarily of automated equipment which is configured as several integrated subsystems, using one common database and computer controls to make, test, and transport specifically designed products at high and uniform quality levels meeting flexible specifications with a minimum of human effort. There are many levels of autofacturing from individual cells, or *islands,* all the way up to a complete and integrated system. Most situations are somewhere in between, but progressing toward a total system.

Autofacturing is a subset of the overall field of automation. It refers to automated manufacturing activities as distinguished from the many other things called automation, such as data processing, office automation, electronic funds transfers, automated banking, central station telephone operations, airline ticket reservations, and the like.

An autofacturing system is composed of several integrated subsystems. Also, along with other systems such as marketing and distribution systems, general accounting and cost accounting systems, management information systems (MISs), and others, it is a subsystem of the corporation's overall operating system. All subsystems should be interconnected, complementary, in balance, and integrated into one total system. Some of the prerequisites and subsystems of an autofacturing system are listed and described below.

DESIGN FOR ASSEMBLY

After management decisions have been made regarding policy, practices, and long-range objectives, decisions must be made about the best combination of a multitude of characteristics of the product. These include the goals to be achieved by it: its salability, functionality, safety, targeted life, repairability, recyclability, ease of use, size, shape, color, and many other considerations, some reinforcing others and some in conflict with others. Designing the product for ease and economy of fabrication, assembly, test, handling, shipping, and installation are some important considerations. Then, the basis of all of the above determinations, decisions can be made regarding the best levels and mix of autofacturing. That includes deciding which portions of the product should be made by people, hard automation, or programmable devices, since the size, shape, and features of its several parts may be different and better suited for one mode over the others.

Human assemblers, for example, have the advantage of three-dimensional vision, color sensitivity, two hands, eye-hand coordination, the facility to jiggle parts together, the ability to back off when things do not go together as they should and look for the reason rather than force them, the ability to pick out patterns which are varied or complex, etc. However, they cannot work in dangerous environments, apply large or exact forces, perform with precision consistently cycle after cycle, or handle very large, very small, or very fragile items, etc.

Hard automation requires the input of a steady, voluminous stream of very uniform parts, all positioned and oriented precisely so that the high-speed machines can operate without jamming or stopping. Machines and instrumentation are better than humans in monitoring and responding rapidly and consistently to many stimuli simultaneously—some beyond the range of human capabilities. They also can be made to do several things at once for long periods of time with little variance or deterioration of performance.

Soft, or programmable, automation permits variations in inputs, as it can sense and adjust for deviations and continue to operate. It is usually faster than manual work but slower than dedicated machines, and its output can permit variations from the standard product so as to achieve marketing and inventory advantages that often more than offset its added cost per unit of output.

In autofacturing, each assembly operation is analyzed to determine whether it is best done manually, with fixed, or with flexible automation. An autofacturing system, then, results in a mixed mode of operation, each with that portion of the product designed to be assembled in the most efficient way available. In it, the actions of conventional machines, automated equipment, human operators, and robots are all interrelated and in balance.

There are over 100 guidelines on how to configure a product and its component piece parts so that they can be assembled as easily and economically as possible. The most important of these rules are as follows:

1. *Minimize and simplify.* Reduce the number of piece parts as much as possible. Determine the essential functions of each part. Transfer functions to other parts. Combine parts. Eliminate as many parts as economically feasible. Simplify before automating.

2. *Modularize.* Design the product such that parts are grouped into modules or sections of the end product. Create modules, like building blocks, so that they can be selected and joined in various ways to make different products. This will aid in assembly, test, repair, and replacement of more manageable subassemblies. It will also make it possible to follow rule 3 below.

3. *Create families of products.* Products should be designed so as to have as much commonality with other products as possible in their modules and piece parts. Products related in a series of escalating capabilities or features can all be built with the same platform (base, power supply, etc.), case, panels, circuit boards, and other common and interchangeable parts. Product distinctions, such as extra features, should be the last components assembled so that the work in process is common for as long as possible. Parts that make a product distinctive should be grouped into the same module. Even seemingly different products can be designed to have the same power supply, switches, gears, and other internal components and modules.

4. *Design parts to have as many different uses as possible,* depending on how they are installed and which portion of them is used. Determine whether the extra cost of the added complexity is more than offset by the benefits.

