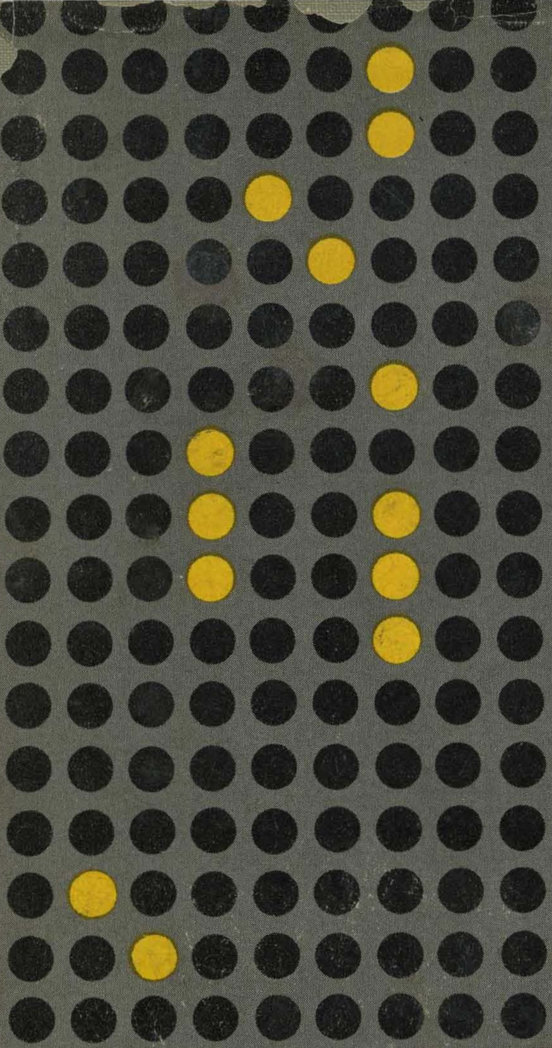
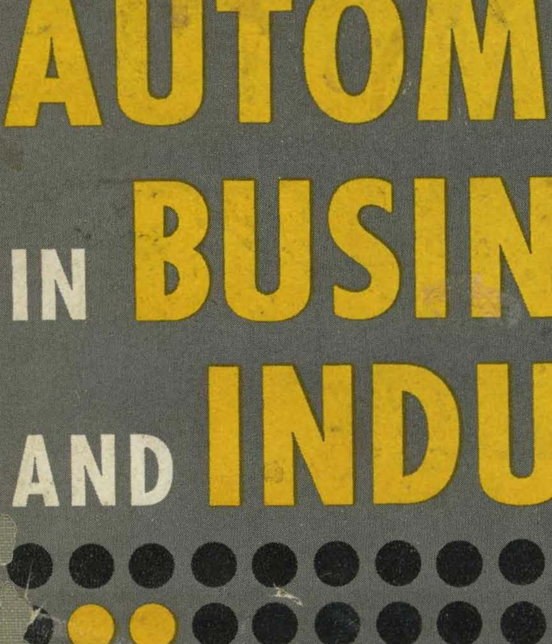


Eugene
M.
Grabbe
Editor

AUTOMATION IN BUSINESS AND INDUSTRY



JOHN WILEY & SONS, Inc.
Publishers

AUTOMATION
in
BUSINESS
and
INDUSTRY

THE AUTHORS

L. M. K. Boelter

Simon Ramo

E. M. Grabbe

John L. Barnes

Harold Chestnut

Herbert W. Ziebolz

Stanley Fifer

Willis H. Ware

John W. Mauchly

Bernard S. Benson

George G. Bower

Walter F. Bauer

Cuthbert C. Hurd

Duane T. McRuer

L. K. Lee

C. G. Laspe

E. L. Harder

John L. Bower

R. W. Bolz

Frank K. Shallenberger

Dean E. Wooldridge

AUTOMATION
in
BUSINESS
and
INDUSTRY

EDITED BY EUGENE M. GRABBE

Senior Staff Consultant
on Automation
The Ramo-Wooldridge Corporation
Lecturer in Engineering
University of California
Los Angeles, California

NEW YORK · JOHN WILEY & SONS, INC.
London · Chapman & Hall Limited

Library of Congress Catalog Card Number: 57–5914

Printed in the United States of America

The Authors

L. M. K. BOELTER, Professor of Engineering, Dean, College of Engineering, University of California, Los Angeles

SIMON RAMO, Executive Vice President, The Ramo-Wooldridge Corporation, Los Angeles, California

EUGENE M. GRABBE, Senior Staff Consultant on Automation, Computer Systems Division, The Ramo-Wooldridge Corporation, Los Angeles, California

JOHN L. BARNES, President, Systems Laboratories Corporation, Los Angeles, California, Professor of Engineering, University of California, Los Angeles

HAROLD CHESTNUT, Project Engineer, Aeronautic and Ordnance Operation, General Electric Company, Schenectady, New York

HERBERT W. ZIEBOLZ, Vice President in Charge of Engineering, Askania Regulator Company, Chicago, Illinois

STANLEY FIFER, President, Dian Laboratories, New York, New York

WILLIS H. WARE, Assistant for Engineering, Numerical Analysis Department, The RAND Corporation, Santa Monica, California, Visiting Associate Professor of Engineering, University of California, Los Angeles

JOHN W. MAUCHLY, Director, UNIVAC Applications Research Center, Remington Rand UNIVAC, Division of Sperry Rand Corporation, Philadelphia, Pennsylvania

THE AUTHORS

BERNARD S. BENSON, President, Benson-Lehner Corporation, Los Angeles, California

GEORGE G. BOWER, Systems Engineer, Riverside Research Laboratory, Motorola, Inc., Riverside, California

WALTER F. BAUER, Head, Digital Computing Center, Computer Systems Division, The Ramo-Wooldridge Corporation, Los Angeles, California

CUTHBERT C. HURD, Director of Electronic Data Processing Machines, International Business Machines Corporation, New York, New York

DUANE T. McRUER, Control Specialists, Inc., Inglewood, California

L. K. LEE, Technical Adviser to Director, Engineering, Research and Development, Mechanical Division, General Mills, Inc., Minneapolis, Minnesota (formerly Manager, Advanced Techniques Laboratory, Stanford Research Institute)

C. G. LASPE, Member of Technical Staff, The Ramo-Wooldridge Corporation, Los Angeles, California (formerly Chief Instrument Engineer at Shell Oil Company, Wilmington Refinery, Wilmington, California

E. L. HARDER, Director, Analytical Department, Westinghouse Electric Corporation, East Pittsburgh, Pennsylvania

JOHN L. BOWER, Engineering Specialist, Autonetics, Division of North American Aviation, Inc., Downey, California

ROGER W. BOLZ, Editor, *Automation,* Penton Publishing Company, Cleveland, Ohio

FRANK K. SHALLENBERGER, President, Shalco Engineering Corporation, Palo Alto, California, Professor of Industrial Management, Graduate School of Business, Stanford University, Palo Alto, California

DEAN E. WOOLDRIDGE, President, The Ramo-Wooldridge Corporation, Los Angeles, California

Preface

An engineering extension course having the same title as this book was given in the spring of 1955 at the University of California in Los Angeles, San Diego, and Berkeley. The course, consisting of a series of lectures by prominent engineers and scientists, proved very successful, having a total attendance of 735. Its purpose was to review for engineers and management from business and industry the present status of developments and applications in the field of automation. It is hoped that publication of these lectures in book form will supply for a much larger audience a wealth of information and details on the fundamentals of automation, new developments in automation techniques, and descriptions of automation systems applications. Each author in answering the questions "What is new?" and "How is it used?" has reviewed automation developments in his special field with liberal use of examples. By having many participants, a broad coverage has been attained; and although the treatment is general rather than technical, much detailed and valuable information is included.

The course was organized to show how the fields of feedback control theory, instrumentation, analog and digital computation, and data processing are now becoming integrated as automation is applied on a broad scale to control systems that encompass the range from top management to individual machines. Emphasis was placed on new developments and applications of control systems that can perform both complex control functions and data processing. As a result considerable attention is devoted to electronics, computers, and data processing.

Components are treated early in the series and systems later. The subject sequence for publication is the same as for the lecture course; the book may therefore be used as a text for similar courses.

The concept of a course of this type came about as a result of organizing material for a *Handbook of Automation, Computation, and Control,* which Simon Ramo, D. E. Wooldridge, and I are editing. Dr. Ramo first proposed the course to Dean L. M. K. Boelter of the College of Engineering, Los Angeles, University of California, and under his general supervision a committee was set up to plan the course. This committee included members of the faculty of the University of California at Los Angeles and Berkeley and interested people from industry. Simon Ramo served as Chairman; others were J. L. Barnes, L. M. K. Boelter, S. Lee Cahn, J. C. Dillon, E. M. Grabbe, Earl Jones, M. P. O'Brien, T. A. Rogers, L. G. Walters, Ernest Wade, and D. E. Wooldridge.

I was requested to coordinate the course and edit the lectures for publication, and arrangements to publish the lectures were made with John Wiley & Sons, Inc., which will also publish the *Handbook.* The arrangements concerned with giving the course were handled by J. C. Dillon of the Engineering Extension and his staff.

In writing on a new subject such as automation, each author feels compelled to give his own definition and views on the development of automation. This results in some repetition but serves as a springboard for the author's treatment of the subject. Uniformity of style and level of discussion cannot be attained in a volume having so many authors. In many instances the authors have covered more material than it was possible to present in the delivered lectures. S. N. Alexander delivered the lecture on "Input-Output Equipment," and W. F. Bauer prepared the chapter on this subject. The chapter on "Automatic Control of Flight" was written by D. T. McRuer; the lecture was given by A. C. Hall. B. S. Benson and G. G. Bower are joint authors of the chapter on "Analog-to-Digital Conversion Units"; the lecture was given by B. S. Benson. Dean Boelter has kindly provided an introduction entitled "Reflections on Automation."

I am grateful to Dean Boelter for giving me the opportunity of editing this lecture series for publication. The lecturers and authors have been very cooperative in preparing material in accordance with the course objectives and in completing the manuscripts. I appreciate the work of Mr. Dillon, Mrs. Park, Mrs. Virginia O. Grant, and Dr. J. F. Manildi in making this course a success.

E. M. GRABBE

January 1957

Foreword

Modern engineering is becoming more and more concerned with the development of large and complex systems for business, industry, and military uses. Automatic control devices and automatic data-processing machines of great flexibility are essential ingredients of such systems, which cut across traditional engineering lines. Engineering developments in one field may have application in another field, so that it is more than ever necessary to provide information flow between management and engineering as well as between the various fields of applied engineering.

To this end the Departments of Engineering, University of California, Los Angeles and Berkeley, arranged a lecture course designed to provide up-to-date information on developments in automation. The series consisted of invitation lectures presented by nationally prominent engineers and scientists from industry, business, and universities.

We are pleased to share this stimulating experience of the *Automation in Business and Industry* lecture series with you, the reader, through the pages of this book.

L. M. K. Boelter
Professor of Engineering
Dean, College of Engineering
University of California
Los Angeles, California

M. P. O'Brien
Professor of Engineering
Dean, College of Engineering
University of California
Berkeley, California

Contents

5. BASIC CONCEPTS OF INDUSTRIAL INSTRUMENTATION AND CONTROL 89

6. ANALOG COMPUTERS 132

Introduction: Reflections on Automation

L. M. K. BOELTER

University of California
Los Angeles, California

HISTORICAL DEVELOPMENTS

The trend toward automatic operation of machinery and processes, indeed the control of all equipment in which motion exists, has been hastened by the attempt to eliminate drudgery from the contribution of the labor component of industry, as well as by the phenomenal developments in instruments, computers, and control devices, both electrical and mechanical, and the inability of man to manually control the product (in its intermediate or final forms) at the greater speeds and complexity resulting from mechanization. The word "automation" stems from "automatization," which is difficult to pronounce and spell—thus the simplification (1).

The transfer of energy, matter, and intelligence and their transformations within and across these three divisions of nature has occupied man from his beginning. The control of the transfer and transformation processes rested within the human (and other life forms) for a long period of time. During this long period, the power available to man per human "controller" increased (through the use of groups

---

[1] From an address presented at the Symposium on Electronics and Automatic Production, jointly sponsored by the National Industrial Conference Board and Stanford Research Institute, San Francisco, California, August 22, 1955.

of men, for instance, manning the oars of a ship; through the use of other natural forces, such as the aerodynamic force on a ship's sail; through the employment of animals as sources of work, their use as beasts of burden and as draft animals).

The domestication of plants and trees (and animals) is also a manifestation of man acting as both a "programmer" and a "controller." Because the early farmer programmed and controlled a segment of the ecological cycle in the face of the uncontrolled and unpredictable forces of weather and of animal intrusion, his stature probably increased rapidly.

The transfer processes available were increased (wheel, combustion, controlled flow of water, and others), but their control for man's benefit rested within the nervous system of man. From another point of view, automation may be considered as removing certain of the elementary control tasks from man and accomplishing them through "external" mechanical and electrical devices.

To summarize, the first step in mechanization was to relieve man of certain of his power-generating duties, and the second was (and is) to relieve him of certain of his mental tasks and the related physical tasks.

Relieving the human of tasks (and the corresponding responsibilities) is in itself not a blessing. There are, however, many substitutions for required tasks (especially those that are repetitive) which can be made at the initiative of the individual and which will result in a richer life for him and for the society of which he is a member.

The functions of control and of producing work are closely interrelated in man both physiologically and psychologically. Certain curious intermediate steps in the automation sequence have resulted from this carryover. The whip socket on early automobiles, imaginary robots which resembled man physically, the first incandescent-bulb glass-blowing unit which simulated the sequence utilized by the human glass blower are cited.

Automation in the ultimate implies that a sequence beginning with an input (say, raw material) and proceeding to an output (say, a finished product) of predetermined properties and characteristics will be accomplished without human labor or direction, other than to design the equipment and the process, initiate and stop the sequence, and repair and maintain the equipment. Brunetti (2) correctly states that no development is underway or is anticipated that will make it possible for industry to operate without workers. Two bases are cited: first, the higher-level administrative decisions must be made by man; and second, man can often, for certain types of tasks, produce more work per unit cost (with the implication that the activity is not deleterious to health or well-being) than the machine.

The "higher-level decisions" will be made by man primarily because he, through the exercise of his mind, possesses the only means of integrating and interrelating data for which rational formulations are not yet possible or for which the formulation would be too expensive. Referring to the second point of the preceding paragraph, there will probably always exist situations in which spatial and temporal densities of transformations and of transfer processes will favor the use of manual labor. The use of the human body for the creation of work is natural and proper, subject always to the limitations imposed by the physiological and psychological requirements inherent in the characteristics of the individual and of his environment.

SYSTEMS ENGINEERING

The modern version of the technique of acquiring the behavior characteristics of systems is called operations research or operations analysis. The design of these systems (and the collection and formulation of the requisite data) has been incorporated into the discipline called systems engineering. Techniques and disciplines are included here that do not necessarily find themselves in the curricula, courses, texts, and literature of engineering. Engineering, to a large extent, has in the last fifty years concerned itself more intimately with the components (or elements) of the system rather than with the system itself.

One of the segments of systems engineering is the optimization of the particular sequence which defines this system (including its boundaries and the "flow" across them) for cost, time, materials, and men (3). Optimization with respect to man cannot be performed analytically but, for that reason, cannot be eliminated as a design desideratum. Design of the automated system must include considerations of the effect on all men who are an integral part of the system (customers, workmen, supervisors, stockholders, families of members of each group, etc.), and their relation to this system must be optimized in a manner similar to that utilized for the nonautomated system.

Another aspect of the problems of systems engineering is that they often lie across boundaries of units established in our society. The size of the subunits of our industrial society and their characteristics and interactions introduce difficult problems for which little data exist and which are not necessarily the immediate concern of any one of the subunits. For instance, air pollution cannot necessarily be considered to be the responsibility of the operators of factories, the individual drivers of automobiles, and/or others. Too often the responsibility for the solution of the "horizontal" systems problems is accepted by government through default. The basic data for solution as well as for enforcement may not be available. Thus the character of government is being changed through the acceptance of the newer horizontal tasks (our forefathers included as governmental responsibilities those tasks of the same kind which were then pertinent). Automation will, in my opinion, still further complicate the tasks of government unless additional and possibly new forms of private enterprise are conceived and established.

SOCIAL IMPLICATIONS

The introduction of automation has been designated as a second industrial revolution by Wiener (4) but has been revealed as a gradual and an old process, with the nervous system of man as central, by Brunetti (2). The point to be made is not that these are opposing views but that an acceleration in the automation of industrial processes is possible, if economical, and is often necessary for safety considerations, through the substitution of electrical and mechanical sensors and effectors and the use of computers. The effect on managers, workmen, their dependents, and on society will be good or bad, depending on the rate of introduction of automation; the manner in which the introduction is related to the existing mores, regulations, etc.; the extent to which long-range plans have been developed; and finally the extent to which the knowledge of the impending events has been disseminated.

For example, one characteristic of our present society is the mobility of many members thereof. However, mobility tends to weaken family and group bonds and, if carried too far, can and does threaten the stability of communities and states. Thus the rate of introduction of automation, if it causes men to seek other forms of employment in other places, should be accomplished in consideration of the forces of population mobility.

On the other hand, if automation will result in the same or greater employment levels but will require that many employees change their jobs and vocations, a reasonable learning and adaptation period should be anticipated. A change from physical tasks to supervisory or "white collar" tasks will often cause difficult if not impossible adjustments; again repair and maintenance tasks do not always have the appeal that results from a contribution to a new commodity. Repair tasks, on the other hand, involve a degree of detective work (trouble shooting), which is an exercise of the logical process and is therefore more appealing to certain men than repetitive tasks such as assignments on an assembly line.

Automation can be applied to small operations and thus may become a positive force in population dispersal. Needless to say two devices introduced within the last half-century, the automobile and the telephone, have not been utilized by man to affect or maintain dispersal. Steinmetz, in the 1910's, proposed the small unmanned automatic generating station as a part of a power network to utilize more completely the first-order power potential of our water resources. The point to be made is that automation may be utilized as a force of dispersal or as a force of concentration and that the decision about the sign of the force lies outside the automation program.

EDUCATION AND AUTOMATION

A very important contributory item of our society is the demonstrable skill that has been acquired by each individual. These skills may be divided into two classes, mental and physical. In certain classifications of tasks there has been underemphasis on the mechanism of transfer of these skills. In the first place, the incentive for acquisition may not exist, and, in the second place, the formulation for effective teaching (transfer) has not been accomplished.

The difference between teaching methods employed in a master-pupil relationship and those employed in the teacher-student relationship requires much investigation and attention. Frequently the area of teaching between instruction for skills required in the crafts and instruction leading to professional engineering has been neglected. Broadly, we may distinguish these two bands of technical instruction by the phrases training-for-the-job and training-for-the-profession. Both bands should contain a strong overlay of general education.

The purpose of the discussion above has been to raise the question whether automation will cause still further maladjustments in the learning and educational process. For instance, the child may feel that he does not need to know arithmetic because the computer will solve all arithmetic problems for him, or that man will cease to walk and substitute sitting (a most unnatural position) for standing, the latter being the zero limit of walking and requiring some of the same muscles.

In the opinion of the writer, no psychological (including educational) and physiological problems need arise on the introduction of automation if the proper precautions are taken at all levels and in all parts of our social structure.

PRODUCTION AND DISTRIBUTION

Automation will tend to reduce the rate of consumption of our natural resources, if only because automation usually forces a change in the direction of decreasing entropy increase in the system and/or process. But, in addition, an increased awareness of the problem of the finite extent of our resources and the almost unlimited absorptive capacity of the world population for the products derived from these resources will be a result of the analyses which underlie design for automation. Needless to say, the present rate of consumption of the natural resources of the world deserves the utmost attention.

The distribution function and its correlative service subfunctions of our industry present an interesting situation consisting of high-speed, apparently efficient operation; but when it is viewed, for example, from a point of concentration (5), both in material and time, of heterogeneous materials, an astounding improvement is seen to be possible. The ratios of retail selling prices to the costs of production (including the costs of raw materials and all costs to the "factory door") reveal the magnitudes of the improvements that can be effected. Although automation is often thought, in the first instance, to be applicable to production, the point will be made that it can serve (and has served) the distribution function admirably, even if applied in the most elementary manner.

The first contribution to the distribution function can be made by facilitating the flow of information. The development of an information flow system involves at least three important considerations: (1) minimization, including selection, of information to be stored and/or transmitted; (2) timing of the transmittal; (3) accuracy of transmission process (a minimum increase of entropy).

The normal tendency as instrumentation and computers become available is to obtain "too much" information. Thus a relatively new discipline must be further developed. The distributed selective process of the ages is not availiable to us in this connection to aid the selection of the "classics." Also, the distribution of information is a two-way process. In a given organization and throughout society a studied effort must be made if all facets of pertinent information are to be available to those who participate in the contributing events.

The transport of man and goods must be attacked as a problem of systems engineering with a view to decreasing the time and cost of transport and increasing its effectiveness.

One general problem is the provision of alternative methods and paths that will keep a system operative although a temporary local dislocation has occurred. The design (which includes economics) of such systems including a temporary alternate or alternates deserves thorough analysis for both the production and distribution functions. The operation of a production line or of a freeway are examples that have pertinence here.

Automation will require the expenditure of large sums for system (including process) redesign as well as for the design of the automatic features. The accumulated man-energy required translated into capital investment can readily serve as a barrier to change and the consequent industrial progress. Already there appears to be evidence that there exists a considerable negative force to rational "advance or change" in commodity items which are mass-produced and/or widely advertised.

CONCLUSION

All technological developments can be used by man for good or for evil. Degrees of good can also be identified in terms of the character of the utilization of the developments.

The manner of utilization and the rate of introduction are ultimate decisions that can only be made after thorough study of all facets of the social impact of the developments. In general, for the larger aspects of automation, the decisions will require "horizontal" organizations (often temporary in nature) in which all parties, groups, and individuals concerned will contribute to the data, its analysis, and the distribution of the resulting information. Through the use of this device, or a better one, automation can be made a force for momentary and ultimate good.

The predicted "second" industrial revolution will then be a gradual acceptable industrial evolution (a significant part of which we are witnessing as contemporaries). Such can be the process of automation provided we act wisely, basing all decisions upon the accumulated, properly formulated, knowledge of past experiences (in particular those of the "first" industrial revolution) and then extrapolating the conclusion into the future, modified by such changes as characterize our present society and its attitudes in terms of its antecedents.

REFERENCES

1. John Diebold, *Automation, The Advent of the Automatic Factory,* D. Van Nostrand Co., Inc., New York, 1952, p. ix.

2. Cledo Brunetti, The Meaning of Automation. Address delivered to 82nd Annual Forum, National Conference of Social Work, San Francisco, May 29–June 3, 1955.

3. L. M. K. Boelter, Some Thoughts on Evaluating Research and Development, I. R. Weschler and Paula Brown, Editors, *Evaluating Research and Development,* Institute of Industrial Relations, University of California, Los Angeles, pp. 7–12.

4. Norbert Wiener, *The Human Use of Human Beings,* Cybernetics and Society, Doubleday Anchor Books, Doubleday & Co., Inc., Garden City, N. Y., 1954, p. 136.

5. The S.S. Warrior, An Analysis of an Export, Transportation System from Shipper to Consignees, Publ. 339, National Academy of Sciences, National Research Council, Washington, D. C., Nov. 30, 1954.

1. Automation in Business and Industry

SIMON RAMO

The Ramo-Wooldridge Corporation
Los Angeles, California

1.1 THE PRESENT STATUS OF OUR TECHNOLOGY

It is becoming a commonplace these days to hear that we are engaged in a second industrial revolution, the era of automation. This time it is not man's muscles that are to be replaced and extended, but rather man's brains. More specifically, what is meant is that in business and in industry the functions accomplished by people are going to be accomplished more and more by complete networks of complex automation systems, including a host of devices spanning large geographical distances and broad functional operations. These systems will employ what might be called "synthetic intelligence" devices that replace or extend man's brains and senses in a wide variety of the pursuits in which humans are engaged. And anyone who is foolhardy enough to challenge the idea that the replacing of man's brains will be the top industry in the nation some years hence is in danger of having his brains among the first to be replaced.

It is easy to see why there is so much talk about another industrial revolution. Nucleonics may give us even greater sources of energy to control, advances in medicine may decrease disease and increase longevity, but neither of these occurrences suggests so radical a change in our civilization as the advent of the mass production of automation devices that can compete with the human brain and senses.

There are some important reasons why something like this can be expected to happen, and there are some important reasons why it will not happen overnight. Indisputable forces are pushing toward rapid and wide development of automatic systems for filing or remembering, sorting, computing, comparing, decision making, and all the other functions usually performed by people.

First of all, the times are technologically right for the development of synthetic intelligence devices for automation. There are scores of aids to business and industry and transportation that present technological art is capable of providing without a single new discovery in basic science. It would be possible for engineers today, on the basis of known pure science, to design and produce devices that could displace a very large fraction of the white-collar workers in business and industry who are doing jobs in which the proportion of their intellectual capabilities used is rather small. These people are engaged in routine paperwork or routine intelligence transfer assignments which can be reduced to simple equivalent thought processes that electronic machines can and do handle better and faster.

Moreover, modern technology has advanced to the point where it is possible for instruments to be designed to measure continuously various conditions and phenomena important in industrial processes and operations. These measurements can then be compared with previously set values, and automatic controls actuated to bring the process or operation closer to the desired condition. A very much wider range of automatic control could be provided in almost every business and industry by utilizing the electronic techniques and components now available.

A perfect example of the status of scientific and engineering art is the military. For various reasons, the military today has a tremendous need for synthetic intelligence devices. To understand this, we need only to consider the enormous importance of control of the air, both to safeguard our own country from enemy bombardment and to make possible retaliation. With tremendously increased speeds and the need for operation under all weather conditions, it has become necessary for the military to consider guided missiles in which the human pilot is omitted and the problem of destroying the enemy in the air is left to electronic brains which will guide the missile-carrying airplane, find the enemy, close in on him, predict where he will be despite his maneuverings, and destroy him. For such an operation, the electronic brains must have memory, stored intelligence, the ability to compute and make decisions at high speeds, and the ability to control aircraft on the basis of the decisions. Machines with these capacities are being developed for the military today because our basic science permits it. The engineering art is developing, in other words, without having to wait for another Einstein to come through with a second and more complex theory of relativity.

The most important ingredient of modern military technology is probably not, as is so commonly supposed, the availability of even more powerful explosive powers contained in a single bomb, but rather the trend of operational war toward more and more complex automatic integrated huge systems. Military applications demand that the available science be reduced to engineering applications in a period of several years. Business and industry usually take a longer period in reducing new ideas to practical applications, but they are benefiting greatly by military know-how which is released to them.

1.2 THE NEED IN BUSINESS AND INDUSTRY

In an important way, business and industry in America today are ripe for the present surge in automation. Business and industrial organizations are complex and are rapidly becoming even more complex. Instead of hundreds, thousands and tens of thousands of individuals' activities are now in need of close coordination to make for a successful industrial operation. This necessitates a tremendous amount of red tape and paperwork.

Whether the operation is an air line or factory, an insurance company, a bank, or a department store, everyone is conscious of the enormity of the red-tape problem. A larger and larger fraction of the personnel seems to be engaged in moving information, papers, directives, and plans about, and a smaller fraction seems to be engaged in the substantive matters with which the business at first glance appears to be primarily concerned.

These are some reasons why this revolution in the replacing of man's brains in business and industry can be expected and is indeed under way. Technologically we are at a point where great strides can be made. The military situation in the world is sparking the program and, in effect, is sponsoring and financing the development of techniques that will have application to nonmilitary systems as well. The growing size and complications of business and industry, and the effect of relationships with government and with labor, increase the need for these devices. Accordingly there is a demand to get on with the job of producing new automation devices and automation systems.

Merely to refer to what is happening as a second industrial revolution is an oversimplification that is about as wrong as it is right. Perhaps, more accurately, it is that, having passed strongly into the age of the utilization of our increasing knowledge of nature, we have succeeded in making a highly technical world. The more technical the world, the more complex all of its operations become in so far as interaction, speed, and controls are concerned. Whether we fight wars or whether we try to supply each other with all of our demands in a peacetime economy, we are advancing so rapidly that the orderly control of the operation is getting beyond us, unless we can increase the effectiveness of our population, not so much any longer by more hands and power, but by more brains and senses, even if the artificial brains and senses are in many ways highly inferior to those with which the average human is endowed.

1.3 TECHNICAL PROBLEMS

Suppose, for the sake of argument, that some years hence we really make a major change in the way American business and industry operate, and then we examine the whole operation and compare it with what we have today. We might discover that 50 per cent of all of the white-collar workers of the United States were engaged in performing different functions as a direct result of the machines' having been brought in to perform some of the functions that people do now, reserving the human workers for a higher type of activity on new projects, and perhaps a more leisurely work week. We have only to consider the number of workers and the number of machines this implies to realize that what we are talking about involves many billions of dollars in new and complex equipment. Such a development would dwarf, for example, the present telephone system or the present television system, including the home receivers in the nation. It cannot come about without a great deal of standardization of parts, and this in turn cannot come about without our understanding the interrelationship of all of these functions of business and industry with practical engineering possibilities. It has taken many, many thousands of man-years to bring into being our telephone system, our power system, our transportation system.

Although such a major development might be theoretically possible from the standpoint both of science and economics, it is going to require a great deal of high-grade systems engineering; and, since systems engineering is in very short supply in the nation, it is going to mean the training of scientists and engineers in substantially different disciplines. The mere relationship of people to the new systems will be a subject requiring a great deal of serious, careful scientific attention for many years before it will be understood well enough for the practical widespread wholesale adoption of techniques that will radically change the operations of business and industry in regard to their use of people.

There is today in the United States a shortage of scientists and engineers. This is particularly true if we speak of the development engineer, the research scientist, the originator—the high caliber scientific worker able to originate the very new. From men everywhere—industrial leaders, military leaders, government spokesmen, university presidents—we hear of the alarming shortage of able technical people in comparison with the nation's needs, and this situation is expected to remain the same for some years before it can possibly be changed. The importance of this shortage in the automation field is, in my opinion, as great if not greater than in any other field to which the applied scientist could devote himself, for the difficulties and the size of the problem are second to none among the technical problems of the day.

So we must create during the next several years new kinds of engineers and scientists. When the revolution has finally taken place—whether it be in five years, which I say is too optimistic, or in twenty or thirty years, which is closer to the truth—we shall note that the universities are graduating control engineers and synthetic-intelligence experts who are products of an entirely new curriculum which combines the study of the human brain, physiologically and psychologically; the physical sciences and mathematics; business, industry, and economics; and the requirements of government rules and regulations.

A shortage of scientists and engineers, the need for a host of new developments, the strong emphasis on the systems problems as against the development of specific specialized components, all of these factors tend to work against the development of automation on a large scale. The net result, of course, will simply be that the field will develop as rapidly as it can, and the pace will be such that at least most of those who are reading this book will be able to notice the development as it progresses. It will not occur one night during some hour lost in changing over from daylight to standard time.

1.4 NEW TOOLS FOR AUTOMATION

There are two fields involved in the automation area. One of them is *process control in the factory.* The evolution of the automatic factory will be gradual. New developments will provide unusual precision-measuring devices and computer devices to monitor the process, adapted to the particular job. It would be foolish to provide, for those narrow applications, involved electronic brains that have broad powers. So the brains will be specially geared to special jobs.

The second major field of automation devices might be characterized as *business data handling*—the handling of paperwork in large organizations, whether it be inventory, production controls, customer's bills, invoices, or credit accounting. The problems encountered concern memory, sorting, and simple computations in general. The goal here is the development of huge standard systems that are logical and reliable and will not break down. With sufficient study of both the businesses and the electronic techniques, we shall evolve general solutions that will have wide applications. Standard black boxes will be cabled together; the situation will be entirely similar to that of our telephone networks, in which it has gradually become possible to design standard equipment that is reliable despite the millions of input and output points.

If we are to attempt in this book to provide the reader with some appreciation of the scope of this field, then of course we must be sure that we cover its main aspects. In particular, we must be certain to show that we are involved not just in the design of gadgets that record something previously written down, or that sense something in a physical process previously handled entirely by human beings. We must include the larger-scale aspects of data handling of large complex operations and of interactions and intercommunications among the elements of the system.

When we begin to focus our attention on these larger engineering problems, we find that certain tools are quite indispensable. We must inevitably learn something about feedback, information theory, statistical and random phenomena, digital computers, and data handling.

Feedback is regarded by many as the most fundamental concept in this field. To illustrate, let us take an exceedingly simple business, that of a newsboy selling papers on a street corner. If the boy tries to be quantitative about his operation, he will tend to obtain for his initial supply a number of papers that will bear some relation to his expected sales. If he sells far fewer than he planned, the difference is going to be used by him as an indicator of the number of papers he should obtain the next day. He could easily place so much emphasis on this profit indicator (i.e., have his gain set so high) that, if he undersells a few papers, he becomes concerned about the surplus and the next day orders very few papers. Then he may find that he cannot begin to supply the demand, whereupon on the third day he may order an even larger number than the first day and then have even more left unsold. This could bring him to such an impasse that on the fourth day he may ask for no papers at all and for all practical purposes be out of business.

It is difficult to imagine any process not involving feedback that is of any interest to us in business and industry. The thermostat of a house is often used as an example of "closed loop" feedback control. Suppose that we deliberately design an "open loop" temperature control for a house. To do this we would estimate the average amount of heat the house ought to have during the winter months and fix a time cycle to turn the heat on and off. The observed result, of course, will be that during the winter it will be either too hot or too cold some fraction of the time, even if we are very competent in our estimating. When it is too hot, the people living in the house are likely to open the windows and let the heat out. When it is too cold, since they are not able to do anything about the furnace, they will put on more clothes. In effect, they are going to take actions based on the difference between what the temperature ought to be according to their indicators, and what it actually is, i.e., they will develop a new feedback loop to control the situation.

With feedback control, of course, we have the possibility of instability, oscillations, runaways, and overshooting. The problem becomes more complex as we relate a large number of factors. Let the newsboy attempt to gage the number of papers to be ordered, for instance, by noting not only the difference between the number he ordered and the number he sold on any specific day, but also whether the day was a rainy one and whether some startling news event was reported. Feedback and instability are part and parcel of complex systems, and they are hardly a specialized and narrow characteristic.

In a similar way, we cannot expect to understand and design for automatic control of an operation without being able to be quantitative about the amount of information that must be handled. Modern communication theory has made it clear to engineers in that field that information can be given quantitative definition, and this becomes an essential parameter in the design of communication systems. But it goes further than communication. If we are storing data, or if we are developing a system for human operators in which they record facts to use later, the system that is set up is very much a function of the quantitative nature of the information that is to be stored. The old problem of how many weighings are necessary to pick the one bad penny out of twelve illustrates this well. The average person confronted with this problem would go through a trial-and-error operation, and when he finished he would not be certain whether he had arrived at the minimum number of weighings. The information theory expert applies a simple formula and knows that three weighings should be sufficient to select the one bad penny out of twelve.

If we wish to design systems that take stored information and can continuously generate new data and arrive at logical conclusions in accordance with preset rules, we soon find that such design is hopeless by trial-and-error method and we must use symbolic logic or its equivalent.

Random and statistical phenomena must be understood by the engineer who would involve himself with complex systems engineering design. First of all, it is true that, in dealing with large-scale operations, we cannot expect to measure everything with absolute definiteness. We must deal with probabilities. Some of the basic parameter data will be given in the first place in terms of probabilities—that is, we can only know some statistics or make some assumptions about the probabilities of certain aspects of the whole operation, and all we ask in the performance of the system is that it achieve certain probabilities of certain results.

There are two concepts to distinguish here. One is statistics of the kind that the insurance companies have caused us to be familiar with. These statistics deal in terms of average life expectancy, average percentage of women who are blondes, etc. Here we are thinking primarily in terms of specifying initial conditions or answers in terms of these probabilities. The other concept has to do with accuracy and noise. Noise is often used as a popular term for the errors in the system and its parts—human operators, machines, communication links, storage devices, and instruments. All will have errors. These fluctuations bounce around a system, sometimes adding, sometimes subtracting, but always putting any observation on a system in some measure of doubt and always interfering with the overall performance. Oftentimes in design we must assume or experiment to determine the average level of the noise. Again we cannot expect to define the noise as a definite quantitative signal. If we could, this would be the same as saying that we do not really have any noise, for we can subtract or add to compensate for known disturbances. We can only then talk about, measure, and deal with the noise as a statistical phenomena, and a good bit of the time we assume that the various noise sources are randomly oriented with respect to each other.

It is obviously important to be able to design quantitatively with respect to noise in a system, and it will affect a system in quite different ways, depending on the nature of the function. Noise in a guided missile system in which the pilot is replaced by some synthetic device will cause the missile to miss at least some of the time. Noise in department store accounting may cause the wrong bills to be sent out or mislead the purchasing agent concerning the inventory. Inaccuracy in a bank's records is often considered intolerable. In any case, a bank may be able to be satisfied with an error of one card in a million, but not, if in listing its million accounts, the names and the monies are shifted by just one place.

1.5 CONCLUSION

So great a part of what can be done and what will be done in the coming years and decades depends on the unusual flexibility and speed of electronics techniques that we certainly must become familiar with certain aspects of the data systems. Accordingly, a glimpse at the problems of automation from the standpoint of the overall systems engineering should afford us with an appreciation of the underlying physical and mathematical concepts, an inspection of some of the typical apparatus and components, and finally a look at some complex systems.

Perhaps the most glaring shortcoming of today's art is our lack of quantitative understanding of human beings as part of a complete system that includes both the machine and the human being. If we go to the other end of the spectrum, however, and look at devices and embryo systems, we find that we are only in the beginnings of tailoring smaller subsystems to specific tasks in business and industry. We do not yet find it easy to extract the main qualitative and quantitative concepts and patterns that should basically control design. We do not yet understand business and industry very well in regard to how electronics might best be used. We are designing equivalents for existing systems or pieces of systems. We do not yet know very well what and how reliable the economics of the new equipment will be.

2. The Language of Automation

E. M. GRABBE

The Ramo-Wooldridge Corporation
Los Angeles, California

2.1 LANGUAGE AND MENTAL IMAGES

In general the object in establishing terminology is to translate a mental image to language so phrased that it will re-create the original concept in the mind of the reader. The image created by a definition depends greatly on the reader's association with the words. A story demonstrates very clearly the concept of a mental image associated with a word. A schoolboy asked his father to tell him the difference between the words "irritation" and "exasperation." The father replied, "It is hard to describe the difference in meaning of these words, but I can show you." So he asked his son to pick a number at random from the telephone book. The son did so, and the father proceeded to dial the selected telephone number. A woman answered, and the father inquired, "Is Maurice there?" The woman politely replied, "There is no Maurice at this number. I'm sorry, you have the wrong number," and hung up. Whereupon the father dialed the same number again and inquired, "Is Maurice there?" This time the woman was very curt in saying, "Young man, I recognize your voice, you called just a few moments ago. You have the wrong number again. I am a very busy woman. My husband is coming home any minute now and I am getting dinner, so I don't want to be bothered any more," and hung up. The father said, "That is irritation. Now, son, I'll show you what exasperation means." So he dialed the telephone number once more. This time, when the woman answered, he said, "This is Maurice. Any telephone calls for me?" And thus a mental image of the word exasperation is created.

The idea of a mental image is a very important aspect of terminology. Even with common words, such as democracy, education, production, each person pictures some aspect of the word with which he is familiar. Such word images may have emotional content; that is, they excite and stimulate the imagination of the individual. For casual communication, this variability in interpretation is workable, but even so frequently gives rise to difficulties in human relations.

In the rapidly expanding technical field of today, we find a rash of new terms—automation, cybernetics, operations research—and the mental images that these words may create can border on "science fiction"—with all the promise they hold for future benefits, and the concern they arouse for the changes which will come. Under these circumstances the emotional content of such new words can be very great. Although this does not preclude the usefulness of such words, it makes their definitions more difficult. For people in science and industry, it becomes essential to narrow down meanings of words; otherwise much loss of time, manpower, materials, and facilities can result from a misunderstanding of a few words or concepts. And, with the advent of broad application of scientific technology to business and industry, the role of communication assumes greater proportions. Never before have we been in a position when there were such strong cross-currents between the office, the factory, and the laboratory. There is a need for more understanding between the businessman, the industrialist, the engineer, and the production man. For one of the characteristics of the complex systems of today is that all of these people are involved, as Dr. Ramo has pointed out, in reviewing the mushrooming growth of automation systems in business and industry. One of the purposes of this book is to bring about a better understanding of the basic concepts of automation to this broad segment of society. We aim to accomplish this not only by definition of words, but by illustrations and practical examples.

Within the field of engineering alone there are complexities. The diverse fields of servomechanisms, feedback control, industrial process control, computers, and data processing have grown up separately; each has developed its own terminology. The convergence of these fields in applied automation has led to the recent appearance of magazines such as *Control Engineering* which in coverage cuts across all areas of engineering. This convergence has also led to conflicts in terminology for which scientific societies are the only arbiters.

As an example of mental images in terminology, take the word "flip-flop." A flip-flop is a computer circuit that has two stable states. The circuit will remain in one state until a suitable signal causes it to change to the other state. Other basic names for this device are "bistable multivibrator" and "Ecles-Jordan" circuit—both are cumbersome. On the other hand, many engineers thought that flip-flop was a rather inelegant phrase. (Some patents on the subject make rather comical reading in the manner in which the flips and flops of the invention are described.) As a result engineers in different laboratories coined new words. "Toggle" is one example of a term being carried over from the description of a two-position switch. "Trigger" is another term that was applied to this circuit, since the state of the flip-flop can be changed by a signal which triggers it. However, although these terms describe certain aspects of the flip-flop, they also have other connotations, and so they fell by the wayside. Today "flip-flop" is the accepted term, approved by terminology committees. It passes the test of providing a good mental image of the function of the unit.

Another example is the word "address"—a computer term. In computing it is necessary to place numbers in storage for later use. This requires knowledge of the location of a number in the storage unit, and the word address is used to designate this. In this case there is a one-to-one correspondence in meaning. Since the word "address" carries the mental picture of a location, it is a very good definition for computer terminology and has been accepted universally.

2.2 THE MEANING OF AUTOMATION

The word automation was first coined by Del Harder of the Ford Motor Company in 1947. Harder shortened the word "automatization" to *automation,* and defined it as the "automatic handling of parts between progressive production processes." During the years since the first use of automation, it has assumed in usage a broader meaning and greater significance. Recently Mr. Harder, now vice-president in charge of manufacturing at Ford, stated that automation "is a concept which embraces our planning of all manufacturing processes." In 1952 John Diebold in his book entitled *Automation* defined it as "denoting both automatic operation and the process of making things automatic."

Since 1954 the word has been in common usage. A few typical examples of definitions are as follows:

K. R. GEISER, General Electric: Automation is the accomplishment of a job by an integrated mechanism with a minimum of human assistance of any kind.

MILTON H. ARONSON, editor of *Instruments and Automation:* Automation is a substitution of mechanical, hydraulic, electronic and electric devices for human organs of decision and effort.

HAROLD MARTIN, Rensselaer Polytechnique Institute: Automation is the entire accomplishment of a work task by a power-driven integrated mechanism wholly without the direct application of human energies, skill, or intelligence.

ANONYMOUS: Automation is the automatic recognition, evaluation, and solution of a processing problem.

Some definitions emphasize the philosophy or purpose of automation. For example, W. E. Brainard of Hughes Aircraft Company has provided the rather lengthy definition: "Automation is more than merely transferring. Nor is it a push-button factory. It is a philosophy that may extend back to the design of the product. It is a new method of manufacture, not necessarily a new way of cutting metal, but a way of controlling the various processes. Automation is a philosophy of design, it is a manufacturing method, and it is control within a machine."

A nation-wide survey was made in 1954 by the industrial division of Minneapolis-Honeywell. Discussion with some 400 industrialists revealed a wide acceptance of the word. Automation, they said, embraces "automatic handling of materials; control of temperature, pressure, and velocity; automatic processing; assembly of parts; measurements of variables; operation of aircraft and missiles; receiving, storing, and shipping functions; computing and data handling; cost accounting; and control of household devices." These 400 industrialists furthermore agreed that "automation signified the automatic performance of a controlling function by mechanisms instead of men," and that "automatic control is obtained through measuring and correcting variables by instruments and mechanisms with no human intervention." The survey revealed that many engineers think of "control" and "instrumentation" separately—both as a part of automation. You can see that all the aspects of this book are covered by the subject matter mentioned in this survey on automation.

These definitions and the survey give us a better understanding of the acceptance of the word "automation," the current usage, and its implications. The basic features of definitions of automation are:

(1) The replacement of the human operator in a step or steps of a process. Man appears outside the process—as a designer, planner, monitor, and maintainer—the *machine* does the work. This concept

of increased mechanization was the first to be associated with automation.

(2) The increased employment of feedback control—both theory and techniques—to the design and operation of automation systems. Mathematics is becoming an essential tool of business and industry.

(3) The use of sensing, decision, and computing elements to replace human operators, implying machines with a higher "intelligence" content and ability to control a process. The computer is becoming a more and more common component.

(4) A broad "systems" approach to new developments, viewing operations as a complex of men, materials, machines, methods, and money, rather than an array of isolated components. This leads to the breaking away from conventional techniques and approaches, both in machine and product design and in operating procedures.

2.3 THE NEED FOR A NEW WORD

The rapid acceptance of automation as a word indicates clearly that the field was ripe for a new term. Magazine feature articles as well as advertisements make liberal use of the word. An editor of a weekly technical news magazine has stated that he cannot put out an issue without using automation at least half a dozen times—there is not another word that has similar connotations. Management and labor are both talking automation, and Congressional hearings on the subject have been held.

In any field there is always the need for words that have rather broad connotations. The word automation fills such a need in this case. Compare it, for example, with the word development. Development is an unfolding process, a growth, and when we speak of development we can refer to study, investigation, breadboarding of equipment, design of production models, or the completed equipment. It is a very broad word, but there is no misunderstanding of what it means in a certain context. It does not describe specifically what the development is concerned with, but we know that it is an unfolding of something new.

Automation fills a similar requirement for a broad word which signifies the use of self-powered, self-sensing, and self-guiding devices, whether it be the automatic washing machine, the electronic processing of data in the office, control of huge factory machines, or control of a chemical process. The availability of new tools and techniques, from both industrial and military research and development, makes possible sensing and decision elements of a different order of magnitude in complexity and speed than was previously attainable in automatic systems. These new developments exceed man's capabilities as an observer, decision maker, and doer, and to many people this signified the need for a new word.

Probably most significant is the need for a broad word to assemble under one umbrella the components, products, systems, and practices of our modern technology directed toward automatic operations and control. As the word automation is used more and more often, people grow accustomed not only to hearing of this newcomer to our society, but also seeing its results; the word is already losing much of its emotional content.

2.4 THE SIMILARITY OF PROCESSES

We note that the word "process" is used in describing automation systems. Let us take a few moments to discuss the implications of the word process. Process is defined by Webster as "a series of actions or operations definitely conducing to an end." We speak of chemical processing, manufacturing processing, food processing, petroleum processing, and data processing. The use of this term process in all of these systems indicates a basic identity. The common characteristics of processes are that they are complex integrated systems, combinations of men, machines, and procedures for carrying out operating plans. In thinking of processes the emphasis is usually placed on the equipment, but the advent of electronic computers now focuses attention on the operating program. In any process the operating program is originally represented by blueprints of products, flow charts, or statement of objectives. These contain the information required to describe the process which must then be carried out by the machines and men.

Every process handles energy, material, or information. The general characteristics possessed by every process, either manual or automatic, are:

(1) Input of materials, energy, or information.

(2) Storage for inputs: materials—spacial storage; energy—storage in materials; information—storage in patterns in energy or materials.

(3) Machine or processor: the device that performs the required work, manipulation, or operation. It shapes, positions, assembles, and treats materials or computes and performs logical operations on information.

(4) Control for directing the machine. In manual operation, man provides the guidance; in automation, the control is automatic.

(5) Output of materials, energy, or information.

Recognition of the basic similarity of processes places emphasis on the operation plan and mode of control, rather than on the equipment. Such emphasis is necessary if radical advances in automation systems are to be made, and is valuable even in "piecemeal" application of automation.

2.5 CROSS CURRENTS BETWEEN OFFICE AND FACTORY

Combined office and factory automation will appear earliest in those operations in which (1) economic factors that may be quantitatively determined can be used as a basis of controlling operating conditions, (2) metering is part of the process and such measurements are used in accounting operations. Examples in which progress has been made are automatic telephone message recording, economic power dispatching in utilities, and petroleum product refining and blending.

Two of the largest problems of modern manufacturers are scheduling of production and inventory control. Here we can readily see the one-to-one correspondence between factory operations and the data processing which parallels it. The greatest part of office data processing in manufacturing industries is concerned with planning, directing, and monitoring factory operations in coordination with sales.

Data-processing machines being used in scheduling operations are really midway between the accounting department and the factory. Such operations have a tendency to exert pressure in two directions. The scheduling and inventory data are of direct value in accounting, sales, marketing, and forecasting functions and will pace these operations. On the other hand, as new techniques are developed for using up-to-date information, there will be a demand for more data of higher accuracy. Consequently pressure will develop to meet schedules, and to further speed up output from the factory: i.e., more automation. Eventually we shall see the computers used for data-processing equipment being applied first to monitor and then actually to control the equipment used in the factory.

2.6 CONCLUSIONS

The concept of automation covers a range of processes that produce products and supply services. These processes are all concerned with basic inputs and outputs, of energy, material, and information, and they also involve storage, processing, and control. We see a tendency for an integration of the varieties of the two extremes of automation—business data processing and the automatic factory. It is likely that integrating office and factory automation will provide greater savings and efficiency than when each is carried out separately. This integration places further emphasis on communication and the need for a broad comprehension of the language of automation.

2.7 GLOSSARY OF TERMINOLOGY

The brief glossary of selected terms presented here has been prepared from sources listed in the bibliography. The terms listed under a number of subject headings have been arranged in a logical rather than an alphabetical sequence, with explanatory block diagrams. The simplest and most easily understood definitions were selected from available glossaries, with emphasis on the newer fields of digital computers and data processing. The progress made in standardization of terminology in the fields of feedback control, automatic process control, and servomechanisms is demonstrated by the tables covering these subjects. This brief glossary should prove useful as a reference for basic terms that will be used in later chapters.

2.7.1 General Definitions

Automation. (1) The technique of making a process or system automatic. (2) Automatically controlled operation of an apparatus, process, or system, especially by electronic devices.

Cybernetics. "The field of control and communication theory, whether in the machine or in the animal [Norbert Wiener]."

Operations Research. "Operations research is a scientific method of providing executives with quantitative basis for decisions regarding operations under their control [P. P. Morse]."

Linear Programming. Linear programming is a mathematical method for sharing a group of limited resources among a number of competing demands, where all decisions are interlocking because they all have to be made under a common set of fixed limitations.

Language. A system consisting of (*a*) a well-defined, usually finite, set of characters, (*b*) rules for combining characters with one another to form words or other expressions, and (*c*) a specific assignment of meaning to some of the words or expressions, usually for communicating information or data among a group of people, machines, etc.

Machine Language. (1) A language, occurring within a machine, ordinarily not perceptible or intelligible to people without special equipment or training. (2) A translation or transliteration of this language into more conventional characters but frequently still requiring special training to be intelligible.

Common Language. The use of identical "machine language," throughout a system, so that machines can communicate directly without translation to ordinary language.

2.7.2 Computers, Simulators, Trainers

Computer. (1) A machine for carrying out calculations. (2) By extension, a machine for carrying out specified transformations on information.

Digital Computer. A computer that operates with information, numerical or otherwise, represented in digital form.

Analog Computer. A type of calculating machine that operates with numbers represented by directly measurable quantities such as voltages, resistances, rotations, etc.

Simulator. The representation of a physical system by computers and associated equipment.

Trainer. The representation of an operating system by computers, associated equipment, and personnel.

2.7.3 Digital Computers (See Table 2.1 and Fig. 2.1)

Digit. One of a set of symbols used to represent numbers.

Arithmetic Unit. That part of a computer which performs arithmetic operations.

Control. Usually those parts of a digital computer that effect the carrying out of instructions in proper sequence, the interpretation of each instruction, and the application of the proper signals to the arithmetic unit and other parts in accordance with this interpretation.

Storage. (1) The act of storing information. (2) Any device in which information can be stored, sometimes called a memory device. (3) In a computer, the section used primarily for storing information. Note: The physical means of storing information may be electrostatic, ferroelectric, magnetic, acoustic, optical, chemical, electronic, electrical, mechanical, etc., in nature.

Write. To introduce information, usually into some form of storage.

Read. To extract information, usually from some form of storage.

2.7.4 Computer and Data Processor Programming

Program. (1) A plan for the solution of a problem. (2) Loosely, a synonym for *routine.* (3) To prepare a program.

Flow Diagram. A graphic representation of a *routine.*

Table 2.1 Representation of Information in Digital Computers and Data Processors

Term	Definition	Example of Information	Binary Digital Representation
Binary	Involving the integer two, as in the binary number system	0, 1, 2, 3, 4, 5, ···, 8, 9, ···, 16, ···	0, 1, 10, 11, 100, 101, ···, 1000, 1001, ··· 10000, ···
Binary digit	A digit of a binary number or code, sometimes abbreviated as "bit"	0 or 1	0 or 1
Binary coded decimal	A system of number representation in which the decimal digits of a number are expressed by binary numbers	159	0001 0101 1001 (binary numbers) 0100 1000 1100 (arbitrary code)
Character	(1) One of a set of elementary symbols such as those corresponding to the keys of a typewriter. (2) A code representation of such a symbol	0 to 9 A to Z .,?; etc.	000000 to 001001 001010 to 100011 100100 to 111111 (64 possible characters)
Word	An ordered set of symbols which is the smallest unit in which information is normally stored and transferred within the computer	(*a*) 643821859 (*b*) J. R. Wellington	(*a*) 0110, 0100, ···, 01000 (10 numbers) (*b*) 010011 100100 011011 100100 100000 001110 010101 010101 010010 010000 010111 011101 011000 010111
Block	A group of words considered or transported as a unit		
Item (unit record)	Blocks of words containing related information	J. R. Wellington (name) 64,322 (payroll no.) 2.89 (pay rate) 3.00 (ins. deduction) etc.	010011 100100 011011 100100 100000 001110 010101 010101 010010 010000 010111 011101 011000 010111 (name) 000110 000100 100101 000011 000010 000010 (payroll no.) 000010 100100 001000 001001 (pay rate) 000011 100100 000000 000000 (ins. deduction)

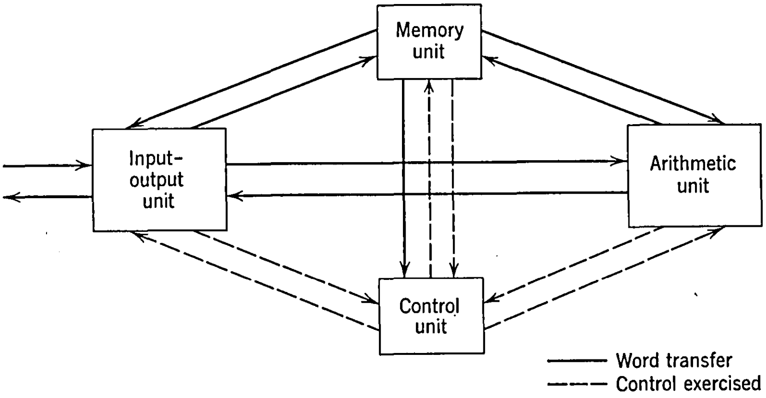


Fig. 2.1 General block diagram of a digital computer

Code. (1) A system of characters and rules for representing information. (2) Loosely, the set of characters resulting from the use of a code. (3) To prepare a *routine* in *machine language* for a specific computer. (4) To encode, to express given information by means of code.

Instruction Code. An artificial language for describing or expressing the instructions that can be carried out by a digital computer. In automatically sequenced computers, the instruction code is used when describing or expressing sequences of *instructions,* and each instruction word usually contains a part specifying the operation to be performed and one or more *addresses* identifying a particular location in storage. Sometimes an *address* part of an instruction is not intended to specify a location in storage but is used for some other purpose.

Operation Code. (1) The list of operation parts occurring in an instruction code, together with the names of the corresponding operations (e.g., "add," "unconditional transfer," and "add and clear," etc.). (2) Synonyms for operation part of an instruction.

Address. An expression, usually numerical, that designates a particular location in a *storage* or memory device or other source or destination of information.

Routine. A set of instructions arranged in proper sequence to direct the computer to perform a desired operation, such as the solution of a mathematical problem.

Subroutine. In a *routine,* a portion that causes a computer to carry out a well-defined mathematical or logical operation.

2.7.5 Data-Processing Operations

File. (1) A sequential set of items (not necessarily all of the same size) (*noun*). (2) To insert an item into such a set (*verb*).

Sort. To arrange items of information according to rules dependent upon a key or field contained by the items.

Merge. To produce a single sequence of items, ordered according to some rule (i.e., arranged in some orderly sequence), from two or more sequences previously ordered according to the same rule, without changing the items in size, structure, or total number. Merging is a special case of collation.

Extract. To remove from a set of items of information all those items that meet some arbitrary criterion.

Collate. To combine two or more similarly ordered sets of items to produce another ordered set composed of information from the original sets. Both the number of items and the size of the individual items in the resulting set may differ from those of either of the original sets and of their sums.

2.7.6 Tabulating Equipment[1] (See Table 2.2)

Tabulating Card. A unit record card designed for the recording of data in the form of punched holes to be sensed by mechanical or electrical (including electronic) means.

Table 2.2 Punched Card Equipment versus Electronic Data Processors

	Punched Card	Electronic Data Processors
Equipment	Card punches, reproducers, collators, sorters, accounting machines (calculating punches and printers)	Processor (computer), magnetic tapes, card ⇄ tape converters, plus preparatory equipment for punched tapes and cards
Program control	External plugboard on each machine	Internally stored program control processor, tapes, and converters
Program steps (instructions)	Up to 100	1,000 to 20,000 typical Limited only by storage capacity
Computing speed	Low (limited by card readers)	High
Storage	80 characters (card)	Up to 20,000 characters (cores)
	80 characters (relays)	Up to 1.8×10^6 characters (drums)
		Up to 5×10^9 characters (tapes)

Tabulating System. Any group of machines capable of entering, converting, receiving, classifying, computing, and recording alphabetic and/or numeric accounting and/or statistical data by means of tabulating cards, and in which tabulating cards are used for storing data and communicating it within the system; provided that "tabulating system" shall not include "electronic data-processing system" as hereinafter defined.

Tabulating Machine. A machine or device and attachments therefor used primarily in a tabulating system.

Electronic Data-Processing System. Any machine or group of automatically intercommunicating machine units capable of entering, receiving, storing, classifying, computing and/or recording alphabetic and/or numeric accounting and/or statistical data without intermediate use of tabulating cards, which system includes one or more central data-processing facilities and one or more storage facilities,

---

[1] Present-day electronic systems make extensive use of tabulating equipment; hence definitions covering this aspect of data processing are of interest. The list of definitions covering tabulating machines has been taken from the Final Judgment of U.S. District Court Civil Action 72–344, U.S.A. vs. IBM.

and has either (1) the ability to receive and retain in the storage facilities at least some of the instructions for the data-processing operations required, or (2) means, in association with storage, inherently capable of receiving and utilizing the alphabetic and/or numeric representation of either the location or the identifying name or number of data in storage to control access to such data, or (3) storage capacity for 1000 or more alphabetic and/or decimal numeric characters or the equivalent thereof.

Electronic Data-Processing Machine. A machine or device and attachments therefor used primarily in or with an electronic data-processing system.

Table 2.3 Feedback and Automatic Control Symbols and Definitions

	Recommended Term				
Symbol	AIEE	ASME	Nonstandard Terms	Nonstandard Symbols	Definition
g_2	Controlled system	Process	Plant; load	y, kg, h, μ, a	The body, process, or machine, a particular quantity or condition of which is controlled (AIEE) The collective functions performed in and by the equipment in which a variable is to be controlled (ASME)
c	Controlled variable		Output; regulated variable; measured variable	θ_o, O, r, o as subscript	The quantity or condition of the controlled system that is directly measured and controlled
h	Feedback elements		Feedback transfer functions	y, kg, g, β	The portion of the feedback control system that establishes the relation between the primary feedback and the controlled variable
b	Primary feedback		Monitoring feedback; feedback	θ_f	A signal which is a function of the controlled variable and which is compared with the reference input to obtain the actuating signal
v	Command	Set point	Input; desired value; con- control point	θ_i, d	The input which is established or varied by some means external to and independent of the feedback control system under consideration
a	Reference input elements		Conversion elements; primary elements; sensing elements		The portion of the feedback control system that establishes the relation between the reference input and the command
r	Reference input		Input; reference standard; desired value; set point	$\theta, v, \theta_i, d, i$	A signal established as a standard of comparison for a feedback control system by virtue of its relation to the command
e	Actuating signal		Error; unbalance; actuating error; correction	$\epsilon, \theta, \theta_e, u$	The reference input minus the primary feedback
g_1	Control elements		Amplifier; controller; servo amplifier; relay; error corrector	y, kg, h, μ, a	The portions of the feedback control system that are required to produce the manipulated variable from the controlled variable
m	Manipulated variable		......	......	The quantity or condition that the controller applies to the controlled system
u	Disturbance		Load disturbance; upset noise; drift	l, d, n	A signal (other than reference input) which tends to affect the value of the controlled variable

Service Bureau Business. The preparation with tabulating and/or electronic data-processing machines of accounting, statistical, and mathematical information and reports for others on a fee basis.

Service Bureau. An organization engaged principally in the service bureau business.

2.7.7 Automatic or Feedback Control Systems (See Table 2.3)

A *Feedback Control System* is a control system that tends to maintain a prescribed relationship of one system variable to another by comparing functions of these variables and using the difference as a means of control (AIEE). See Fig. 2.2.

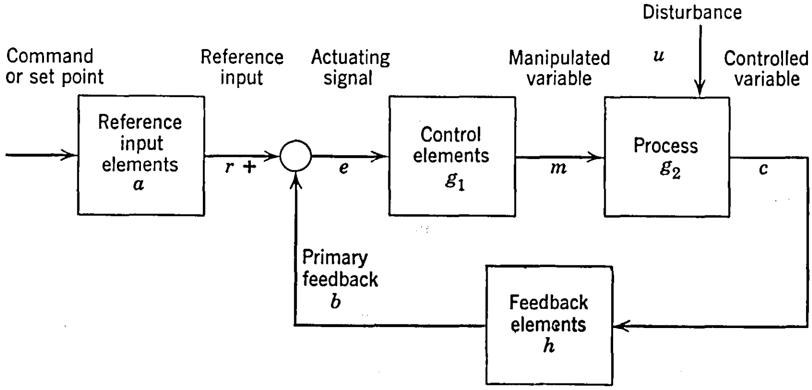


Fig. 2.2 Block diagram of feedback control system

An *Automatic Control System* is any operable arrangement of one or more automatic controllers connected in closed loops with one or more processes (ASME).

An *Automatic Controller* is a device that measures the value of a variable quantity or condition and operates to correct or limit deviation of this measured value from a selected reference. An automatic

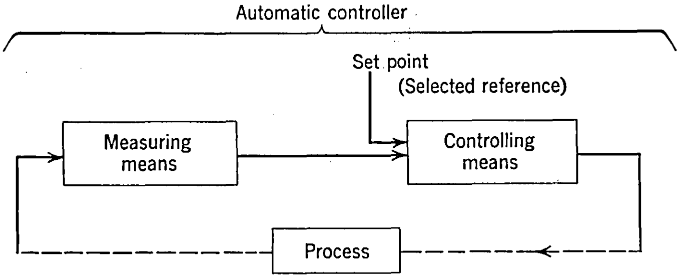


Fig. 2.3 Automatic controller and its feedback loop

controller includes both the measuring means and the controlling means. See Fig. 2.3.

A *Servomechanism* is an automatic control system in which the controlled variable is mechanical position (ASME). It is a feedback control system in which the controlled variable is mechanical position (AIEE).

A *Transducer* is a device for converting a signal or physical quantity of one kind into a corresponding physical quantity of another kind.

2.8 REFERENCES ON TERMINOLOGY

2.8.1 Control Systems

Institute of Radio Engineers (IRE), Standards on Terminology for Feedback Control Systems 55 IRE 26.S2. Standards on Graphical and Letter Symbols for Feedback Control Systems 55 IRE 26.S1.

American Society of Mechanical Engineers, Instruments and Regulators Division, Automatic Control Terminology, 1954.

American Standards Association, Letter Symbols for Feedback Control Systems, ASA Y10, 1955.

American Institute of Electrical Engineers, Proposed Symbols and Terms for Feedback Control Systems, in *Electrical Engineering,* October 1951.

British Standards Institution, Glossary of Terms Used in Automatic Controlling and Regulating Systems, British Standard 1523.

2.8.2 Computers and Data Processing

Institute of Radio Engineers (IRE) Standards on Electronic Computers, Definitions of Terms 50 IRE 8.S1, 55 IRE 8.S2.

Association for Computing Machinery, First Glossary of Programming Terminology, June 1954.

Glossary of Terminology in the Field of Computers and Automation, *Computers and Automation,* January 1956.

2.8.3 Magazines

Automation, Penton Publishing Co., Cleveland, Ohio.
Automatic Control, Reinhold Publishing Corp., New York, N. Y.
Computers and Automation, E. C. Berkeley Assoc., New York, N. Y.
Control Engineering, The McGraw-Hill Publishing Co., New York, N. Y.
Instruments and Automation, Instrument Publishing Co., Pittsburgh, Pa.
Journal of the Instrument Society of America, Pittsburgh, Pa.

3. Fundamentals of Automation

JOHN L. BARNES

University of California
Los Angeles, California

3.1 INTRODUCTION

It is not difficult to get general agreement on the value of stressing fundamentals. However, when a census is taken of just what are the fundamentals of a large subject, such as *automation,* then a wide spectrum of answers usually appears. In the following I shall give you my views on the fundamentals of automation.

What is automation? Chapters 1 and 2 have introduced this subject and have reviewed some of the language of the field. Before giving my answer to the meaning of automation, let me first relate an incident in automatic banking.

A young man from a small town was sent to a technical college by his rich uncle. On returning home after graduating, he convinced his father, who managed the local bank, that the bank should install some recently developed banking equipment. Unfortunately the equipment was not completely debugged when it arrived. Well, as you may guess, there were initial troubles. To begin with, something went wrong with the high-speed check sorter. After sorting the first batch of checks in the twinkling of an eye, it shredded the second batch into confetti. Not long after this, the day of reckoning came, The father, who was teller as well as manager of the bank, had to inform the next customer desiring to withdraw some of his funds, "I'm sorry, sir, but our new automatic banking device shifted into reverse, and your interest ate up the principal."

Automation holds forth a brilliant future for you and for me, I believe, but not without a few difficulties such as those suggested by this banking story. If an automatic computer can turn out a million correct results in a day, then if improperly designed it might turn out errors at a similar rate.

Let us return to the question, "What is automation?" My answer is: Automation is the use of a nonliving system to control and carry out an operation.

A term with a broader realm of application is feedback control. This applies to living as well as nonliving systems. For example, "The rate of human breathing is controlled by eleven different feedback loops," according to Norbert Wiener.

How old is feedback control? My answer is that examples of feedback control go back to the beginning of life on this planet, say one or two billion years.

How old is automation? The best answer that I can give you is that automation goes back to James Watt's invention of the fly-ball governor for his steam engine which was patented in 1769. Of course, mechanisms without self-control go back to the invention of animal traps, and to later inventions such as the wheel. By the time of the Romans there had been invented elementary computing mechanisms such as the water clock and the chariot wheel revolution counter. The latter was arranged to have a marble drop through a hole with each revolution of the chariot wheel. At the end of the trip, the number of marbles that had dropped through gave a measure of the distance traveled by the chariot. This digital revolution counter was probably the first taxi meter. Today it is cited in analog-to-digital converter patent suits.

Why is there at present such a great interest in automation for science, engineering, industry, and business? It is my belief that the reasons for this recent great upsurge of interest are (1) the accelerated availability of electronic and magnetic computer and actuator equipment; (2) growth of automation "know how"; coupled with (3) the accelerated realization of our needs for automation brought on by World War II and subsequent military activity. In short, now we can get automation! Not that we do not already have a fair amount of automation, but rather that now (1) we have considerable knowledge on this subject; (2) we have developed a reasonable understanding of the design of linear automatic systems; (3) we have made a start on applying automatic systems to industry and business; and (4) we have lots of hardware and system design "know how" waiting to be applied.

A new category of engineer is now appearing, the BUSINESS ENGINEER. As with earlier two-field experts, the business engineer often receives the greatest acclaim as an engineer from the businessmen, and as a businessman from the engineers.

In earlier years it was standard practice to introduce historical perspective in talks on automatic computers by references to the contributions of ideas on automatic mathematical-table computers by Babbage, and to ideas on differential analyzers by Lord Kelvin. Through their ideas these men were laying the conceptual foundations for computers and consequently for automation. As you recall, an automatic system consists basically of a computer coupled to a set of actuators. During the early and middle nineteenth century when Babbage and Kelvin were generating these ideas for computers, proper means for reduction to practice were not at hand. Hence, practical versions of such computers did not appear until the 1920's and 1930's. Now in the 1950's with good network and feedback engineering theory and components, with good analog and digital theory and engineering, we are arriving! The goals of Watt, Babbage, and Kelvin, however, are now well behind us.

What are our new goals? What does automation, present and future, offer business and industry? Let us first go over some familiar fundamentals, look at the new goals, and then discuss the means for obtaining them. We cannot repeat too often that the first phase of the industrial revolution has been replacing the muscle work of man and beast by power from coal, oil, other chemicals, flowing water, and now from atomic nuclei, and that the second phase of this revolution is replacing the elementary brain work of man by the work of automatic electromechanical and electronic computers. These two phases combine to give us automation. *And what net benefits are we to derive from additional automation in business and industry?* The answer is—a better life! Just as the standard of living in the United States has kept pace with the increased use of nonmuscle power, so, barring wars, it will continue to rise with the combined use of nonliving power and information processing in automation.

Many of the labor leaders are aiding the extension of automation in their fields, since the broad result is an upgrading of the kinds of work done by people with a consequent net increase in the standard of living! Monotonous routine work is delegated to automatic information processors. Each person generates a greater profit for the concern and hence can receive greater pay. The expansion of needs for new varieties of goods and services is growing more rapidly than the supply of people to turn them out. Automation can help to bridge the gap.

3.2 THE ROLES OF SCIENCE, MATHEMATICS, AND ENGINEERING

Now let us turn to the question, namely, "What are the roles of (1) science, (2) mathematics, and (3) engineering in helping us to extend automation?"

(1) Science supplies us with essential information about the physical world, including the people in it. The scientific method of hypothesis, error-corrected by a feedback loop using quantitative experiment, is often difficult to apply to industry and business. Part of the difficulty is assignable to the lack of control over conditions for proper experiments. Putting a simplified version of the problem on a computer with actuators, as a simulator, may help in the application of the scientific method.

(2) Mathematics includes reasoning which can be used on the simplified models of items in the physical world. More advanced automation is possible with better models. These better models are often more complicated. Hence, we need the large automatic computers to deal with them.

(3) Engineering is the application of science and some mathematics to solve man's problems. In this instance they are problems concerning improved or extended automation of industry and business. Prominent in the more recently emphasized aspects of this engineering is the systems approach. Problems of how to split up a large system into parts with minimum interaction, and how to synthesize a large system from subsystems, arise and call for greater attention. Also, in the large systems we have in mind there are both people and automata. This excites the question: *"How can we divide the systems job between people and automata?"* A few comparative characteristics are as follows. People are usually more variable in their performance and slower than modern electronic-magnetic automatic systems. A person's brain, viewed as a computer, is very advanced, extremely compact, and very efficient in the use of power. A person has an enormous memory. His brain is subject to saturation in a short time. A person can usually work on one problem at a time, and it is difficult to couple brains for series, parallel (tandem) bridge, or other systems operation. On the other hand, automata have extremely elementary "brains" as of today. They do not saturate in a short time. They can work 23 or 24 hours a day and do not need week ends off, for the most part. They can be more reliable than people. They can be much less variable than

people, and usually are. Automata can be now best applied to fixed-policy repetitive problems needing rapid solution. Later variable-policy problems will be solvable.

3.3 DESIGN OF AN AUTOMATIC SYSTEM

3.3.1 Military Contributions

The most advanced applications of automation to date have been in military engineering. Examples of such applications appear in antiaircraft gun directors, in increasingly automatic fighter planes, and in guided missiles. It is the carry-over of the physical and mathematical techniques evolved in military research and devolopment to the solution of industrial and business automation problems that will be emphasized in this section.

3.3.2 Automatic Subsystems

As mentioned above, people will be present as subsystems in our overall system. However, it is with automatic subsystems that we shall be concerned. In many instances these automatic subsystems will perform operations that humans have never done and cannot do. At the same time they will permit humans to carry out operations with their brains that are far beyond the capabilities of automatic computers now and in the foreseeable future.

3.3.3 Elements

As the elements of an automatic industrial or business subsystem we shall take: (1) the subsystems input equipment or sensory pickups, (2) the subsystem output equipment or controlled actuators, and (3) the processors or combination computer-amplifiers which connect input and output. Of course the environment of the automatic subsystem is present in the form of (4) a set of driving signals, disturbances, and initially stored energies, and (5) externally coupled input and output parts of the whole system which in special cases would be described as source and sink admittance levels.

3.3.4 A Design Method

Details and examples will be given to answer the following general question: *What is a typical procedure for designing an automatic business or industrial system?*

(1) Decide on the class of problems to be solved by the automatic system in the time era during which the system is to be operational. Allow for changes in this class of problems with the passage of time.

(2) Estimate the class of environments to be encountered by the system during its operation. Allow for improved environmental data resulting from better measurements.

(3) Rough out blocks of the whole system, using estimates based on pertinent past experience and on realistic assumptions and calculations for weighting or transfer functions of the subsystems. Check the consistency of assumed development or acquisition times of proposed components and subsystems. As the development progresses, recheck for this consistency at appropriate intervals. Take the required corrective action.

(4) Decide on measures of effectiveness of the system in the attainment of its various objectives.

(5) Define each type of error by an appropriate criterion that measures the amount by which the actual-system effectiveness fails to attain each desired-system objective.

(6) Make error analyses assuming a linear invariant approximating system for each system objective. Use the Laplace transformation method, together with "rout locus" aids to treat the approximating system. Use a real-time simulator to study the whole system. That is, make a full-scale dynamic model of the system by means of a computer-simulator and then run through the set of operational modes under the set of estimated environmental conditions. Information learned from simulation should be fed back into the system design to improve (*a*) system stability margin, (*b*) equality of response to typical sets of input signals under expected operating environmental conditions, and (*c*) reliability of operation.

(7) In the simulator, replace linear computer approximations to components and subsystems of the system under design by nonlinear components and subsystems. Again check the various modes of the simulated system.

(8) As the actual linear and nonlinear hardware components of the system become available, insert them in the simulated system in place of the approximating computer equipment. As before, check the modes of the now partially actual and partially simulated system.

(9) Successively refine the system error analysis using statistical ensembles for drives, boundary conditions, and disturbances. Adjust the system design connections and parameters—until the system stability, quality, and reliability are satisfactory.

(10) By the same process used in the development of the main system, develop monitoring instrumentation for locating failing components in the main system.

(11) Run a set of tests on the main system to determine appropriate statistics on the reliable life of components and subsystems.

(12) Collect appropriate statistics on the system during its era of actual operation and use the results to further improve the design of later editions of the automatic system.

3.3.5 Automation for Control

Many businesses now use automation in obtaining financial information for records which are used for tax purposes, financial reports, and (historical) records. This use does not place strong emphasis on speed. However, if the financial information is made available in a short enough time, it can be used for essentially continuous control of the business. In obtaining the required speed for high-speed control it is often necessary to exchange some of the recording operations for solution time.

3.4 CONCLUSION

Although the foregoing remarks merely touch on some of the fundamentals of automation in business in industry, it is my hope that they may stimulate the reader to deeper thought on the principal objectives of automation and to a practical approach to the attainment of more extensive and better automation.

For those who would go into more detail a bibliography is attached.

3.5 REFERENCES

1. G. R. Andlinger, *Operations Research; Challenge to Modern Management,* Operations Research Associates, Cambridge, Mass., 1954.

2. W. R. Ashby, *Design for a Brain,* Chapman & Hall, London, and John Wiley & Sons, Inc., New York, 1952.

3. D. A. Bell, *Information Theory and Its Engineering Applications,* Sir Isaac Pitman & Sons, Ltd., London, 1953.

4. D. A. Bell, *Statistical Methods in Electrical Engineering,* Chapman & Hall, Ltd., London, 1953.

5. R. Bellman, Some Applications of the Theory of Dynamic Programming, A review, *J. of Operations Research, 2:* 275–288 (Aug. 1954).

6. D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions,* John Wiley & Sons, Inc., New York, 1954.

7. R. Carnap and Y. Bar-Hillel, An Outline of a Theory of Semantic Information, *MIT, Research Lab. Electronics, Tech. Rept., 247,* 1952.

8. A. Charnes, Lectures on the Mathematical Theory of Linear Programming, Part II of A. Charnes, W. W. Cooper, and A. Henderson, *An Introduction to Linear Programming,* John Wiley & Sons, Inc., New York, 1953.

9. A. Charnes, W. W. Cooper, and A. Henderson, *An Introduction to Linear Programming,* John Wiley & Sons, Inc., New York, 1953.

10. J. F. Coales, Application of Information Theory to Data-Transmission Systems, and Use of Binary Coding to Increase Channel Capacity, IEE, Prob. 5, part 3, 291–302, Sept. 1953.

11. J. Diebold, *Automation: The Advent of the Automatic Factory,* D. Van Nostrand Co., Inc., New York, 1952.

12. R. Dorfman, *Application of Linear Programming to the Theory of the Firm; Including an Analysis of Monopolistic Firms by Nonlinear Programming,* University of California Press, Berkeley, 1951.

13. J. E. Edison, L. S. Morse, R. A. Pizetz, and G. Sternlieb, Electronics—New Horizon in Retailing; The Application of Electronics in Retailing; The Application of Electronics and Electromechanical Systems to Retail Control, AER Associates, Clayton 5, Mo., 1953.

14. A. D. Franklin and E. V. Hankam, Bibliography on the Use of IBM Machines in Science, Statistics, and Education, Watson Scientific Computing Laboratory, IBM, New York, 1952.

15. S. Goldman, *Information Theory,* Prentice-Hall, Inc., New York, 1953.

16. T. C. Koopmans, *Activity Analysis of Production and Allocation,* John Wiley & Sons, Inc., New York, 1951.

17. J. C. C. McKinsey, *Introduction to the Theory of Games,* McGraw-Hill Book Co., Inc., New York, 1952.

18. B. McMillan, *Information Theory,* University of Pittsburgh Press, Pittsburgh, 1953.

19. P. M. Morse and G. E. Kimball, *Methods of Operations Research,* Technology Press and John Wiley & Sons, Inc., New York, 1951.

20. H. Quastler (Ed.), *Information Theory in Biology,* University of Illinois Press, Urbana, 1953.

21. R. K. Richards, *Arithmetic Operations in Digital Computers,* D. Van Nostrand Co., Inc., New York, 1955.

22. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication,* University of Illinois Press, Urbana, 1949.

23. H. S. Tsien, *Engineering Cybernetics,* McGraw-Hill Book Co., Inc., New York, 1954.

24. A. Tustin, *Mechanism of Economic Systems; an Approach to the Problem of Economic Stabilization from the Point of View of Control System Engineering,* Harvard University Press, Cambridge, Mass., 1954.

25. J. von Neumann, *Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components,* California Institute of Technology Lectures, 1952.

26. J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior,* Princeton University Press, Princeton, 2nd ed., 1947.

27. N. Wiener, *Cybernetics; or Control and Communication in the Animal and Machine,* John Wiley & Sons, Inc., and Technology Press, New York, 1948.

28. N. Wiener, *The Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications,* John Wiley & Sons, Inc., and Technology Press, New York, 1950.

29. N. Wiener, *The Human Use of Human Beings; Cybernetics and Society,* Houghton Mifflin Co., Boston, 1950.

30. J. D. Williams, *Compleat Strategyst: Being a Primer on the Theory of Games of Strategy,* McGraw-Hill Book Co., Inc., New York, 1954.

31. P. M. Woodward, *Probability and Information Theory, with Applications to Radar,* McGraw-Hill Book Co., Inc., New York, 1953.

4. Feedback Control Systems

HAROLD CHESTNUT

General Electric Company
Schenectady, New York

4.1 INTRODUCTION

Automation is a recently coined word that is used to describe "continuous automatic production." In its ideal form it signifies the accomplishment of a desired end or product without the use of manual power or direction during the process itself. Although the word automation is new, many of the processes—the tools, techniques, and components—have been developed over the past twenty or more years. From a purely technical point of view, the challenging aspect of automation is the ability to perform the desired tasks without human participation. In many cases the engineering and physical problems involved are difficult and their solution is not readily apparent. The way of doing something automatically may be quite different operationally from the way it is done manually, although the same functional result is obtained.

From a practical point of view, one of the major problems of automation is deciding whether a product or job is one for which automation is justified. The nature of the product, the volume or quantity of labor involved in the process, the cost of automatic production versus manual production, the quality of the automatically produced product contrasted with the manually produced one—these are some of the factors that must be considered before embarking on a program of automation. In other words, the question "how to automate" comes before the question "whether to"; the method of automation must be determined before its adoption can be considered.

Although many processes when so scrutinized may be found not to warrant automation, the pressures on our economy are in the direction of more and more automation. As evidence of this, statistics show mounting labor costs, more national product per worker, increasing numbers of automatic control components available at reasonable prices, and an improved general level of technology. Certainly we are justified in assuming that the technical aspects of the problem are important and worthy of our attention.

4.1.1 Nature of Problem

Let us attempt to formulate the general technical problem of automation, i.e., automatic production. Figure 4.1 illustrates in block diagram form the flow nature of the automation problem as an open-loop control starting with raw materials as inputs and continuing on through to the finished product. The raw materials are acted upon in the fashion of a flow process by a number of production operations in series, perhaps joined together at some point, then further operated upon, finally emerging as the finished product or output. Each of the production operations may itself be a controlled process, such as shaping, cutting, transportations, or positioning of the material. The

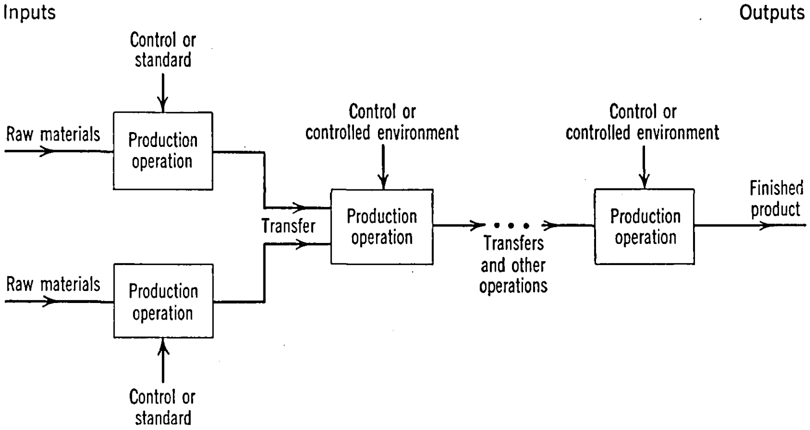


Fig. 4.1 Schematic representation of general automation problem

production operation may take place in a controlled environment of temperature, humidity, acidity, or some other desired condition that the particular process requires. Controls or standards may be applied during or after many of the production operations so that the product passing from one operation to the next is determined to be satisfactory for the purpose intended. The system as shown in Fig. 4.1 does not indicate directly any means of "closing the loop" to insure that the finished product or output will continuously meet the values desired for these quantities.

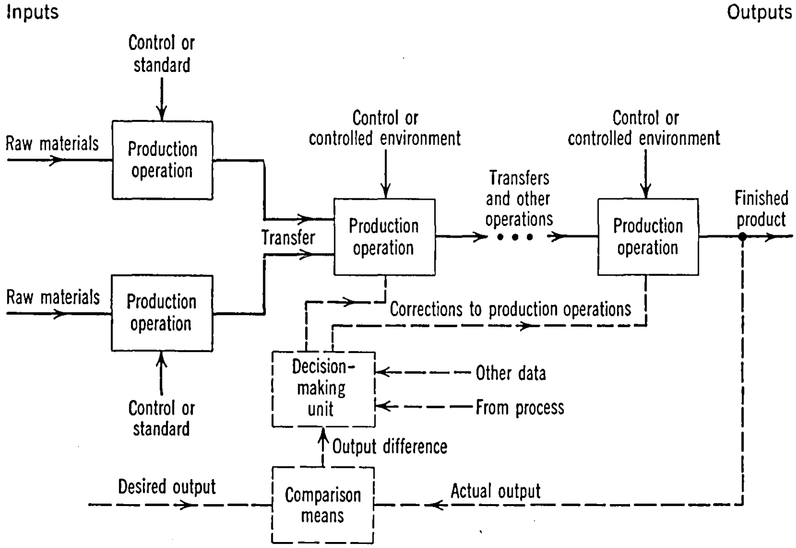


Fig. 4.2 Schematic representation of general automation problem with closed-loop control

"Closing the loop" can be accomplished by the addition of a means for comparing the actual output with its desired value and using the difference or differences to modify the controlled environment or the production operation so that the actual output will more nearly equal the desired output (Fig. 4.2). The term automation could be and is employed to describe either the open-loop or closed-loop version of the continuous production process described. It is of course apparent that the closed-loop production scheme is more automatic and should require less manual attention, provided reliable operation can be maintained.

The preceding description of the general automation problem has served its point if it has indicated that control systems are used extensively as a vital and important part of the automation process. The reason why a human being is not needed in the production process is that some automatic means is available for controlling the position, temperature, voltage, or other conditions required in the process that formerly demanded manual attention.

Without automation, individual isolated controls have been performing automatically but on a batch basis—merely performing one job or part of a job. The use of automation makes possible the tying together of a number of automatic controls into an overall flow process in which the manufacture is continuous without human handling from one automatic operation to the next. Although the complexity, expense, and magnitude of the control problems are greatly expanded by the requirements of automatic production, the basic control problems of a single control system are in many important ways similar to those of the more complex automation process. A fitting starting point for considering the automation control problem is, therefore, to consider the control of a single quantity and to describe the feedback control principles involved in such a control. Next, some of the problems associated with multiple control systems operating in conjunction with one another will be reviewed. Finally, a number of specific examples of automation in industry will be described.

4.1.2 Description of Feedback Control System

One of the building blocks of the automation process is the feedback control system, a simplified form of which is shown in Fig. 4.3. A feedback control system as referred to here is a means of automatic control in which the difference between the reference input and some function of the controlled variable is used to supply an actuating error signal to the control elements and the controlled system. The amplified actuating error signal endeavors to reduce to zero the difference between reference input and controlled variable. A supplemental source of power is available in such systems to provide amplification at one or more points in the feedback control system so that the possibility exists for self-sustained oscillations or instability.

In addition to the principal variables described above and shown by the solid-line portions of the diagram, the desired value, the indirectly controlled quantity, and the disturbance function are shown. The desired value represents the value that the control system is supposed to reproduce, and it differs from the actual reference input by the characteristics of the reference input elements. The indirectly controlled quantity represents the quantity that is the actual system output. It differs from the controlled variable by the characteristics of the indirectly controlled system elements. The disturbance function represents an unwanted input to the system that tends to cause the controlled variable to differ from the reference input. The disturbance function elements shown by dotted lines are intermediate between disturbance function and the controlled system itself. The dotted lines associated with the reference input and indirectly con-

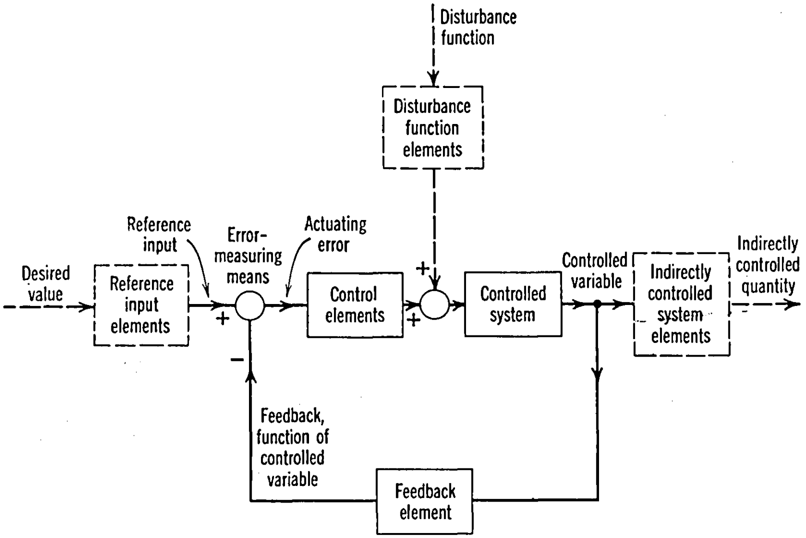


Fig. 4.3 Block diagram of simplified feedback control system

Table 4.1 Listing of Various Forms of Feedback Control System Quantities

Desired Value	Reference Input	Feedback	Error-Measuring Means	Controlled Variable	Indirectly Controlled Quantity
Voltage	Potentiometer voltage	Voltage	Voltage detector	Terminal voltage	Load voltage
	Spring tension	Current	Galvanometer element		
Position	Angular shaft position	Angular shaft position	Differential gears	Machine-tool position	Dimensions of metal being cut
	Electric field	Angular shaft position	Selsyn control transformer		
Temperature	Thermostat contact setting	Bimetal position	Bimetal thermostat	Oven temperature	Process temperature
	Potentiometer voltage	Thermoelectric voltage	Thermocouple		
Speed	Standard voltage	Voltage	Voltage detector	Governor speed	Alternator speed
	Calibrated spring position	Governor position	Linkage position		

trolled system elements serve to indicate that these elements, when equal to unity, have a limited effect on the feedback control problem.

The principles of feedback control operation may be used effectively for any one of a number of different physical kinds of control problems, whether they employ electrical, mechanical, thermal, or other forms of control elements. Table 4.1 indicates some common forms that these variables and elements take for a few types of feedback control systems. This table shows the similarity in function of a number of different controls, although they differ appreciably in operation.

4.1.3 Requirements of Stability and Accuracy

The basic principle of feedback control or closed-loop operation tends to make for accurate performance since the control system endeavors continually to correct any error that exists. However, this corrective action can give rise to a dangerous condition of unstable operation when used with control elements having a large amount of amplification and significant delays in their time of response. An unstable control system is one that is no longer effective in maintaining the controlled variable very nearly equal to the desired value. Instead, large sustained oscillations or erratic control of the controlled variable may take place, rendering the control useless.

If stable feedback control system performance is like that of a manually controlled system with a capable and well-trained operator, unstable feedback control system performance may be compared to that of the manually controlled system with an untrained and irresponsible operator. Rapid and destructive response of the system, in which adequate control is impossible, may result, and destructive action of the controlled variable may occur.

If, in an effort to increase the accuracy of the control system, the amplification of the control is increased without adequate steps being taken to insure stable operation, the advantages of the feedback control principle prove illusory. Furthermore, it is necessary to do more than have a system that is stable; a system must have an adequate margin of stability and be able to recover rapidly and smoothly from the shocks of irregular inputs or of severe disturbances.

The requirements of stability and accuracy are mutually incompatible. The higher the desired accuracy, the smaller is the actuating error that can be allowed to perform the proper corrective action. Thus high accuracy requires high amplification. With high amplification, small actuating errors, if applied for too long a period of time, can produce too large values of the controlled variable which may later cause larger actuating errors. With lower-gain systems, the time during which the actuating error is applied is not so critical, for less controlled-variable motion takes place in the system having the same time rate of response. Hence the tendency for increasing actuating errors is reduced with a low-gain system. Although less rapid performance and higher errors result from the lower-gain system than might appear possible from a higher-gain system, the lower-gain system is generally the more stable.