5. *Use the best overall method to make each part.* That is, consider not only its fabrication costs, but the relative costs of handling, subsequent operations, assembly, inspection, scrap, etc.

6. *Design in layers.* Where possible, design the product so that it is made up of layers of components, such that making it requires the adding of one layer on top of another, built up from the base to the cover or outer case. See Fig. 17.7.2.

7. *Assemble in short, straight vertical strokes.* If rule 6, above, is observed, then this should be possible. Avoid the need for lateral, curved, or complex motion paths in assembling components. Minimize the number of directions of assembly. Minimize the amount of lifting, rotating, and other handling of the components, subassemblies, or product. See Fig. 17.7.2.

8. *Present parts appropriately.* Bring parts to the machines, robots, and human operators in the position and orientation best suited to use without the need for rehandling or repositioning. Hold them in trays, on reels, in connected strips, etc. rather than disoriented in a tote box or bin. Parts stamped from strips or molded should be left on their webs as long as possible, for they are orderly. Parts to be fed to assembly stations by vibratory bowl feeders should be designed so as to feed through easily and not jam, snag, tangle, nest, shingle, bridge, or interlock.

9. *Design for symmetry and orientation.* To the degree possible, design parts to be symmetrical so that no matter what position they are in, it is proper for assembly. If perfect symmetry is not possible, strive to have the parts go together in as many ways as possible. Where that is not possible, accentuate the asymmetry and give the parts an orientation feature, so that a person or robot can easily sense it and orient the part correctly. See Fig. 17.7.2.

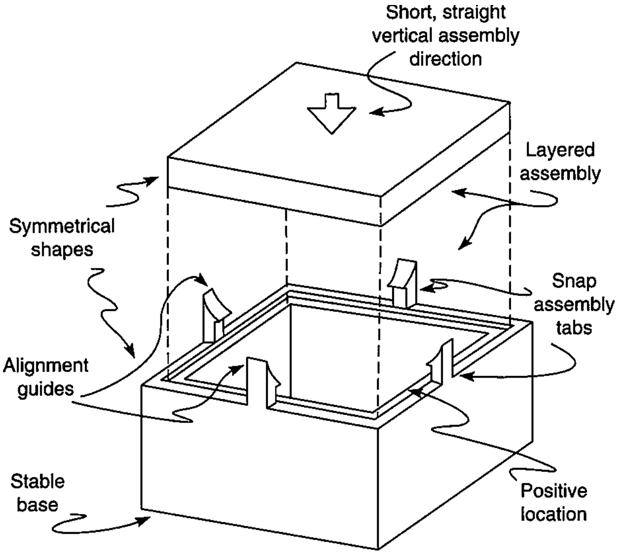


Fig. 17.7.2 Selected design for assembly guidelines.

10. *Make errors impossible.* Design the product and its parts so that the components physically cannot be assembled wrong. If a part has internal or hard-to-sense features, add a feature that can be sensed, even if it has no operational function. Do not have the dimensions of a part be too similar if the part's proper orientation and assembly depend on discerning a small difference.

11. *Give each part a positive location.* Through the use of shoulders, recesses, guides, tabs, and the like, create an exact place where each component is to go. See Fig. 17.7.2.

12. *Minimize fasteners.* Eliminate or reduce to the lowest essential number the product's discrete fasteners. Nuts, bolts, screws, and the like are difficult and expensive to handle and assemble. They should be used only where future disassembly, adjustments, or other needs make them necessary. Very short screws, bolts, and rivets are more difficult to position than are their longer counterparts, as the weight of their heads often makes them fall out of the holes. It is more difficult to place the blade of a driver in screws and bolts with round heads and slotted tops than in those whose tops are flat with closed patterns for the driver's matching blade tips.

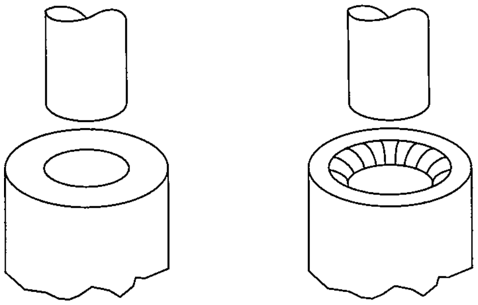
13. *Snap in.* Where possible have the parts snap together instead of being held with discrete fasteners. Plastic tabs can be added to a part while it is being molded at negligible extra cost. The operations of placing the part in position for assembly and of assembling it become a single action when it is snapped into place. See Fig. 17.7.2.