Time delays in the various control elements and the controlled system that were not significant in a low-gain system may become appreciable for the system with high amplification. After the corrective action is started and the need for correction has ceased, the inherent time delay of the system elements may prevent stopping the action of the control elements in time to prevent an overshoot by the controlled variable. The overshoot may be greater than that which initiated the control motion, and the process of continued corrective action, building up to violent oscillations, is thus started.

Following this line of reasoning, we see that the time delays present in the control elements cause the instability. By appropriate use of "anticipation" means to compensate for the inherent time "delays" in the control elements, it is possible to obtain a high-gain system with satisfactory stability. Although improved accuracy and adequate stability can be obtained, they are generally achieved only at the expense of additional equipment or complexity.

4.1.4 Mathematical Basis for Stability

Essential and valuable as is the physical picture of stability, mathematical definitions provide more useful and exact means of describing system performance. The principal mathematical means for determining stability of linear control systems are the following: (1) locating by analytical or graphical means the actual position, on the complex plane, of each of the roots of the characteristic equation of the system; (2) applying Routh's stability criterion to the coefficients of the system's characteristic equation; (3) applying Nyquist's criterion to a graphical plot of the open-loop response of the system as a function of frequency for a sinusoidal driving function.

The labor involved in locating the exact position of the roots of the characteristic equation or in calculating their values is such as to limit the use of this method. A graphical method of locating the locus of the characteristic equation roots, known as the "root locus method," has been developed by Evans and has permitted more extensive use of the location of the roots themselves as a means for determining system stability and to a certain extent as a measure of performance. The Routh criterion involves the use of a brief, simple algebraic process and permits the ready determination of the system stability. The graphical data necessary for applying Nyquist's criterion provide quantitative information on the accuracy of the system, the degree of system stability, as well as the system stability itself. Hence it is the Nyquist criterion in one or more of its modified forms that is used extensively to determine system stability.

The Nyquist stability criterion places on a firm mathematical basis the well-known physical fact that, when the feedback signal to a control element is equal in magnitude and in phase with the actuating signal producing it, instability will result. Thus the Nyquist criterion establishes the necessary conditions for stability in terms of the ratio between a sinusoidal actuating signal and the feedback signal. The ratio is expressed by an amplitude and a phase relationship as a function of frequency. This ratio can be determined even when the feedback is not connected to the error-measuring element; thus the system need not be a closed loop when its stability as a feedback control system is being evaluated. As such, the analysis of the problem is reduced somewhat in complexity, although the results are valid for the more complicated feedback control condition of system operation.

4.1.5 Features of Feedback Control System Performance

The two principal advantages of feedback control over control without feedback are that lower tolerances and greater time delays can be permitted for the control elements. To appreciate some of the advantages of the feedback control system, a comparison will be made between the open-loop and closed-loop (feedback) control systems. Figures 4.4 and 4.5 are block diagrams showing how the controlled

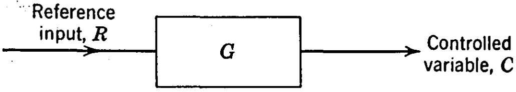


Fig. 4.4 Open-loop control system

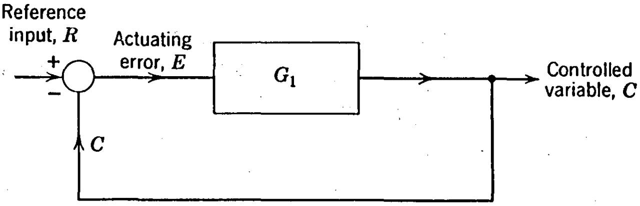


Fig. 4.5 Feedback control system with direct feedback

variable C is related to the reference input R for each of these two systems. The ratios of the controlled variable to the reference input for the open-loop and feedback control systems are, respectively,

$$\frac{C}{R} = G \tag{4.1}$$

and

$$\frac{C}{R} = \frac{G_1}{1 + G_1} \tag{4.2}$$

where $C/E = G_1$ and E is the actuating error.

The terms G and G_1 represent the transfer function of the control elements. In addition to gain or constant terms, the transfer functions may contain time functional relationships having a wide range of values from 0 to ∞ under varying input conditions.

In contrast with the feedback control system in which the controlled variable is compared directly to the reference input to provide the error signal that actuates the control system elements, the open-loop control system makes no direct comparison of these two variables. Thus it is assumed that the transfer function G is known and fixed so that the value of the controlled variable is known for each and every value of the reference input and its time variation. An example of this is a meter element: the deflection of a meter pointer is calibrated against a standard input, for example, a voltage. Subsequently it is assumed that the meter response is identical with its calibration figure and that the same input voltage will produce the same pointer deflection. However, a change in the characteristics of the transfer function G of the meter element may cause the same input to produce a value of output different from the calibrated value. The change in the value of the output, in terms of a change of the transfer function alone, is

$$\begin{pmatrix}\text{Change in}\\ \text{controlled}\\ \text{variable}\end{pmatrix} = \begin{pmatrix}\dfrac{\text{change in } G}{\text{proper value of } G}\end{pmatrix} \times \begin{pmatrix}\text{proper value}\\ \text{of controlled}\\ \text{variable}\end{pmatrix} \tag{4.3}$$

From a knowledge of the value of the controlled variable alone, it is impossible to distinguish between a change in the reference input and a change in the transfer function. Thus very close tolerances in manufacture and constancy of controller characteristics with time are required to obtain high performance of an open-loop system.

For the closed-loop system, there is less need for maintaining the transfer function constant, provided the value of the transfer function

G_1 is large. This may be seen from eq. 4.4 where the expression for the change in the value of the controlled variable in terms of its proper value and the transfer function G_1 of a feedback control system are shown.

$$\begin{pmatrix}\text{Change in}\\ \text{controlled}\\ \text{variable}\end{pmatrix} = \left(\frac{1}{1+G_1}\right)\left(\frac{\text{change in } G_1}{\text{proper value of } G_1}\right) \times \begin{pmatrix}\text{proper value}\\ \text{of controlled}\\ \text{variable}\end{pmatrix} \tag{4.4}$$

Although the change in the controlled variable is proportional to the change in the transfer function G_1, there is a greatly reduced multiplying factor of $1/(1 + G_1)$ that decreases the size of the actual change in the controlled variable caused by changes in G_1. If G_1 has a value of the order of 10 or more, the advantages of feedback control operation in this respect are at once evident. Thus even with the use of lower-precision components or ones subject to wider variation under use, it is still possible to maintain high-precision feedback control performance.

Two additional items are worthy of note in this brief comparison of open-loop and feedback control systems. First, it is not always physically possible to perform a direct comparison of the controlled variable to the reference input as shown in Fig. 4.5. For example, the reference input may contain present-input data for a computer that has as its controlled variable some functional relationship of the reference input, such as the future value of the controlled variable after some period of time, for instance, T_F. Only by performing a duplicate calculation or by allowing the time T_F to transpire is it possible to know whether the performance of the system T_F seconds ago was correct. Thus it is clear that for certain control problems it is difficult to provide an error-sensing device that can compare the value of the controlled variable to the reference input and modify the controlled variable properly.

Second, a feedback control system may be designed to have the controlled variable very nearly equal to the reference input for only certain types of input signals. For other signals the feedback control system may purposely be designed to make the controlled variable reproduce little or none of the input. As an example of this, position servomechanisms are frequently designed to possess low-band-pass characteristics. By this it is meant that reference input signals having low frequency or constant values are transmitted without appreciable error. The system transmission for higher-frequency signals, which generally represent extraneous inputs, is purposely designed to be small, and the controlled variable has a high error for this type of input.

4.2 FEEDBACK CONTROL SYSTEM PROBLEMS

The preceding material has been of a generally descriptive nature to give us an overall picture of the use of feedback control as it might be applicable as a part of the automation problem; but this material has not served to emphasize the nature of the details of feedback control systems. The material that follows will bring out the nature of the elements that may make up the control system and will show that the system error characteristics are a function of the nature of the elements. System stability, frequency response, and transient response will next be presented. Finally the effects of disturbances in the input or in other locations will be described.

4.2.1 Mathematical Nature of Control System Elements

A control system element may be defined as one of a number of parts that, when connected together, form the feedback control system. Ideally, each control system element performs some specific function that is necessary to the overall performance of the system. In general, a control system element receives an input signal from another element or group of elements and transmits its output to another control system element or group of elements.

The transfer function of a control system element is a mathematical expression that indicates the dynamic characteristics of the element in terms of the ratio of the output to the input of the element. This transfer function, $G(s)$, is expressed in the form of the Laplace transform of the output to input ratio in which all the initial values are set equal to zero. Thus, from the relationship

$$O(s) = G(s)\,I(s) \tag{4.5}$$

where $I(s)$ = Laplace transform of the input
$G(s)$ = transfer function
$O(s)$ = Laplace transform of the output

the transfer function is obtained as the ratio

$$G(s) = \frac{O(s)}{I(s)} \tag{4.6}$$

These two equations show that, physically, the transfer function is the dynamic characteristic that modifies the input in transferring it to become the output quantity.

When the input to a control system element is varying sinusoidally, the transfer function may be determined by replacing s by $j\omega$. The transfer function is then said to be in its complex or sinusoidal form. For example the sinusoidal form of the equation above is

$$G(j\omega) = \frac{O(j\omega)}{I(j\omega)} \tag{4.7}$$

which serves to indicate that sinsuoidal excitation has been impressed as the input. Figure 4.6 is a block diagram used to indicate the relationship of eq. 4.7.

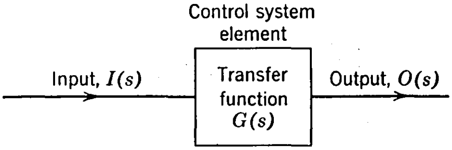


Fig. 4.6 Block diagram representation of control system element transfer function

Starting with the physical relationships of proportionality, differentiation, integration and/or other effects present in the control element itself, the transfer function of a control element can be expressed. A brief tabulation of the transfer functions as well as their schematic representation for some representative control elements is contained in Table 4.2. From this table it is evident that although the physical appearances and methods of operation for many of these devices are different, the transfer functions and therefore the time responses for many of them are similar. By considering the various control system elements from the point of view of their transfer functions, we are able to determine the performance of the control system of which they form a part.

Although the mechanical, electrical, and hydraulic elements listed are shown to have linear coefficients, transfer functions can also be used to describe the characteristics of nonlinear devices. The describing function of a nonlinear element is a function of both amplitude and frequency; thus the same sort of analysis and synthesis techniques can be used with nonlinear systems as are used with linear systems. As implied in Table 4.2, the describing function of a nonlinear element assumes that the input to the element is sinsuoidal and that the fundamental component of the output at the frequency of the input is ade-

Table 4.2 Transfer Functions for Representative Elements

Element	Transfer Function
Mechanical	
Rotary spring inertia damper	$\frac{\theta_1(s)}{\theta_m(s)} = \frac{1/n}{(J/K_s)s^2 + (K_D/K_s)s + 1}$
Translatory spring mass damper	$\frac{X(s)}{Y(s)} = \frac{1}{(M/K)s^2 + (D/K)s + 1}$
Spring-dashpot (lag)	$\frac{X(s)}{Y(s)} = \frac{1}{(D/K_s)s + 1}$
Spring-dashpot (lead)	$\frac{X(s)}{Y(s)} = \frac{(D/K_s)s}{(D/K_s)s + 1}$
Electrical	
D-c motor (speed control)	$\frac{S(s)}{V_a(s)} = \frac{1}{K_e(T_m s + 1)}$
D-c generator-motor (speed control)	$\frac{\theta(s)}{E_c(s)} = \frac{K_g/K_e R}{S(T_f s + 1)(T_m s + 1)}$
Stabilizing network (lead)	$\frac{E_o(s)}{E_{in}(s)} = \frac{T_s}{T_s + 1}$
Stabilizing network (lag)	$\frac{E_o(s)}{E_{in}(s)} = \frac{T_2 s + 1}{T_1 s + 1}\ T_1 > T_2$
Stabilizing network (lead-lag)	$\frac{E_o(s)}{E_{in}(s)} = \frac{T_1 T_2 s^2 + (T_1 + T_2)s + 1}{T_1 T_2 s^2 + (T_1 + T_{12} + T_2)s + 1}$
Hydraulic	
Valve-piston (negligible reaction)	$\frac{X(s)}{Y(s)} = \frac{C_1}{s}$
Valve-piston (spring dominant)	$\frac{X(s)}{Y(s)} = C_2$
Hydraulic motor	$\frac{\theta(s)}{Y(s)} = \frac{S_p/d_m}{s\left(\frac{VJ}{B{d_m}^2}s^2 + \frac{LJ}{{d_m}^2}s + 1\right)}$
Valve-piston linkage (lag)	$\frac{X(s)}{Y(s)} = \frac{b/a}{T_v s + 1}$
Valve-piston linkage (lead)	$\frac{X(s)}{Y(s)} = \frac{d}{1 + b}\left[\frac{T_3 s + 1}{T_v s + 1}\right]\ T_3 > T_v$
Nonlinearities	
Time delay	e^{-sT}
Describing function $G'(a)$	$G'(a) = \frac{\frac{1}{\pi}\int_{-\pi}^{+\pi} O_a(t)e^{-j\omega t}\,d\omega}{I_a}$

quate to represent the actual output with sufficient accuracy. For many nonlinearities in practical control systems this is a reasonably good assumption.

Nonlinearities may be of an unavoidable or of an intentional type. Examples of unavoidable nonlinearities might be backlash, dead band, time delay, saturation, or other effects that are inherent in the equipment or element itself. Presumably these nonlinearities can be designed to be small enough to be acceptable or steps must be taken to minimize their effect.

Intentional nonlinearities such as nonlinear gain controls, clamping, and limiting devices are being used more extensively to perform decision functions which can be very valuable and important in automatic controls. In many cases, the requirements of the control vary with different conditions of such quantities as position, speed, error, and other parameters. Use of this supplemental information in a nonlinear fashion can be valuable in improving system performance.

A special form of nonlinearity is present in sampled-data systems where the output of an element is not continuous but provided only at uniform instants of time. Generally some sort of "holding" device is provided to average the output during the times between sampling instants. Since digital computer output data and other important input devices to control systems are digital in form, approximate analytical methods have been developed for representing these nonlinear portions of otherwise linear control systems.

Although the material that follows emphasizes the analytical approach to the "paper-and-pencil" solution of feedback control problems, the use of problem-solving machines to perform much of this work is gaining much greater acceptance as more of this equipment is available. General-purpose computers, both analog and digital, permit the more rapid and accurate solution of design problems, especially those of a complicated or involved nature. With these tools, the physical problem of what takes place in the process or mechanism assumes much greater importance and more emphasis can be placed on the physical process, which is where it rightfully belongs.

Regardless of whether the actual form of the feedback control is electric, electronic, hydraulic, or pneumatic, the same basic principles of design and application can be applied. In each case, power amplifications ranging from one to two up to the millions can be obtained. The nature of the power available, the requirements of the automation process, and the importance of such features as size, weight, reliability, maintenance, and performance will all influence the selection of the most suitable form of control for the job.

4.2.2 Controlled-Variable Response from Constant Actuating Error

An important measure of the performance of a feedback control system is the actuating error that is required to produce a desired controlled-variable response. An appreciation of the nature of the relationship between the controlled-variable response and the actuating error can be obtained by considering a simple feedback control system with direct feedback as shown in Fig. 4.7.

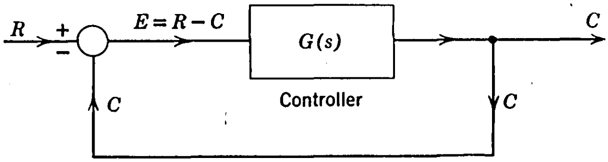


Fig. 4.7 Simple closed-loop control system with direct feedback

The actuating error E is the difference between the reference input R and the controlled variable C.

$$E(s) = R(s) - C(s) \tag{4.8}$$

The transfer function $G(s)$ relates the controlled variable to the actuating error.

$$G(s) = \frac{C(s)}{E(s)} \tag{4.9}$$

For control systems with power elements, such as motors with their associated amplifiers and other required stabilizing means, the transfer function has the general form

$$G(s) = \frac{K(1 + a_1 s + a_2 s^2 + \cdots)}{s^n(1 + b_1 s + b_2 s^2 + b_3 s^3 + \cdots)} \tag{4.10}$$

where K is the gain of the overall transfer function, a_1, a_2, b_1, b_2, etc., are constant coefficients, and n is an integral positive number indicating the number of series integrations in the transfer function.

The numerator and denominator of this equation can be factored into a series of single or quadratic terms of the form $(1 + Ts)$ or $(1 + 2\zeta Ts + T^2s^2)$ where the T's are real numbers having the dimensions of time. For a typical case, $G(s)$ might appear as

$$G(s) = \frac{K(1 + T_a s)(1 + T_b s)}{s^n(1 + T_1 s)(1 + T_2 s)(1 + T_3 s)} \tag{4.11}$$

To rewrite this equation in terms of the controlled variable and the actuating error,

$$\frac{s^nC(s)}{K} = \frac{(1 + T_as)(1 + T_bs)\,E(s)}{(1 + T_1s)(1 + T_2s)(1 + T_3s)}, \tag{4.12}$$

For a constant value of the actuating error, E_0, the value of the control variable is found to be

$$\frac{1}{K}\frac{d^nc(t)}{dt^n} = E_0 \tag{4.13}$$

A physical interpretation of this equation is as follows.

With $n = 0$, i.e., with no series integration, a constant value of the controlled variable is obtained for a constant actuating error. Conversely, a constant actuating error is needed to produce a constant value of the controlled variable; the greater the gain K, the lower is the error required.

With $n = 1$, one series integration, a constant value of rate of change of controlled variable is produced by a constant actuating error. No error is required to maintain the controlled variable equal to a constant reference input. A constant actuating error is needed to produce a constant rate of change of controlled variable; the greater the value of gain, the lower is the error required.

With $n = 2$, two series integrations, a constant value of the second derivative of the controlled variable is produced by a constant actuating error. No error is required to maintain the controlled variable at a constant rate of change with its position equal to that of the reference input. A constant actuating error is needed to produce a constant acceleration of the controlled variable; the greater the value of gain, the lower is the error required.

Although the condition of constant actuating error is unquestionably a unique one, the preceding material serves to provide some general indication of the kinds of transfer function characteristic that make for an accurate control system. Certainly, increasing the overall gain tends to decrease the value of actuating error. Were more general dynamic conditions considered, however, it would be found that the presence of a number of series integrations tends to require larger errors for a given amplitude of controlled-variable motion of any appreciable variation with time. In addition, the presence of large time constants in the denominator of the transfer function also tends to require larger errors for a given amplitude of controlled-variable motion. The conditions of high gain and numerous integrations that appear attractive from the point of view of permitting low steady-state errors are, unfortunately, not ones that are conducive to stable operation.

4.2.3 Stability of Feedback Control Systems

The purpose of any feedback control system is to have the controlled variable or output of the system bear a definite and known relationship to the desired value or reference input. Therefore it is necessary that the response of the system to any temporary disturbance be a decaying one that vanishes a reasonable length of time after the cessation of the disturbance. Systems in which the motion of the controlled variable is random or erratic and is not responsive to the reference input function and systems in which undesired self-sustained oscillations of the controlled variable are present are said to be unstable. It is not merely necessary that a system be designed to be stable. It is essential that the system be sufficiently stable so that transient disturbances will decay quickly and there will be rapid recovery by the controlled variable.

Nyquist's stability criterion. The Nyquist stability criterion is simple and can be employed to advantage to determine the stability of a feedback control system. Consider the generalized block diagram shown in Fig. 4.8 where $G(s)$ represents the transfer function of the forward loop and $H(s)$ represents the transfer function of the feedback loop. The following relationships exist:

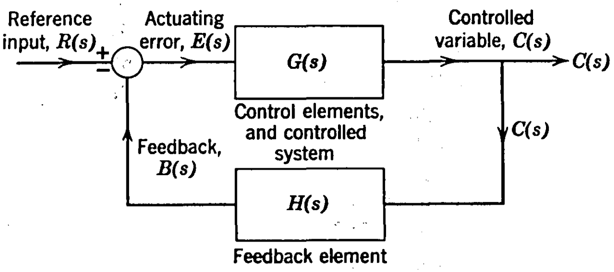


Fig. 4.8 Feedback control system with control and feedback elements

$$\begin{aligned} C(s) &= G(s)\,E(s) \\ R(s) - B(s) &= E(s) \qquad\qquad (4.14) \\ B(s) &= H(s)\,C(s) \end{aligned}$$

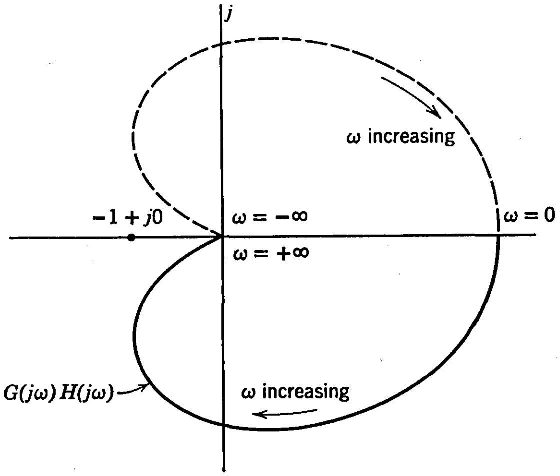


Fig. 4.9 Transfer function of stable regulator-type system

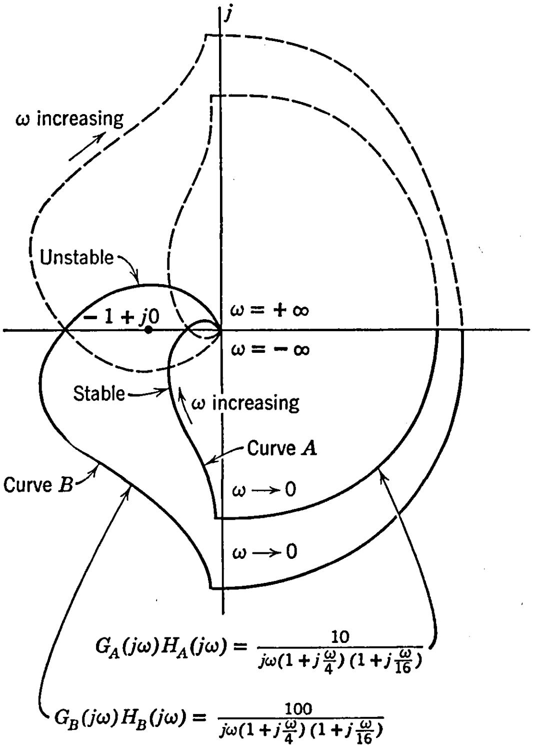


Fig. 4.10 Position control system for stable and unstable operating conditions

From these the resultant expression for $C(s)$ in terms of $R(s)$ is

$$C(s) = \frac{G(s)\,R(s)}{1 + G(s)\,H(s)} \tag{4.15}$$

The Nyquist stability criterion for such a feedback control system can be applied with the aid of a plot of the $G(s)\,H(s)$ function in the complex plane for $s = j\omega$ with all values of ω from $-\infty$ to $+\infty$. Draw the vector from the $-1+j0$ point to a point on this curve and observe the rotation of this vector as ω varies from $-\infty$ to $+\infty$. The simplified form of the Nyquist criterion then states that the net counterclockwise rotation of this vector must be zero for the system to be stable.

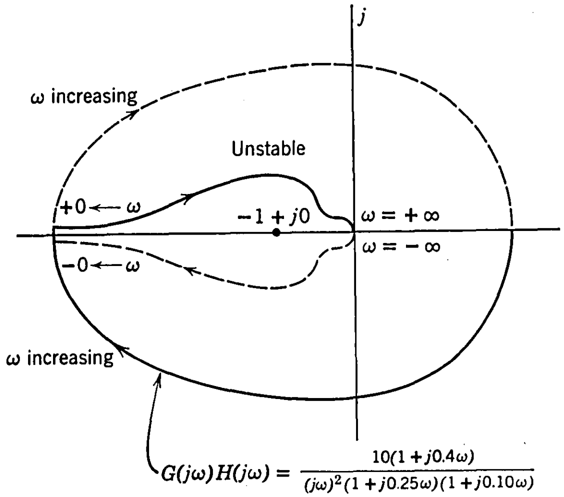


Fig. 4.11 Position control system containing torque motor shown for unstable condition of operation

The examples shown in Figs. 4.9, 4.10, 4.11, and 4.12 serve to illustrate some of the forms which the $G(j\omega)\,H(j\omega)$ functions take when plotted on the complex plane. Both stable and unstable systems are indicated in these illustrations.

Attenuation-phase versus frequency stability representation. An alternate representation of the feedback control system stability criterion is that obtained by a separate plot of the magnitude and of the phase angle of the $G(j\omega)\,H(j\omega)$ function obtained above as a function of frequency. This form of presentation is a result of the work of Bode and lends itself to simplified design synthesis techniques.

To appreciate the significance of the attenuation-phase method of determining stability, consider the transfer functions shown in Figs. 4.13 and 4.14 in the vicinity of the unit circle. The same frequency-varying function is illustrated for each of the three transfer functions in each figure; however, different values of gain are used for each case with the gain increasing from K_a to K_c. With the gain K_a there is a margin of phase angle in each case when the gain of the transfer function is unity and the systems are stable. With the gain K_b, in each case there is no margin of phase angle when the gain of the transfer function is unity and the systems are on the border line of stability. With K_c, the phase angle margin is negative when the transfer function gain is unity and the systems are unstable.

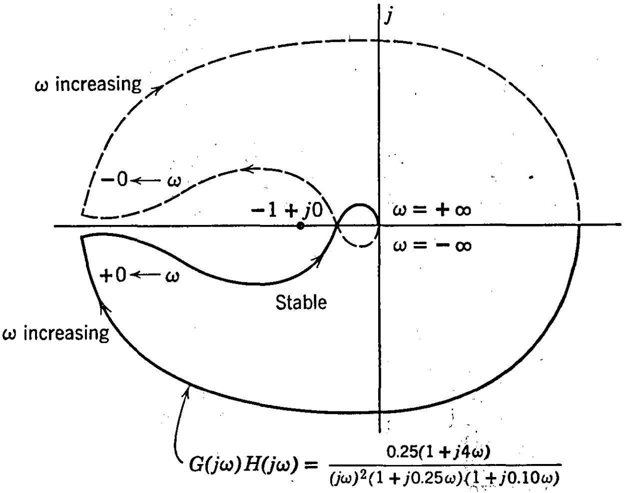


Fig. 4.12 Position control system for stable condition of operation

By reasoning as is indicated in these illustrative examples, the following generalization may be made for systems in which the abbreviated form of the Nyquist criterion applies: Systems having positive phase margin when their transfer function $[G(j\omega)\ H(j\omega)]$ crosses the unit circle are stable, whereas systems with negative phase margin when their transfer function crosses the unit circle are unstable.

By utilizing this abbreviated stability criterion and taking advantage of the relationships between the gain (or attenuation) and phase characteristics as a function of frequency as described by Bode, the stability of a system may be determined from a plot of the attenuation and phase of the open-loop transfer function as a function of frequency. The significant region of interest, stability-wise, is in the vicinity of unit gain, 0 decibels. Figures 4.15 and 4.16 show such plots for stable and unstable control systems respectively. The convenience of this method of representation in system synthesis has resulted in its extensive use in control system design.

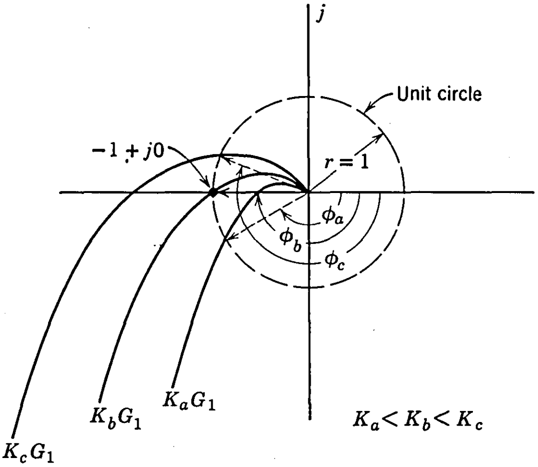


Fig. 4.13 Similar transfer functions on complex plane, showing significance of phase margin at unit circle

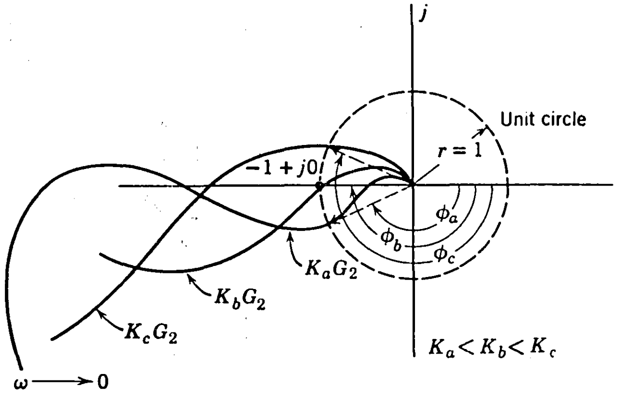


Fig. 4.14 Similar transfer functions on complex plane, showing significance of phase margin at unit circle

Root locus method. The root locus method of describing the stability of a system indicates directly the location of the roots of the closed-loop feedback control system. As such it specifies as a function of the open-loop gain, the decrement rates, and the oscillation frequencies for any disturbance to the system. The basis for the determination of

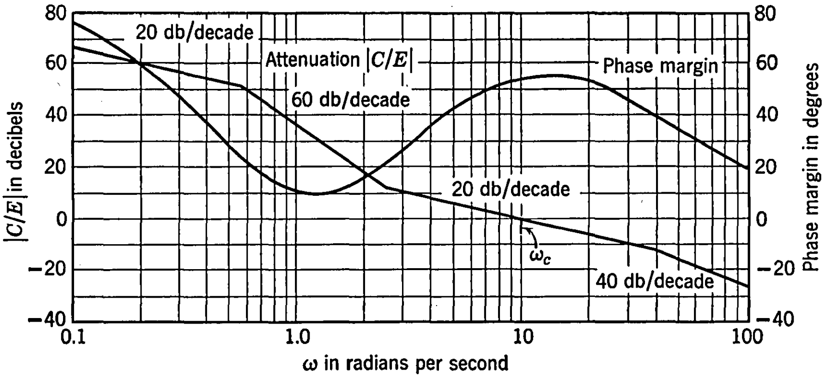


Fig. 4.15 Attenuation and phase margin characteristics as a function of frequency for a stable system

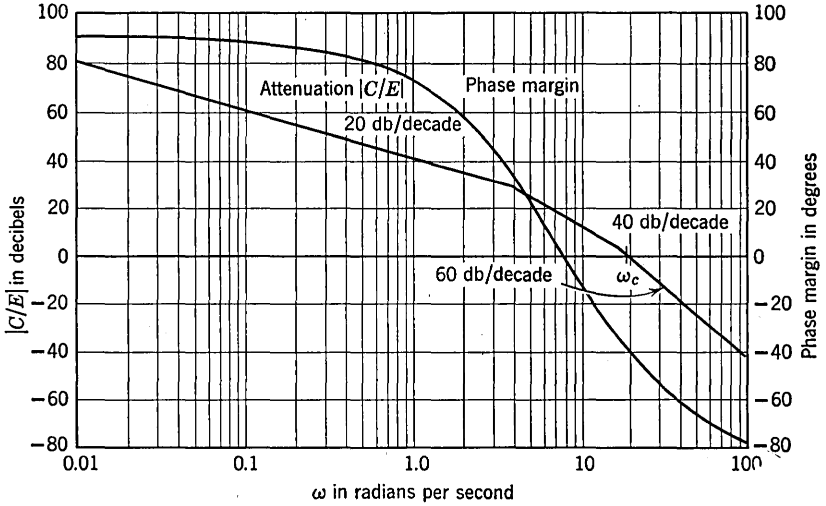


Fig. 4.16 Attenuation and phase margin characteristics as a function of frequency for an unstable system

$$\frac{C}{E} = \frac{100}{j\omega(1 + j0.25\omega)(1 + j0.0625\omega)}$$

the loci of the roots is the characteristic equation, although here again the actual open-loop transfer function is used as a starting point for the graphical analysis.

From eq. 4.15 the denominator of the expression for the controlled variable is $1 + G(s)\,H(s)$ and the condition for which this is equal to zero is $G(s)\ H(s) = -1$. For this equality to be realized, the phase of the term on the left must be 180 degrees and its magnitude must be equal to unity. A plot of the locus of all the points for which the phase is 180 degrees, or odd multiples thereof, is known as the root locus. It is along this locus that the roots of the closed loop of the feedback control system lie.

Figure 4.17 shows such a locus and indicates the specific location of the roots for increasing values of K from K_1 to K_4 as shown. The locus has its origin at the roots of the open-loop transfer function and the roots gradually move along the locus as the gain is increased until a gain, such as K_3 in the example cited, is reached where the system has zero damping and oscillations of a frequency ωc (as shown) are sustained. Another root corresponding to the same gain K_3 lies along the negative real axis and has a relatively high decrement rate.

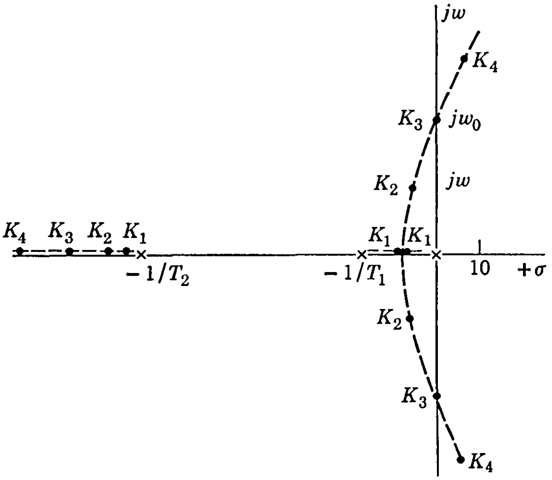


Fig. 4.17 Effect of loop gain on the position of the roots along the locus

Since the locations of the roots of the closed-loop system are known, the actual values of the time for the transients to decay are known directly from the plot of the root locus. Likewise the oscillation frequency of the transient can be determined, in radians per second, for the value of gain selected. It is also possible to indicate, by straight lines through the origin, the loci of positions of roots having a constant ratio of decrement rate to oscillation frequency.

The process of determining the locus of the roots of the closed-loop feedback control system has been greatly speeded up as a result of the work of Evans and the use of his spirule. In addition to these graphical methods, analytical methods have been developed to facilitate the exact calculation of the closed-loop roots of the system.

4.2.4 Frequency Response

The frequency response of a control system is a plot of the magnitude of the ratio of its output to its input as a function of frequency for a sinusoidal input. Figure 4.18 shows the magnitude and phase angle

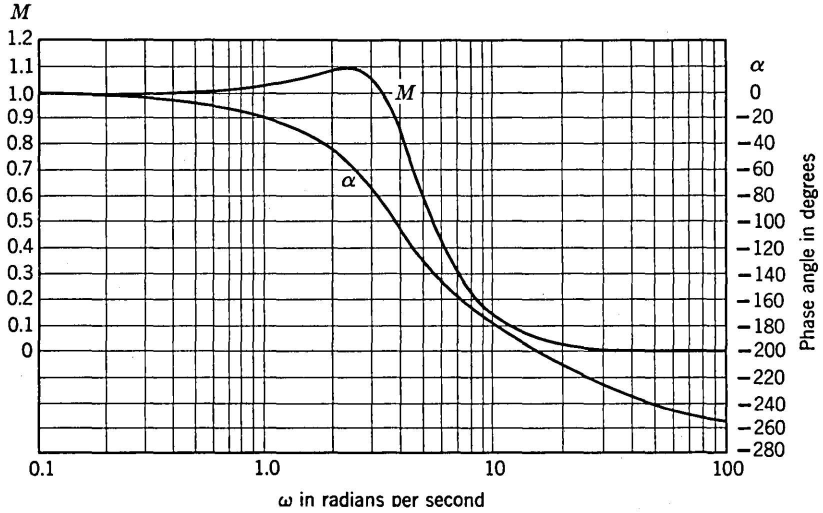


Fig. 4.18 Magnitude and angle of closed-loop frequency response

of the closed-loop frequency response of a simple feedback control system. Although control systems in practical applications may seldom be subjected to pure sinusoidal inputs, nevertheless, the ease with which stable and accurate systems may be designed and checked makes the use of the frequency response approach a convenient method of describing system performance.

From eq. 4.15, the frequency response of a feedback control system with direct feedback, i.e., $H(s) = 1$, can be written as

$$\frac{C}{R} = \frac{G(j\omega)}{1 + G(j\omega)} = Me^{j\alpha} \tag{4.16}$$

where $M = \left|\frac{C}{R}\right|$ = magnitude of ratio of controlled variable to reference input

α = angle between controlled variable and reference input

Figure 4.19 shows how the complex-plane plot used to determine the system stability by Nyquist's method can also indicate the ratio of C/R. The distance from the $-1+j0$ point to the transfer function plot is proportional to R/E at the particular frequency. As such it is inversely proportional to the error function E/R. The distance from the origin to the same point on the transfer function is proportional to C/E for the same frequency. Hence the ratio of $C/R = (C/E)/(R/E)$ for each point on the complex plane is unique, and frequency response information as well as an indication of stability are available from the single complex-plane plot. Figure 4.20 shows a plot on the complex plane of the loci of constant values of the magnitude $C/R = M$.

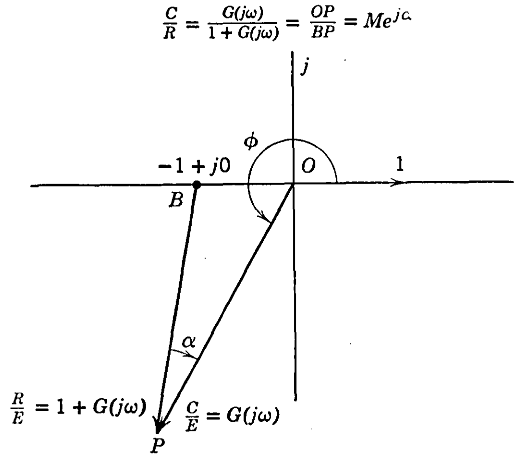


Fig. 4.19 Conventional form of complex-plane diagram for control system

By utilizing the basic frequency response concepts, a number of ways of representing the stability and performance problems have been derived to facilitate the design and synthesis of control systems. The reader is referred to texts devoted to the subject of feedback control for more detailed information on such design techniques (1, 2, 6).

For many control systems it is possible to measure experimentally the frequency response. In these instances specifications on the control system performance may be made in terms of such quantities as $C/R|_m$ $(= M_m)$, the maximum of the frequency response ratio, and ω_m, the frequency at which $C/R|_m$ occurs. For most control system applications, satisfactory values of $C/R|_m$ are less than 1.4. Since in a number of control system applications the input can be described in terms of an equivalent sinusoid or sinusoids, the magnitude and phase shift of the frequency response of the system can be used to determine the system performance under these input conditions.

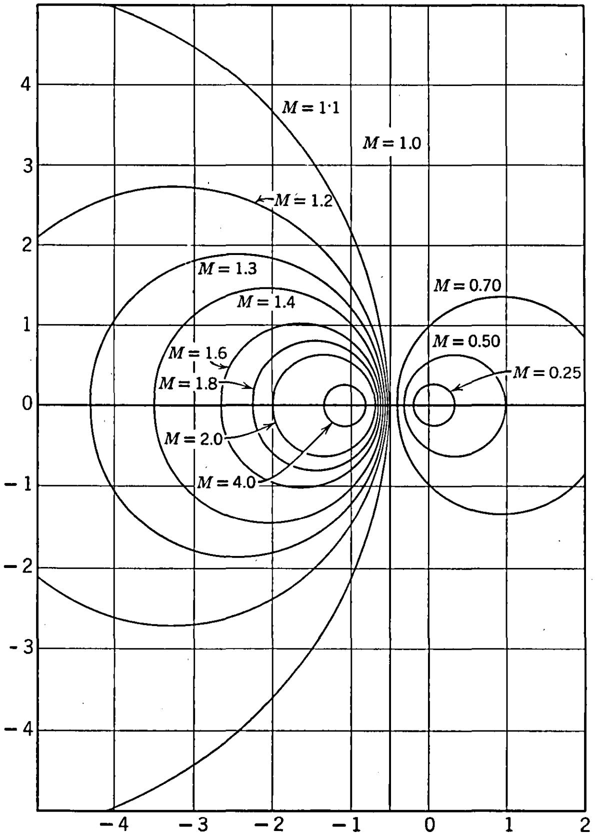


Fig. 4.20 Loci of constant values of the magnitude $C/R = M$

4.2.5 Transient Response

The performance of a control system is frequently judged on the basis of its response to a transient input consisting of a step function of position. Such an input is frequently encountered in practice and is generally easy to obtain experimentally. Further, the system response to such an input is frequently stipulated in the control system specifications. Figure 4.21 indicates a comparison of the frequency response and the transient response to a step input for a particular system. Performance values that are of interest for a transient step input are the ratio of the magnitude of the maximum of the transient output to the value of the input, $C/R|_p$ $(= M_p)$, and t_p, the length of time it takes, from the start of the transient, for the maximum of the transient output to occur.

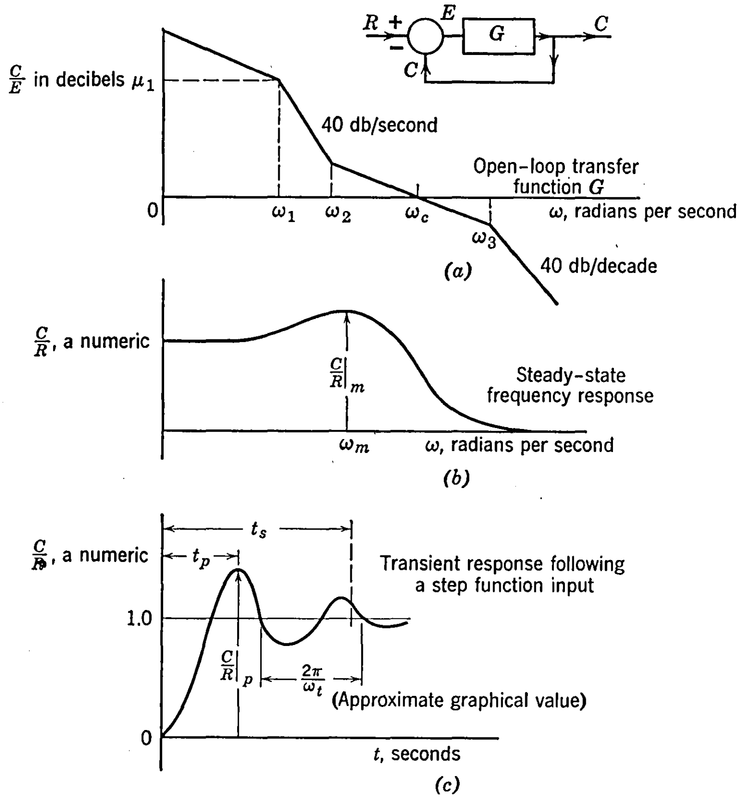


Fig. 4.21 Sketches showing nomenclature used to describe various characteristics of servomechanism performance

Also of interest are t_s, the length of time it takes, from the start of the transient, for the output to continue to differ from the input by less than a small fixed per cent, and ω_t, the lowest frequency of oscillation of the transient response.

For feedback control systems in which the transient response is of principal importance, determination of the transient performance may, as in the root locus method, take the form of finding the decrement rate and the frequency of the oscillations of the system. The actual magnitude of the transient response as a function of time can also be established when the values of the roots of the system are known. The time response $f(t)$ may be expressed as

$$f(t) = C_1e^{s_1t} + C_2e^{s_2t} \cdots + C_ke^{s_kt} \cdots + C_ne^{s_nt} \tag{4.17}$$

where

$$C_k = \left[(s - s_k) \frac{A(s)}{B(s)} \right]_{s=s_k}$$

and $A(s)/B(s)$ is the response function of the control system subjected to a step function input. Graphical means may be employed utilizing the root loci, or, when sufficient accuracy is required, exact calculation of the complete transient response can be performed.

Approximate methods of determining the transient response from the frequency response characteristic have been developed by Floyd for use when accuracy of a few per cent is sufficient. These methods are particularly valuable where frequency response methods of system design have been used and a refined estimate of the transient response to a step input is sought. The inverse transformation of transient response data into frequency response form has also been reduced into a routine process. This operation is particularly useful in permitting experimental transient response measurements to be expressed in terms of their equivalent frequency response.

For approximate, rule-of-thumb purposes it is convenient to be able to relate a few significant performance criteria for frequency response and transient step inputs. For typical position controls, in the range $1.0 < C/R|_m < 1.4$,

$$\left.\frac{C}{R}\right|_m \geq \left.\frac{C}{R}\right|_p \tag{4.18}$$

As $C/R|_m$ approaches unity, in some cases $C/R|_p$ may exceed $C/R|_m$ by a few per cent. When $C/R|_m$ approaches 1.4 or greater, $C/R|_p$ is less than $C/R|_m$. Thus by limiting the value of $C/R|_m$ in design, a satisfactorily small value of $C/R|_p$ tends to be obtained.

The frequency response and the transient response characteristics are somewhat reciprocal. As the range of frequencies covered by the frequency response is extended, the time required for the control to reach its peak transient value is decreased. For many representative position controls the ratio $\omega_c t_p \simeq 3.5$ is valid. The frequency at which the open-loop transfer function has a magnitude of unity is ω_c and t_p is the time it takes for the transient response to a step input to reach its maximum. Thus, for example, if the time required for the transient response to reach its peak must be less than 1 second, ω_c must be 3.5 radians per second or greater.

Although these empirical relationships between frequency response and transient response characteristics are of use only for approximation purposes, they serve to indicate some helpful relationships between these two different ways of describing system performance.

4.2.6 Effect of Disturbances to Control Systems

The primary function of a feedback control system is to cause the controlled variable to be equal to the reference input, subject to the independent action of a number of external conditions. During the course of operation of the control, these external conditions may vary from their nominal values over a fairly wide range, and the effect on the performance of the system must be determined and made to fall within acceptable limits. Varying external-load torques in a speed control and air gusts in an airplane altitude control are examples of such disturbances. The effect of these disturbances is to introduce in the system an extraneous input that also affects the value of the controlled variable. Figure 4.22 illustrates the general case of a feedback control system having two inputs, R_1 the one the controlled variable C is supposed to follow, and R_2 a disturbance input independent of R_1. The transfer functions of the control system elements indicated are G_1, G_2, and H.

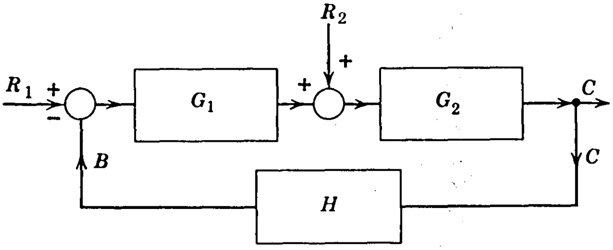


Fig. 4.22 Block diagram of a servomechanism having more than one input

If the two inputs are applied simultaneously, the expression for the controlled variable is

$$C = \frac{G_1G_2R_1}{1 + G_1G_2H} + \frac{G_2R_2}{1 + G_1G_2H} \tag{4.19}$$

Thus it will be noted that equal amounts of the inputs R_1 and R_2 may produce markedly different effects on the output C, depending on the value of G_1. With R_1 the input that must be transmitted,

C/R_1 will be approximately unity for the low-frequency region. This fact will help establish the ratio of G_1G_2 to $1 + G_1G_2H$. The ratio of C/R_2 can be related to that for C/R_1 as

$$\frac{C}{R_2} = \frac{1}{G_1}\frac{C}{R_1} \tag{4.20}$$

If the R_2 input is an undesirable disturbance that should be eliminated from the output, then at low frequencies it is necessary for G_1 to be as large as possible.

Paying attention to considerations such as these—the reference inputs the controlled variable should follow and the disturbances the controlled variable should not follow—permits a more acceptable overall control system design.

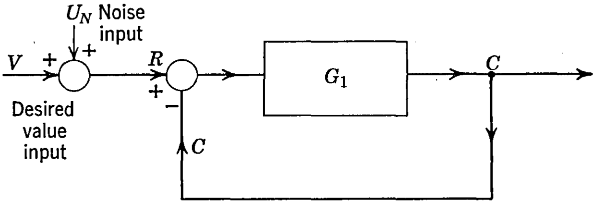


Fig. 4.23 Block diagram showing desired input and noise input to a feedback control system

Disturbance inputs in the form of noise or other extraneous signals may also appear in the reference input to a control system. Figure 4.23 illustrates how a disturbance signal or noise, U_N, and the desired input value, V, may combine to form the reference signal, R, which serves both as the resultant input to the system and as the value which the controlled variable C endeavors to follow. Since the desired and the disturbance signals may differ in their magnitude and/or frequency characteristics, it may be possible to affect one more than the other with the proper selection of control system performance characteristics. Cognizance of the presence of the extraneous signals must be taken in design of the feedback control system so that the effect can be minimized in the output.

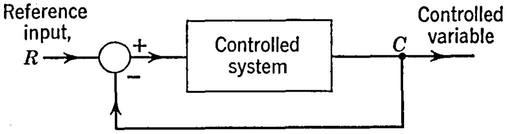
Feedback controls used in conjunction with automation processes may have to accept a wide range of disturbance inputs. Either the control must be designed to operate satisfactorily in spite of these inputs or the overall system design must be modified to reduce the amount of the disturbance inputs initially present.

4.3 MULTIPLE CONTROL SYSTEMS

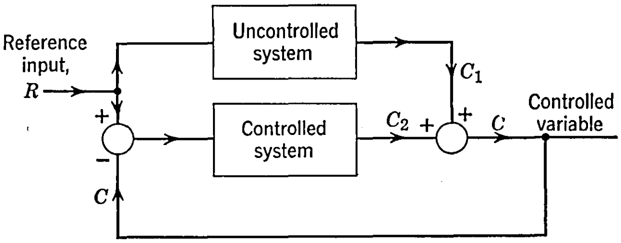
The preceding material has served to point up some of the techniques and considerations that are an important part of feedback control per se. Since in many automation processes individual feedback control systems play an important part, it was worth while to place this emphasis on feedback control apart from automation as such. In the material that follows, more attention is placed on the feedback control problems associated principally with automation.

4.3.1 System Synthesis

The repetitive nature of the automation process and the high level of performance obtainable with automatic control may permit unusual and novel arrangements of controls that are not used under manual or semiautomatic production methods. Since the control requirements of feedback control influence appreciably the nature of the equipment to do the job, considerable thought and attention should be given to various ways of mechanizing a given automation process.



(a) Closed-loop system



(b) Closed-loop plus open-loop system

Fig. 4.24 Closed-loop control contrasted with combined closed-loop and open-loop control

In some feedback control systems the desired controlled-variable performance is obtained by using only a closed-loop system such as is shown in Fig. 4.24*a*. In this system the entire controlled-variable power is handled by the controller, and an accuracy of 1 per cent for the control system output means an accuracy of 1 per cent in terms of the output of the controlled system.

Another solution to this problem, shown in Fig. 4.24*b*, is to use an uncontrolled system (i.e., an open-loop control) to produce C_1, the major portion of the controlled-variable output, with a controlled system (i.e., a closed-loop control) to produce C_2, the remaining portion of the controlled-variable output. In a system of this sort, the better a job the uncontrolled system can do, the less exacting are the demands of the controlled system. However, in many cases the cost of improving the performance of the uncontrolled system beyond a certain point becomes much more expensive than would be the cost of the auxiliary controlled system.

If the uncontrolled system can supply 90 per cent of the resultant system output from a power and from an accuracy point of view, the controlled system needs to be designed to be capable of handling only 10 per cent of the power and motion. To obtain 1 per cent accuracy for the overall controlled variable, an accuracy of only 10 per cent of the maximum value of the controlled-system output, C_2, is needed.

J. R. Moore has written an interesting paper describing a number of combined open-loop and closed-loop control systems and has indicated how improved system performance can be obtained by utilizing open-loop components in conjunction with closed-loop controls having less exacting performance requirements (16).

In addition to the synthesis of the general system configuration, there are other important considerations that should be given to the control system design. By modification of the reference input to account for systematic errors caused by the nominal characteristics of the control system, it may be possible to reduce appreciably the systematic repetitive errors in the process. The location of the error-sensing means, the rigidity of its mounting, and its associated dynamic response also influence the degree of difficulty of the control problem. The presence of backlash or hysteresis in either the power or measuring portion of the control is also to be avoided. In automation applications, especial emphasis must be given to supplemental judgment controls such as limits, stops, and safety features that in a manual system might be supplied by the operator. Frequently, these considerations described above are decisive in establishing the worth of a feedback control system.

4.3.2 System Integration and Interconnection

In addition to the problems associated with each control in an automation process are the problems of all the controls acting together.

A consistent overall objective in terms of product output, accuracy, reliability, and other criteria of performance must be established and adhered to throughout. When closed-loop operation of the complete automation process is employed, the stability and accuracy of this loop must be maintained. Design methods for establishing satisfactory system performance, similar to those employed in the individual controls, should be used for the overall loop.

A system design approach known as "schedule and trim," which has been useful in regulator design, lends itself to application in some automation processes. The idea underlying schedule and trim is that for a certain desired output condition there are nominal values for a number of the control elements that make up the control process. These nominal values are therefore scheduled directly along with the reference input. However, owing to the lack of uniformity of the controls and the process, the nominal values are not adequate and it is necessary to trim the scheduled values to account for the discrepancies between the scheduled values and those necessary to achieve the overall desired result. Therefore, the trimming control is done automatically in a closed-loop control. Figure 4.25 shows in schematic form an example of the schedule-and-trim approach to the control system problem. The trim effect is shown as a summation of feedbacks from the element directly involved as well as from other elements in the system.

The nominal control selector may contain reference, schedule, and even feedback functions that vary with the desired conditions of operation. As such it tends to provide a good deal of the counterpart of the human intelligence that is essential to satisfactory manual operation of a control system. The manpower and material required to determine and provide the control selector function tend to be of a greater degree of complexity than those required for the corresponding manual control equipment.

The interdependent nature of the various portions of the automation process makes it possible to interrelate, in terms of cause and effect, the results in one portion of the control process with action preceding or following it. Cross-connecting of control signals or automatically utilizing gain or speed-changing devices may permit a smoother operation of the overall control system without requiring complete closed-loop operation of the automation process. Figure 4.26 indicates schematically how cross-connecting of control signals can be accomplished on a particular cutter control.

System integration also implies the proper selection of power supplies and the choice of the most suitable form of power controls,

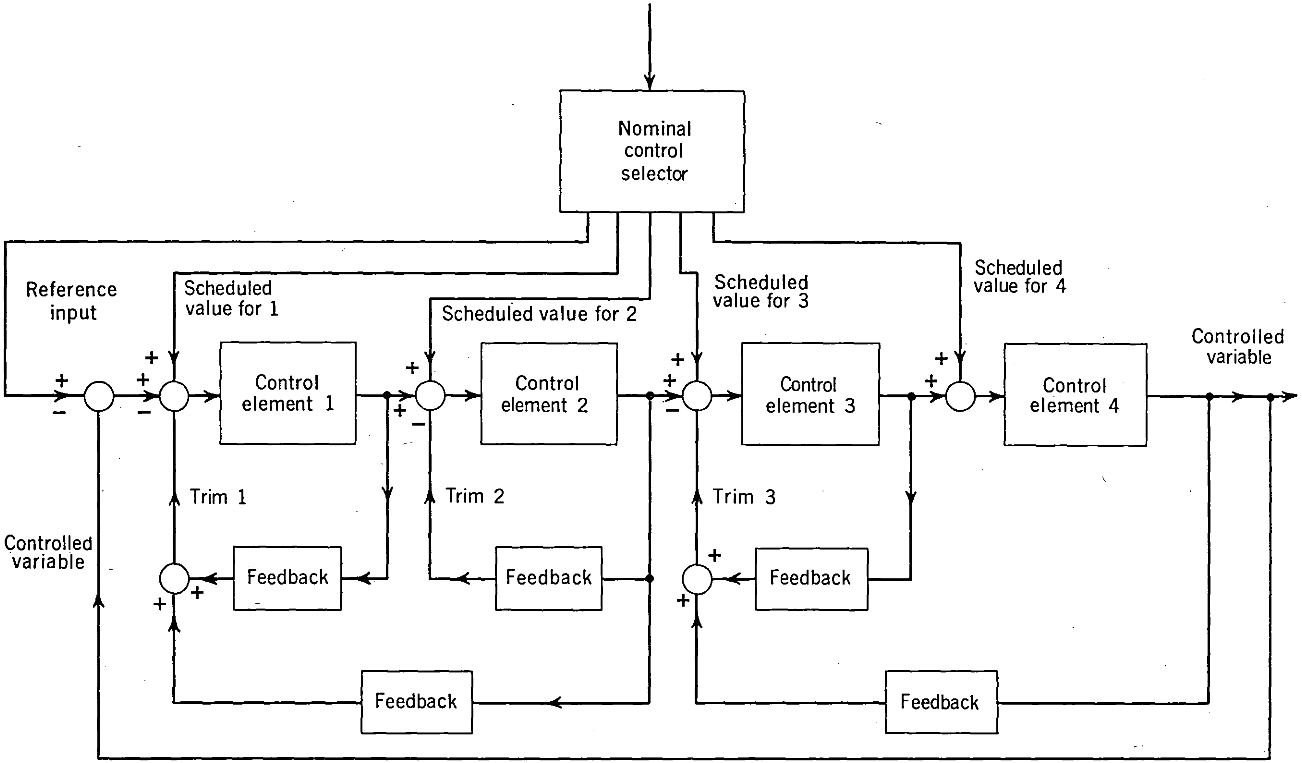


Fig. 4.25 Feedback control system using schedule and trim

whether they be electric, electronic, hydraulic, magnetic, pneumatic, or combinations thereof. The design of interchangeable units for minimizing the spare-part requirements and the provision for easy maintenance and checking means are likewise part of the system integration problem. These practical considerations represent an important phase of the design of an automation system.

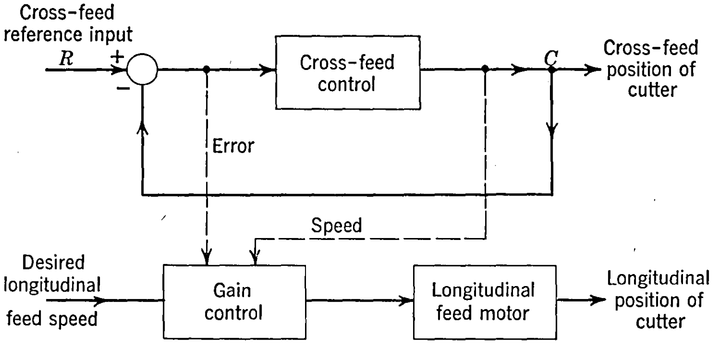


Fig. 4.26 Schematic diagram showing cross-connection of control signals

4.4 EXAMPLES OF AUTOMATION IN INDUSTRY

Many people in industry view automation as the end objective of an evolutionary process in manufacturing that consists of three major phases—manual production, mechanized production, and automation. The principal phases in which advanced industrial activity is taking place at present are those of mechanized production and automation. In mechanized production the machinist sets up his work, the machine under some controlled program turns out or cuts the work automatically, and then the worker takes out the work which is now ready for transfer to the next step or operation. Examples of mechanized production include position tracer controls in one, two, or three dimensions, photoelectric tracer controls, and record playback controls.

In the automation phase the worker is not a direct part of the production process. The transfer of the work from one operation to the next takes place automatically. Such industries as rubber, steel, paper, printing, foods, textiles, the electric industry, and the automotive industry, where means have been developed for transferring the work from one operation to the next, are ones that lend themselves to automation. Examples of automation from the steel industry will be cited as indicative of existing installations.

4.4.1 Position Tracer Controls

Position tracer controls are ones in which a model or template of the desired form is traced by a stylus and the cutting tool follows a corresponding path to shape the actual work. These controls vary in complexity from a simple single-dimension control to more complicated controls for two- and three-dimensional shapes. Hydraulic, electric, and combinations of both have been used for both signal and motive power.

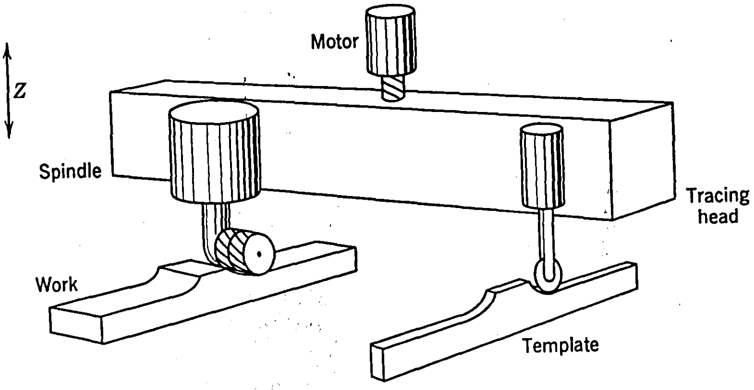


Fig. 4.27 Single-dimensioned control employing tracing head and mechanical tie to spindle motor

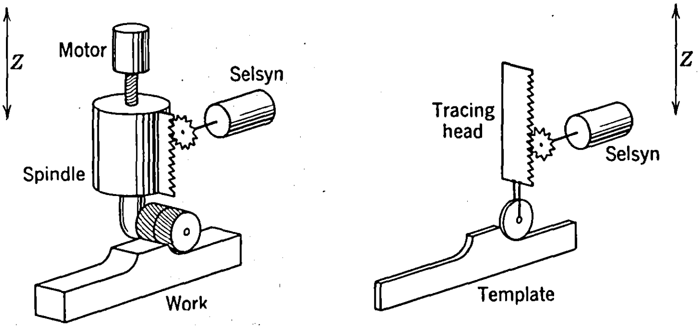


Fig. 4.28 Single-dimensioned control employing selsyn tie between tracing head and spindle motor

In the single-dimension control, a single direction of feed motion is controlled from the template while the longitudinal direction of machine motion is independently controlled. The two motions combine to produce the shapes required. This is illustrated in Fig. 4.27. In such a system, expensive power cams are eliminated and the machine construction is simplified.

Single-dimension tracers depend upon the relative displacement of the tracing stylus with respect to the tool, i.e., the actuating error, to provide the necessary feed speed to obtain a given slope on the work. This displacement along the axis of the controlled feed motion results in an error in the work. In addition to the velocity error that is proportional to the feed speed, other errors proportional to the rate of change and higher derivatives of feed speed occur. Thus, as the independent feed speed is raised or as increased feed slopes are encountered, both requiring higher tracer-controlled feed speed, the error is increased. When the slope of the work, the tolerances, and the feed speeds are known, this information can be used with machine performance data to determine what control system performance is required to meet the desired error tolerance in the finished part.

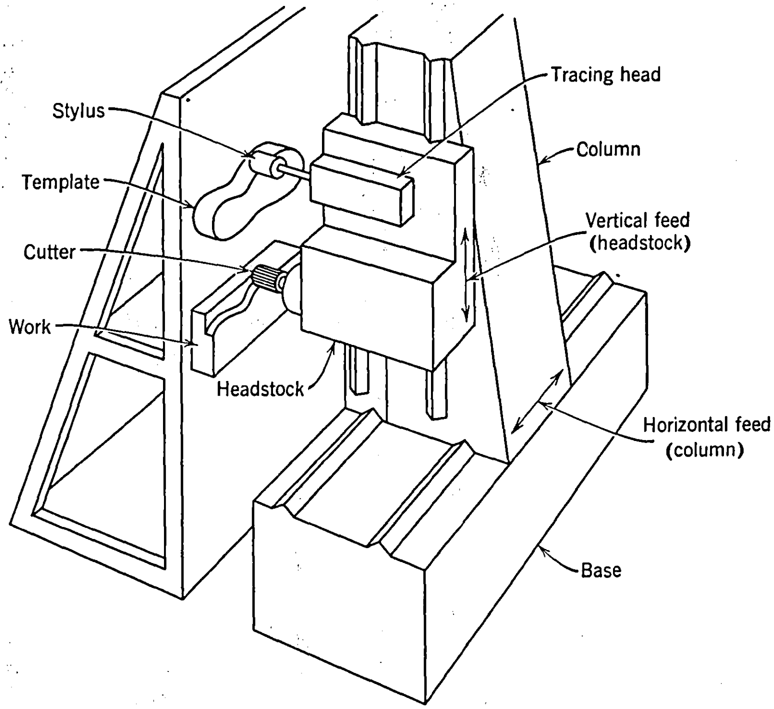


Fig. 4.29 Two-dimensional machine for profiling applications

As the size or mechanical complexity of the machine increases, there may be problems in mounting the template or tracing head conveniently. In such cases the single-dimension electric tracer control system can be modified to permit remote mounting of the tracing head. This is illustrated in Fig. 4.28 where a selsyn generator control-transformer system is used between the spindle and the mechanical tracing head. The spindle is moved in or out to match its selsyn position to the one driven from the tracing head. Although this system has greater flexibility in mechanical design, it is subject to the same sort of control system analysis as was the previous single-dimension control.

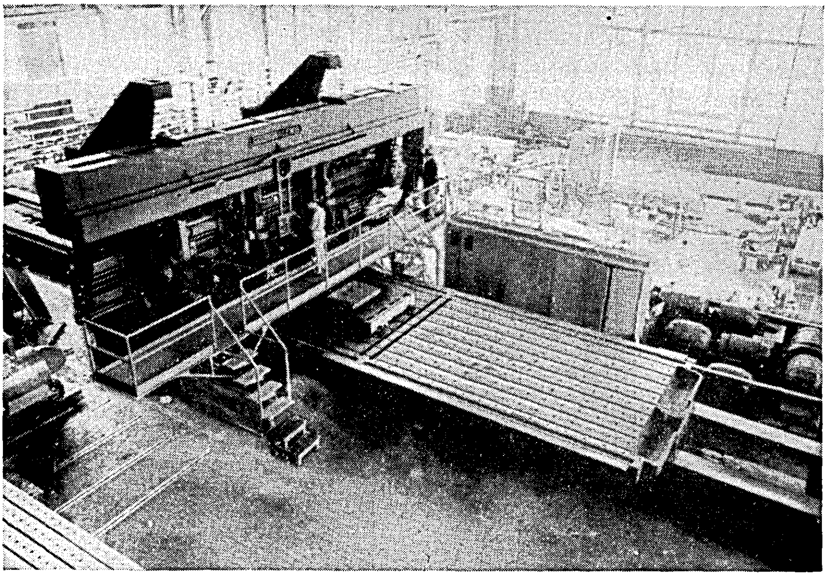


Fig. 4.30 Large milling machine having two-dimensional tracer control

For profiling work, where the tool may be required to trace completely around the work, two-dimensional tracer systems are used. As illustrated in Fig. 4.29, two motions of the machine at right angles to each other are controlled automatically from a single tracing head. The signals from the tracing head are combined in the electronic control panel to provide for constant feed speed of the cutter past the work, regardless of direction. This system operates with a constant deflection of the tracing head so that velocity errors such as were present in the single-dimension system do not occur. However, errors dependent on the rapidity of change in the direction of feed may result. Applications of this two-dimensional type of tracer control have been made to vertical boring mills, small milling machines, and large milling machines, such as that shown on Fig. 4.30 where 25 horsepower is required for this control alone.

For machining cavities such as dies and forged parts, the "selective two-dimension tracer" is used. With this system, each of three coordinate motions is powered, but only a selected pair are controlled simultaneously from a single tracing head. This system is illustrated in Fig. 4.31. Any pair of axes can be selected to machine in the desired plane. The control is similar to the two-dimension system described previously. The difference is in the type of tracing head used. This must be responsive to motion in three coordinates; hence it is more complex mechanically than the two-dimension head.

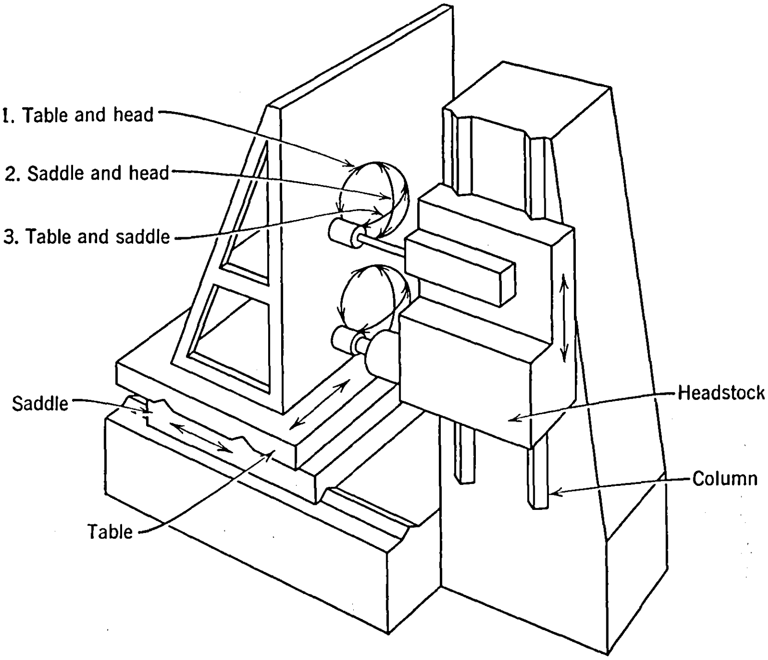


Fig. 4.31 Three-dimensional machine, selective two-dimensional operation in three mutually perpendicular planes

The tracer control systems that we have looked at have been ones using two- or three-dimension master shapes traced by a tracing head using a mechanical stylus to contact the master. Where tolerances permit, a photoelectric tracing head can be used to follow a line on a drawing. Such an application is shown in Fig. 4.32. In this particular case the machine was used to make masters for aircraft jet-engine blades. A small milling machine was used and the motions of the machine were driven from selsyn motors which were in turn supplied with power from selsyn generators driven by the feed motions of the line follower machine. A gear reduction is provided at each motion on the milling machine so that reduction in size between this line drawing and the machined part is possible, with corresponding improvement in accuracy. A number of variations in the type control to be used in photoelectric tracing suggest themselves for different applications, depending on the requirements of the job.

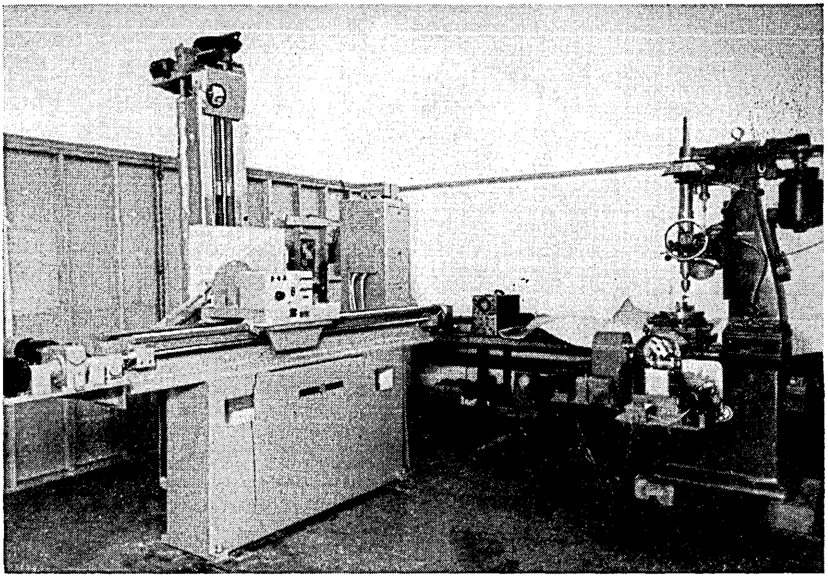


Fig. 4.32 Developmental tracer control using photoelectric tracing head with small milling machine

4.4.2 Record Playback Control

The position controls described above are ones in which the reference input information controlling the machine is mechanically stored. In the record playback control, the information used to control the motion of the machine is stored on magnetic tapes in the form of recorded signals, in much the same fashion as magnetic tape is used to record sound. Figure 4.33 shows a machine with two milling heads that can be controlled with such a system. When under tape control, the motions of the table, the head cross-feed and the rise and fall of one head are all automatically controlled. The tape-handling equipment and control for the various motions are shown in Fig. 4.34. Approximately a one-sixteenth-inch width of tape is required for each motion recorded. A 1-inch-wide tape is used; 4800 feet of tape in a 14-inch-diameter reel are suitable for 1 hour of machine operation.

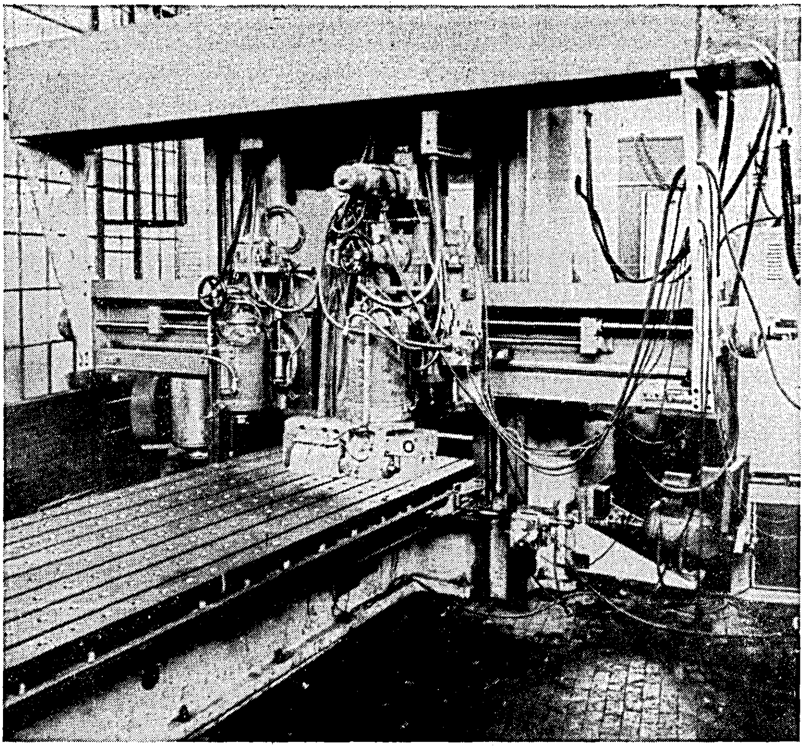


Fig. 4.33 Record playback installation in which table and head motions are controlled

When making a recording of the machining of a desired part, the angular positions of the various feed motions are continuously recorded. This is done by means of signals produced by small selsyns which produce voltages proportional to their angular positions. These selsyns are directly connected to the feed drive motors. A reference signal is also recorded on the tape at 200 cycles per second. It is this reference signal, and not the length of the tape, that is the true base against which the revolutions of the various feed motions are compared. Thus any dimensional changes in the length of the tape due to temperature, humidity, or other causes do not in any way affect the accuracy of playback.

On playback, the reference signal is used to create excitation voltages which are then used to excite the same selsyns connected to the motors. The position of each feed motor is checked against the recorded program 400 times per second, and any error in its position from the originally recorded position initiates corrective action in the control. The driving motors are supplied with signals from amplidyne generators which in turn are controlled by the tape signals. Thus the motors are made to follow the same program with respect to the ref-

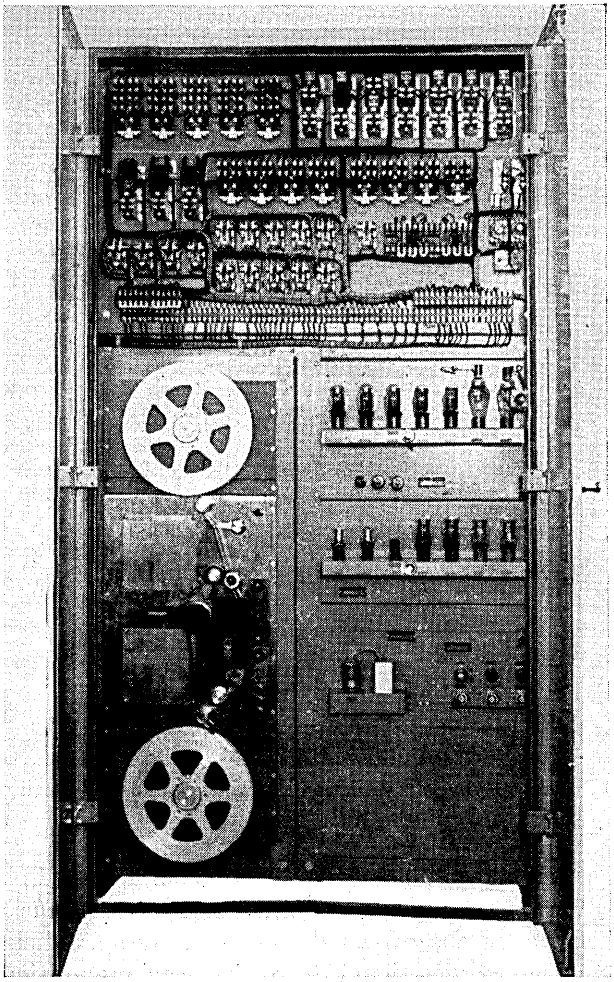


Fig. 4.34 Record playback tape-handling equipment and control

erence signals as they originally went through during the recording operation and the desired part is thus produced.

The feed speeds and accuracies obtained are of the same order of magnitude as those obtained with a good tracer-controlled machine. In addition to controlling the feeds of the machine, the record playback system also controls on-off functions such as starting and stopping spindles, turning on coolant, and shifting from one milling head to another, all functions normally controlled by the operator. This system has resulted in increased production and reduced spoilage in the machining of complex parts. Only useful productive motions need be recorded and any erroneous motions may be "erased" and rerun. Costly templates are replaced by the magnetic-tape reference.

4.4.3 Steel Mill Controls

The steel industry has been a leader of the basic trend in American industry of making and handling more and more material in strip form. Other industries making and processing brass, copper, aluminum, textiles, rubber, plastics, and paper have also made good use of this means of increasing production and decreasing unit costs. Mills in these industries through the years have increasingly gone to "continuous automatic production." Speeds of the order of 7000 feet

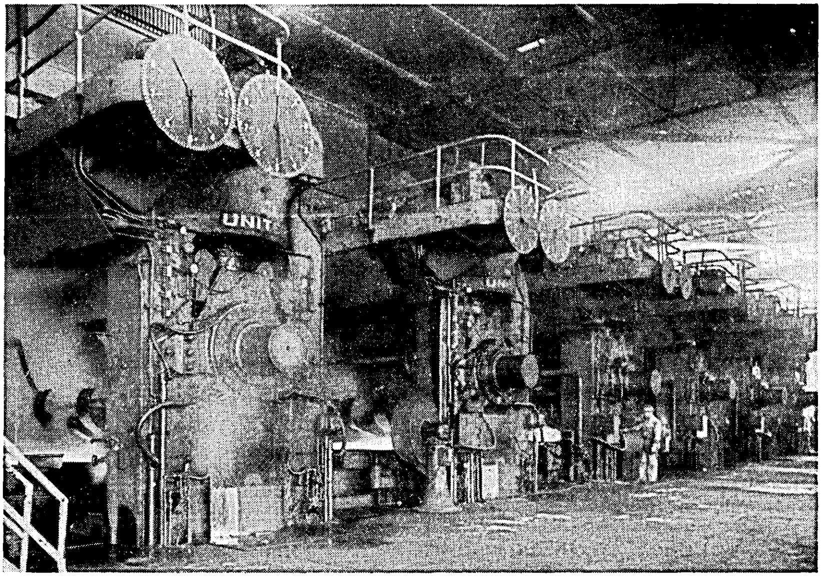


Fig. 4.35 Six-stand hot-strip steel mill

per minute and powers as high as 28,000 horsepower have been used to supply some steel mills. Although automation at this power level is not what we think of first when automation in business and industry is mentioned, nevertheless the problems of economy, reliability, maintenance, and performance are common to both forms of automation.

Figure 4.35 shows a hot-strip steel mill in operation and gives an idea of the size of equipment involved in one portion of the process. The conversion equipment and drive motors required to power large blooming mills is also impressive, as demonstrated in Fig. 4.36. With the finished value of the product output of such a mill amounting to $1,000,000 per 24 hours, it is apparent that high performance, good reliability, and ease of maintenance are all essential.

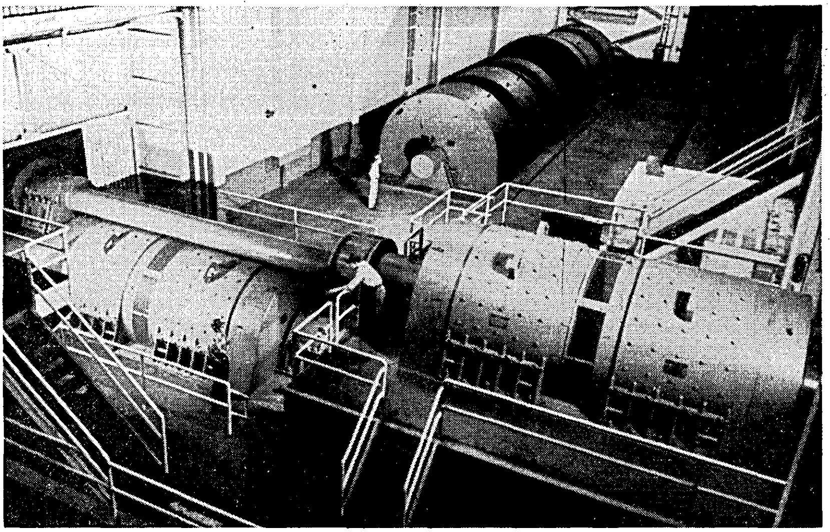


Fig. 4.36 Conversion equipment and drive motors for large blooming mill installation

4.4.4 Voltage Regulation

Extensive use is made of the basic form of the Ward Leonard system for control of d-c motors in these mill applications. Voltage, current, speed, tension, and loop position are some of the quantities that may be controlled in mills of this sort. Figure 4.37 shows a block diagram of a stand voltage regulator for a tandem cold-strip mill and indicates the multiple-loop feedback control system that is required to control only one variable in a continuous automatic production process. The compensation term for voltage drop due to resistance (*IR*) represents in part the effect on this particular control of other portions of the process. As such it serves as an example of an interconnected control, which was mentioned previously as being beneficial in integrated control systems. The generator field exciter voltage is a nominal value of voltage that is used to establish a nominal output voltage for the generator, a scheduled value in terms of the schedule-and-trim philosophy. The regulator exciter (RE) amplidyne provides the trim feature and is operated closed-loop to provide the output voltage control for the generator. The ohmic drop (OD) amplidyne serves as a power amplifier to raise the voltage and power of the feedback signals that control it.

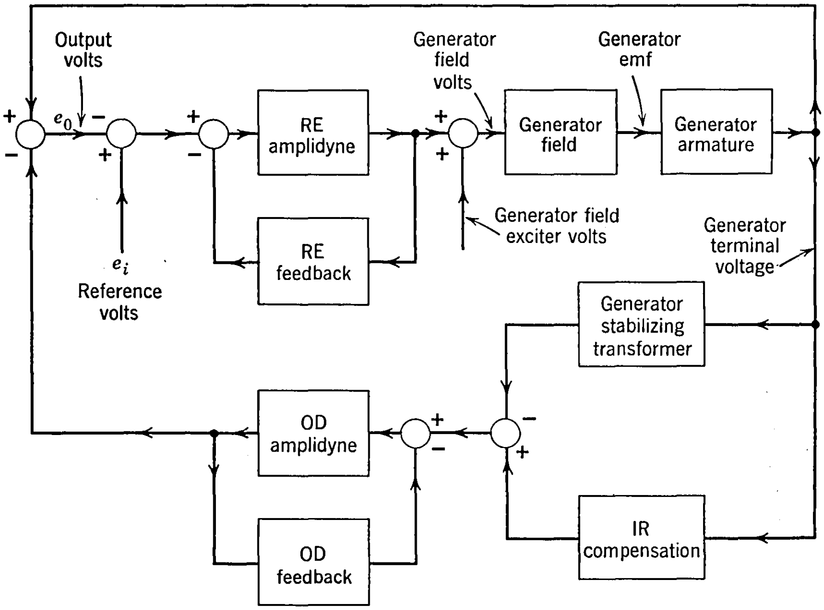


Fig. 4.37 Block diagram of stand voltage regulator for tandem cold-strip mill

4.4.5 Magnetic Loop Control

The magnetic loop control was developed to fill the need for a faster, more dependable, easily maintained control for loop position in steel mill pickling lines. A schematic diagram of such a control system is shown in Fig. 4.38. An acid-proof magnetic pickup rests in the bottom of the tank. It keeps the strip near the tank bottom for maximum pickling efficiency but does not allow the strip to scrape the bottom of the tank. The output of the pickup unit is amplified through an amplistat and amplidyne to control a buck-boost generator in series with the armature of the pinch roll drive motor. This system controls the speed of the pinch rolls to regulate the position of the strip loop.

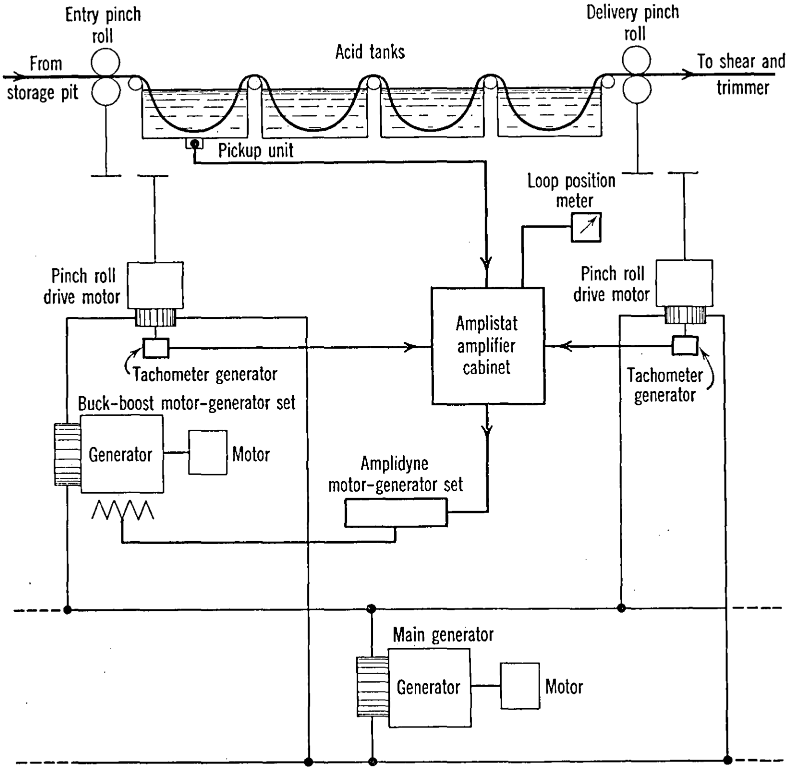


Fig. 4.38 Schematic diagram of magnetic loop control system for steel mill pickling lines

In the amplistat, the loop position signal from the pickup unit is compared electrically with a voltage reference. The reference is set during installation adjustment, and usually no further change is necessary. The reference determines the normal position in which the control will hold the steel strip. The difference signal is amplified by the amplistat and the output is fed into the amplidyne.

The magnetic loop control is a position regulator which, to provide stability, utilizes a signal proportional to the rate of change of the position being regulated. This signal is the derivative of position or of the strip speed. Tachometer generators are used on the motors at each end of the tank. The two voltages are compared and the resultant relative speed signal tells the amplifier whether the strip is moving up or down. When the relative speed signal is zero, the loop height is not changing. When the pickup unit signal is normal, the amplistat uses the tachometer difference signal to hold the entry pinch roll speed in synchronism with the delivery end-roll. Thus accurate synchronization of pinch roll speed tends, within limits, to hold the strip loop at normal height, regardless of strip speed. The pickup unit detects and causes the control to correct any tendency of the loop to drift up or down.

The loop control is automatic. The operator has only to look at the loop position meter whenever he wants to know the loop height, and even this may be remotely located. The process indeed is one of "controlled continuous automatic production" and well qualifies as part of an automation process.

4.5 SUMMARY

The economic pressures on our economy are forcing us in the direction of automation. The process, the tools, the techniques, and the components are available to do a far greater job of automation than is currently being done. In many instances design methods for feedback control systems are available to handle the individual controls as well as the overall automation process.

The challenge and the opportunity for engineers in the field of automation are the need for ingenuity and originality in conceiving better ways to perform these processes continuously and automatically. Inspired design of new devices, intelligent usage of existing devices, and creative thinking in putting the systems together, all these and more are asked of the engineer.

Especial emphasis is placed on the need for reliable performance, ease of maintenance, and simplicity at a cost that is economically acceptable.

Although automation tends to relieve man of many monotonous and arduous tasks, it does not appear to supplant him in the need to create, design, build, and maintain the equipment that will, through automation, provide man with a more abundant life.

4.6 REFERENCES

1. H. M. James, N. B. Nichols, and R. S. Phillips, *Theory of Servomechanisms,* McGraw-Hill Book Co., Inc., New York, 1947.

2. G. S. Brown and D. P. Campbell, *Principles of Servomechanisms,* John Wiley & Sons, Inc., New York, 1948.

3. I. A. Greenwood, J. V. Holdam, Jr., and D. MacRae, *Electronic Instruments, Pt. II,* McGraw-Hill Book Co., Inc., New York, 1948.

4. R. J. Kochenburger, Frequency Response Method of Analyzing and Synthesizing Contractor Servomechanisms, *Trans. AIEE, 69,* Pt. I: 285–292, 1950.

5. W. R. Ahrendt and J. F. Taplin, *Automatic Feedback Control,* McGraw-Hill Book Co., Inc., New York, 1951.

6. H. Chestnut and R. W. Mayer, *Servomechanisms and Regulating System Design,* Vol. I, John Wiley & Sons, Inc., New York, 1951.

7. A. Tustin, *Automatic and Manual Control,* Academic Press, Inc., New York, 1952.

8. E. Calvin Johnson, Sinusoidal Analysis of Feedback Control Systems Containing Non-Linear Elements, *Trans. AIEE, 71,* Pt. 2: 169–181, June 1952.

9. W. E. Miller, The Magnetic Amplifier and Its Application in the Steel Industry, *G. E. Rev., 742* (also *The Iron and Steel Engineer,* May 1953).

10. A. M. Fuchs, A Bibliography of the Frequency-Response Method as Applied to Automatic Feedback Control Systems, ASME Paper 53-A-13, Dec. 1953.

11. W. E. Miller and F. S. Rothe, Selection and Analysis of Regulating Systems for Steel Mill Drives, *G. E. Rev., 743,* 1953.

12. W. R. Evans, *Control-System Dynamics,* McGraw-Hill Book Co., Inc., New York, 1954.

13. C. M. Rhoades, Jr., Contributions of Electric Equipment to Automation in the Aircraft Industry, *Soc. Automotive Engrs., Special Rept., 309:* 63–76, 1954.

14. W. E. Miller, Electric Systems for Main Drives, *G. E. Rev., 1003* (also *The Iron and Steel Engineer,* July 1954).

15. H. Chestnut and R. W. Mayer, *Servomechanisms and Regulating System Design,* Vol. II, John Wiley & Sons, Inc., New York, 1955.

16. J. R. Moore, Application of Servo Systems to Aircraft, *Aeronaut. Eng. Rev., 8,* 32–43, Jan. 1949.

Also see such magazines as *Automation, Automatic Control,* and *Control Engineering,* each of which was first published during 1954.

5. Basic Concepts of Industrial Instrumentation and Control

HERBERT W. ZIEBOLZ

Askania Regulator Company
Chicago, Illinois

5.1 BASIC CONCEPTS OF INDUSTRIAL INSTRUMENTATION

The subject to be discussed in the following chapter deals, as indicated by the title, with two basic industrial devices: (1) instruments and (2) automatic controllers.

5.1.1 Definition of Industrial Instruments

In order to clarify the subject under discussion, we shall define an "industrial instrument" as a device which represents the magnitude of a process variable by a corresponding displacement of an index mark (analog) or a number (digital). Emphasis in this definition is placed on the translation of the variable into a common language, i.e., a displacement, s, or a number. Examples of such instruments are: dial gages, recorders, U gages, watches, micrometers, and oscilloscopes.