14. *Reduce variations.* Where fasteners must be used, have as few variations as possible. Seven or eight different screw sizes can often be reduced to two or three. The added cost of overdesign where a slightly larger size than needed is used is more than made up in purchasing, handling, inventory, record keeping, and tooling savings.

15. *Chamfer.* Assembly can be speeded up and less expensive tooling can be used if holes are chamfered and pins are slightly pointed and rounded. See Fig. 17.7.3.

16. *Avoid problem parts.* Flimsy, amorphous, and long flexible parts such as wire, fabric, insulation, and the like are difficult to handle and should be avoided or provided for in the process. Some coil springs tangle because of their helix and wire diameter or if they are open at their ends.

17. *Help the robot.* Give different parts a common feature so that the same gripper can be used to handle them all. Changing grippers takes

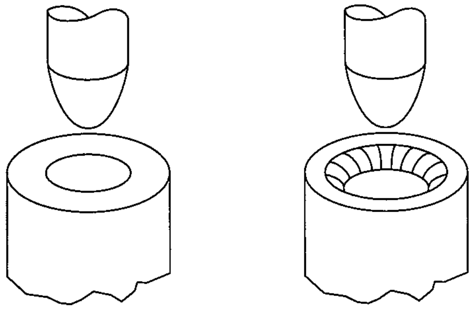


Peg not chamfered
Hole not chamfered

Not Good

Peg not chamfered
Hole chamfered

Better



Peg chamfered
Hole not chamfered

Better

Peg chamfered
Hole chamfered

Best

Fig. 17.7.3 The use of chamfers to accommodate slight misalignments.

time. Extra or complex grippers add to expense. Allow space around components so that the gripper can get in to assemble the part. Design the grippers to fit the contours of the parts or to wrap around them rather than merely hold them with pressure and friction. This will minimize slippage and movement.

AUTOFACTURING SUBSYSTEMS

Material Holding, Feeding, and Metering In the progression from raw materials and purchased parts to the finished product, the first autofacturing system subsystems are those whose functions are to hold, feed, and meter the correct amount of input to the operating stations at the correct time. In designating the holding system, the material must be described in detail. Then the required holding status must be defined. Is it to be held in bulk form or as discrete items? Is each item unique or the same as all the others? Must each item be segregated from the others or may they all be commingled at this stage? Then, the best-suited holding devices must be specified, e.g., bins, hoppers, reservoirs, reels, spools, bobbins, trays, racks. Subsequent to holding, the material or parts are to be fed into the next stage of the process. Again, a series of determinations must be made. Are parts to be fed individually or en masse, in metered amounts or at any rate, separated or connected, oriented or unaligned, selectively or randomly, pretested or untested, interruptible or nonstop? Then the most suitable feeding and metering devices must be specified, e.g., dispensers, pumps, applicators, vibratory feeders, escapements. See also Sec. 10.1.

Operations The operating stations of the system include those machines, tools, and devices that convert the raw materials to parts or finished products, or bring parts together and assemble them into finished products. It is at these stations that the casting, molding, bending, machining, soldering, welding, riveting, bolting, snapping together, painting, and the like are done. The operators of the machines may be humans or robots or the machines may be designed to operate automatically or under the direction of a computer.

Transferring When more than one operation is required, means of transferring the work in process from one machine or station to the next must be provided. In many companies, the time that material is on a machine and being processed is much less than the time it is being unfixtured from the machine, moved to the next machine and fixtured in it, with several delays, storages, and handlings in between. Work can also be transferred while on a machine, but to other tooling or test stations of it. Some such machines are the in-line (or straight line) pallet (or platen) and plain (or nonpallet) transfer machines and the dial (or turret) rotary table machines. These may move continuously or intermittently (indexing). They may be of the trunnion, center column, or shuttle-table type and may be purchased as standard or special custom designs. Robots and other mechanisms are often used to transfer work from station to station within a machine and between machines. See Secs. 10.5 and 10.6. It is important that the inputs and outputs of the individual stations and machines be in balance, in order to avoid idle equipment and work-in-process buildups.

Sensing There will be many points within an autofacturing system at which the sensing of conditions will be required. Sensing is a function which usually must be done prior to measurement, inspection/test, counting, sorting, computing, actuating/moving, feedback/control, and correction. Sensors can be used to measure position, presence or proximity, characteristics, or conditions of the parts, product, machines, tooling/fixture, equipment, people, and environment. Contact-type sensors include mechanical, electromechanical, electrical, chemical, and air jet. Noncontact-type sensors include permanent magnet, electromagnet, capacitance, inductance, reluctance (magnet and variable), sonic/ultrasonic, photoelectric/optoelectronic, laser, radio frequency, eddy current, thermal and relative expansion, Hall effect, air (pressure and fluidic), and inertial/accelerometer.

In deciding whether to use a sensor, where, and what type, the questions to ask are

1. What *should* happen?
2. What *could* happen?
3. What measurable phenomena, conditions, data, features, properties, changes, etc. will be present to distinguish item 2 from item 1?
4. When is the earliest that item 3 can be detected?
5. Where? How?
6. What is the best sensor or combination of sensors to do this?
7. What should be done with the information sensed?

Measurement, Inspection, and Test In autofacturing, it is essential that the product be inspected while on the machines and while it is being made. The objective is to make only good products. To do this, critical parameters are sensed and measured in real time. If they begin to drift beyond preset statistical limits, error signals are fed back and adjustments in the machine are made automatically to bring the variables back in control before any bad products are made. It is important, therefore, that the product be designed so that there is access to inspection and test points during processing. It also means that the machines should include sensors and inspection stations amid the processing stations.

Equipment Monitoring and Performance Maintenance In addition to the product, all of the machines and equipment must be monitored constantly to assure continued peak performance and to avoid breakdowns. Again, a multitude of sensors, clocks, and counters is a must for the measurement, feedback, correction, and control of the status of the system. For tooling, where wear is certain, automatic tool changers that pull the used tool out and replace it with a new one can be used. An operator or maintenance person then removes the used tool from the device and loads it with another new one so the tool changer is ready again. In some cases, duplicate modular workstations are kept on hand so that a malfunctioning one can be replaced rapidly.

Sorting, Counting, and Marking After the product is made, its characteristics can be sensed and measured and it can be sorted according to whatever decision rules are set. Deflection devices—such as conveyor belt gates, air jets, and the like—can be used to divert each class of product to its own receptacle. They can then be counted (with the data going to the computer) and marked or labeled, as desired.

Wrapping and Packaging The wrapping and packaging of the product is just as much an integrated part of the overall system as is assembly, and should be just as automated. Even the packing cartons should be designed to facilitate automated operations. Instead of conventional flaps that tend to close and block the robot's path in trying to put products in, boxes should have straight sides and a lid so that both the products and the lid can be handled in straight, vertical motions.

Automated Material Handling and Automated Warehousing An aim of autofacturing is to have the raw material and purchased parts flow from the receiving department to the shipping department as smoothly and with as few stops, temporary storages, and handlings as possible as they are gradually converted to finished products. To do this, automated material handling and warehousing equipment must be installed whose capacities are in balance with the outputs of the production machines. The equipment available ranges from forklift trucks and automatic guided vehicles (AGV) to robots and automatic storage and retrieval systems (AS/RS). See Sec. 10.7 and the other sections cited above for descriptions of such equipment.

Order Picking, Packing, and Shipping The automation of the factory does not end with the production and storage of the products. As orders are received, the correct number of the correct items must be picked, packed, and shipped from inventory. A perfect autofacturing system would be one wherein the production and demand for items is so balanced that products are shipped directly from the final production and inspection station. In all cases, the receipt of orders should be entered into the same computer system that schedules production so that the match between what is being sold and what is being made is as close as possible.

Operator/Machine Interfaces Even the most automatic machines occasionally require the attention of a human attendant or maintenance person. The fields of study that examine the human/machine interfaces and the degrees of effort and responses of people at work are called *human engineering, human factors engineering, biomechanics, ergonomics,* and *cybernetics.* While each concentrates on a particular aspect of the subject, they are all concerned with matching the needs and capabilities of humans to the designs of the machines (and total environment) with which they interface. Humans control machines via buttons, knobs, handles, switches, keyboards, voice recognition devices, and the like, and these must all be designed, colored, located, etc. to minimize human effort and errors and to maximize effectiveness. Machines output information to humans via lights, sounds, dials, gages, video screens, and the like. These, too, must be designed to maximize response and minimize errors. At all times, the human/machine interface must be designed with safety in mind. See Sec. 17.6.