5.1.2 The Concept of "Translators"

Common to all of these devices is the output "displacement" or "stroke." This class of instruments—"translators"—may be represented by a box with an "input," a variable to be measured, and an "output," a stroke, i.e., the displacement of a pointer relative to a scale (Fig. 5.1). (As used here the term "translator" includes "transducers" as well as other devices such as a lever or a scale.)

In order to avoid the time-consuming effort of drawing boxes or pictures as in Fig. 5.1, we shall use a symbolic shorthand, which can be handled by typewriters, with the input over the output variable,

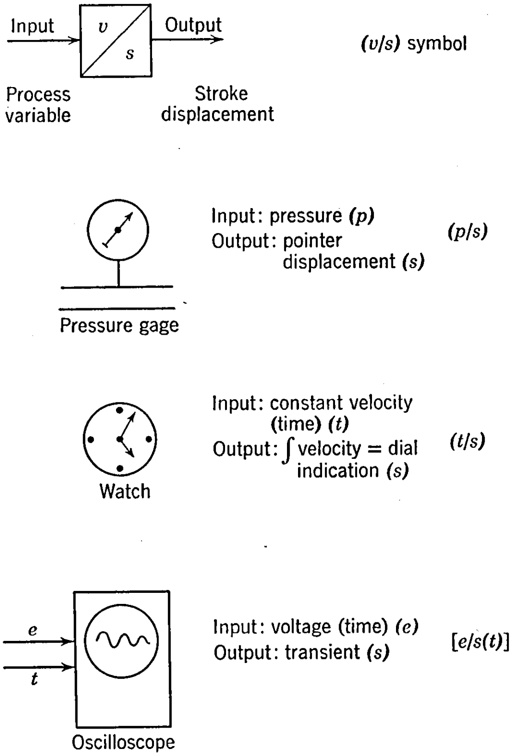


Fig. 5.1 Typical examples of "instrument-type" translators: pressure gage, watch, oscilloscope

for example, p/s for a "pressure gage." As the subject under discussion is further limited to "industrial" instrumentation as compared with the broader subject of scientific instrumentation, it is submitted that the difference between these two types of instruments lies in the ultimate purpose for which they are designed.

5.1.3 Industrial versus Scientific Instruments

The purpose of an *industrial instrument* is to furnish information to an observer that will enable him to make *decisions* relative to the plant variable that the industrial instrument represents. The purpose of a scientific instrument is to enable the observer to establish *facts.*

If this interpretation is accepted, it will be noted that the operator as the receiver of the "intelligence" or the "communication" becomes an important link in the process of measurement, and even the closed loop, which, if mechanized, means an "automatic controller," is implied in the expected decision of the operator.

From the above, a number of far-reaching conclusions follow, conclusions which, if overlooked in the design or in the specifications, result in mismatching of process, instrument, and operator. For this reason the "human engineering" approach has emphasized in recent years the need for designing instruments having both language and display which the particular operator can understand and use under the particular conditions of his environment. Examples of poor design are the domestic-type fever thermometer and the poorly illuminated miniature dial gages which are read by a furnace operator who is temporarily blinded after having looked into a furnace.

5.1.4 Instrument Application and Accuracy

Another consequence of this basic definition is a call for a re-examination of the questions (1) what to measure, and (2) what accuracy to demand of the instrument. Perhaps the most difficult question to decide upon is what to measure in a process. This question, which is one of establishing a correlation between desired end-product qualities and process variables, will be dealt with later on.

With regard to accuracy it can be stated that an accuracy specification beyond the need of the particular plant operation can be a handicap rather than a desirable feature. It would be useless and confusing to give the janitor of a building a temperature-measuring bridge circuit which measures with an accuracy of 1/1000° F—even if it were available at the time as surplus or at the same price as an ordinary thermometer with a maximum of 1/2° F or even coarser subdivisions. A liquid-level indicator measuring with the accuracy of a vernier gage would be just as useless in most cases.

The question, what is the desired or required magnitude of the instrument divisions, is not always easy to answer; since the complications of design, space requirements, cost, and maintenance are vitally affected, much fundamental cooperative work by human engineering groups, plant designers, instrument manufacturers, and plant operators is needed. The amount of effort necessary in particular cases is best illustrated in following the evolution of the flight cockpit instrumentation, a task of a remarkable magnitude, in both man-hours and cost.

Some good examples of matching operator requirements and instrument outputs are to be found in "high-low-normal" gages of automobiles and "ready-to-defrost" indicators of refrigerators. These latter examples, however, should not mislead us to assume that "error type" indicators are in all, or even in most, cases best for the needs of an operator. In analyzing the problems of an operator, it will be necessary to study his particular requirements carefully. It is only natural to expect that, for instance, the operator of a prototype plant with a scientific interest in the plant behavior may have requirements quite different from those of the accounting operator or of the production people.

5.1.5 Different Needs for Different Operators

Mr. D. Boyd, who has contributed so much in basic thinking to the field of instrumentation, illustrates this point[1] by means of the diagram shown in Fig. 5.2. On the left-hand side different types of instrument users are enumerated, and on the top of the columns different

× Preferred

● Less desirable

		Analog-type multiple-pen intermittent single-chart recorder	Analog-type multiple-pen continuous single-chart recorder	Deviation single-pen recorder	Recorder-controller
The user	Research, accounting	×	×	●	×
	Instrument department	●	×	×	×
	Plant operator	●	●	●	×
		With separate controller			With built in controller
		Type of instrument			

Fig. 5.2 Preference of various users for various instruments (Courtesy D. Boyd)

---

[1] Personal remarks to the author.

types of instruments are tabulated. In all of the first three instruments listed it is assumed that an automatic controller is added as a separate item. The fourth instrument, a recorder-controller, of course, incorporates both features in a single unit.

It will be noted that the information provided by certain instrument types is satisfactory to the research engineer or the accountant; however, in the particular case analyzed, the operator preferred the combined recorder-controller for the following reasons: (1) closeness of the record, representing a memory storage, to the adjustment of the control setting; (2) continuous and complete record and indication of the specific variable; (3) one recording scale factor only, one which can be easily correlated to the setting of the control dial adjustment.

It will also be noted that the instrument service engineer, because of his different background and purpose of action, is easier to satisfy than this particular operator. The objection of the service engineer was mainly directed against intermittent-type multiple recorders which make his job of adjusting controls difficult; but the research engineer may prefer this type of instrument for his purpose of logging average values.

It was stated that the diagram of Fig. 5.2 applies only to a particular plant, and it is necessary for the product engineer to examine carefully and analyze the specific requirements of the production process as well as the background and needs of the operators who are to run the plant. An interesting problem in this connection is presented by plants which have to be operated by unskilled workers or by people who speak foreign languages, with technical supervision by people who speak only English.

In order to find our way through the multitude of instruments and controls now available to the user and to find a more rational approach to their analysis and design, it appears promising to go back to fundamental concepts.

Basically any device with dynamic characteristics can be considered as a box with an "input" and an "output," as we have seen before.

In order to catalog such "translators" of "inputs" into "outputs," it was found to be convenient to design a simple reference chart with various inputs and outputs.

5.2 THE TRANSLATOR CHART

In Fig. 5.3 the vertical column represents the inputs, starting with the building blocks of the mechanical engineer, i.e., stroke, pressure, and force, and continuing further down with electrical variables like voltage, current, frequency, etc. Any one of the pigeonholes in the chart represents a reference filing cabinet for a "translator" of any given input variable into the same or any other output variable. For instance, the translator, stroke/stroke = s/s, in box number 1 represents devices that translate one stroke into another. As shown in Fig. 5.4, a number of basic units will be covered by this symbol.

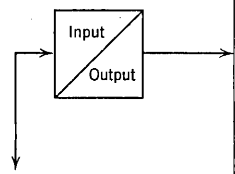
		Stroke (or angle α)	Pressure	Force	$\frac{ds}{dt}$ speed $\left(\text{or } \frac{d\alpha}{dt}\right)$	$\frac{d^2s}{dt^2}$ acceleration $\left(\text{or } \frac{d^2\alpha}{dt^2}\right)$	Rate of flow	Direct current	Alternating current	Voltage, d-c	Voltage, a-c	Temperature	Light intensity	
		s	p	F	V	α	W	$\underline{\underline{i}}$	$\underset{\sim}{i}$	$\underline{\underline{e}}$	$\underset{\sim}{e}$	T	Q	
Stroke (or angle α)	s	1	2	9	10	25	26	49	50	81	82	121	122	
Pressure	p	4	3	8	11	24	27	48	51	80	83	120	123	
Force	F	5	6	7	12	23	28	47	52	79	84	119	124	
$\frac{ds}{dt}$ speed $\left(\text{or } \frac{d\alpha}{dt}\right)$	V	16	15	14	13	22	29	46	53	78	85	118	1	
$\frac{d^2s}{dt^2}$ acceleration $\left(\text{or } \frac{d^2\alpha}{dt^2}\right)$	α	17	18	19	20	21	30							
Rate of flow	W	36	35	34	33	32								
Direct current	$\underline{\underline{i}}$	37	38	39	40	41								
Alternating current	$\underset{\sim}{i}$	64	63	62	61									
Voltage, d-c	$\underline{\underline{e}}$	65												
Voltage, a-c	$\underset{\sim}{e}$	10												
Temperature	T	1												
Light intensity	Q	1												
Resistancee	R	14												
Inductancee	L	1												

Fig. 5.3 The "translator chart"

The indices in Fig. 5.4 indicate the class, A, B, C, and D, to which the device belongs. Thus:

"Class A" symbol $(s_1/s_2)_a$ is a direct linear or nonlinear translator.

"Class B" symbol $(s_1/s_2)_b$ is an indirect translator with two or more translation stages and with more than one variable, with a not necessarily repeatable relationship between input and output.

"Class C" symbol $(s_1/s_2)_c{}^{\mathrm{steam}}$ is a translator employing outside power without a necessarily repeatable relationship.

"Class D" symbol $(s_1/s_2)_d^{\mathrm{hydraul}}$ is a servoloop using outside energy and feedback, and a definite input-output relationship.

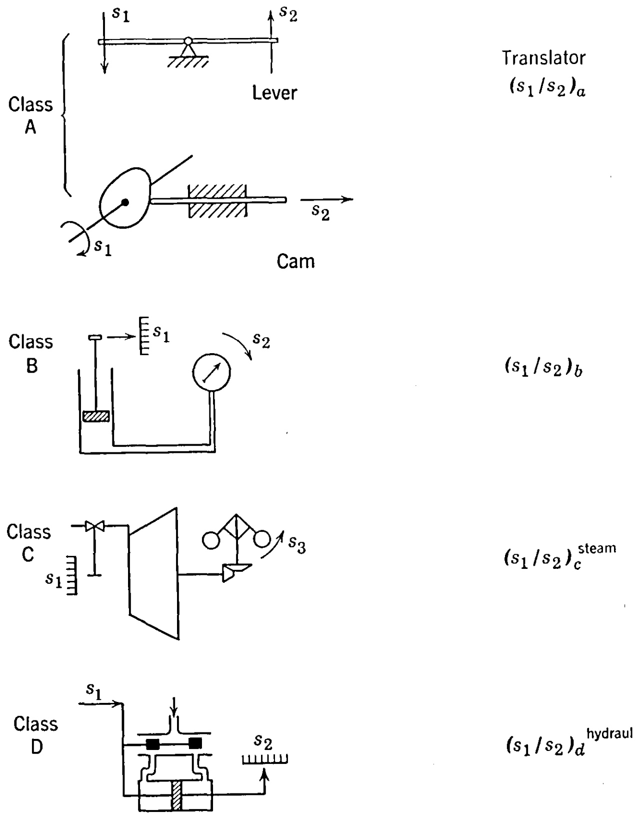


Fig. 5.4 Four basic types of s/s translators: A, direct proportional; B, proportional; C, proportional with outside energy supply; D, direct proportional with outside energy supply and feedback

5.2.1 Symbolic Translator Equations

The device B of Fig. 5.4 can be described in a pseudo equation as:

$$(s_1/p)_b + (p/s_2)_a = (s_1/s_2)_b$$

In this equation $(s_1/p)_b$ (the bicycle pump) is of class B; $(p/s_2)_a$ of class A, and the final device s_1/s_2 consequently of class B. Although at first glance this combination may appear unusual outside the transportation field, it is sometimes used to transmit valve stem travel in a "proportional plus reset" controller, and is therefore not so farfetched as may appear on the surface.

The example above shows a "two-stage translator" in which two components or variables s and p are used. Obviously, multistage translators can be developed and may take this form:

$$s_1/p + p/\underline{\underline{e}} + \underline{\underline{e}}/\underset{\sim}{e} + \underset{\sim}{e}/s_2 = s_1/s_2$$

The chain in this case can again be expanded or contracted by substitutions as, for instance, by introducing:

$$s_1/p + p/\underline{\underline{e}} = s_1/\underline{\underline{e}}$$

or

$$p/\underline{\underline{e}} + \underline{\underline{e}}/\underset{\sim}{e} = p/\underset{\sim}{e}, \qquad \text{etc.}$$

Thus, the translator chart offers to the engineer a systematic refence system for storing known instrument and controller components and for creating new instrument and controller components. But it can do even more.

5.2.2 Instruments and Controls in the Translator Chart

An analysis of the significance of certain columns of the chart shows one vertical column with all known variables as the input and stroke = s as the common output variable. We recognize such trans-

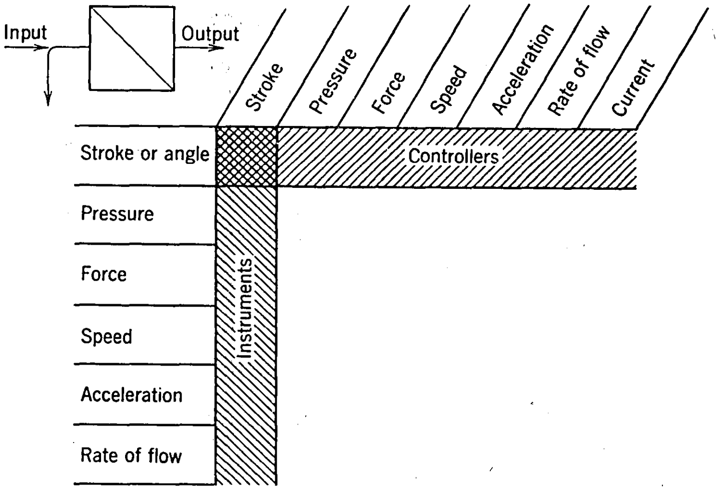


Fig. 5.5 The place of instruments and controllers in the translator chart

lators as the "instruments" defined in the beginning of this chapter (Fig. 5.5). In each case the variable is represented by a displacement, s, or in our symbolic form, an "instrument" is given as (variable/s).

On the other hand, the horizontal column starting with strokes as inputs and all variables as outputs represents translators of the type (stroke/variable), which we shall call "controllers," for controllers perform the function of such translators with s representing the setting, even if they are permanently built in or are not adjustable from the outside. (Note that this broader definition of "controllers" covers even an s/F translator, i.e., a "spring," which would not ordinarily be classified under this term.)

It is now easy to show that if it is possible to demonstrate that at least one solution is available for each translator in both the "instrument" and the "controller" columns, then at least *one* solution is available for *any* translator in the chart. This follows from the simple symbolic equation

$$\text{variable}_1/s + s/\text{variable}_2 = \text{variable}_1/\text{variable}_2$$

On the other hand, if we arrange the translators in the opposite sequence

$$s_1/\text{variable}_1 + \text{variable}_1/s_2 = s_1/s_2$$

we have *as many solutions for s_1/s_2 as variables!*

5.2.3 Electrical Inputs or Outputs

By extending the meaning of an "instrument" so that it includes translators with any other variable—for instance, a-c voltage $\underset{\sim}{e}$ as output, we can repeat our procedure of filling the individual translator boxes with additional translators or transducers. This appears to be the significance of the present trend away from the displacement parameter s in favor of electrical outputs (not necessarily only voltage or currents, but also phase displacements, frequencies, pulse duration, etc.).

5.3 THE OPERATOR CHART

An obvious question may arise at this point: Why should one go to these seemingly more complicated solutions?

In order to answer this question, we prepare a second reference chart of another basic device which is used in instruments, controllers, and computers. This device performs a mathematical operation on one or more input variables and is called for this reason the "operator." A simple example of such a device is a summarizer. In Fig. 5.6, two mechanical summarizers and a pneumatic one of the flow rate type are shown. Other operators are: multipliers, integrators, differentiators, and function generators such as sine or cosine generators.

In Fig. 5.7, the "operator chart," we find again in the vertical column the input variables, stroke, pressure, force, etc., but now in the horizontal column the respective operators are listed. An attempt to fill this chart completely has been unsuccessful so far. For instance, there is at present no operator available for direct summation of pressures. This means that the designer, faced with this problem of summarizing pressures, has to look for suitable summarizers to other variables, i.e., "strokes" or "voltages," and for convenient translators from pressures into such variables, which he can find in his translator chart.

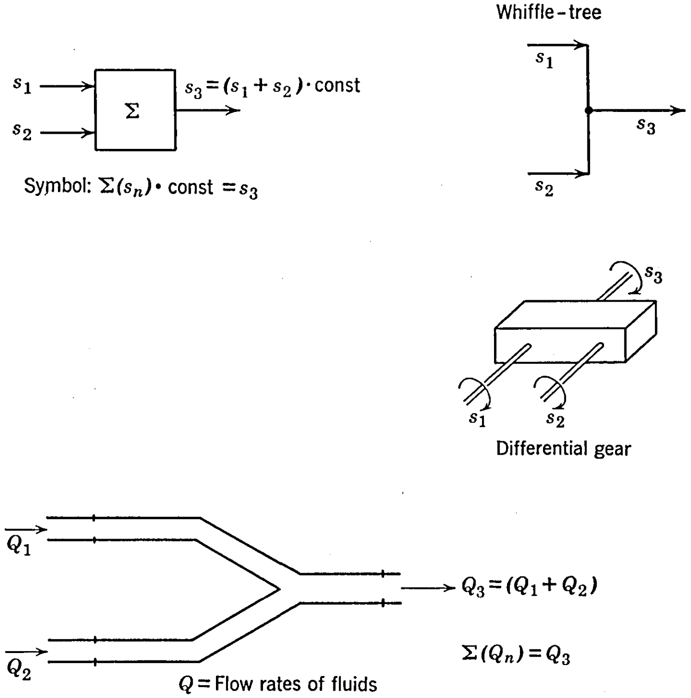


Fig. 5.6 Typical mechanical summarizer for strokes and flow rates

The choice of the particular design to solve his problem can thus be based on a systematic evaluation of alternatives rather than on accidental "inventions." Cost, reliability, simplicity, accuracy, space requirements, dynamic behavior, ambient conditions, and many other considerations enter into the final decision, which of the available alternate paths to choose.

The tremendous impetus industrial "automation" has taken in the last few years has been mirrored in a not always sound development race, in which alternate solutions were often introduced for the sake of novelty or patent circumvention, rather than on the basis of superiority and need.

Operator		Summarizer	Multiplicator	*n*th power	Differentiator	Integrator	Splitter	Sine generator		
Symbol		Σ	M	x^n	$\frac{dx_1}{dx_2}$	$\int$	Sp	sin		
Type		A	B	C	D	E	F	G		
Stroke	s	A_1	B_1	C_1	D_1	E_1	F_1	G_1		
Pressure	p	A_2	B_2	C_2	D_2	E_2	F_2	G_2		
Force	F	A_3	B_3	C_3	D_3	E_3	F_3	G_3		
$\frac{ds}{dt}$ Speed	V	A_4	B_4	C_4	D_4	E_4	F_4	G_4		
$\frac{d^2s}{dt^2}$ Acceleration	a	A_5	B_5	C_5	D_5	E_5	F_5	G_5		
Rate of flow	W	A_6	B_6	C_6	D_6	E_6	F_6	G_6		
Direct current	$\underline{\underline{i}}$	A_7	B_7	C_7	D_7	E_7	F_7	G_7		
Alternating current	$\underset{\sim}{i}$	A_8	B_8	C_8	D_8	E_8	F_8	G_8		

Fig. 5.7 The "operator chart"

However, the same dynamic forces have produced a number of new basic components in the last few years that permit the designer of new instruments and controllers to go far beyond what was possible in the past.

5.4 NEW PACKAGED TOOLS OF THE MODERN DESIGNER

First of all, there has been a steady growth in the number of computer elements ("operators" in our language) which are directly applicable to instrumentation and controller designs. These computer elements are either (1) mechanical: differential gears, multipliers, integrators, multiple-bar linkages, etc.; (2) pneumatic: integrators ("reset"), summarizers (with *p/F* or *p/s* translators); or (3) electric: analog high-speed and real-time scale, digital, memory storage systems (punched cards, magnetic tape, punched tape, film trace, etc).

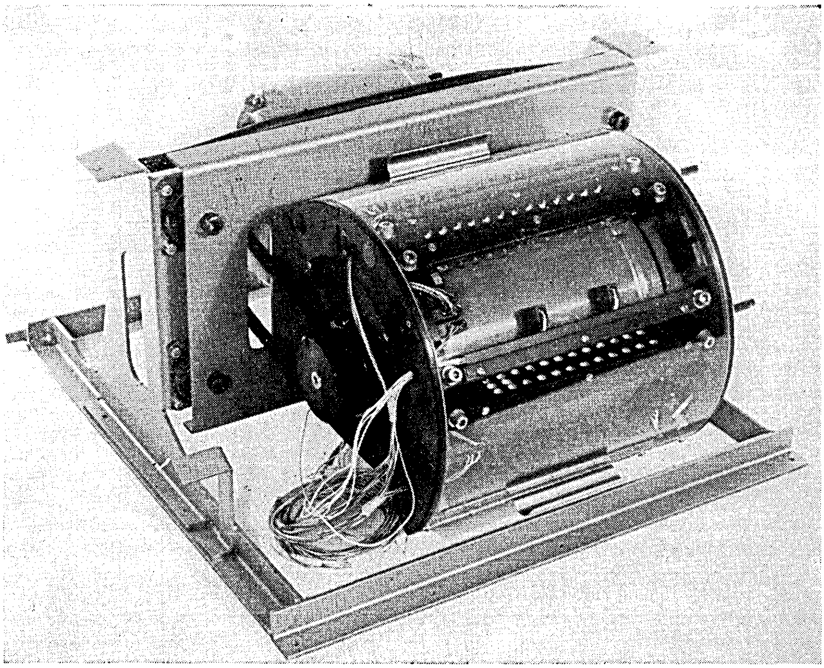


Fig. 5.8 Magnetic memory drum (Courtesy Librascope Inc.)

As indicated in the last type of computer elements, a very important, but greatly improved, class of *memory devices* have become available to the industrial designer. Although used for many years in the textile industry and in the entertainment field from the "pianola" (player piano) of yesteryears to the "hi-fi" records of today, it has been only very recently that these techniques have been applied to machines, computers, and instrumentation (Fig. 5.8). One of the reasons for the delay is the previous lack of interpreters or translators to make these multilingual instruments compatible, i.e., to make them understand each other, or to use a common language in communicating with each other.

However, such translators are now available, and they open to the instrument designer, in addition to improved process instrumentation, new and very promising territories to conquer, i.e., the field of office operations and the supervisory loops of a production plant, from entering and handling of accounting information to manufacturing a metal part. Since the language of the accounting department and the economic common denominator—i.e., the dollar—are digital, it is necessary to have translators from digital to analog, and vice versa (Fig. 5.9). Present efforts of designers concentrate on an increase in speed and life expectancy of such devices; however, there are a number

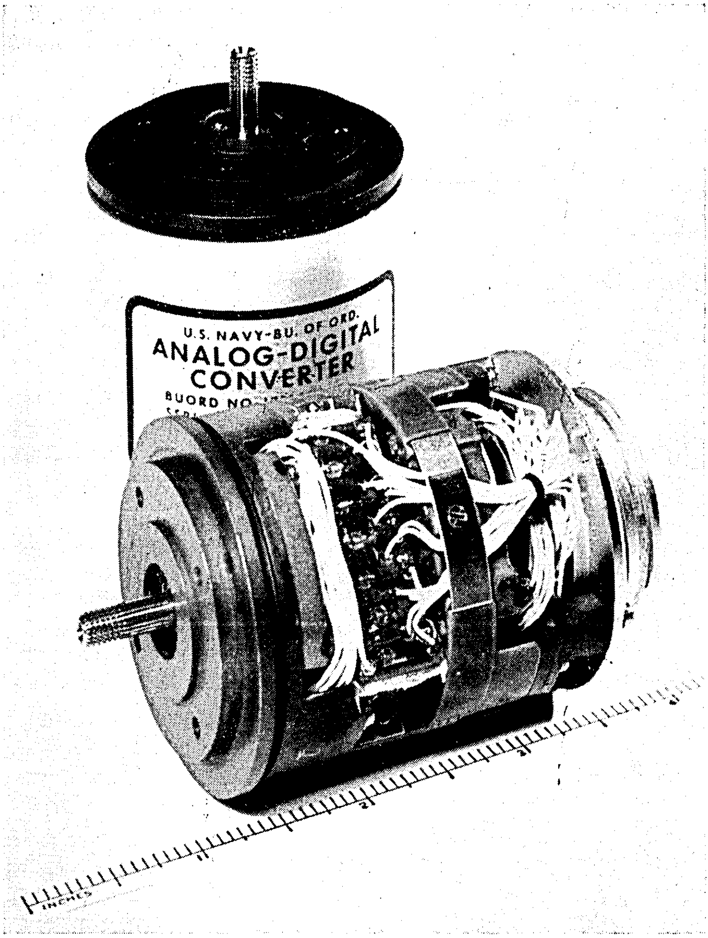


Fig. 5.9 Analog-to-digital converter (Courtesy Librascope Inc.)

of various designs either on the drawing board or already available to take care of the general present requirements.

The next important link, a development which is still in a state of flux, is the (variable/electrical signal) transducer. Industry has not as yet sufficiently crystallized its requirements in this field, with the result that these devices still differ widely in output variables and energy levels. Typical of recent designs are outputs of 4 to 8 milliamperes direct current and 0.5 to 5 milliamperes direct current, as used in two of the leading miniaturized electronic controllers. Other instruments used in telemetering have frequencies or pulse intervals as their common language.

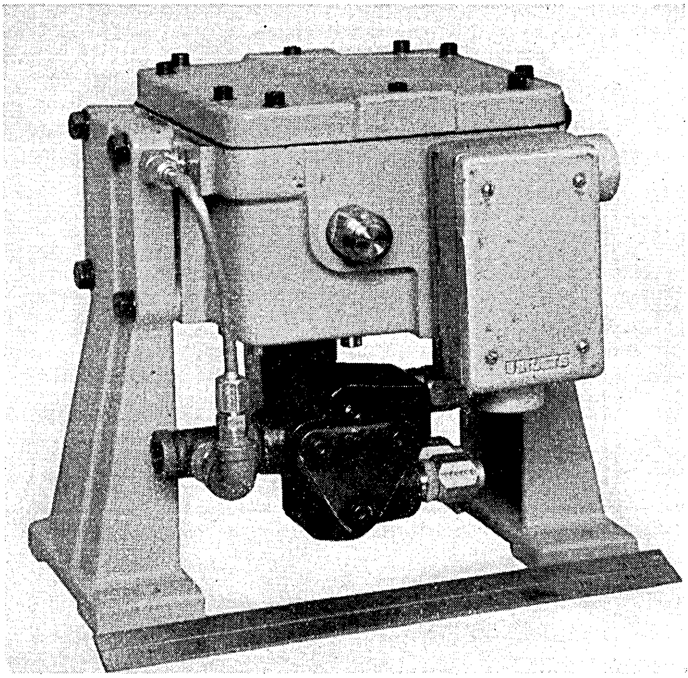


Fig. 5.10 Electrohydraulic relay (Courtesy Askania Regulator Co.)

It is to be expected and hoped, at least by the writer, that in the next few years a common signal type and level will be arrived at, which will do away with the present multitude of intermediate translators or multipliers, i.e., "amplifiers," which are now needed to assure compatibility. In many of the computer elements, operators, and translators, servomechanisms or stabilized amplifiers have been used, either to enforce linearity or repeatability, or to increase energy levels (class D translators). The understanding of their dynamic behavior, based on the fundamental work done by the MIT Radiation Laboratory during World War II, has produced a great multitude of servo components and amplifiers which are available to the designer of instruments and controls as package units. Typical examples are the electrohydraulic amplifier shown in Fig. 5.10, the pneumatic amplifier-computer of Fig. 5.11*a* and Fig. 5.11*b*, and the pneumatic hydraulic power booster in Fig. 5.12.

In addition, magnetic amplifiers, and recently also transistor-controlled servomechanisms have become the common tools of the modern designer.

In telemetering, as well as in digital techniques, pulse techniques, because of their higher "signal to noise" ratio, have been developed sufficiently to make the respective transducers or translators reliable tools for present industrial applications. To illustrate this point, it is only necessary to visualize the obvious advantages of a dial telephone over one using an analog method of transmitting the desired number. The ease of translation of pulse duration into digits is one of the reasons for its wide use in modern computers—and it will be the reason for its broad application to telemetering in industry.

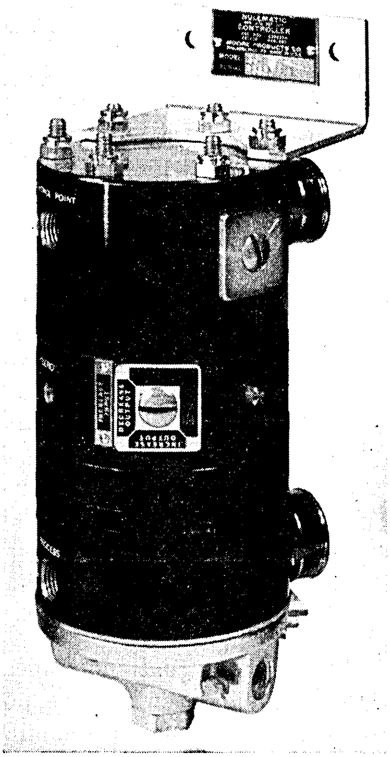


Fig. 5.11*a* Pneumatic controller (Courtesy Moore Products Co.)

Furthermore, the high frequencies and broad frequency bands permissible in electronic circuits permit "time sharing" of intelligence on the same wire or radio beam, thus permitting a multitude of signals to be transmitted on one carrier in sequence or simultaneously.

This high potential increase in speed opens the door to an important new method of measurement which is repetitive and selective, i.e., the "scanning technique."

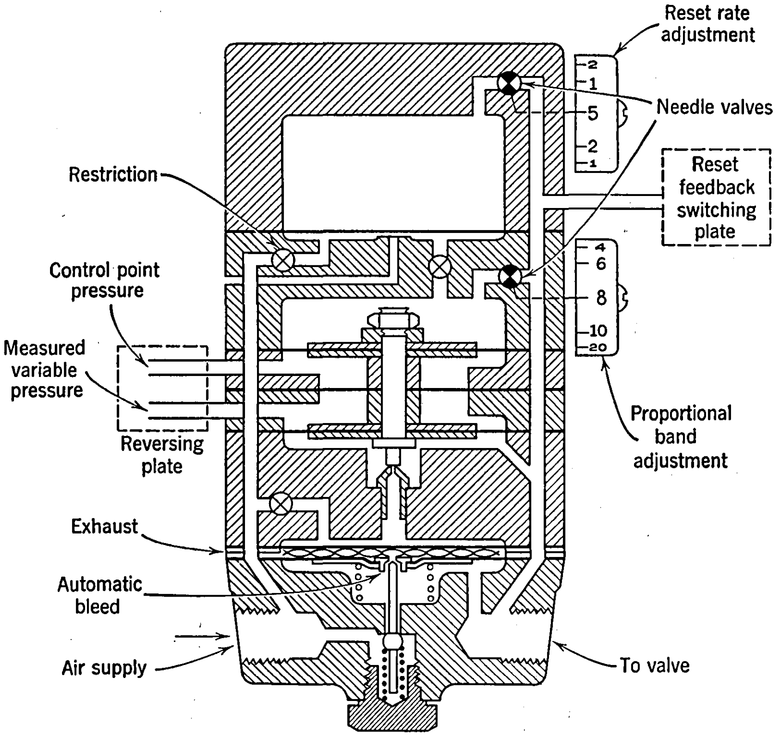


Fig. 5.11*b* Schematic of controller in Fig. 5.11*a*

5.4.1 Scanning Techniques

Instead of a continuous measurement of a variable at one point, it is possible to switch the point of measurement automatically at some high frequency of scanning (high relative to process and human observer time constants). Thus the "field"—or distribution of a single variable in space—is communicated to the operator, rather than to an individual point reading, without the advantage of apparent continuity being lost (Fig. 5.13); alternately a number of different variables may be scanned and displayed. This method, which has until now only been used for measuring purposes, could and will be logically extended to control applications, as will be discussed later on.

As a matter of fact, the human operator's present practice of manual and semiautomatic control is to apply this scanning procedure to plant control, at a slower speed of course, readjusting at intervals of time the positions of control valves or set points of con-

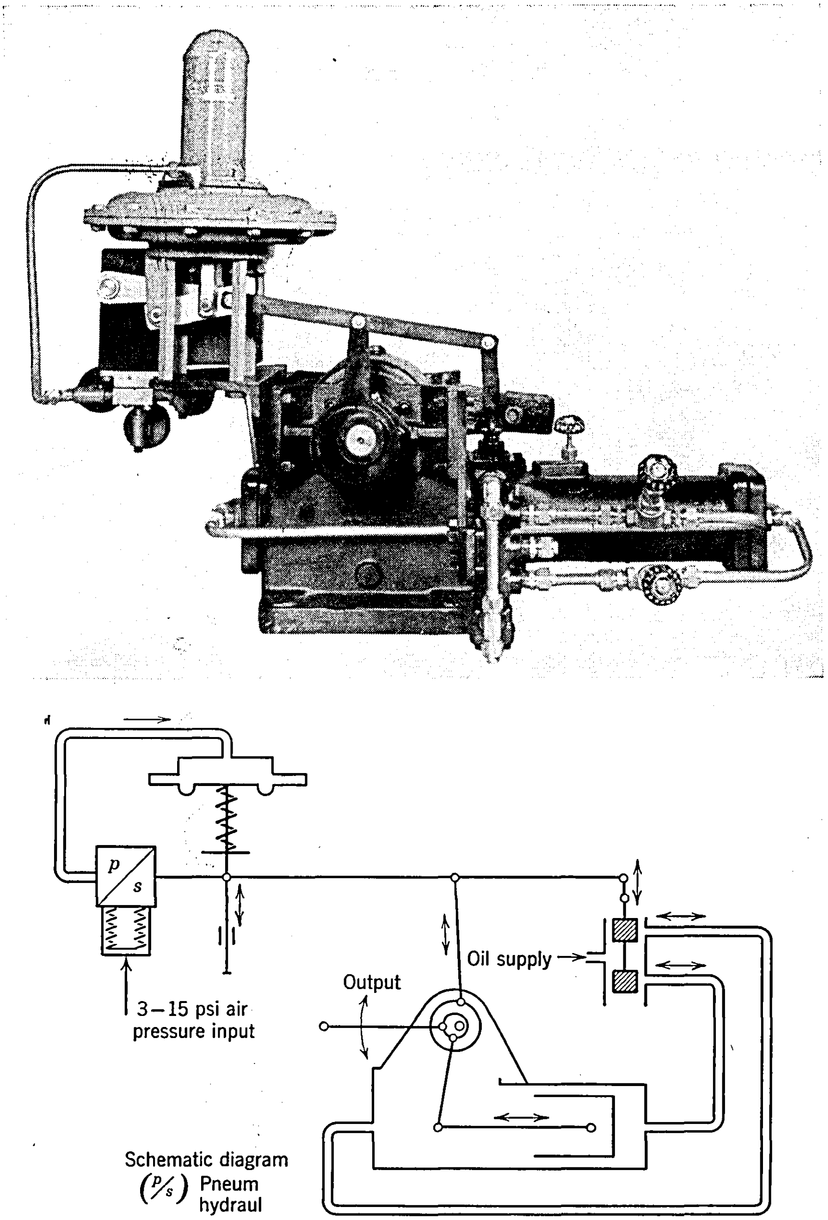


Fig. 5.12 Pneumatic hydraulic relay (Courtesy Askania Regulator Co.)

trollers. There is no reason for not using one control with many variables and corresponding gains and many control valves if the chosen scanning frequencies are high enough and the inputs and outputs are made compatible.

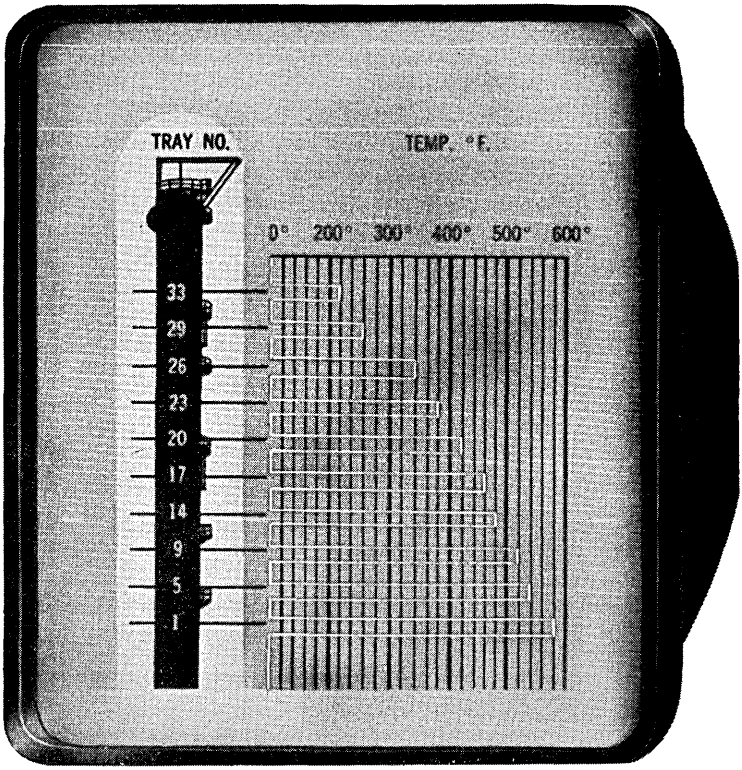


Fig. 5.13 A scanning system which presents "simultaneously" several variables or the same variable at different points of the process (Courtesy The Kybernetes Corp.)

5.4.2 Decision Elements

Perhaps the most important or significant contribution to instruments and controls is the recent introduction of *decision elements,* which, taking advantage of the theoretical advances in symbolic logic, permit a control instrument to choose alternate courses of action. Figure 5.14 shows one version of these building blocks which are basically switches without moving parts. The two inputs are, for instance, of the nature of "yes-no" intelligence, or in the binary code, 0 or 1. A third input is a clock voltage which triggers the answer "yes" or "no," depending on the inputs. These devices, working at high frequencies (megacycles), can be used to play games, or, more important, to establish a strategy of procedures for acting under pre-established emergency patterns (the old safety interlock), or to choose optimum procedures (automatic telephone scanners for open connections between cities), or even to choose different modes of control to satisfy predetermined safety or stability specifications.

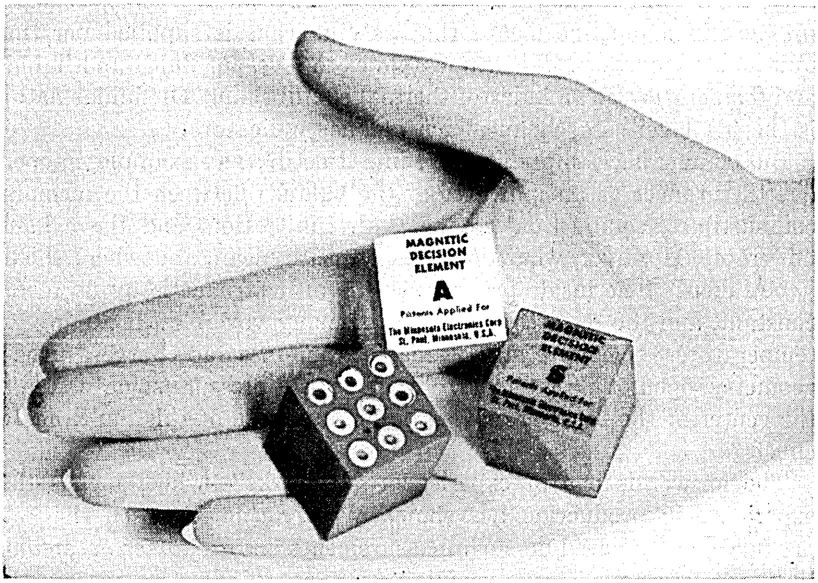


Fig. 5.14 Magnetic decision elements (Courtesy Minnesota Electronics Corp.)

With such a formidable array of new tools at his disposal—the list is by no means complete—the instrument designer of today is better equipped than ever before to tackle the problems of industry. In passing, the absence of one still unavailable basic device should be mentioned in the hope that this may stimulate thinking and trigger inventive ideas. There is no instrument as yet that can identify patterns, faces, signatures, or the identity of landscapes in winter and summer. We shall see later on that the lack of this basic building block is one of the major handicaps in the further development of full automation and in particular of automatic inspection. (For example, signatures must still be identified by a human checker in banks.)

5.5 THE SIGNIFICANCE OF MEASUREMENTS

In deciding on the measurement of variables in the industry to control a process, some basic questions must be asked. One, for instance, is what does this variable mean?

The most frequently measured and controlled variables like level, pressure, and temperature are actually energy balance indicators which give some information on the difference of input and output energy flow rates and the prevailing energy level. Thus, *constant pressure* in a pipeline means that as much gas is supplied per time unit as is withdrawn. The same applies to *level* for fluids and liquids, and *temperature* as an index of thermal equilibrium. Of similar nature is the *p*H level as a chemical ion balance indicator.

But coming back to pressure, we find it used, as an example, in open-hearth furnaces as an indicator of the balance between the incoming combustibles, combustion products, air infiltration, and the exhaust flow rate. However, what does this pressure mean, and where should it be taken? The inside pressure profile of a furnace is by no means constant, and the result of the effect of many variables, including gas temperatures, velocity distributions, location of scrap charges, furnace geometry, open or closed doors, etc. Why is a pressure signal taken at the center of the furnace roof acceptable in spite of this multitude of effects?

The answer is one of experience: It "works"—which means that experience in production has shown a *correlation* between the end-*product quality* based on such measurements and *customer acceptance* that is great enough to be useful.

5.5.1. Correlation between Measured Variable and Desired Property

In general, such practical and seemingly arbitrary choice of indices for somewhat remote measuring targets is very widely used by experienced operators. Its soundness should not be underestimated by the academic purist who will, for instance, shudder at the idea of measuring air flow in a boiler as a pressure drop in the combustion chamber—a method violating the most fundamental laws of any beginner's textbook on flow measurements.

To systematize this choice a group of German engineers under the leadership of Dr. Ing. K. Daeves have given this correlation problem much thought; they have developed rather simple techniques to make such a correlation evaluation a general tool for industry. Unfortunately, no one has taken the time so far to translate Dr. Daeves' most useful and stimulating book, *Praktische Grosszahlforschung* (see reference 8).

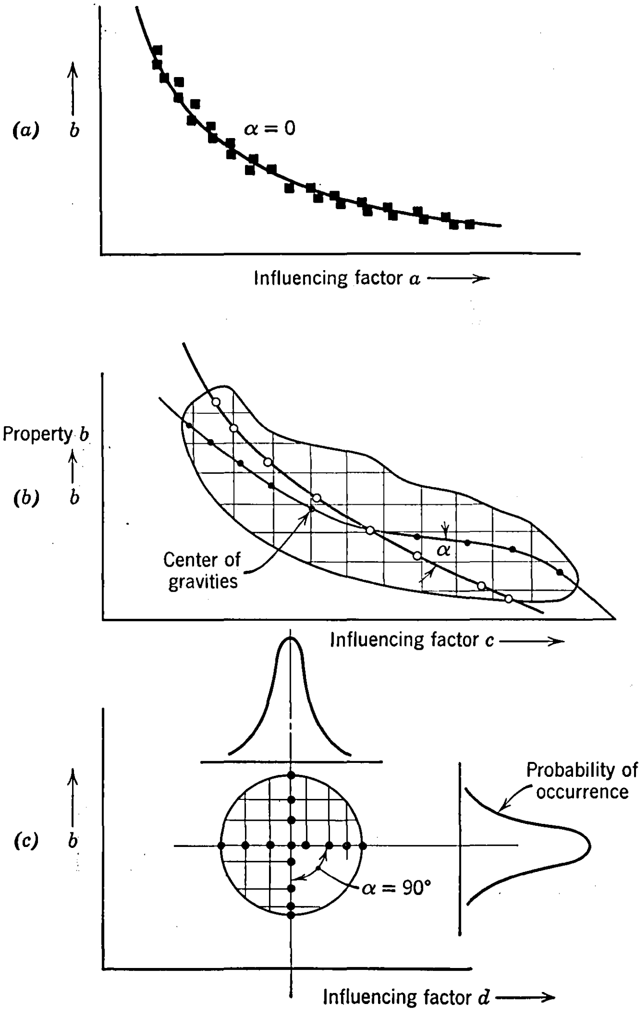


Fig. 5.15 Evaluation of correlation of various variables and the desired property of an end-product (After Daeves)

Although the conventional engineering approach hopes for the perfect-law relationship between two variables a and b (Fig. 5.15a) that is expressed by a narrow band curve (preferably a line), there are many cases where the observed data cover a broad area (Fig. 5.15b). To handle these data, Daeves connects the outer rim of the data area by a continuous border line and then establishes the center of gravities of sections parallel and perpendicular to the abscissas. Connecting their respective centers of gravities, he obtains two curves with an intersection and an intersection angle (α).

If $\alpha = 0$, we obtain (Fig. 5.15*a*) the classical law (100 per cent correlation), or if $\alpha = 90$ degrees in Fig. 5.15*c*, zero correlation.

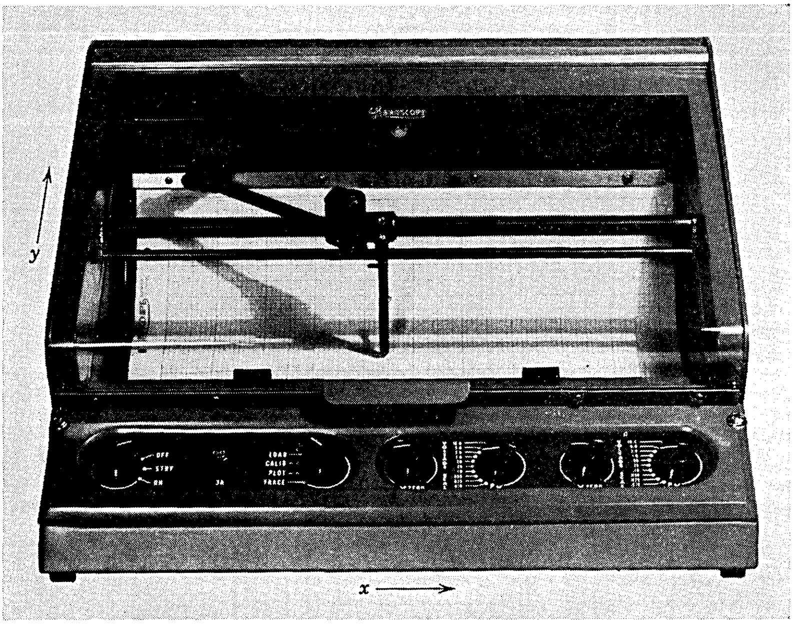


Fig. 5.16 An x/y recorder (Courtesy Librascope Inc.)

The important, and at present mostly overlooked implication (known to the old-timer expert—but often overlooked by the young engineer) is that adjusting not only b but also c and other influencing factors for the optimum of the desired property b, is likely to give better results, often with less-complicated procedures, than concentrating on the usually preferred 100 per cent correlation factors.

For the purpose of simplifying or mechanizing this type of study, "x/y recorders" become almost a necessity. Figure 5.16 shows a late model of such an x/y recorder which can handle digital as well as analog input data directly from transducers, plotting continuously or in batches, the correlation diagrams of $b = f$ (a, c, or d) of Fig. 5.15, or data fed into it from tape records or from punched cards. This approach of systematically exploring the influence of variables on the desired property of the end products is a very powerful tool which will gain more importance in the future. It will become standard practice in process analysis as soon as all instruments speak the same output tongue, or common language.

5.5.2 Statistical Instruments

With the advent of digital and electronic techniques, it should be pointed out that devices and instruments for statistical analysis will become everyday tools. Statistics in the past have been used only to a limited degree in plants, except in quality control, i.e., for the usual inspection procedures of plant products. One of the reasons for this delay in a more general acceptance of this powerful tool might be that statistical terminology and charts commonly used are difficult for the uninitiated to understand. The tediousness of the job of continuously recording "standard deviations" for instance, or of recording a "control chart," is not conducive to enthusiastic acceptance in the shop or even in the laboratory. The automatic computing tools now available should overcome this handicap and offer continuously significant—and thus valuable—intelligence to the operator.

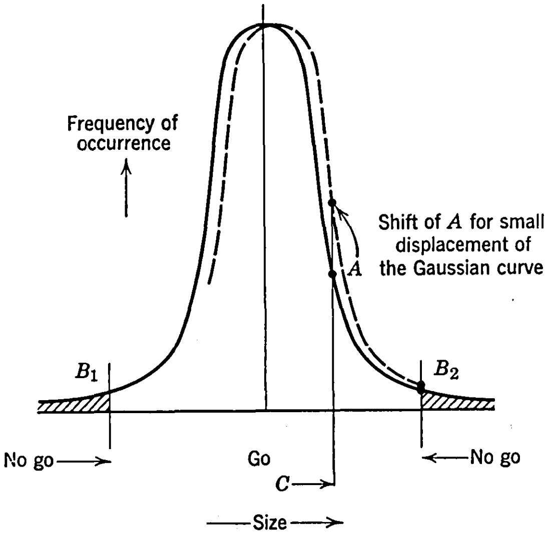


Fig. 5.17 Inspection method using Gaussian distributions

As an example, let us take the case of the inspection of the size of an article made by hand or by a machine. If the frequency of occurrence of a size within a given range is plotted over the size, we obtain, in general, a Gaussian distribution (Fig. 5.17). Present "go"–"no go" philosophy tries to reduce the "no go" samples to a minimum, or zero, and is thus faced with the problem of trying to measure the occurrence frequency of an event that should not happen in the first place. It becomes a real problem, for instance, to decide whether the rejects have increased from 1 per mil to 2 per mil—or, even worse, to 1.5 per mil.

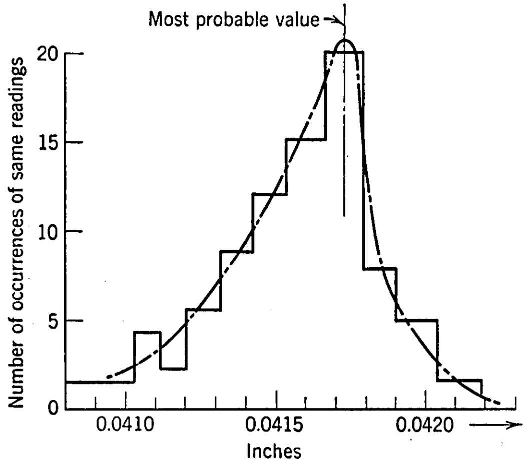


Fig. 5.18 Various readings taken with the same instrument on the same workpiece (Courtesy Sheffield Corp.)

Standard deviation methods and similar statistical tricks solve this problem; however, it might be worth while to change the basic approach by measuring the occurrence frequency A in the range C, rather than the values B_1 and B_2. An instrument to graph the diagram above and to show shifts of the mean can also directly indicate the value of A for a given sample lot which has a close correlation to the rejects B and can thus *anticipate them* at a greatly enlarged scale. Such an instrument can easily be built with memory and indicator devices now available.

Of course, the hope of all instrument builders for a *"customer acceptance indicator"* is the ultimate target—but in most cases it is still a dream and probably will be for some time to come. In the meantime we may rely on hunches, the opinion of our relatives or of our secretary, the Sales Department, and the Gallup poll—the latter being based again on the more promising statistical approach. All of them can boast of a certain degree of reliability—for otherwise our economy would have collapsed long ago—but the failure of these methods could be, and has been in many cases, catastrophic.

Superimposed on the probability curve of the quality of the end product are the probability curve of the instrument measuring the correlated measurable variable, the probability curve of the observer noting the instrument reading (Fig. 5.18), and the probability curve of a distortion of the message during transmission to the man or the machine, that must initiate an action based on the information received.

As most practical production machines have, when in action, the property of staying for a reasonable length of time within established distortion levels, the trend of replacing human beings by the process of "automation" will successfully continue.

5.6 AUTOMATION IN PROCESS CONTROL

Let us therefore examine how far automation has gone in process control and how far we are likely to extend its use in industry! For the purpose of this analysis we shall consider a typical plant as shown in Fig. 5.19. In this diagram we recognize the well-known basic

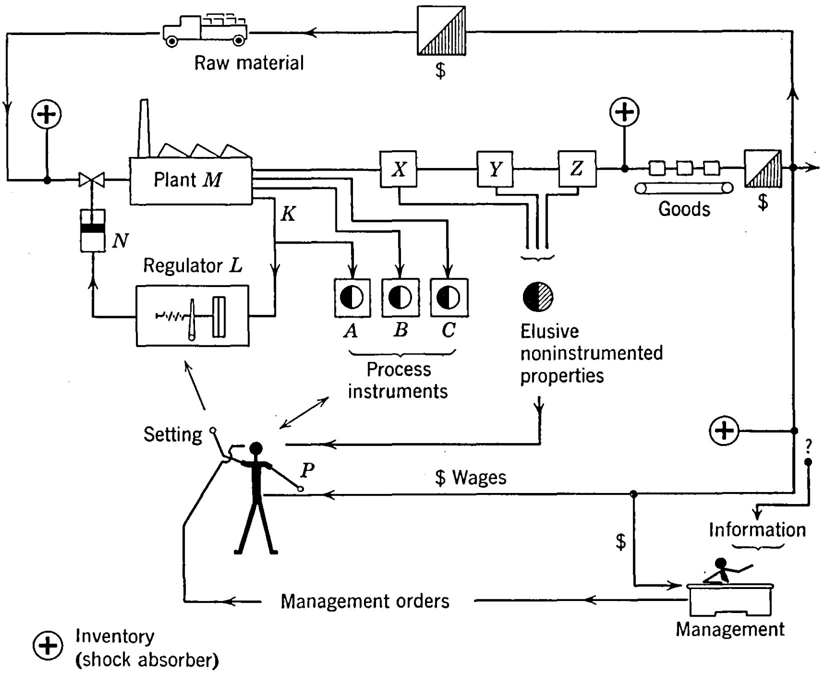


Fig. 5.19 Typical industrial-plant control diagram

plant controller loop. However, it contains a few disturbing items that are usually left out. First of all, and this appears to shock many, it indicates at the right as the goal of an industrial plant a dollar sign, the symbol of the economic justification of all the efforts, machinery, and equipment invested in the endeavor. At least, this goal seems to be dominant in our American industrial peace economy. In case of government efforts and war, other targets may replace the dollar sign. "Customer acceptance" in the broadest sense may be symbolized by this dollar sign, even if the economic transaction is not as obvious as, for instance, in the case of the atom bomb explosion where other targets in the most literal sense have to be reached.

The most common major loop at present is the combination of the plant M with a regulator L and a valve operator N. Various energy balance indices or process variables are transmitted as inputs to the regulator L and finally result in action on the manipulated input. The important but usually forgotten operator P is shown to change the settings of the regulators L in accordance with his experience, knowledge, hunches, or the indications of other instruments, A, B, C, remotely correlated indices, X, Y, Z, and the orders of his boss or of management.

The computation and correlation of all these factors influencing the operator's decision can, and are, being handled more and more by computers; but this can be done now only in cases where the correlation between incoming information and required action is mathematical or at least reasonably well defined. Unfortunately, this is not always the case—or perhaps it is more correct to say that it is very seldom the case. We have previously discussed the A, B, and C type of process variable instruments, such as pressure, temperature, flow rate, level, etc.

5.6.1 The Problem of Measuring Customer Acceptance

Before going further into the controller or regulator subject, it may be worth while to look at the meaning of the X, Y, and Z symbols. These "instruments" are designed to measure customer acceptance, i.e., dollar return for investment—a ratio which must be larger than one (negative entropy!) if the whole venture is to survive or be at least self-sustaining.

A few examples will illustrate this point. In the petroleum industry the only reliable index for the "octane number" is the satisfaction of the ultimate consumer, the "engine." Thus a high-precision "standard engine" is used to establish its value, for, as I understand it, no other sufficiently close correlation between this number and a more easily measured variable is known at present.

Another example of a problem of great economic importance, for which no solution is at present available, is a standard for "cleanliness" of paper or fabrics. It is possible, and this has been done, to scan the paper with a photocell device and to obtain repeatedly the same number. However, this number is *not* the required answer, for it only represents an integral or average surface light reflection value, and it cannot distinguish between various dirt patterns which would easily be judged or classified by a trained observer.

Again another example is the ultimate customer acceptance of alcoholic beverages or perfumes, in particular of those calling for a prolonged aging period. Here the decision of whether to consider a production run acceptable or to reject it poses some extremely difficult problems.

In the first example of the octane number, the solution was to test the product under actual operating conditions. The instrument man's target is thus clearly established, i.e., either to produce a continuous test of a suitable kind with the answer continuously available to the operator, with a minimum time delay, or to find another closely correlated variable and to control it with or without a computer or controller. The paper-scanning problem is of a nature which, although it gives immediate answers to an observer, does not locate any of the many possible causes for rejection or suggest action to avoid their recurrence. In this area, statistical instruments taking into account dirt distribution pattern (the subject of Gestalt psychology) or the evaluation of statistical data obtained from inspectors may be an immediate answer but perhaps only a partial one.

The last example of the alcoholic beverage or perfume evaluation shows how much fundamental research is still necessary in an instance where time delays and changes of habits may cause the taste of the consumer to diverge from a product whose quality could have been within specifications at the time of its manufacture. Again statistical methods offer a guide for determining causes of output variations, customer acceptance, or for the interpretation of observable and correlated properties. The systematic use of x/y plotters and statistical computers, not as a laboratory device but as a continuous production tool, will materially assist the operator of the future in gathering pertinent data under such trying conditions.

5.6.2 The Raw-Material and the Accounting Loop

In the general diagram we notice, in addition to the supervisory "operator" loop, the "management" loop, and the "raw material" loop. The raw-material loop is a feedback of part of the output into the plant input, after its translation into dollars and then back into material, thus permitting the process to continue. Its value must be less than the output in economic terms to guarantee continuation of the process. Its supervision is the province of management, accounting, and purchasing; and its communication method for the last few thousand years has been the recording of data on paper, usually in digital form and in the letters of the alphabet. The read-in and read-out time of this process is notoriously slow and even IBM or similar methods are still bulky, complex, and spacebound.

Inventory figures representing continuous drain rates are available to plant operators with considerable delay; sometimes inventory figures are not available for months. Rate of dollar flows to individual process branches is often unknown or only approximately certain. Here an important change is likely to take place in the near future. With magnetic drums or magnetic tapes replacing or supplementing book entries, and by the use of simplified computers with decision element circuits that furnish, at high speed, data in a tabulated or graphed form, this information will become less intermittent and should be available with little delay to the operator and to management. Thus it can be foreseen that this type of advanced automation will greatly speed up the office, administration, and accounting loops, in which the present instrumentation and control are far behind minimum process standards. It will be noticed, incidentally, that part of the economic return goes into the pocket of the operator and part into the pocket of management. These are important loops; they have a bearing on efficiency, continuity of the process, and the quality of the end product.

5.6.3 The Management Loop

The last loop mentioned above is the "management" loop, which determines the rate of production, scheduling, and quality and supervises the flow of raw material, money, energy, new facilities, and manpower for the venture. Here we find at present a minimum of instrumentation and a maximum need of assistance, with unbearable time delays of badly needed information and a surprising lack of interpretation of past records and of extrapolation of present and past data into the future. This calls for an unusually skilled manager with an astounding willingness to take risks and with an almost uncanny feel for the correct thing to do.

The cost of top management, its scarcity, its high turnover, and its migratory habits dramatically illustrate a very real need for better instrumentation and better communication as tools for better control.

5.6.4 The Inventory Control Problem

It will be noted that at several strategic points of the diagram plus symbols are added to indicate storage or inventory facilities. In the broadest sense, such storage facilities are accumulators which are needed because there are discrepancies between the integral of input and output rates; or they become necessary as filters or equalizers for product quality variations. Inventory appears on a dollar rate flow diagram, which, incidentally, does not coincide with the energy flow diagram, as a source of dollar loss in capital charges, rent for space, storage facility write-offs, and storage maintenance. Inventory controls to maintain constant levels for existing facilities or to schedule products and raw materials to maintain minimum levels become extremely important. This applies to a refrigerator-producing plant in which it was discovered that the buying rate of the ultimate user was steadier than the retailer order rate. It also applies to the general trend to reduce in-process and in-transit inventories to a minimum.

A good example of the trend toward a decrease of storage capacity is the old pressure steam boiler whose immense water drums are disappearing and are being replaced by low-capacity, high-pressure tube arrangements which—and this is significant—can no longer be manually controlled unless at a constant load. This means that man as the bottleneck in strength, speed, and accuracy is being, *and has to be,* replaced by controllers to get maximum efficiency of dollar investments. He still represents, however, the limiting link in emergencies and in instances where supervisory action is necessary.

5.7 THE BASIC CONTROL LOOP

This brings us back to the first loop, the one of the plant, the regulator, and the power actuator. How far are we in industry with the automation of this fundamental or basic circuit?

Of all the loops described and engineered, it is the one furthest developed, but its limitations should be pointed out to trigger still-missing basic solutions which are badly needed. The example of the airplane will illustrate the point, for an airplane is a rather definite self-contained universe or plant-operator unit. So far, our control engineers have rather satisfactorily solved (1) the problem of constant-level flight.

What this means, really, is that we have succeeded in eliminating time as a parameter, all our efforts being concentrated on returning deviations of attitude or of maneuver *as quickly as possible* to the desired value. This means transients are to be reduced to zero as quickly as possible; several criteria of quality of control action are based on this concept. In an industrial plant, the same requirement may apply to temperatures, pressures, flow rates, or any other variables.

The second airplane control problem of major transients is still unsolved (outside of the guided-missile field): (2) the problem of becoming airborne, of landing, and of evasive maneuvers.

The obvious parallels are getting the plant "on the line," "shutting it down," and making major changes in products or production rates. Anybody having been present under such operating conditions, regardless of the length of time required, is only too well aware of the mental hazards involved as well as the dangers of spoilage, of destruction of facilities, and of delays.

The third airplane problem, which is of a very serious nature, is the emergency situation, or: (3) the problem of survival for operators, passengers, and plane in an emergency.

In many "automated" plants of the present—from the automatic home refrigerator to some chemical plants—the serviceman (no further away from you than your phone) or the operator are firemen in disguise who know, or are expected to know, how to act in emergencies. The methods of coping with emergencies may vary from a decision to shut down the refrigerator or the plant, to emergency measures to save the plane, to a decision to abandon ship by parachute, the latter in some cases automated by ejection seats. We shall see later how all these problems could be solved automatically.

The fourth problem of the airplane is one of scheduling: (4) the problem of deciding where to go, how fast, how high, and when.

Although scheduling was previously left entirely in the hands of the ground crew, a certain freedom, at least in principle, is at present given to the flight captain. His decision is influenced by a great multitude of intelligence reaching his brain, including (*a*) the time table, (*b*) the weather, (*c*) the condition of the plane, (*d*) the condition of passengers and crew, (*e*) his position in space and time, (*f*) his fuel reserve, (*g*) the political situation, etc. Here we are back to our top-management loop; the best we can hope for is that the information on weather, plane condition, and whereabouts is reliable and meaningful and its significance understandable without major computations.

We have said that the basic problem of constant-level flight is more or less solved. Let us therefore take a look into the tool chest of the design engineer and apply our previously gained knowledge of a systematic design approach to the problem of building a controller.

5.7.1 Typical Amplifiers

First of all, to supplement our knowledge, we shall investigate some of the energy amplifiers or multipliers now in general use. For fluid power media, i.e., for air or hydraulic amplifiers, either three basic devices are commonly used or modifications of their basic design can be traced back to these three units (Fig. 5.20). In Fig. 5.20*a*, a four-way valve, consisting of a spool and a sleeve, is displaced relative to two ports which connect to a double-acting cylinder. A displacement of the spool by an amount s to the right will, for instance, open the right-hand side of the cylinder to the supply, and simultaneously the left-hand side of the cylinder to the drain. As a result, the piston in

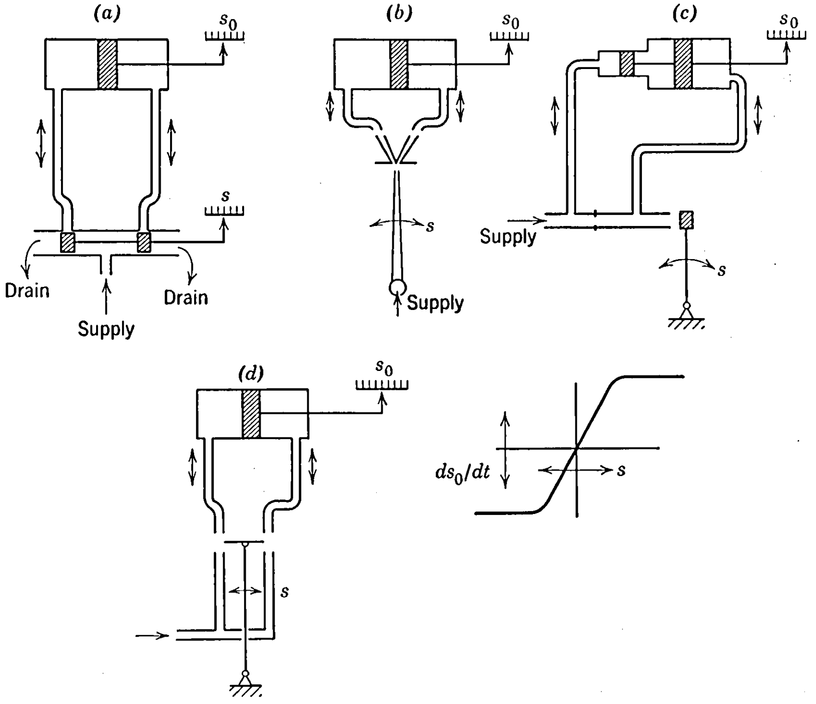


Fig. 5.20 Typical fluid-type relays: (*a*) four-way spool valve, (*b*) jet pipe, (*c*) double throttle, (*d*) jet interceptor

the cylinder will move to the left at a rate ds_0/dt proportional to the displacement of the spool, s.

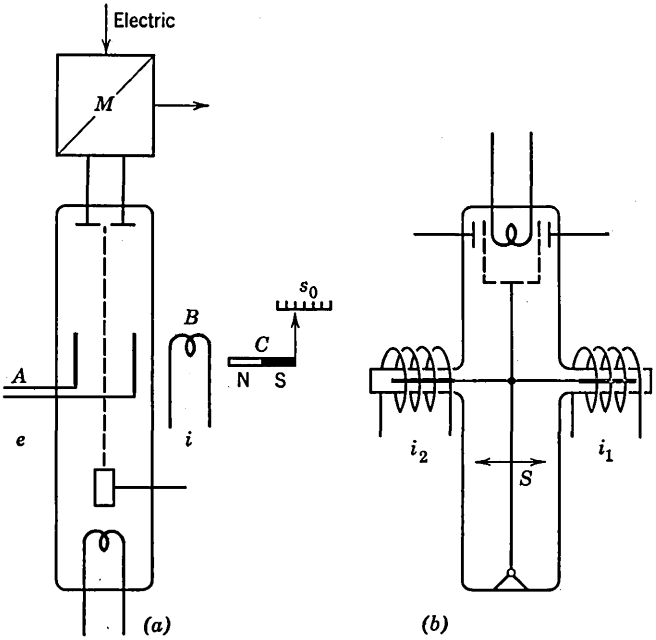
In the next amplifier of Fig. 5.20*b*, the "hydraulic jet pipe," the spool is replaced by a hollow jet whose fluid is delivered into two cylinder lines at high velocities. Thus, we find again a characteristic that establishes proportionality between cylinder speed, ds_0/dt, and amplifier displacement, s.

The third example shown in Fig. 5.20*c* uses a "double throttle" as an amplifier element. This design element has become almost universally accepted in air-operated instruments because of its simplicity of design and its minimum space requirements. It is based on the observation that the pressure between two orifices in series connected to a pressure supply and discharging to atmosphere (or any other given level) will vary as a function of the change in the ratio of the sizes of the supply and relief orifices or restrictions. In the most frequently used design, the first orifice is fixed and the second variable. In Fig. 5.20*c* the supply pressure is connected to the small area of a two-stage piston assembly and the modulated pressure between the two orifices to its larger area. Evidently, there is a position of the outlet orifice, s, which establishes a balance of forces on the piston with a resulting zero speed. Any movement of the flapper (the outlet orifice) to the left will increase the intermediate pressure and thus force the piston to the left. This again will happen at a rate proportional to the displacement s.

In pneumatic controllers of the conventional type, the two-stage piston is replaced by a diaphragm responding to the intermediate pressure and a spring which takes the place of the smaller piston. It will be noted, however, that in this instance the spring force is proportional to the diaphragm stroke, and the double piston, as shown, has a fixed biased force independent of stroke. There are a great number of variations of designs, depending on particular applications and the whims of the individual designer.

One additional variation of design that should be mentioned may be considered as a crossbreed between those of Figs. 5.20*a* and 5.20*b*. In this case the jet (or jets) stands still relative to the receiving ports, and an intermediate vane deflects or intercepts part or all of the fluid streams which determine the equilibrium condition of the cylinder. These amplifiers are basically "push-pull" amplifiers, for their output is duodirectional. (See Fig. 5.20*d*.)

A "translation" of the amplifier (Fig. 5.20*b*) into equivalent electrical terms resulted in an electronic amplifier, as shown in Fig. 5.21*a*. The hydraulic or fluid energy stream which is deflected by the displacement of the jet pipe in Fig. 5.20*b* finds its analogy in a beam of electrons in a beam tube of the cathode-ray type (see reference 18). The space displacement of this electric current relative to two target plates corresponds to the displacement of the fluid stream relative to the two orifices.



Electron beam tube

A–electrostatic deflection, e
B–electromagnetic deflection, i, or flux, ϕ
C–field displacement deflection, by permanent magnet, s_0

Variable grid tube

Electromagnetic deflection, i

Fig. 5.21 Electronic relays

Such a displacement of the beam can be accomplished by (1) an electrostatic field, (2) an electromagnetic field, or (3) a displacement of either field relative to the beam. Following the same systematic approach (Fig. 5.21*b*), a variation of this theme was developed in which the displacement of a lever carrying two grids, and thus varying the impedance of two current paths proportionally to the lever displacement, results in a push-pull circuit equivalent to that of Figs. 5.21*a* or 5.20*d*.

The reason for giving these somewhat unusual examples rather than the conventional electronic push-pull circuits for (1) electronic tubes,

(2) magnetic amplifiers, and (3) transistors is that these examples illustrate the usefulness of the *basic* philosophy of approach previously discussed; and furthermore, standard amplifier circuits are to be found in every one of the many texts already available.

5.7.2 The Design of a Typical Proportional Controller

Equipped with these building blocks or translators and operators, we shall now develop or synthesize three versions of a proportional controller in order to acquire a "feel" for the methods outlined above and to get acquainted with the method of designing new devices. The three solutions chosen are by no means the only possible ones, nor do they claim to be the simplest or the most logical ones to be devised. But before going into the design, i.e., into the "synthesis of the problem," we shall start with an "analysis" of what a "proportional controller" consists of. Figure 5.22 shows the basic block diagram and uses the symbols with which we have previously become acquainted.

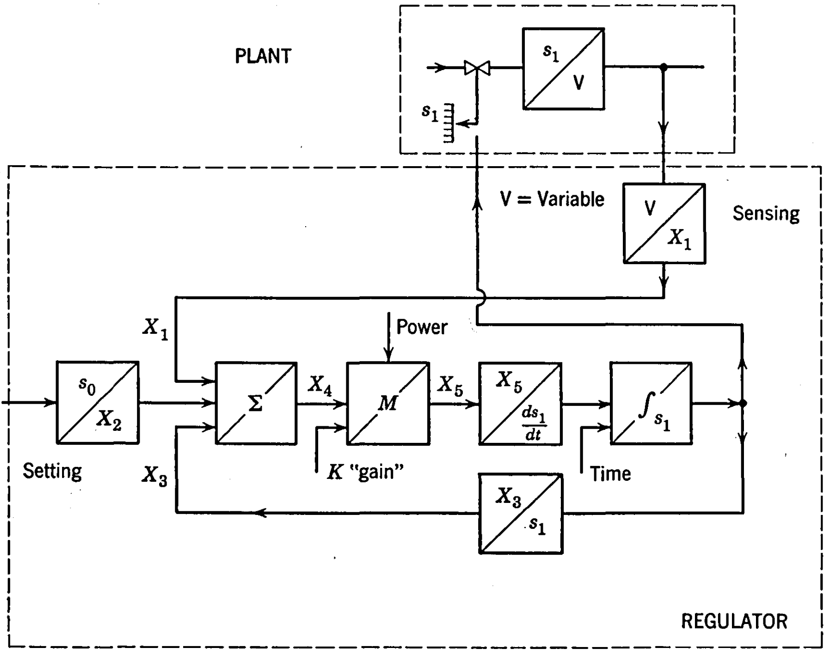


Fig. 5.22 Translator analysis of a basic proportional control loop

Starting with a plant, we find that a plant, in our way of looking at it, is simply a device which translates an input change of *s*, which may be a valve motion, into a corresponding change of a variable. It is evident, of course, that there are many variables in a real plant, which may be affected by a number of inputs. For the purpose of this single-variable controller discussion, however, we shall assume the simplest case of one input (stroke) and one output (variable) only. Thus, the symbol for the plant is s_1/variable.

In a "proportional controller" the valve position s_1 is to be made proportional to the "error," i.e., the difference between a "setting," i.e., the "desired" and the actual variable. The setting s_0 and the variable as well as the valve position s_1 must therefore be translated into a common parameter X so that these three can be added up or compared in a summarizer, our "operator" Σ.

Any output of this summarizer different from zero has to be amplified, i.e., multiplied by means of an operator M with a "gain factor" K to be chosen in line with stability considerations of the whole loop. Such an amplifier will obviously use additional outside sources of energy. Choosing amplifiers as previously discussed in Fig. 5.20, we obtain outputs that are valve operator velocities ds/dt which, integrated in time, produce the necessary or required corrective valve motions s_1 as outputs.

The symbolic equations describing this chain can be written as (1) plant: s_1/variable; (2) regulator:

$$\Sigma(s_0/X_2 + \text{variable}/X_1 + s_1/X_3) + M(K \cdot X_4) + \frac{X_5}{ds_1/dt} + \int \frac{ds_1}{dt} \cdot dt$$

A "force balance" controller. As our first solution we shall try to use a force F for the common denominator $X = F$. Figure 5.23*a* shows one of the many possible solutions. The translation of s_0/F_2 is accomplished by means of a calibrated spring. So is the translation of the valve motion s_1 into F_3.

The plant variable in this chosen example is a pressure p and it is translated into a force F_1 by means of a diaphragm. The summarizer is a lever arrangement which is used at the same time as the amplifier of the hydraulic jet as shown in Fig. 5.20*b*.

The output of this amplifier is a fluid velocity and piston speed ds_1/dt which is proportional to the unbalance of the forces $\Sigma(F_n)$ and the corresponding jet displacement (proportional displacement due to spring restraint, F/s). The integration is accomplished by the hydraulic cylinder with the piston rod movement as the output, s_1.

This completes the design in principle, and the particular choice of the variable force F in the summarizer makes it a representative of the so-called "force balance" type controller (Fig. 5.23*b*).

Stroke-compensated controller. Our next attempt in Fig. 5.24 is based on the choice $X = s$, which is typical of the "position balance" or "stroke-compensated" type controllers. The summarization is accomplished by a mechanical summarizer, the so-called "whiffle-tree linkage" (see Fig. 5.6). No translation of the setting (s_0) is necessary, for it is already available as a stroke.

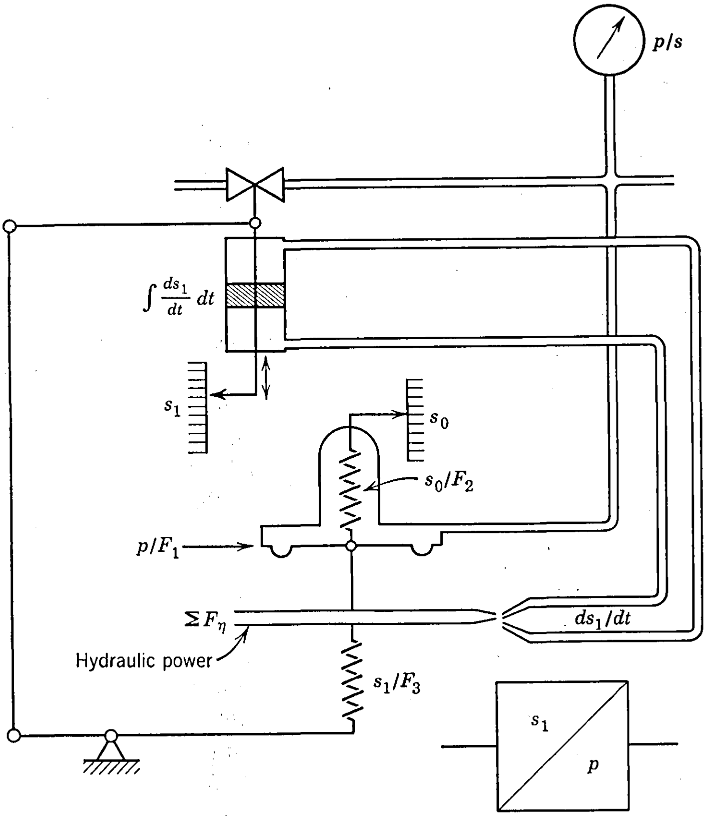


Fig. 5.23*a* The hydraulic amplifier solution for a proportional controller with force feedback $X = F$

The translation of the pressure p into a stroke s_2—p/s_2—is accomplished by a spring-loaded bellows. The translation of s_1, the valve stroke, could have been obtained by means of a mechanical linkage, as done in many cases in the form of s_1/s_3 translators or by an intermediate translator chain

$$s_1/p_3 + p_3/s_3 = s_1/s_3$$

This design would mean that the valve would be equipped with a "pneumatic loader," basically a pressure regulator transmitting its position s_1 as a pressure p_3, and a spring-loaded bellows at the summarizer end to translate p_3 into s_3. The output of the amplifier p_A can be so chosen that it produces a proportional valve movement, $s_3 = s_1$. This pressure, p_A, then represents s_1 and can be used for proportional feedback. The "regulator" consists of two stroke summarizers, one that compares the setting, s_0, with a displacement of a spring-loaded bellows responding to the input p, the plant output. The other summarizer compares this error signal with the valve position represented by p_A and translated into s_3 or s_1, which are identical. The circuit thus incorporates an additional design trick which avoids the addition of another s_1/p_A translator.

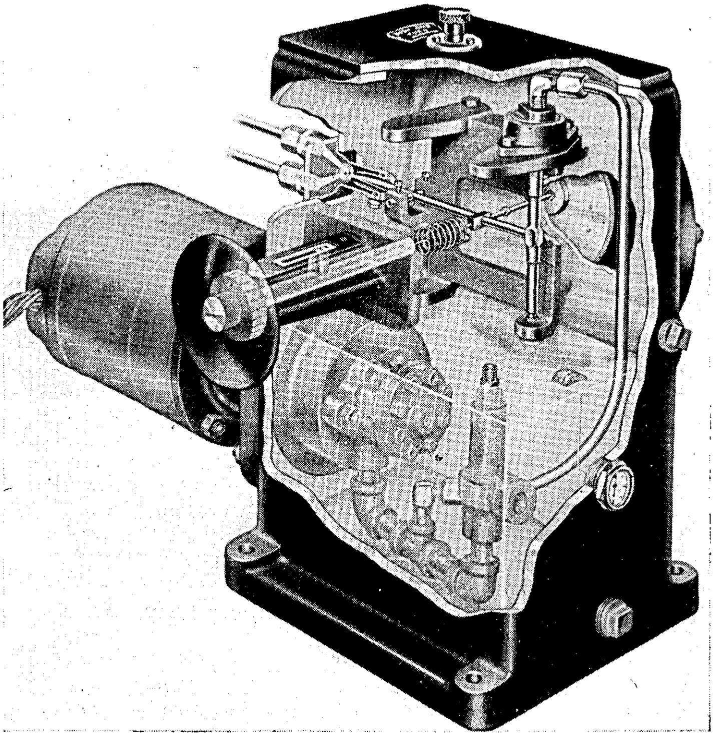


Fig. 5.23*b* A modified design, incorporating the basic elements shown in Fig. 5.23*a*

The corresponding equation change for the regulator would read:

$$\Sigma(s_0 + \text{variable}/s_1 + p_A/s_3) + M(K\cdot s)\text{ pneumatic} + s_5/p_A + p_A/s_{\text{valve}}$$

with $s_3 = \text{const}\cdot s_{\text{valve}}$

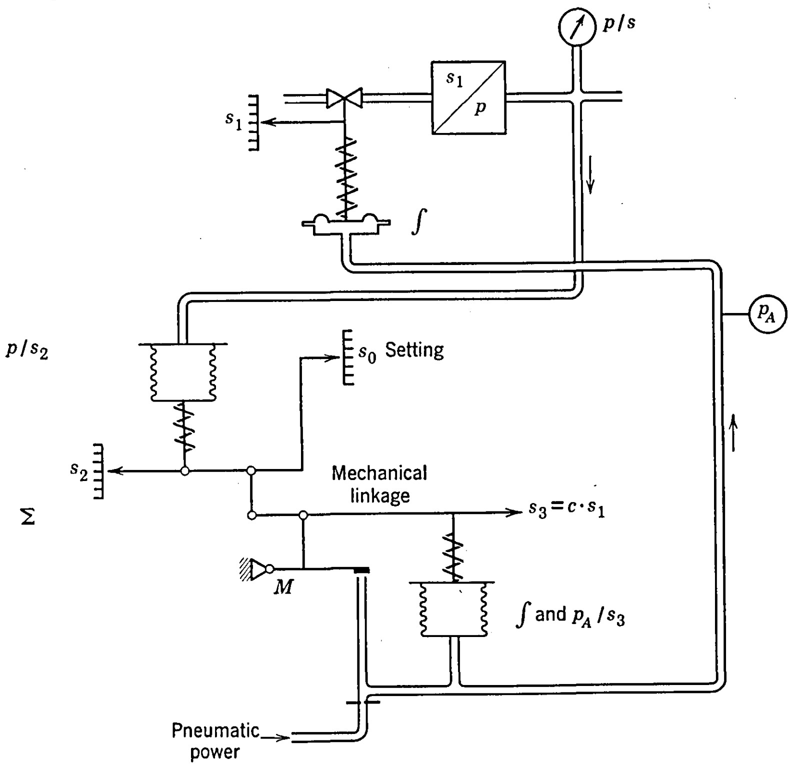


Fig. 5.24 A pneumatic amplifier solution for a proportional controller with position feedback $X = s$

The electrical controller. In Fig. 5.25 the same problem is solved with the use of the amplifier of Fig. 5.21*a*, the electron beam tube. The summarization of X, representing currents, is accomplished by adding the three fields produced by the current coils representing s_0, the setting; s_1, the valve position; and p, the process variable.

As the output of the amplifier which controls the motor does not necessarily produce a motor speed ds_1/dt proportional to the output signal of the summarizer amplifier, a tachometer current feedback loop measuring ds_1/dt is added in opposition to the error signal and represented by it. Its value disappears as soon as the motor, which serves as an integrator, produces a value, s_1, of sufficient magnitude to balance the error, i.e., the difference between the setting and the actual pressure.

Obviously, the designer skilled in circuit developments would not use independent batteries and would most likely replace potentiometer translators by strain gages or inductive pickups with correspondent translations of voltages into currents, or he would choose partial summarization on a voltage basis before final application of an error signal to the tube. But these are variations in detail, not in concept, and would not add any significant insight to this discussion of basic approach principles. The addition of feedback loops, including reset or derivative modes of controls, does not offer any basic difficulties in principle—although it may tax the ingenuity of the designer in his choice of "hardware" to meet price, dynamic requirements, and other specifications.

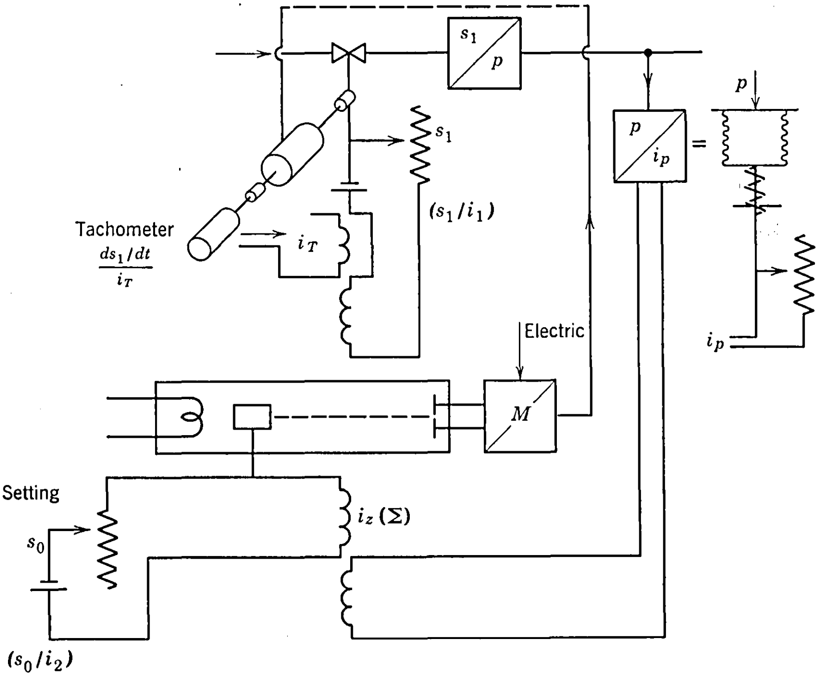


Fig. 5.25 An electronic amplifier solution for a proportional controller with current feedback $X = i$ (electron beam tube used as summarizer)

The "proportional mode" of control given above represents the majority of industrial controls now being used, if not in hardware at least in the basic philosophy of approach.

5.7.3 Program Control

Adding a program device to the setting adjustment gives some dynamic scheduling of the plant performance, assuming again that the setting transient can be paralleled or duplicated by the performance of the plant. This is the required behavior of a guided-missile control in which the transient is enforced by a program of attitude versus time.

In ballistics the dynamics of the process (its trajectory) is known or computed for a given missile and a selected target, and eventual coincidence with the target is expected after the missile release, leaving the rest to "natural" behavior. The same could be done to plants where a computer could, for instance, continuously establish the rate of heat input into a process or the load reduction rate in such a way that certain chosen critical variables would stay within safe limits. This is the "becoming airborne and landing" problem previously discussed.

Similarly, the problem of handling emergency conditions could be better taken care of by the operator and by fully automatic controls if a computer operating at a fast time scale (note the parallel to tracer bullets and fire controls in artillery problems) would explore the to-be-expected results of alternate decisions on actions to be taken, and choose the one that falls within safety limits. By following this approach in reasoning, the use of a scanning type of computer suggests itself; this type of computer can anticipate the future, just as ballistics does. A further analysis indicates that every human operator works on this basis. In intervals of time he scans the available intelligence—instrument readings, plant conditions, and his background of experience—projects the expected effects of his actions into the future, arrives at a decision, checks its validity by watching the effect of his action, and corrects or supplements his action, depending on results obtained. It will also be noted that he does this faster than the process and repeatedly.

5.7.4 The Two-Time Scale Computer

This suggests a controller as shown in Fig. 5.26. The process A is duplicated by a computer, D, which simulates the dynamic behavior of the original plant; however, it does so repeatedly at a higher rate. For this purpose the attitude of plant A is scanned at frequent intervals and coincidence of computer conditions is enforced with plant conditions at the moment of the start, T_0, of computation. Instead of controlling plant *A*, a high-speed dynamically matched controller regulates the computer simulation of the plant, *D*. Its corrective action is sampled and servoed through *F* to make the valve motion, i.e., the corrective device, follow the pattern of the controller *E* at the real-time scale rate of the plant. This action closes the loop.

It will be noted that the computer starts continuously with new initial conditions which are those of the actual plants at that particular moment. As the computer can and does look into the future and works at a much higher time rate, it accomplishes the task of a tracer bullet, or of a ballistic table as far as plant dynamics are concerned, and by adding the high-speed controller it also predicts the future of the effect of controller action on such a plant. As the time scale is reduced and as all intelligence in this computer is translated into a common language, volts or digits, this device has in addition the advantage of being able to add intelligence to the controller which ordinarily would not be available (as for instance, temperature transients at inaccessible points of the process or load disturbances) and which may greatly simplify the stability problem. In addition, the controller working with voltages or currents in an analog fashion can use rate and higher derivative operators at an economical level which would be unbearable at real-time scale values.

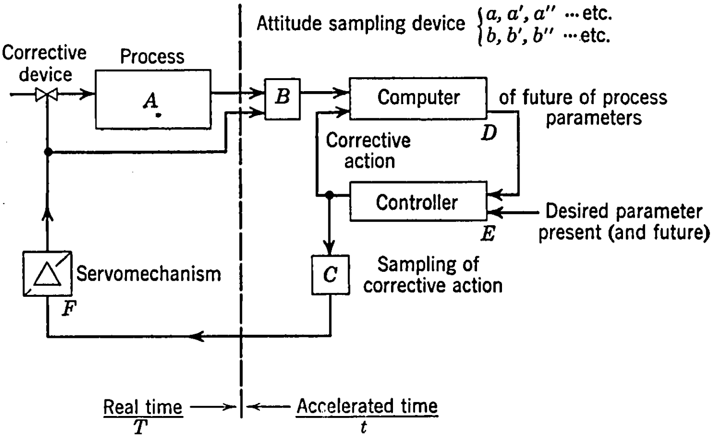


Fig. 5.26 The two-time scale method of process control—repetitive scanning and computing

By adding to this type of control interlocks in the form of high-speed decision elements, the scanning process can be extended also to modify adjustments of the chosen regulator in order to obtain stability, or it can scan the effects of different modes of control to meet safety specifications in case of emergencies. This design therefore could handle the emergency problems of controls which at present have not as yet been vigorously attacked. The scheduling or managerial loop, which also has been discussed previously, could use similar techniques. Such managerial information can be incorporated if it lends itself to the translation of the data into a language that can be handled by a computer, and the theory of games or the methods of operational analysis may be used to arrive at answers, which in turn can be used to initiate control action.

5.8 CONCLUSION

We have reviewed the present status of instrumentation controls and computers from the standpoint of the designer of mechanisms faced with the problem of finding a "hardware solution" for the basic problems in this field. We have noted trends of technology and some of the new tools available to solve the needs for a continuously expanding technological universe.

But we have also outlined certain areas in which man with his dynamic and environmental limitations prevents the complete solution of the full automation so much in the limelight today. These barriers will be pushed back as long as the challenge of "no available solution at present" is tagged on such problems.

Where all this leads to is unknown to us just as the ultimate shape of the growing tree cannot be determined from its seed. The dominating forces compelling man to exert his energies in the conquest of the potential of his habitat may be as fundamental a force as the procreation of his race—whose ultimate goal is also shrouded in mystery.