Communication, Computation, and Control The system's robots and other equipment have a multitude of microprocessors embedded in them. That, and the fact that they also have many sophisticated sensors and heuristic programs justifies calling them "smart" machines. They, in turn, are controlled by more powerful microcomputers that report to minicomputers, thence to a powerful central host computer. The host computer does not control each detail of the system; it merely integrates and balances a distributed and hierarchical communication, computation, and control capability that uses the same database and is linked by a local-area network (LAN).

Auxiliary Systems There are several ancillary systems that can be employed to enhance automated manufacturing and the autofacturing system. Some of these are group technology (GT), concurrent engineering or simultaneous engineering, and computer-aided design and computer-aided manufacturing (CAD/CAM) in the product design and prototyping stages; just-in-time (JIT) delivery, material requirements planning (MRP), and manufacturing resources planning (MRP II) in the production planning and inventory control stages; computer-aided testing (CAT) and total quality control (TQC) in the production, inspection, and testing stages; computer-integrated manufacturing (CIM) in the communication, computation, and control stages; and management information systems (MISs) in the reporting and analysis stages.

MICRON AND NANO AUTOFACTURING

As some products get smaller for operational, energy, space, and economic reasons, they can reach the point of size reduction to where human workers cannot see them, let alone manipulate them for fabrication, assembly, and test. In such cases, microelectromechanical systems (MEMS) and nanotechnology methods and processes, rather than human methods, must be used, and an autofacturing mode is essential. These very small, very thin, very elemental parts, components, sensors, activators, devices, and products are becoming prevalent in computers, electronic circuits, communications, medicine, warfare, automotive, appliances, robots, toys, games, and other products. Micron technology deals with things in the size of a millionth of a metre (0.00004 inches). Nanotechnology deals with things in the size of a billionth of a metre (0.00000004 inches).

MEMS processes are similar to those used to fabricate integrated circuit "chips" or microprocessors, such as photolithography, thin film deposition, and etching. But where only one side of a wafer is typically processed to create integrated circuits, MEMS fabrication typically processes both sides of a wafer and sometimes a stack of them to create 3-D items. The MEMS processes are still in development, but they may be divided into four technologies: (1) bulk, a subtractive process in which material is selectively removed until the desired configuration is achieved; (2) surface, an additive process in which layers of various materials are added to a substrate; (3) hybrid, a combination process in which bulk techniques are used on surface processed substrates; and (4) high-aspect ratio, which uses deep ultraviolet lithography and x-ray techniques to create micron-sized items out of polymers, metals, and ceramic materials. Some MEMS products in existence combine sensing capabilities with information processing to detect, interpret, activate, and control systems. They are being used in automobile air bags, robots, aircraft guidance systems, missiles, and elsewhere.

In nanotechnology, the goal is to produce items much much smaller than those micron sized, striving for the range of 10 to 100 nanometres. The dot at the end of this sentence can hold approximately 350 million nanometre-sized objects. The measurement of objects in the nanometre range is not new; what is new is nanomanipulation—the creating of objects in that range. To create new, better, and useful materials and devices of this size requires mastering the science and art of observing, moving, manipulating, rearranging, and positioning individual atoms and molecules of matter. To obtain the item sought, it is built atom by atom, molecule by molecule, and causing them to arrange themselves in the structure desired. To accomplish this not only is a full autofacturing mode essential, but special apparatus such as scanning probe microscopes (SPM), scanning tunnel microscopes (STM), atomic-force microscopes (AFM), in addition to photolithography and soft lithography having smaller feature sizes, and others.

There are two approaches being pursued: nanofabrication, also called nanoengineering, which is the building of the item with man-made tools; and self-assembly, which is causing the atoms and molecules to move together and adhere in the desired way and shape, based upon their size, shape composition, and physical properties—much like nature causes soil, water, air and the energy of the sun to combine to form trees and vegetation. Some areas where nano items may find application are in pharmaceuticals, chemicals, energy, agriculture, food, bar codes, communication, electronics, avionics, transportation, motors, military, materials, and a wide range of everyday products. One problem with creating such small objects is that atoms do not always act according the laws of classical physics; at the nano level quantum mechanics phenomena may occur.