5.9 REFERENCES

1. A New Approach to Design, *Machine Design,* June 1947.

2. F. W. Cunningham, Employing Computer Components in Machine Control, *Machine Design,* May 1950.

3. Macon Fry, Designing Computing Mechanisms, *Machine Design,* Aug. 1945 to Feb. 1946.

4. G. S. Brown and D. P. Campbell, *Principles of Servomechanisms,* John Wiley & Sons, Inc., New York, 1949.

5. I. M. Embree and S. P. Higgins, A Systematic Design Approach Using Block Diagram Analysis, ASME Paper 52-SA-23, June 1952.

6. K. Daeves, *Praktische Grosszahlforschung,* Chemie, VDI, Berlin, 1933.

7. W. Oppelt, *Kleines Handbuch Technischer Regelvorgaenge,* Chemie, Weinheim, 1954.

8. K. Daeves and A. Beckel, *Grosszahlforschung und Haeufigkeits Analyse,* Chemie, Weinheim, 1948.

9. H. W. Ziebolz, Designing Pneumatic and Electric Servos, *Machine Design,* Sept. 1947.

10. H. W. Ziebolz, Designing Hydraulic Servos, *Machine Design,* July 1947.

11. H. W. Ziebolz, Ratio and Multiple-Fuel Controls in the Steel Industry, *Trans. ASME,* Nov. 1944.

12. H. W. Ziebolz, *Relay Devices and Their Application to the Solution of Mathematical Equations,* Askania Regulator Co.

13. H. W. Ziebolz, *Analysis and Design of Translator Chains,* Askania Regulator Co., Sept. 25, 1946.

14. H. W. Ziebolz, Systematic Design of Mechanisms, *Machine Design,* Dec. 1955.

15. H. W. Ziebolz, Methodical Approach to Mechanisms, *Cleveland Eng.,* March 5, 1953.

16. H. W. Ziebolz and H. Paynter, *Possibilities of a Two-Time Scale Computing System for Control and Simulation of Dynamic Systems,* Askania Regulator Co., Oct. 1953.

17. Millard H. LaJoy, *Industrial Automatic Controls,* Prentice-Hall, Inc., New York, 1954.

18. Paul Glass, Deflection Beam Tube, *Electronic Ind.,* Aug. 1944.

6. Analog Computers

STANLEY FIFER

Dian Laboratories, Inc.
New York, New York

6.1 INTRODUCTION

The purpose of this chapter is to familiarize the reader with the types and functions of analog computers, their components, and the manner in which these components are combined in the simulation of physical systems. Particular emphasis is placed on the general-purpose functional computer, in view of the great popularity enjoyed by this type of analog. The basic components, e.g., the feedback amplifier, potentiometer, multiplier, passive network, diode, relay, function generator, and recorder are described. Applications to linear and nonlinear differential and algebraic equations are discussed, and the programming and mechanization of typical problems are illustrated. The scope of analog computer design and application is extremely broad; this chapter is intended to serve only as a somewhat elementary introduction to this subject. An understanding of the basic characteristics and philosophy of analog computation given here is necessary if the reader is to appreciate and evaluate the manner in which analog computers are applied to more complicated and difficult problems, for example to those in automation and flight control, which constitute the subjects of later chapters. Here we can touch only lightly on what are believed to be the more significant aspects of analog computers as related to automation; the reader is referred to the bibliography for a more extensive treatment of the subjects that are of particular interest or importance to him.

The analog computer is distinguished, in general, by the fact that the equations which describe the relations between the computer variables are the same as those governing the relation between the problem variables whose solution is sought. The computer variables may, for example, be voltages, currents, or shaft rotations; the given physical variables may be displacements, velocities, accelerations, pressures, temperatures, or angles. In the course of the solution of a problem on an analog computer, we learn to interpret the computer configuration and the computer outputs in terms of the given physical situation itself, rather than as some abstract analog of it. It is this feature which endows the analog computer with its great value in design and analysis. The computer variables themselves are, in general, continuous quantities and are measured, rather than counted, as on a digital computer. This characteristic, of course, corresponds more closely with physical reality. There is such diversity among analog computers, however, that we cannot make too many generalizations about their characteristics. Moreover, there are some areas in which it is even difficult to effect a sharp delineation between analog and digital methods and components. For example, digital techniques are employed in some analog multipliers, and digital devices are often used for read-out purposes on analog computers. Analog-to-digital and digital-to-analog converters (Chapter 9), important elements in the field of automation, are examples of components in which the two forms of computation are brought together.

We may classify analog computers according to use. They may be employed as computers, simulators, trainers, and control devices. The computer, as the name implies, is employed to solve mathematical problems. The commercially available d-c machines, Beckman, Boeing, Electronic Associates, Goodyear, Mid-Century, Philbrick, and Reeves, are representative of this class. Virtually every engineering, research, and aircraft firm of moderate size has an analog computer as a standard tool to solve problems arising in the course of its work that may be formulated in mathematical terms.

A simulator involves the combination of a computer and actual physical hardware whose response is to be tested and optimized. For example, consider the problem of designing an autopilot, which is to roll-stabilize a missile in flight by means of a roll gyro. The gyro detects the missile angle of roll and feeds a signal proportional to this angle to the aileron surfaces which are caused to deflect so as to reduce the roll angle to zero. If this problem is analyzed on a computer, then the dynamics of the gyro, the control surfaces, as well as the missile, are all expressed in terms of mathematical equations. In a simulator, on the other hand, the actual gyro is mounted on a roll table driven by the roll angle signal as calculated on the computer, which solves the equations of motion of the missile. The output of the gyro, in turn, is returned to the computer at the point in the mechanization where the control surface deflection is generated. In the simulator, then, the actual hardware is subjected to realistic tests as it is in flight. Tremendous savings in time and money are effected. Multiple-axis tables, loaders, and servojigs all play a role in this program.

The purpose of a trainer is to teach operating personnel in the laboratory how to guide some moving craft, such as an airplane, ship, submarine, or tank. The computer again solves the dynamic equations of motion of the body. Actual hardware is again employed; for example, in the case of a flight trainer, the cockpit, including the instrument panel and the joy stick, is integrated with the computer and the pilot maneuvers the controls as he would in actual flight. Link and Curtiss-Wright trainers for commercial and military aircraft are examples of this application. Take-off, cruise, landing, and emergency conditions are simulated.

Finally, analog computers serve as control devices, in industrial and military applications. Automatic pilots, for example, are analog computers designed to stabilize or guide an aircraft along a prescribed flight path. Analog computers are used in industrial control applications, for example, to control the operation of lathes.

The block diagram of an analog computer system for the training of submarine personnel is shown in Fig. 6.1.

(1) The blocks in the trainer include a table representing a section of the submarine. The table is mounted on gimbals and is free to roll and pitch precisely as does the simulated submarine. Submarine personnel on the table experience these motions and also observe instruments which communicate such information as the forward speed u, the depth z, and the heading ψ of the submarine. On the basis of these signals, the personnel adjust control surface deflections (δ), vary the propeller revolutions per minute (Q), and fill or blow tanks (W), in order to alter the depth.

(2) Also included is an analog computer which receives these signals and solves the three force (drag, side, and lift) equations and the three moment (roll, pitch, and yaw) equations, which determine the orientation of the submarine. The outputs of this

computer are the three components of velocity along the body axes of the submarine, u, v, and w, and the three angular rates of rotations about these axes, p, q, and r. These equations are extremely nonlinear in nature.

(3) Another analog computer accepts the angular and velocity rates and, on the basis of theoretical and experimental data, computes the hydrodynamic forces and moments acting on the submarine, which are, in turn, introduced to the force and moment equations.

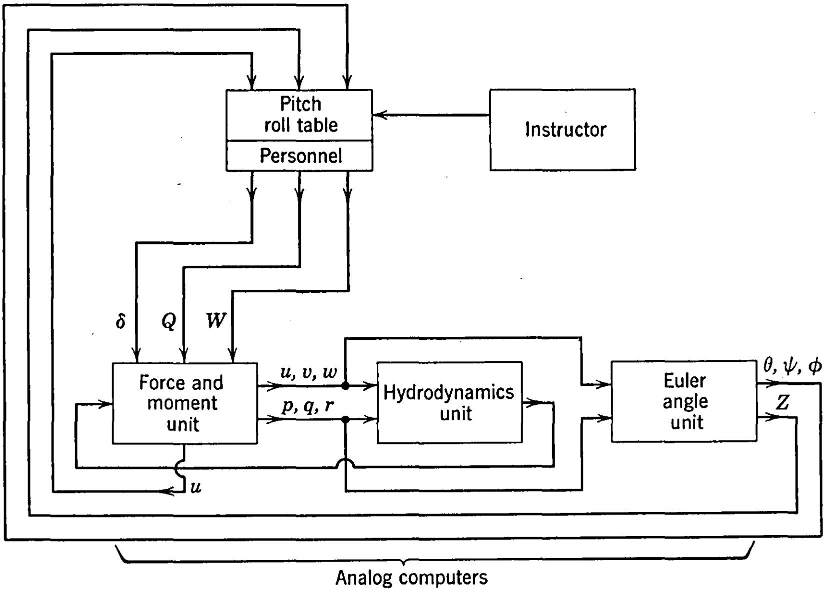


Fig. 6.1 Submarine simulator block diagram

(4) The rates are also taken to another analog computer which calculates the Euler angles θ, ψ, and ϕ, as well as the depth z. These signals are returned to the table to close the loop. The forward speed is also fed back from the first computer.

(5) Included in the simulation is an instructor who observes the progress of the personnel in the operation of the submarine, and who is free to change the operating conditions by calling for the firing of torpedoes or for a change in depth or heading, or by producing "casualties" in some part of the vessel.

The trainer may also play the role of a computer if it is used to study the effect of a change in the submarine design or in the power controls, and as a simulator if the response of control instruments and levers mounted on the table are analyzed and optimized. Automation of the submarine by replacement of some or all of the personnel in the closed loop with automatic controls is another possible application of the trainer. Inasmuch as the operation of a submarine is, in some aspects, a most delicate and sensitive task, the analog computer provides a safe and relatively inexpensive method for testing the feasibility of automation without endangering human lives or incurring the loss of submarines. This is precisely the role which the analog computer has played with outstanding success in the automation of guided missiles and piloted aircraft.

A second means of classifying analog computers is according to whether the analog is direct or functional. In the direct analog, a one-to-one correspondence exists between the given parameters and variables and the computer parameters and variables. The foremost example of the direct analog is the network analyzer, in which a correspondence is established between Kirchhoff's current and voltage laws for a network, and Newton's Equations for a dynamic system. In this type of analog, each component in the computer, i.e., inductor, resistor, or capacitor, is the direct equivalent of a physical parameter, e.g., mass, damper, or spring. After sufficient insight has been gained into this direct equivalence, it becomes possible to proceed from the physical system under study to the appropriate computer mechanization without the necessity of formulating equations to describe the system. This has proved to be a valuable tool, for example, in the study of aircraft flutter, which leads to a great number of coupled linear differential equations. In Section 6.2, we shall see how the direct analog is formulated for a relatively simple vibrating system.

Other more specialized examples of the direct analog are the conducting sheet, the electrolytic tank, the membrane, and the soap bubble, which have found extensive application to the problems in fluid dynamics, electrodynamics, etc., that are describable by Laplace, Poisson, and associated partial differential equations. The potential in a conducting medium and the displacement of a membrane or a soap bubble, under suitable restrictions and boundary conditions, satisfy precisely these equations.

The functional analog, with which we are most concerned here, is predicated on the correspondence between computer components and specific mathematical operations. This is the most flexible and versatile of all analog computers. The heart of the d-c functional computer is the feedback d-c amplifier, which carries out the operations of addition, subtraction, multiplication by a constant, integration, and differentiation. Multiplication and division by variables are performed by a wide variety of devices, the most accurate and successful of which, at the present time, is the servomechanism. Linear potentiometers introduce constants or parameters, and function generators, those terms which vary in some arbitrary manner.

Analog computers may also be classified as to whether they are mechanical, electrical, electromechanical, or electronic. The slide rule is the classical example of the mechanical analog computer. In the electrical group fall the network analyzer, which has been discussed previously, as well as many special-purpose computers such as simultaneous- and secular-equation solvers. The electromechanical analogs include mechanical differential analyzers, fire control computers, and many d-c and a-c computers which contain servomechanisms. A distinguishing component of some of these computers is the mechanical ball-and-disk, or wheel-and-disk, integrator which is a component of high precision and accuracy. Historically, the mechanical differential analyzer is one of the earliest of analog computers. It had its origin in the work of Lord Kelvin about 1876; its application to the solution of differential equations was pioneered by Hartree in England and by Vannevar Bush in this country. The electronic computer is the most recent development in the analog computer field. The functional, commercially available, d-c machine is basically electronic, inasmuch as the feedback amplifier is the essential component.

Analog computers may also be classified according to the speed of operation. There are high-speed, real-time, and extended-time computers. This is a most important grouping and often determines which analog computers are most suitable for particular applications. The real-time computers, as the name denotes, are those in which solutions occur over the same period of time as do the physical phenomena which are being simulated. When analog computers are used for the solution of mathematical equations, the programmer generally has the option of selecting a time scale that is most convenient. He may "extend" the time scale so that one second of real time is converted to 10 seconds of computer time, or he may "contract" the time scale so that one second of real time is converted to one-tenth of a second of computer time. High-speed computers, such as the network and the repetitive analyzers, solve a given problem many times a second. The solution is displayed on a cathode-ray oscilloscope as a stationary curve by synchronizing the scope sweep with the repetitive input. We can vary a parameter on such a computer and observe the effect of the variation immediately on the cathode-ray tube display. The most notable example of the extended-time computer is the mechanical differential analyzer which, in general, must operate slowly if errors due to mechanical inertias are to be minimized. Solution times may be in the order of hours in this case.

Finally, analog computers may be distinguished as either special-purpose or general-purpose instruments. In the first group are those computers which are designed to carry out one specific task, such as harmonic analyzers and autopilots. In the second group fall the functional computers, network and repetitive analyzers, which are capable of being programmed to solve a wide variety of problems.

Analog computers may be evaluated on the basis of the following criteria: reliability, speed of operation, accuracy, flexibility, applicability, and cost. The experience of general-purpose analog computer installations over the past five or ten years has demonstrated that a reliability figure of 95 per cent to 100 per cent can be achieved; i.e., in any given work week the computer operates successfully without breakdown at least 95 per cent of the time. In many weeks, a figure of 100 per cent is realized. Special-purpose computers should have even less difficulty in satisfying high reliability standards.

The distinction among high-speed, real-time, and extended-time computers has been discussed. It should be emphasized that the greatest demand in industry, in military applications, and even in purely mathematical analyses is for real-time machines. The advent of the electronic multiplier (Section 6.3), with an accuracy approaching that of the servomechanism and with a frequency response that is at least two orders of magnitude greater, has brought the goal of an all-electronic computer closer to realization. This is an area in which aspiring engineers can make notable contributions. The nonlinear operation of resolution (Section 6.3) is carried out today primarily by electromechanical means. The development of an electronic resolver would be another significant contribution to the art of analog computation.

It is difficult to make generalizations about the accuracy of various types of analog computers. However, to give the reader some idea of the relation of accuracy to the speed of solution, we cite the following accuracy figures: high-speed computers, 1 to 10 per cent; real-time computers, 0.1 to 1 per cent; extended-time computers, 0.01 to 0.1 per cent.

Note that degree of accuracy appears to vary inversely with speed of operation. With regard to these figures, it should be borne in mind that there are many instances in which the figure of accuracy given here for any one group has been exceeded by at least one order of magnitude, and that skill in programming coupled with a knowledge of the problem and an understanding of the capabilities and limitations of the computer are prerequisites to the realization of the maximum accuracy.

In selecting an analog computer, we must also weigh whatever requirements exist for flexibility, and the need for a general-purpose versus a special-purpose computer, with the cost of the machine. The availability of computing elements, automatic digital and punched-tape read-in and read-out features, and removable problem boards all increase the applicability of the computer at a corresponding increase in cost. The cost of a typical d-c computer containing 20 to 30 amplifiers is in the order of \$15,000. Computing installations with some 200 to 300 amplifiers plus associated nonlinear and recording equipment costing in the neighborhood of \$500,000 to \$1,000,000 are becoming prevalent.

6.2 FUNDAMENTALS OF ANALOG COMPUTATION

Before analyzing the theory of operation of representative analog computer components, we shall describe the fundamental computer operations and the philosophy behind the mechanization of some simple examples.

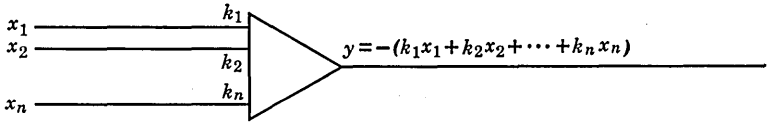


Fig. 6.2 Amplifier block diagram

As stated before, the feedback amplifier, shown symbolically as a triangle in Fig. 6.2, performs the addition of voltages x_1, x_2, $\cdots$, x_n, multiplied, respectively, by constants k_1, k_2, $\cdots$, k_n. The amplifier is comprised of three stages with negative feedback and gives a sign inversion in the output. The result is

$$y = -k_1x_1 - k_2x_2 - \cdots - k_nx_n = -\sum_{i=1}^{n} k_ix_i \qquad (6.1)$$

Each constant k_i, called the gain, represents the ratio of the feedback impedance of the amplifier to the corresponding input impedance. In the case of the summing amplifier the two impedances are resistances. The value of k_i is generally taken to be some integer such as 1, 5, 10, or 50.

The integrating amplifier, Fig. 6.3, is represented by a triangle with a double line. Each input, in this case, is multiplied by a gain k_i; the result is summed, integrated, and inverted in sign to give

$$y = -\int (k_1x_1 + k_2x_2 + \cdots + k_nx_n)\,dt = -\int \left(\sum_{i=1}^{n} k_ix_i \right) dt \qquad (6.2)$$

Electronically, the sole difference between the summing and integrating amplifiers is in the feedback impedance. The capacitor replaces the resistor as the feedback element in the integrator. In most analog computers, switches are provided by means of which an amplifier may be interchangeably a summer or an integrator. In contrast to the integrating amplifier which always integrates with respect to time, the mechanical integrator is distinguished by the fact that it can perform integration with respect to any variable, independent or dependent. In this respect the mechanical differential analyzer is a more versatile instrument than is the electronic computer. If the input impedance to the amplifier is a capacitance and the feedback impedance is a resistance, then the amplifier becomes a differentiator.

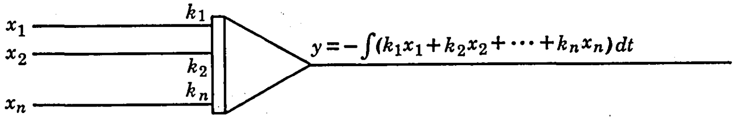


Fig. 6.3 Integrator block diagram

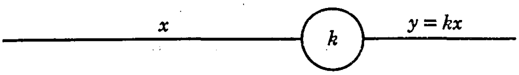


Fig. 6.4 Potentiometer block diagram

Multiplication by a constant k of magnitude less than or equal to unity is performed by a linear potentiometer. Represented symbolically by a circle, the potentiometer accepts the input x, Fig. 6.4, and yields the product

$$y = kx \qquad (6.3)$$

The combination of a potentiometer and an amplifier produces nonintegral gains greater than unity. For example, the relation

$$y = -7.45x \qquad (6.4)$$

is achieved by a potentiometer set at the value of 0.745 in series with an amplifier with an input gain of 10, as shown in Fig. 6.5. A typical analog computer has 30 amplifiers and 40 to 50 potentiometers (Fig. 6.6). Together, the two components suffice to solve linear differential equations with constant coefficients. Problems in servomechanism design, aircraft design, and chemical analysis, for example, are describable, in the first approximation at least, by this class of equations.

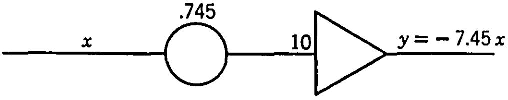


Fig. 6.5 Potentiometer amplifier combination

Passive networks are also employed to produce linear relationships, when expressed in transfer function form, which is a common engineering practice. For example, the simple time delay given by the relation

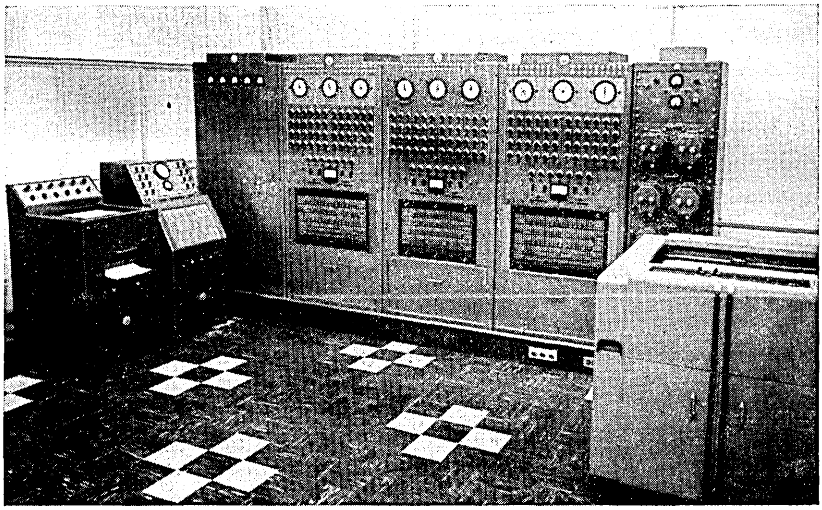


Fig. 6.6 Analog computer installation (Courtesy Dian Laboratories, Inc.)

$$\frac{y}{x} = \frac{1}{1 + RCp} \tag{6.5}$$

is simulated by the RC network of Fig. 6.7. The operator p denotes differentiation with respect to time. High-order lead and lag networks are synthesized by combinations of RC components. In the network analyzer, inductors are employed as well as resistors and capacitors.

The next group of components, servomechanisms (which are representative of the class of multipliers), resolvers, diodes, relays, and function generators enter into the computation picture when the problem contains nonlinearities of some sort. The servomechanism (Section 6.4) is represented by the symbol of Fig. 6.8, where ±100 volts is placed across the follow-up potentiometer and $\pm w$ across a second multiplying potentiometer ganged to it. The output at the arm of the multiplying potentiometer is the product of the input voltage x to the servomechanism and the voltage across the servo pentiometer w divided by the voltage across the follow-up potentiometer, 100, i.e.,

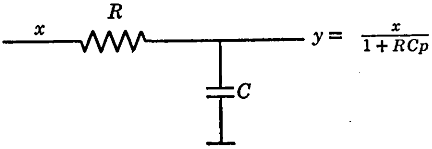


Fig. 6.7 *RC* network for eq. 6.5

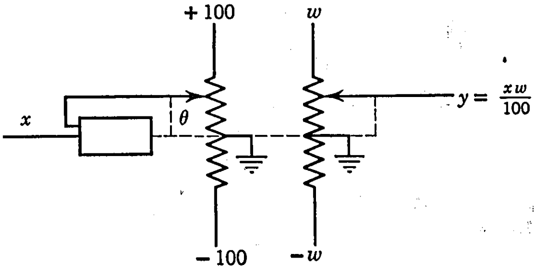


Fig. 6.8 Servomechanism block diagram

$$y = \frac{xw}{100} \tag{6.6}$$

Interchanging the voltages w and $-w$ across the multiplying potentiometer gives

$$y = -\frac{xw}{100} \tag{6.7}$$

The resolver is a special type of nonlinear device which carries out transformations from rectangular to polar coordinates and conversely. Its most general function is shown in block form in Fig. 6.9. The resolver is mounted on a servomechanism whose shaft turns through the angle θ. If A and B, the rectangular coordinates of a point P, are inserted in the resolver and θ is the angle through which the coordinate system is rotated, the two outputs of the resolver

$$\begin{aligned} \bar{A} &= A\cos\theta + B\sin\theta \\ \bar{B} &= -A\sin\theta + B\cos\theta \end{aligned} \tag{6.8}$$

are then the coordinates of the point with respect to the rotated system.

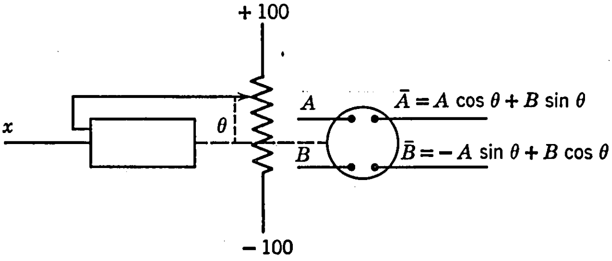


Fig. 6.9 Resolver block diagram

Diodes are employed in many applications to generate straight-line characteristics or linear approximations to curvilinear functions. In the particular arrangement shown in Fig. 6.10, voltages V_1, V_2, $\cdots$, V_n are inserted across the cathodes of successive diodes, each of which conducts when the input x exceeds the corresponding cathode voltage. The output y is composed of a series of straight-line segments, the breakpoints of which are determined by the voltages V_i and the slopes of which are a function of the resistors R_i.

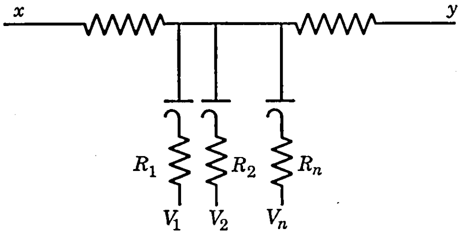


Fig. 6.10 Diode arrangement

The relay amplifier shown in Fig. 6.11 is a "logical" device which controls the operation of one or more single-pole, double-throw relays. Two voltages w and x are introduced at the plus and minus inputs, respectively, to the amplifier; the voltage V_1 is inserted at the normally closed contact and V_2 at the normally open contact of the relay. When $x \leqq w$ the relay is de-energized and the output of the relay arm (A) is V_1. When $x > w$ the relay is energized and the output is V_2. Each of the four signals may be either constant or variable. Such phenomena, characteristic of control systems, as dead space, backlash, hysteresis, and coulomb friction can be simulated by means of relays.

The function generator, shown schematically in Fig. 6.12, represents any analog device for producing the function $f(x)$, given x. It may involve a combination of diodes or relays, as described above, a photoelectric tube–cathode-ray tube unit, a mechanically driven cam, a tapped or nonlinear potentiometer, or a curve follower. The function generator is necessary for the introduction of arbitrary functions, which are customarily synthesized from experimental data. Typical illustrations are the lift and drag curves of an aircraft expressed as functions of Mach number and angle of attack.

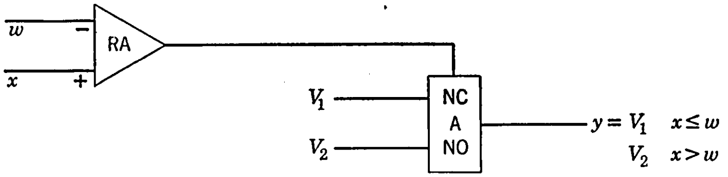


Fig. 6.11 Relay amplifier block diagram

Having reviewed the basic units of the functional computer, let us consider the mechanization of a few simple differential equations, arising from the description of some common physical phenomena. The object of the following exposition is to demonstrate the manner of proceeding from the differential equations to the computer circuit in a straightforward manner. The "patching" of the computer according to the circuit diagram is virtually "one-to-one," i.e., there is one patch cord for every connection on the diagram.

Fig. 6.12 Function generator block diagram

The first equation

$$m\ddot{y} = mg \tag{6.9}$$

describes the trajectory of a falling body subjected to gravity $y = 32.16$ ft/sec^2, where y represents vertical distance. The initial displacement is taken to be $y(0) = -60$, and the initial velocity $\dot{y}(0) = -100$. The recommended procedure is to write each differential equation with the derivative of highest order as the sole left-hand member, and with the remaining terms in the right-hand member. Equation 6.9 is already in this form; division by the common factor m gives

$$\ddot{y} = g \tag{6.10}$$

The constant g is generated (Fig. 6.13) by inserting 100 volts across a potentiometer 1 set at the value $g/100$. Its output is therefore $100(g/100) = g$ which, by virtue of eq. 6.10, is equal to $\ddot{y}$. This term is introduced to integrator 1 whose output is simply $-\dot{y}$, which in turn becomes the input to integrator 2, the output of which is the desired trajectory y. Initial conditions are inserted across each integrator by means of an associated linear potentiometer. A switch places either +100 or −100 volts across potentiometer 1, which is adjusted to give the proper initial value. Inasmuch as the output of integrator 1 is $-\dot{y}$, the correct initial condition is $-\dot{y}(0) = +100$. The initial condition inserted across integrator 2 is $y(0) = -60$. The voltage outputs of integrators 1 and 2 are taken to a recorder which displays the velocity $\dot{y}$ and the displacement y as functions of time. Note that there are three external connections in the mechanization; the computer circuit accordingly requires three cords.

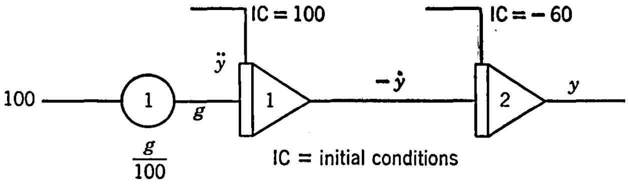


Fig. 6.13 Schematic for solution of $\ddot{y} = g$

If we consider the free vibrations of a mass m suspended from a support by a linear spring k and subject to viscous friction r, the differential equation describing the motion is given by

$$m\ddot{y} + r\dot{y} + ky = 0 \tag{6.11}$$

where y denotes linear displacement, and where the initial coordinates are given by

$$\begin{aligned} y(0) &= A \\ \dot{y}(0) &= B \end{aligned} \tag{6.12}$$

Isolating the highest derivative $\ddot{y}$, and dividing by m, we have

$$\ddot{y} = -\frac{r\dot{y}}{m} - \frac{ky}{m} \tag{6.13}$$

Let us assume, for the moment, that $\ddot{y}$ exists as the input to integrator 1 in Fig. 6.14. This may appear strange since $\ddot{y}$ is patently unknown. The philosophy, however, is to create $\ddot{y}$ from the two terms in the right-hand member of eq. 6.13, which are proportional to successive integrals of $\ddot{y}$, namely $\dot{y}$ and y. We therefore integrate $\ddot{y}$ in integrator 1 to form $-\dot{y}$ which is multiplied by r/m by means of potentiometer 1. The output is returned to integrator 1 since the term $-r\dot{y}/m$ is precisely one of the two contributions to the term $\ddot{y}$. Integrating $-\dot{y}$ in integrator 2 produces y which is converted to $-y$ by means of amplifier 3. This term is multiplied by k/m, potentiometer 2, and is fed back as the term $-ky/m$ to integrator 1. The loop is now closed and the original assumption that $\ddot{y}$ exists as the input to integrator 1 has been validated. The principle of feedback occurs in virtually all analog computer circuits (the first example, Fig. 6.13, is an exception). Thus, feedback characterizes analog computation in two distinct respects: in the design of individual components such as amplifiers and servomechanisms, and in the programming of computer circuits. The initial condition $-B$ is inserted across integrator 1 and A across integrator 2. The damping of the system may be varied by potentiometer 1 and the stiffness by potentiometer 2.

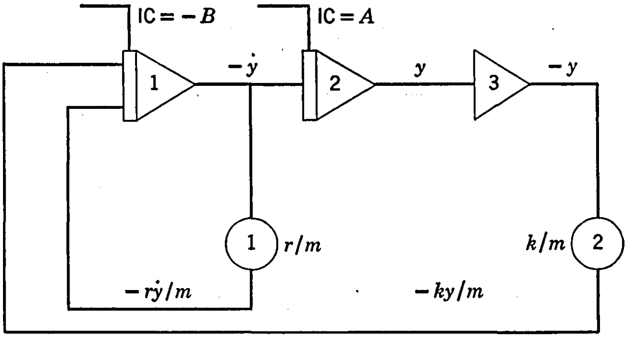


Fig. 6.14 Schematic for solution of $m\ddot{y} + r\dot{y} + ky = 0$

The first two examples have been linear in nature, requiring only amplifiers and linear potentiometers for their solution. As a nonlinear example, consider the trajectory of a falling body subject to nonlinear resistance proportional to the square of the velocity. The equation of motion is

$$m\ddot{y} = mg - c\dot{y}^2 \tag{6.14}$$

The initial conditions are taken equal to zero. We write

$$\ddot{y} = g - c\dot{y}^2/m \tag{6.15}$$

As before, the acceleration $\ddot{y}$ is assumed to be the input to integrator 1 in Fig. 6.15. The gravitational term is generated as in the first example.

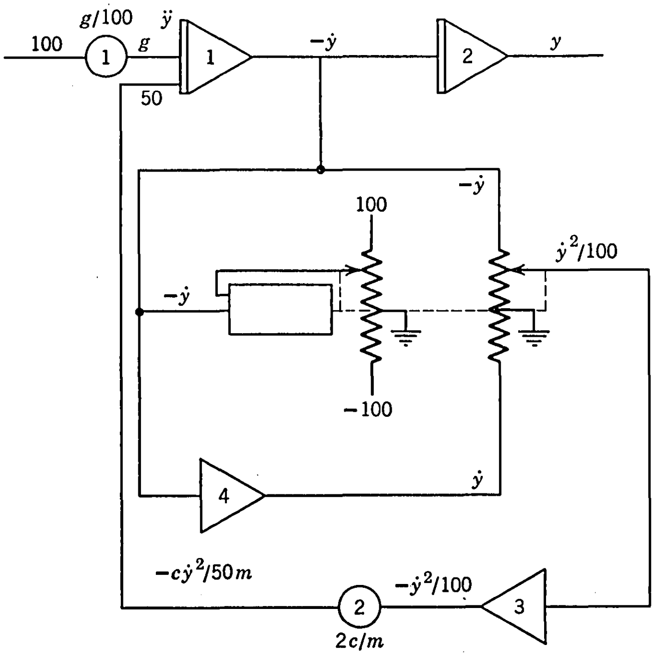


Fig. 6.15 Schematic for solution of $m\ddot{y} = mg - c\dot{y}^2$

The series combination of integrators 1 and 2 produces y. In this example there is a feedback loop due to the presence of the term $-c\dot{y}^2/m$, which is generated by means of a servomechanism. If the servomechanism is driven by $-\dot{y}$ and if $\mp\dot{y}$ are inserted across one of the servo potentiometers, the output of this potentiometer is $\dot{y}^2/100$. Sign inversion by means of amplifier 3 yields $-\dot{y}^2/100$ and multiplication by $2c/m$ (potentiometer 2) gives $-c\dot{y}^2/50m$. Since the contribution to $\ddot{y}$ is $-c\dot{y}^2/m$, this term is introduced to integrator 1 on a gain of 50. This closes the loop.

It is pertinent to observe how the direct analog would treat eq. 6.11 with a forcing function $F(t)$

$$m\ddot{y} + r\dot{y} + ky = F(t) \tag{6.16}$$

Many electrical analogs are possible, two of which are listed in Table 6.1.

Table 6.1

Mechanical Quantity	Analogous Electrical Quantity	
	Kirchhoff's Voltage Law	Kirchhoff's Current Law
Velocity	Current	Voltage
Force	Voltage	Current
Mass	Inductance	Capacitance
Reciprocal spring stiffness (compliance)	Capacitance	Inductance
Damping coefficient	Resistance	Reciprocal resistance (conductance)

In the first case, Kirchhoff's voltage law is used, current being the analog of velocity. Write

$$\dot{y} = v \tag{6.17}$$

Then eq. 6.16 becomes

$$m\dot{v} + rv + k\int v\,dt = F(t) \tag{6.18}$$

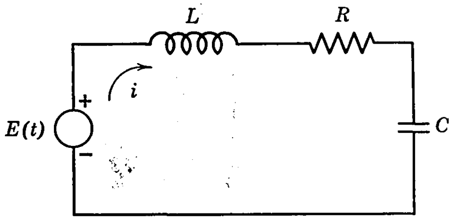


Fig. 6.16 *LRC* circuit for eq. 6.19

Consider the loop equation of Fig. 6.16. Equating the applied voltage to the voltage drops in the loop, we have

$$L\frac{di}{dt} + Ri + \frac{1}{C}\int i\,dt = E(t) \tag{6.19}$$

The correspondence between eqs. 6.18 and 6.19 is clear:

$$v = i, \qquad F(t) = E(t), \qquad m = L, \qquad \frac{1}{k} = C, \qquad r = R \tag{6.20}$$

Thus, for each mechanical parameter, a corresponding electrical parameter is obtained. The vibrations of the mechanical system are studied by observing the oscillations of the electrical system.

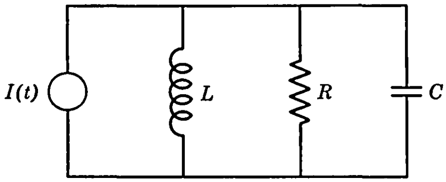


Fig. 6.17 *LRC* circuit for eq. 6.21

Alternately, Kirchhoff's current law may be used, Fig. 6.17, where the analog of velocity is voltage. Equating the applied current $I(t)$ at a node to the currents emanating from the node, we have

$$\frac{1}{L}\int E\,dt + \frac{E}{R} + C\frac{dE}{dt} = I(t) \tag{6.21}$$

The analogy between the mechanical and electrical quantities is

$$v = E, \qquad F(t) = I(t), \qquad m = C, \qquad \frac{1}{k} = L, \qquad r = \frac{1}{R} \tag{6.22}$$

Transformers and amplifiers are introduced in the analyzer in more complicated examples—to achieve physical realizability, for isolation, and for impedance matching.

6.3 COMPUTING EQUIPMENT

6.3.1 The Feedback Amplifier

The feedback d-c computing amplifier consists of a high-gain d-c amplifier with gain $-\mu$, an input impedance Z_i, a feedback impedance Z_f, and an a-c chopper amplifier with gain K, Fig. 6.18. The input is designated as x and the output as y. The first of the three stages of the d-c amplifier is a dual-grid tube with the summing junction of the input and feedback currents at the first grid. The voltage at the grid e_g is magnified by the a-c chopper amplifier and introduced at the second grid of the first stage. It will be shown that the effect of the chopper amplifier is to increase the open-loop gain of the amplifier by a factor K, which makes it a more accurate computing component and which decreases the effect of drift or unbalance, e_d. The grid current, ideally

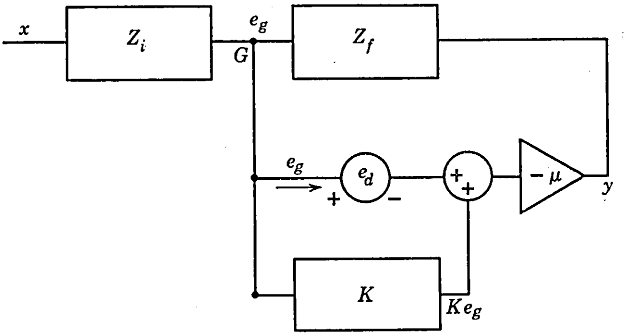


Fig. 6.18 D-c amplifier circuit

equal to zero, is denoted by i_g. Applying Kirchhoff's current law at the node G, we have

$$\frac{x - e_g}{Z_i} = \frac{e_g - y}{Z_f} + i_g \tag{6.23}$$

The output of the amplifier is the product of its gain, $-\mu$, and the sum of the voltages appearing at the first stage, namely the grid voltage e_g, the unbalance e_d, and the chopper output Ke_g; that is

$$y = -\mu(e_g + e_d + Ke_g) \tag{6.24}$$

By eliminating e_g between eqs. 6.23 and 6.24, the result is

$$y = -\frac{Z_f x}{Z_i}\left(\frac{1}{1 + \dfrac{1 + Z_f/Z_i}{\mu(1 + K)}}\right) - \frac{e_d}{1 + K}\left(\frac{1 + Z_f/Z_i}{1 + \dfrac{1 + Z_f/Z_i}{\mu(1 + K)}}\right) + \frac{Z_f i_g}{1 + \dfrac{1 + Z_f/Z_i}{\mu(1 + K)}} \tag{6.25}$$

In the ideal case,

$$y = -Z_f x/Z_i \tag{6.26}$$

That is, the ratio of output voltage to input voltage is the negative of the ratio of output impedance to input impedance. If both Z_f and Z_i are resistances, the amplifier is a summer; if Z_f is a capacitance and Z_i is a resistance, the amplifier is an integrator; if Z_f is a resistance and Z_i is a capacitance, the amplifier is a differentiator. This ratio is the gain k_i of Section 6.2.

Consider the three terms in the right-hand member of eq. 6.25. The first term reduces to the ideal relation if the product $\mu(1 + K)$ is infinite. Typical values are

$$\mu = 3 \times 10^4, \qquad K = 2 \times 10^3, \qquad \mu(1 + K) = 6 \times 10^7$$

The chopper reduces the error term in this expression by the factor $(1 + K)$. For zero unbalance and zero grid current, this term describes the performance of the feedback amplifier. The second term in the right-hand member of eq. 6.25 may be reasonably approximated by the expression $e_d/(1 + K)$. Thus, the chopper reduces the error due to unbalance also by the factor $(1 + K)$. The third term is essentially equal to $Z_f i_g$ and reveals that the error due to grid current is approximately equal to the product of the grid current and the feedback impedance. A typical value for i_g is 100 micromicroamperes. In the case of an integrator, $Z_f = 1/pC$, which shows that the error due to grid current is integrated. Some recent developments in amplifier design are aimed at reducing the error due to grid current also by a factor approximately equal to K. As this analysis demonstrates, one of the most significant advances in the electronic analog computer field since its inception was the invention of the chopper amplifier, for it makes long solution times possible without the deleterious effects of drift.

The functional computer generally has four modes of operation, which are related to the states of the summing amplifiers and integrators. (Differentiators are used rarely since they magnify the effect of noise.)

(1) *Balance.* All inputs to amplifiers and integrators are disconnected. Any initial unbalance is removed so that there is zero output for zero input in each amplifier.

(2) *Reset.* All amplifier inputs are connected, but the integrator inputs remain open. Initial conditions are inserted, and the amplifier outputs assume their proper initial values, but integration does not take place.

(3) *Operate.* All connections are made and the solution is obtained.

(4) *Hold.* All inputs to integrators are disconnected so that the solution is halted. The output of every amplifier may be inspected, and then the solution may be continued by switching once more to the operate state.

6.3.2 Passive Networks

Passive networks consisting of resistors and capacitors are often employed in functional and direct analog computers precisely as they are in automatic control systems—namely to synthesize transfer functions. For example, the simple delay

$$\frac{y}{x} = \frac{a}{b + cp} \tag{6.27}$$

is simulated by the network of Fig. 6.19. Equating the ratio of output voltage y to input voltage x to the ratio of their respective impedances to ground, we have

$$\frac{y}{x} = \frac{(R_1/pC)/(R_1 + 1/pC)}{R + (R_1/pC)/(R_1 + 1/pC)} = \frac{R_1}{R + R_1 + RR_1Cp} = \frac{a}{b + cp} \tag{6.28}$$

with

$$a = R_1, \qquad b = R + R_1, \qquad c = RR_1C \tag{6.29}$$

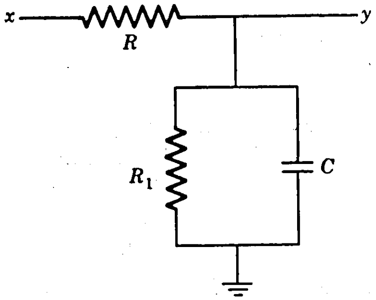


Fig. 6.19 *RC* delay network for eq. 6.27

In Fig. 6.20 is shown a circuit for synthesizing a somewhat more complicated transfer function. The relation

$$\frac{y}{x} = \frac{1/pC}{(1/pC) + \dfrac{R(R_1 + 1/pC_1)}{R + R_1 + 1/pC_1}}$$
$$= \frac{1 + (R + R_1)C_1p}{1 + (RC + R_1C_1 + RC_1)p + RR_1CC_1p^2} = \frac{1 + ap}{1 + bp + cp^2} \tag{6.30}$$

holds, with

$$a = (R + R_1)C_1, \qquad b = RC + R_1C_1 + RC_1, \qquad c = RR_1CC_1 \tag{6.31}$$

This method effects a savings in equipment, but it may suffer from inaccuracies due to the loading of the circuit by the impedance of the element that follows it, e.g., the input resistance of an amplifier. The d-c feedback amplifier, whose output impedance is virtually zero, may

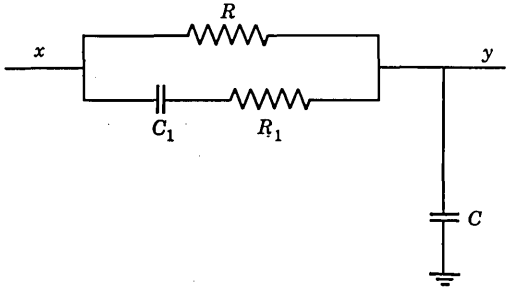


Fig. 6.20 *RC* network for synthesizing the transfer function of eq. 6.30

serve to generate complex transfer functions if we recognize that in the ideal expression

$$\frac{y}{x} = -\frac{Z_f}{Z_i} \tag{6.26}$$

the impedances Z_f and Z_i need not be simple resistances or capacitances, but may indeed be *RC* networks. In this role, the amplifier is called an operational amplifier. In Fig. 6.21 the input impedance consists of

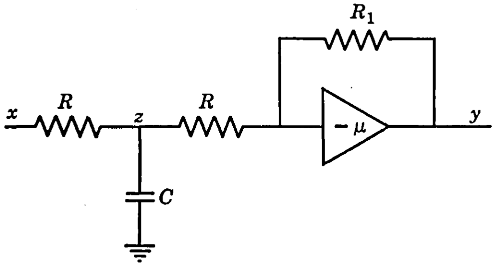


Fig. 6.21 Operational amplifier with transfer function specified by eq. 6.34

a combination of two resistances, each of value R, and capacitance C. The feedback impedance is R_1. Recalling that the grid is effectively at zero potential, we write

$$\frac{z}{x} = \frac{\dfrac{R/pC}{R + 1/pC}}{R + \dfrac{R/pC}{R + 1/pC}} = \frac{1}{2 + RCp} \tag{6.32}$$

and

$$\frac{y}{z} = -\frac{R_1}{R} \tag{6.33}$$

or

$$\frac{y}{x} = \frac{y}{z} \cdot \frac{z}{x} = -\frac{R_1}{R(2 + RCp)} = -\frac{a}{b + cp} \tag{6.34}$$

This expression is similar to eq. 6.28 (within a minus sign) with

$$\begin{aligned} a &= R_1 \\ b &= 2R \\ c &= R^2C \end{aligned} \tag{6.35}$$

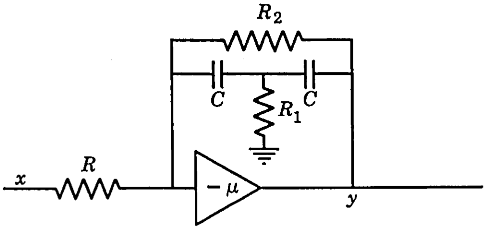


Fig. 6.22 Operational amplifier with transfer function specified by eq. 6.36

Figure 6.22 shows a second operational amplifier with the feedback impedance a more complicated network in this case. The relation of output to input is given by (see eq. 6.30)

$$\frac{y}{x} = -\frac{R_2}{R} \frac{1 + 2R_1Cp}{1 + 2R_1Cp + R_1R_2C^2p^2} = -A \frac{1 + ap}{1 + bp + cp^2} \tag{6.36}$$

with

$$\begin{aligned} A &= R_2/R \\ a &= 2R_1C \\ b &= 2R_1C \\ c &= R_1R_2C^2 \end{aligned} \tag{6.37}$$

Convenient tables exist for selecting the proper *RC* networks for a given transfer function by either of the methods above. In Section 6.6 we shall treat a more flexible method, although a less economical method equipment-wise, for generating transfer functions, using conventional d-c feedback amplifiers.

6.3.3 Linear Potentiometer

The linear potentiometer is shown symbolically in Fig. 6.23 where x is the input, y the output, R the potentiometer resistance, a the fraction of the total resistance contacted by the wiper or arm, and R_1 the load resistance. We derive

$$y = \frac{aRR_1/(aR + R_1)}{aRR_1/(aR + R_1) + (1 - a)R}\,x = \frac{ax}{1 + (R/R_1)a(1 - a)} \tag{6.38}$$

The desired output is $y = ax$; the actual output is somewhat less, owing to the nonzero value of the ratio R/R_1. Loading curves are available to give the correct setting. In most cases, the potentiometers are set manually (or automatically by means of a digital keyboard) by inserting 100 volts across it and observing the output on a servomechanism. The potentiometer is adjusted (with the loading in effect) until the actual output is the desired fraction of the input. The final correct setting is then generally larger than the value a.

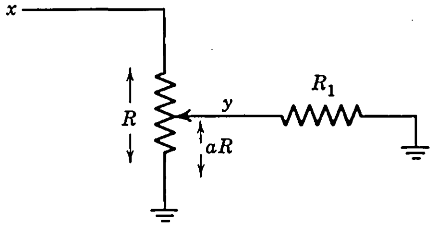


Fig. 6.23 Linear potentiometer circuit

6.3.4 Multiplier

The computing servomechanism is essentially a device for converting an electrical voltage into a mechanical shaft rotation. Its principal function is to perform the multiplication of variable quantities, as shown in Fig. 6.24, where x is the input voltage, θ the shaft rotation, $\pm z$ the voltages across the follow-up potentiometer, and $\pm w$ the volt-

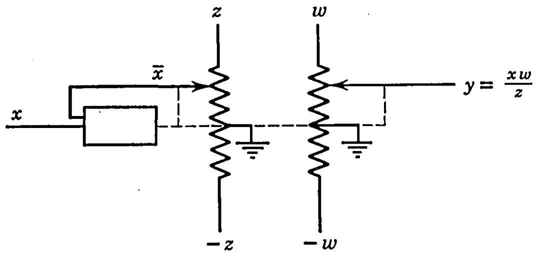


Fig. 6.24 Servomechanism circuit

ages across the multiplying potentiometer. The follow-up and the multiplying potentiometers are ganged together mechanically, so that the arms of both potentiometers are displaced through the same angle θ. It follows then that

$$\frac{\bar{x}}{z} = \frac{y}{w} \tag{6.39}$$

or

$$y = \frac{w\bar{x}}{z} \tag{6.40}$$

where $\bar{x}$ and y are the voltages at the arms of the follow-up potentiometer and multiplying potentiometer respectively. As described in Chapter 4, the servomechanism moves in such a manner as to null continuously the difference between the input and the follow-up signals (if the velocity and acceleration limits of the servomechanism are not exceeded), i.e., we assume

$$x = \bar{x} \tag{6.41}$$

combining eqs. 6.40 and 6.41 gives

$$y = \frac{wx}{z} \tag{6.42}$$

It is customary to set $z = 100$. The servomechanism loses its effectiveness when the signal has a high-frequency content, for in this case the mechanical shaft motion tends to lag behind its correct value, invalidating eq. 6.41. Products of x with other variables may be formed by ganging additional potentiometers on the shaft.

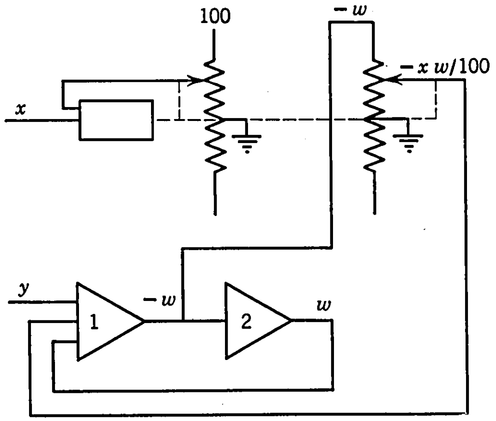


Fig. 6.25 Division circuit using potentiometers

Division is accomplished by placing the multiplying potentiometer in the feedback path of an amplifier circuit. One scheme is illustrated in Fig. 6.25. Let the quotient

$$Q = \frac{y}{x} \tag{6.43}$$

be sought where x is positive. The output of amplifier 1, designated as $-w$, is inserted across the multiplying potentiometer, the output of which, $-xw/100$, is returned to amplifier 1. The signal $-w$ is inverted in amplifier 2 and the output w is also fed back to amplifier 1. The equation satisfied at the input to this amplifier is

$$w = \frac{-xw}{100} + w + y \tag{6.44}$$

or

$$w = \frac{100y}{x} \tag{6.45}$$

which is proportional to the desired quotient. In Section 6.5 a division scheme which makes use of the implicit-function technique and which involves only a single high-gain amplifier is described.

Time division. The first of the electronic schemes for multiplication of two variables to be described bears the general name of time division. A symmetrical square wave is generated with the duration of each pulse in the wave train proportional to one of the variables and the amplitude proportional to the other. Filtering yields the average value of the square wave which is proportional to the product of the two signals.

Amplitude and frequency modulation. A second, more recent, development in the search for an electronic multiplier is one based on the amplitude and frequency modulation of a high-frequency carrier signal, Fig. 6.26. Frequency modulation of the carrier produces a frequency deviation which is proportional to the variable x. The amplitude of the resulting FM signal is then set according to a constant voltage w (which plays the role of the follow-up voltage in the servomechanism). A discriminator-detector extracts a signal $a_w - b_w$ proportional to the FM content which is fed back to match the input x, and a signal $a_w + b_w$ proportional to the AM content which is fed back to match w. The FM feedback assures that the frequency deviation in the carrier is truly proportional to x. The AM feedback serves to insure that the FM signal is of constant amplitude, namely w, inasmuch as frequency modulation can introduce some undesired amplitude modulation. The AM feedback loop, or automatic-gain circuit, has its servo analog in the voltage regulation necessary to keep the voltage across the follow-up potentiometer constant. The FM signal is now fed to an amplitude modulator, the second input to which is the sum $z + w$. The output of this modulator is a signal whose frequency deviation is proportional to x, and whose amplitude is proportional to $z + w$. The AM output of the second discriminator-detector $a_y + b_y$ is fed back to match $z + w$. The second output is designated as $a_y - b_y$. The following relations hold, in which k is a constant, and Δ_f frequency deviation.

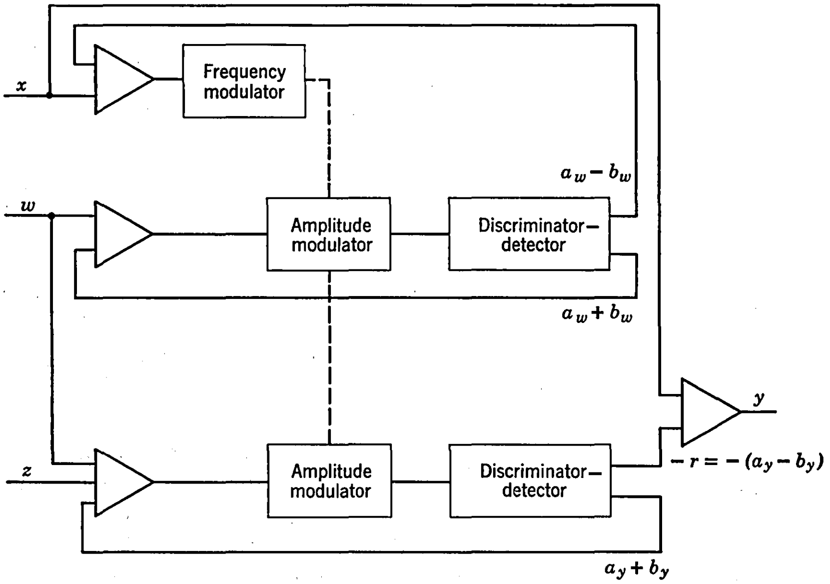


Fig. 6.26 AM-FM multiplier block diagram

$$a_w - b_w = k(w)(\Delta f) \tag{6.46}$$

$$a_y - b_y = k(w + z)(\Delta f) \tag{6.47}$$

$$a_w - b_w = x \tag{6.48}$$

$$a_y - b_y = r \tag{6.49}$$

From eqs. 6.46–6.49

$$\frac{w}{w + z} = \frac{a_w - b_w}{a_y - b_y} = \frac{x}{r} \tag{6.50}$$

or

$$r = \frac{xz + xw}{w} = \frac{xz}{w} + x \tag{6.51}$$

In practice the quantity $-r$ is formed and x is added to it. The product xz/w is produced at the output of the summing amplifier.

Several additional multiplication schemes, which demonstrate the diversity in the approach to this problem, are described briefly. The "quarter-square" method is based on the equation

$$(a + b)^2 - (a - b)^2 = 4ab \tag{6.52}$$

Given the quantities a and b, the sum and difference of the two can be formed; these are then squared and subtracted to yield the desired product of a and b. Thus, the burden of multiplication is thrown on the generation of squares. Electronically, this has been carried out by means of: vacuum tubes in which a parabolic relation exists between plate current and grid voltage, diodes arranged to give a straight-line approximation to a parabola, and more recently by means of thyrite and the photoelectric function generator. The success of these methods depends strongly on producing identical square-law curves.

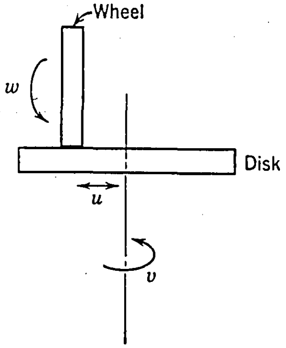


Fig. 6.27 Wheel-and-disk integrator

Another method makes use of the logarithmic relations

$$\log xy = \log x + \log y \tag{6.53}$$

and

$$xy = \operatorname{antilog} (\log xy) \tag{6.54}$$

That is, given x and y, the logarithms of the two quantities can be generated, the results summed, and then the antilogarithm formed. Any of the standard function generators may be applied to this end.

On mechanical differential analyzers, where integration is carried out with respect to any variable, multiplication is reduced to a problem in integration by virtue of the formula for integration by parts

$$xy = \int x\,dy + \int y\,dx \tag{6.55}$$

The mechanical wheel-and-disk integrator is shown in Fig. 6.27. If the wheel is driven by the variable w, the disk by v, and if the wheel is displaced from the center of the disk through a distance u, then equating corresponding elements of arc on disk and wheel gives

$$u\,\Delta v = r\,\Delta w \tag{6.56}$$

where r is the radius of the wheel. Passing to the limit and setting $r = 1$ for convenience, we have

$$w = \int u\,dv \tag{6.57}$$

To form the product xy according to eq. 6.55 therefore requires two mechanical integrators.

Dual amplitude modulation is another electronic method of multiplication, finding particular application in high-speed computers, Fig. 6.28. A carrier, sin ωt, is first amplitude-modulated according to x to give x sin ωt. The result is then amplitude-modulated according to y to give xy sin ωt. Demodulation of this signal yields the product xy.

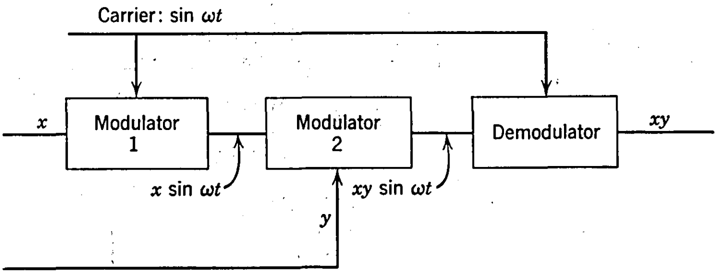


Fig. 6.28 Dual amplitude modulation

6.3.5 Resolvers

Resolvers are generally capable of three modes of operation:

(1) *Polar to rectangular transformation.* Given polar coordinates r and θ, to generate the rectangular coordinates

$$\begin{aligned} A &= r\cos\theta \\ B &= r\sin\theta \end{aligned} \tag{6.58}$$

(2) *Rectangular to polar transformation.* Given rectangular coordinates x and y, to generate the polar coordinates

$$\begin{aligned} r &= \sqrt{A^2 + B^2} \\ \theta &= \tan^{-1} B/A \end{aligned} \tag{6.59}$$

(3) *Rotation of axes.* Given the rectangular coordinates A and B, of a point with respect to one set of orthogonal axes x, y, to find the coordinates of the point with respect to a second set of orthogonal axes which are rotated through the angle θ with respect to the first set

$$\begin{aligned} \bar{A} &= A\cos\theta + B\sin\theta \\ \bar{B} &= -A\sin\theta + B\cos\theta \end{aligned} \tag{6.60}$$

The most popular and accurate method for performing resolution on an a-c computer is the a-c induction resolver, mounted on a servomechanism, in which resolution is achieved by magnetic coupling between two stators and two rotors. Voltages proportional to A and B are inserted across the two stators which are perpendicular to one another. Voltages proportional to $\bar{A}$ and $\bar{B}$ (eqs. 6.60) are then induced in the two rotors which are perpendicular to one another and are free to rotate through the angle θ. If the voltage B is equated to zero and the voltage A is equated to r, then from eqs. 6.60, we have

$$\begin{aligned} \bar{A} &= r\cos\theta \\ \bar{B} &= -r\sin\theta \end{aligned} \tag{6.61}$$

which are analogous to eqs. 6.58. As before the conversion from rectangular to polar coordinates is effected by feeding A and B into the stator coils. In this case, however, the signal B is fed back as the input to the servomechanism in place of the customary feedback from the follow-up potentiometer. Since the servomechanism drives to a null, we have

$$-A\sin\theta + B\cos\theta = 0 \tag{6.62}$$

or

$$\theta = \tan^{-1} B/A \tag{6.63}$$

and the second output of the resolver becomes

$$r = A\left(\frac{A}{\sqrt{A^2+B^2}}\right) + B\left(\frac{B}{\sqrt{A^2+B^2}}\right) = \sqrt{A^2+B^2} \tag{6.64}$$

The use of the a-c resolver with the functional d-c computer necessitates modulation and demodulation to make the two compatible. Lately, d-c resolution has been accomplished by means of a tapered wire-wound potentiometer, as in Fig. 6.29. If $\pm B$ are inserted across the first potentiometer, its outputs are $B\sin\theta$ and $B\cos\theta$. The arms of the second potentiometer, mechanically ganged to the first, are displaced by 90 degrees from them. Accordingly, if $\pm A$ are inserted across it, the outputs are $A \cos \theta$ and $-A \sin \theta$. Amplifiers are required to form the necessary additions for realizing the functions in eqs. 6.60.

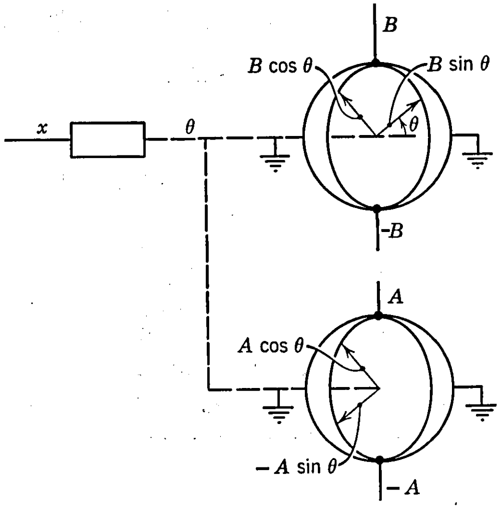


Fig. 6.29 Resolver circuit

6.3.6 Function Generators

The term function generator describes an extremely broad class of analog computer devices whose purpose it is to introduce arbitrary functions into the mechanization. These functions may, for example, be simply sinusoidal or exponential in nature; they may represent the nonlinear characteristic of some device, such as backlash in a gear train or dead space in a servomechanism, or they may simulate the lift and drag curves of a missile in terms of its speed or angle of attack. Among the types of function generators are: (1) electronic diodes; (2) electromechanical relays; (3) nonlinear, tapped, and loaded potentiometers mounted on servomechanisms; (4) photoelectric followers; (5) input tables; (6) two- and three-dimensional cams.

Where speed of response is of the essence, e.g., in real-time operation, diodes and the photoelectric follower prove applicable. The electromechanical relay, which is driven by a relay amplifier and has a closure time in the order of milliseconds, may also serve the purpose. A simple diode configuration for generating a function consisting of two straight-line segments is shown in Fig. 6.30*a*. The diode is placed in the feedback path of a d-c amplifier whose input resistance is R_i and whose feedback resistance is R_f. A battery is placed in series with the diode and the resistor R_1, such that the voltage presented to the cathode of the diode is equal to $y + e_c$, where y is the voltage output of the amplifier and e_c is the battery voltage. When y exceeds $-e_c$,

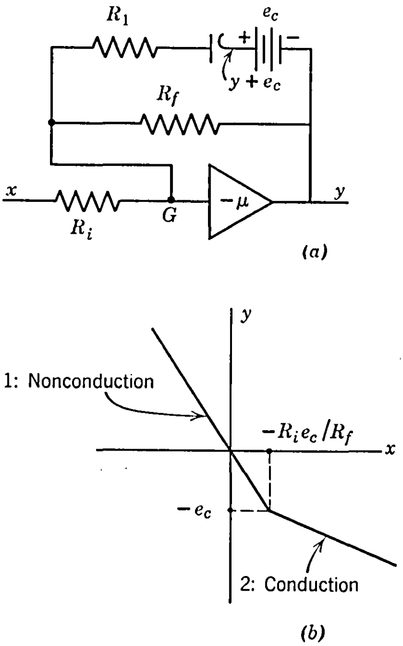


Fig. 6.30 Diode circuit

the voltage $y + e_c$ is positive; the diode does not conduct and the output is given simply by

$$y = -\frac{R_f}{R_i}x \tag{6.65}$$

which is the equation of segment 1 in Fig. 6.30*b*. When y decreases below $-e_c$, the diode conducts, and we have by summation of currents at the grid G

$$\frac{x}{R_i} = -\frac{y}{R_f} - \frac{(y + e_c)}{R_1} = -y\left(\frac{R_1 + R_f}{R_1R_f}\right) - \frac{e_c}{R_1} \tag{6.66}$$

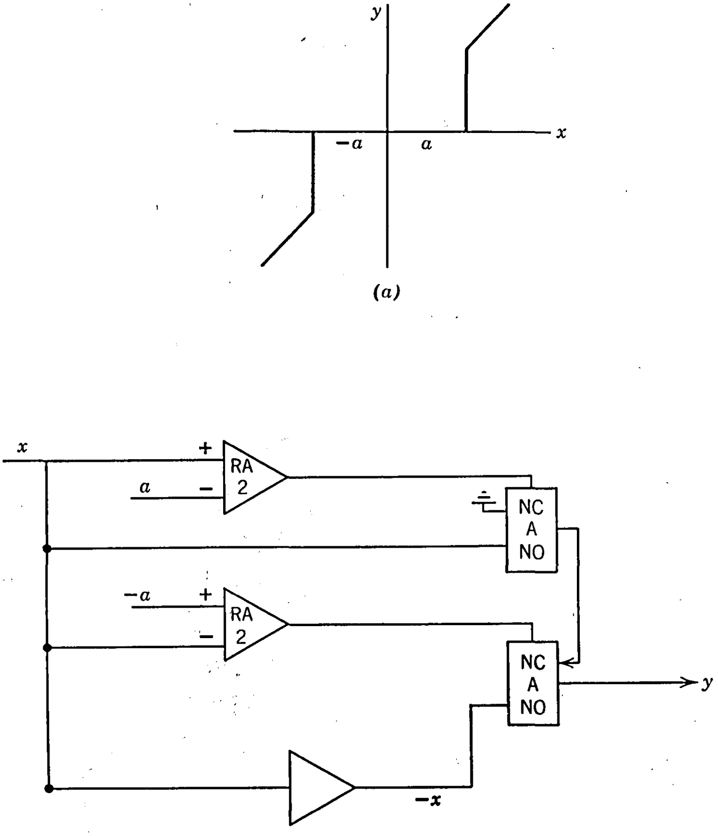
or

$$y = -\frac{R_1R_f}{R_i(R_1 + R_f)}x - \frac{R_f}{R_1 + R_f}e_c \tag{6.67}$$

which is the equation of segment 2. There are many possible arrangements of diodes, in series and in parallel, in the input and feedback paths of the amplifier for simulating various characteristics.

A simple relay circuit for simulating the dead-space characteristic of Fig. 6.31a is shown in Fig. 6.31b. The equations that apply are

$$\begin{array}{ll} x > a & y = x \\ -a < x < a & y = 0 \\ x < -a & y = -x \end{array} \tag{6.68}$$



(a)

(b)

Fig. 6.31 Relay amplifier circuit

Recalling that the relay amplifier energizes its associated relay (normally open position) when the plus input exceeds the minus input, it follows that relay 1 is energized when $x > a$ and relay 2 is energized when $x < -a$. The voltage x is also inserted in the normally open contact of the first relay and $-x$ is introduced in the normally open contact of the second relay. The normally closed contact of relay 1 is grounded and its arm (A) is connected to the normally closed contact of relay 2, the arm of which yields the desired output. The output of the first relay is then either x or 0, and that of the second is x, 0, or $-x$, as summarized in Table 6.2.

Table 6.2

State	Relay 1	Output	Relay 2	Output
$x > a$	NO	x	NC	x
$-a < x < a$	NC	0	NC	0
$x < -a$	NC	0	NO	$-x$

The photoelectric curve follower, Fig. 6.32, consists of a cathode-ray tube, a mask, and a phototube. The given curve is shaded black below its boundary and is interposed between the cathode-ray tube and the phototube. The horizontal deflection of the cathode-ray tube beam is controlled by the x input, and the vertical deflection by the phototube and the y amplifier. The circuit constants are selected so that the correct amount of light is seen by the photocell when the spot is approximately half shielded by the shaded portion of the mask. If the spot becomes more exposed, the additional light produces a voltage which is reflected to the vertical deflection plates of the cathode-ray tube and which drives the spot down behind the mask until the light intensity is again correct. Conversely, if the photocell sees insufficient light, the spot is driven in an upward direction. Thus, as the beam is deflected in the x direction, the spot is forced continually to ride on the edge of the mask. The voltage controlling the vertical deflection is $y(x)$.

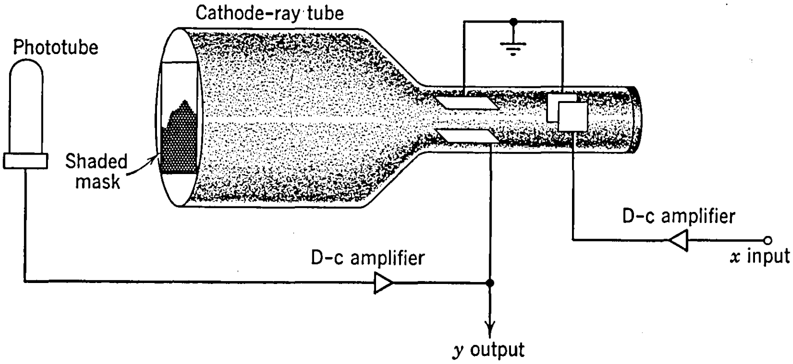


Fig. 6.32 Photoelectric function generator block diagram

6.3.7 Recorders

The results of an analog computer analysis are continuously available, either for visual observation or for a permanent record. The choice of a suitable recorder depends on several considerations: (1) the required speed of response, (2) the required accuracy, and (3) whether the computer outputs which are to be plotted are restricted to being functions of time or whether they are more generally functions of some dependent variable. High-speed and repetitive computers display the results on a cathode-ray tube which is pulsed in synchronism with the initiation of the solution so that the results appear stationary to the viewer. The effect of changing a parameter, by means of a linear potentiometer, say, is immediately apparent to the observer. Permanent records may be obtained by using a camera in conjunction with the cathode-ray tube. Optical recorders are employed for functions which contain frequencies as high as 15,000 cycles per second, the high-frequency response being achieved by focusing a beam of light on moving film. Limited accuracy is the price paid for the high-frequency response of these recorders.

In general-purpose analog computer installations, multichannel direct-writing recorders are used for plotting functions of time with a frequency response up to 100 cycles per second. Where higher accuracy is desired, the plotting board is employed. It consists of a servo-driven carriage, moving in one direction, on which rides a pen that is servo-driven in a perpendicular direction. Accuracies of 0.1 per cent are realizable on plotting surfaces of 30 inches by 30 inches. The frequency response is limited by the velocity and acceleration capabilities of the servomechanisms. Families of curves plotting any one variable versus any other with a third function as parameter are conveniently obtained on the plotting board.

Magnetic tape recorders are finding increased application in analog computer installations, particularly in those problems where a "memory" is required. Analog-to-digital converters are introduced to furnish accurate static measurements. Their utility as dynamic recorders is limited, however, because of the restricted speed with which these results may be recorded.

6.4 AN AUTOPILOT PROBLEM

The mechanization of some simple differential equations was developed in Section 6.2. We consider here the analysis of a typical autopilot system which is described by the following set of equations:

$$\ddot{\theta} = -0.200\dot{\theta} - 0.053\dot{\alpha} - 4.32\alpha - 15.2\delta_c \tag{6.69}$$

$$\dot{\gamma} = 0.812\alpha + 0.364\delta_e \tag{6.70}$$

$$\theta = \alpha + \gamma \tag{6.71}$$

$$\delta_e = a\dot{\theta}/(1 + T_p), \qquad p = \frac{d}{dt} \tag{6.72}$$

where θ = pitch angle of the aircraft
γ = flight path angle
α = angle of attack
δ_e = elevator deflection
T = time lag of the elevator servomechanism
a = proportionality constant, taken as a parameter

The first equation expresses the fact that the pitching moment to which the aircraft is subjected is proportional to the pitch rate $\dot{\theta}$, the rate of change of angle of attack $\dot{\alpha}$, the angle of attack α, and the elevator deflection δ_e. The second equation describes the lift force exerted on the aircraft, the resulting angular rate $\dot{\gamma}$ being proportional to α and δ_e. The third equation defines the pitch angle θ to be the sum of α and γ. The last equation describes the behavior of an autopilot which is mechanized so that the elevator deflection is made proportional to the pitch rate $\dot{\theta}$ within the time delay T. This is an example of automation in which the function of the autopilot is to minimize the pitching motion of the aircraft. The purpose of such analysis might be (1) to determine whether the autopilot carries out successfully the function for which it is designed; (2) to determine the optimum value of the proportionality constant a; and (3) to find the minimum value of the time constant T necessary to insure rapid system response.

The first two differential eqs. 6.69–6.70 are already in a form for solution on the computer, i.e., with the highest-order derivative terms isolated in the left-hand members. The last equation written in the form of a transfer function is replaced by the differential equation

$$\dot{\delta}_e = -5\delta_e + 5a\dot{\theta} \tag{6.73}$$

where we have set $T = 0.2$.

The pitch acceleration $\ddot{\theta}$ is formed at the input to integrator 1, Fig. 6.33, the output of which is $-\dot{\theta}$. After being multiplied by the factor 0.200, potentiometer 1, this term is fed back to integrator 1. The term $-\dot{\gamma}$ is formed at the input to amplifier 5 and is then integrated in 3 to give $-\gamma$. The result of combining θ and $-\gamma$ in amplifier 6 is $-\alpha$, by virtue of eq. 6.71. This term is multiplied by the factor 4.32/5, potentiometer 2, and is fed back to integrator 1 on a gain of 5. This is the term -4.32α in eq. 6.69. The term $-\alpha$ is also multiplied by the factor 0.812, potentiometer 7, and is fed back to 5 to contribute to the summation, $-\dot{\gamma}$. The outputs of amplifier 5, $\dot{\gamma}$, and of integrator 1, $-\dot{\theta}$, are combined in amplifier 7 to give $\dot{\alpha}$, which is inverted in amplifier 8. The result is multiplied by the factor 0.053, potentiometer 3, and the output is returned to integrator 1. In this manner, we have mechanized all of the terms in the first three equations other than the contributions of the control surface deflection δ_e. In accordance with eq. 6.73, we multiply the output of integrator 1, $-\dot{\theta}$, by the factor 5 in amplifier 9 to give $5\dot{\theta}$. Setting the input to integrator 4 equal to $\dot{\delta}_e$ and its output to $-\delta_e$, we form the terms $5a\dot{\theta}$, potentiometer 4, and

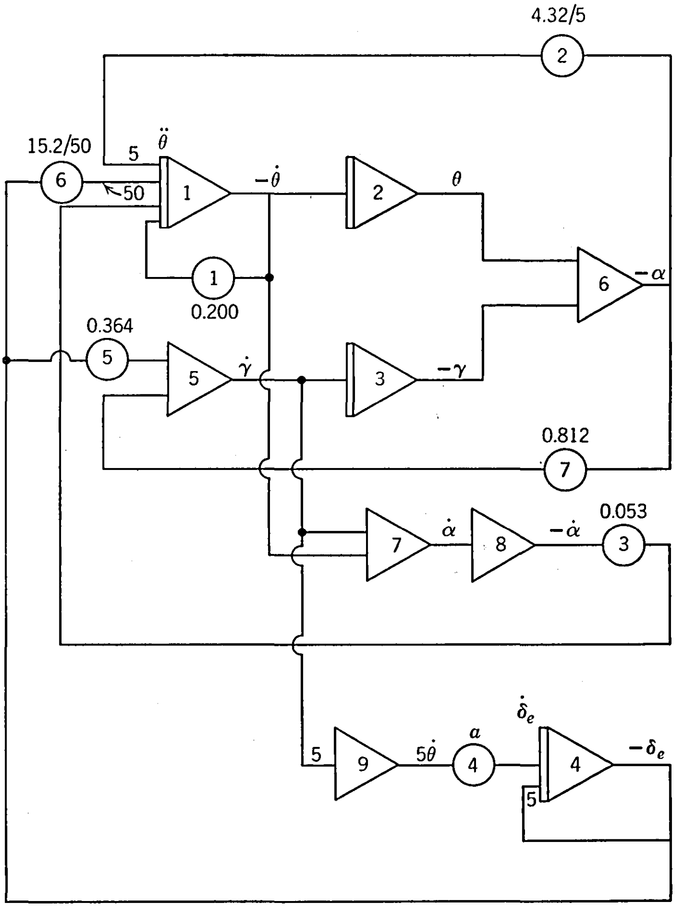


Fig. 6.33 Autopilot schematic

$-5\delta_e$ by means of the input gain of 5 on integrator 4. The term $-\delta_e$ is multiplied by the factor 15.2/50, potentiometer 6, and is returned to integrator 1 on a gain of 50. It is also multiplied by the factor 0.364, potentiometer 5, and returned to amplifier 5. This completes the mechanization.

This linearized analysis of the pitch stability of an aircraft is a vast simplification of the fundamental and more general problem which concerns itself with the intercoupled six degrees of freedom of a moving body (in contrast to the two included here): pitch, roll, and yaw moments, lift, drag, and side forces. If we add to these equations the relations expressing the dynamic behavior of the radar, power plant, and guidance systems, we arrive at the type of problem which involves several hundred amplifiers, some 25 to 50 multipliers, some 20 to 30 function generators, plus resolvers, recording equipment, etc. Many analog computer installations are concerned with the analysis of this type of system.

6.5 SELECTED TECHNIQUES

The application of the analog computer to problems in such fields as partial differential equations, polynomial equations, the secular equation, the potential equation, network synthesis, the analysis of noise and random variables, and flutter involves highly specialized techniques. We treat here a few programming techniques which are illustrative of the versatility of analog computers and which are more directly applicable to problems in automation.

6.5.1 Transfer Functions

In Section 6.3, two different methods for synthesizing transfer functions were discussed: (1) by means of *RC* networks, and (2) by means of operation amplifiers in which *RC* networks are inserted in the input and/or feedback paths of a feedback amplifier. We now investigate a third method which involves standard amplifiers. Although this technique entails more expensive equipment than do the first two, the mechanization which results is the most convenient one to use when any of the coefficients in the transfer function is to be treated as a parameter which is to be varied from run to run in order to find its optimum value. Consider, for example, the transfer function

$$\frac{V}{E} = \frac{a_0 + a_1 p + a_2 p^2}{b_0 + b_1 p + b_2 p^2 + p^3}, \qquad p = \frac{d}{dt} \tag{6.74}$$

Equation 6.74 may be replaced by the equivalent set of two equations

$$\bar{V} = \frac{E}{b_0 + b_1 p + b_2 p^2 + p^3} \tag{6.75}$$

or

$$\dddot{\bar{V}} + b_2\ddot{\bar{V}} + b_1\dot{\bar{V}} + b_0\bar{V} = E \tag{6.76}$$

and

$$V = a_0\bar{V} + a_1\dot{\bar{V}} + a_2\ddot{\bar{V}} \tag{6.77}$$

Assuming that E is a given input, eq. 6.76 is solved for the highest derivative

$$\dddot{\bar{V}} = -b_2\ddot{\bar{V}} - b_1\dot{\bar{V}} - b_0\bar{V} + E \tag{6.78}$$

as shown in Fig. 6.34. The term $\dddot{\bar{V}}$ is formed at the input to integrator 1, and successive integrations yield $\ddot{\bar{V}}$, $\dot{\bar{V}}$, and $\bar{V}$, which constitute the feedback terms to integrator 1. Having generated the lower-order derivatives of $\bar{V}$, we obtain V by forming the appropriate linear combination of these three terms in accordance with eq. 6.77. The mechanization makes use of three integrating and two summing amplifiers. Note that each coefficient in the given transfer function may be varied by simply changing the setting of one potentiometer.

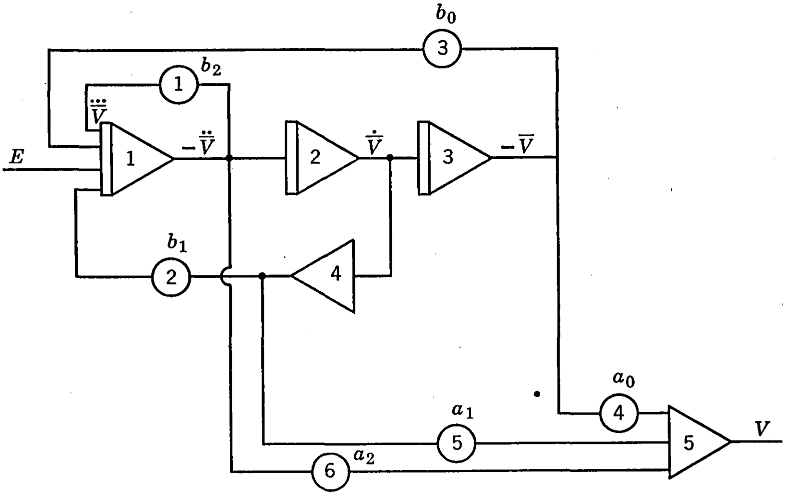


Fig. 6.34 Transfer function schematic

6.5.2 Implicit-Function Technique

In the discussion of the multiplying servomechanism (Section 6.3) a method for performing division is described that requires two am-

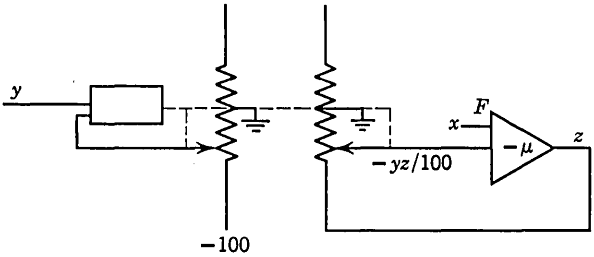


Fig. 6.35 Division circuit with a high-gain amplifier

plifiers in a feedback arrangement. Figure 6.35 shows a similar mechanization which involves a single high-gain amplifier, i.e., an amplifier in which the customary resistance feedback is removed. We write the desired quotient expressed explicitly,

$$z = 100x/y, \quad y \text{ negative} \tag{6.79}$$

as the implicit relation

$$F(x, y, z) \equiv x - \frac{yz}{100} = 0 \tag{6.80}$$

This is replaced on the computer by the approximate relation

$$F(x, y, z) \cong -\frac{z}{\mu} \tag{6.81}$$

Equation 6.81 is mechanized by inserting F as the input to a high-gain amplifier (with gain $-\mu$) and denoting its output as $-z$. Equation 6.81 is a reasonable approximation to eq. 6.80, if we recognize that the value of μ at d-c is in the order of 10^7. The choice of the negative sign in the right-hand member of eq. 6.81 is necessary to insure stability and is dictated by the fact that y is assumed to be negative in the present example.

The implicit-function technique is a powerful tool for generating functions by the expedient of forming their inverses instead, an operation that may be simpler to carry out. For example, the square root

$$z = \sqrt{A} \tag{6.82}$$

is obtained by squaring; i.e., we write

$$F(z, A) = z^2 - A \cong -\frac{z}{\mu} \tag{6.83}$$

which is mechanized as in Fig. 6.36.

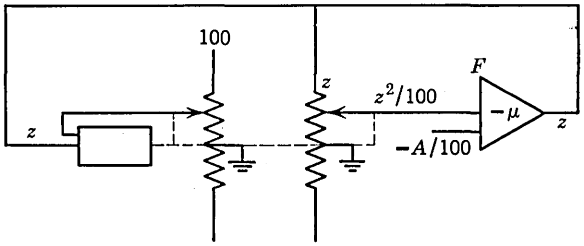


Fig. 6.36 Square-root circuit

6.5.3 Differential Equations

We have investigated the manner in which function generators are employed for the introduction of arbitrary functions into an analog computer. Wherever possible, it is generally wise to consider the feasibility of creating the functions, or close approximations to them, as the solution of differential equations which can be mechanized in the standard manner. For example, the harmonic function

$$y = A\cos\omega t \tag{6.84}$$

may be generated as the solution of the differential equation

$$\ddot{y} = -\omega^2 y \tag{6.85}$$

with the initial conditions

$$y(0) = A, \qquad \dot{y}(0) = 0 \tag{6.86}$$

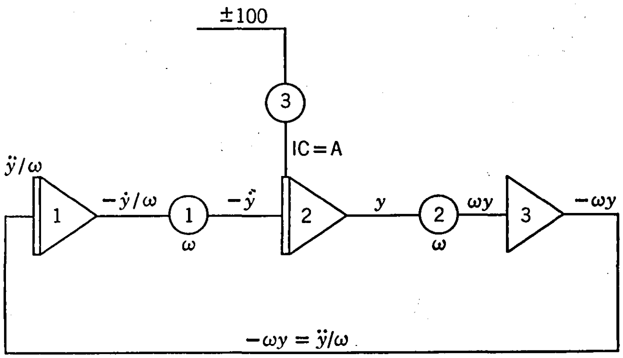


Fig. 6.37 Harmonic-oscillator circuit

This is mechanized as shown in Fig. 6.37. The outputs of integrators 2 and 1 are, respectively,

$$y = A\cos\omega t \quad \text{and} \quad \frac{-\dot{y}}{\omega} = \frac{A\omega\sin\omega t}{\omega} = A\sin\omega t \tag{6.87}$$

The angular frequency ω of the harmonics is adjusted by means of potentiometers 1 and 2.

6.5.4 Linear Simultaneous Algebraic Equations

The solution of a set of simultaneous algebraic equations

$$\sum_{j=1}^{n} a_{ij}x_j = b_i \qquad i = 1, 2, \cdots, n \tag{6.88}$$

on an analog computer is, strangely enough, fraught with more difficulty than is the solution of linear or nonlinear ordinary differential equations. Limitations of space prohibit a detailed analysis of the reasons for this phenomenon; we must content ourselves with the statement that if algebraic equations are mechanized by means of standard d-c amplifiers, the setup will, in general, be unstable because of the fact that the open-loop amplifier gain $(-\mu)$ is frequency dependent and is not a constant. This problem deserves particular attention here in view of the fact that such equations occur frequently in automatic processes, e.g., in the oil and chemical industries. The stability problem is not a simple one to solve; there are many approaches to its solution, one of which is illustrated here. As an example we take the case of three simultaneous equations. It can be shown that stability is insured if the given set of algebraic equations is replaced by a combined set of algebraic and differential equations where the ϵ_i represent the errors in the solution, namely

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 - b_1 &= \epsilon_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 - b_2 &= \epsilon_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 - b_3 &= \epsilon_3 \end{aligned} \tag{6.89}$$

$$\begin{aligned} \frac{dE_1}{dt} &= K\epsilon_1 \\ \frac{dE_2}{dt} &= K\epsilon_2 \qquad K > 0 \\ \frac{dE_3}{dt} &= K\epsilon_3 \end{aligned} \tag{6.90}$$

$$\begin{aligned} x_1 &= -a_{11}E_1 - a_{21}E_2 - a_{31}E_3 \\ x_2 &= -a_{12}E_1 - a_{22}E_2 - a_{32}E_3 \\ x_3 &= -a_{13}E_1 - a_{23}E_2 - a_{33}E_3 \end{aligned} \tag{6.91}$$

The first set is solved for the errors ϵ_i, the second for the variables E_i, and the third for the unknowns x_i. The philosophy behind the mechanization is that the differential equations, if stable, possess steady-state solutions which satisfy the given algebraic equations. For if steady-state values are attained, then

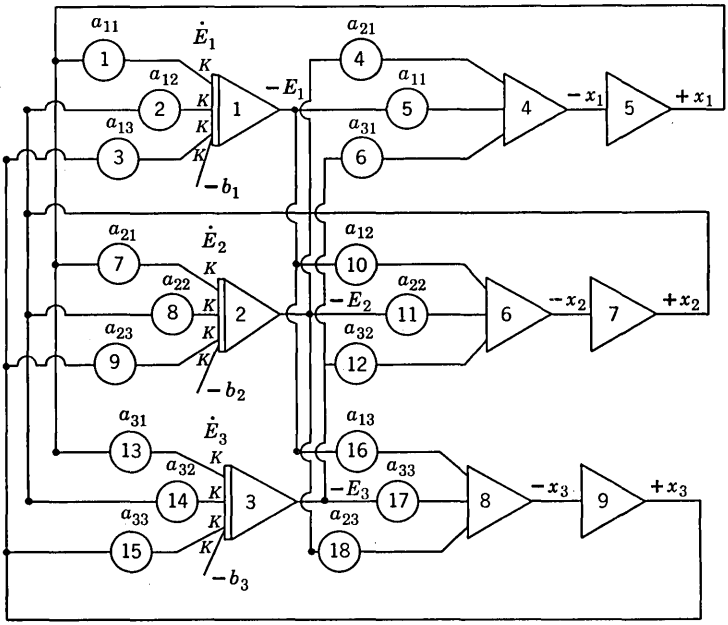


Fig. 6.38 Schematic for linear simultaneous algebraic equations

$$\frac{dE_i}{dt} \rightarrow 0$$

which implies that $\epsilon_i \rightarrow 0$, and that eqs. 6.89 reduce to eq. 6.88. This is an automatic procedure; the time delay before steady-state values are reached may be minimized by selecting the constant K large. A mechanization, with all a_{ij} assumed to be positive, for simplicity, is shown in Fig. 6.38.

6.6 DEVELOPMENTS AND REQUIREMENTS

In conclusion, we shall indicate some of the areas of analog computer development which exhibit the greatest activity, or in which the greatest need for design and development exists. With the trend toward automation in business and industry, and the concomitant requirements of real-time computer operation, it is not surprising that the goal of a great deal of this effort is an all-electronic computer which achieves high-frequency response with no significant reduction in accuracy.

(1) The one component which is the subject of the greatest engineering effort is the electronic multiplier. The time-division, diode, and AM-FM multipliers, discussed in Section 6.3, are results of this program. Although they are capable of high-frequency response (in the order of several hundred cycles per second), their accuracy (about 0.2 per cent) has not quite equaled that of the servo multiplier (between 0.02 per cent and 0.1 per cent). The complexity of these electronic multipliers is considerably greater than that of the servomechanism. The search for a simple multiplier which obeys some physical law has produced some results but, unfortunately, these multipliers are not too accurate.

(2) Whereas success is being approached in the design of an electronic multiplier, the same cannot be said of the electronic resolver. Such a device constitutes one of the important needs in the analog computer field today.

(3) The third in the series of nonlinear components which is described in this chapter is the function generator. Here the need is for a unit capable of generating functions of several variables (such as the three-dimensional cam). The most fruitful approach to this problem has been in the analytic decomposition of the given function into sums and products of functions of a single variable, each of which can be synthesized by standard techniques.

(4) In Section 6.3 it was observed that recorders which possess high-frequency response are generally incapable of high accuracy, and conversely. It is often ironic that analog computer solutions stemming from mechanizations containing hundreds of components are generally more accurate than are the means of recording them. The need exists for recorders whose accuracy and frequency response approach those of the computer itself.

(5) Associated with the problem above, as explained before, is the development of analog-to-digital converters for recording analog computer solutions accurately, although slowly. More generally, both analog-to-digital and digital-to-analog converters are the object of tremendous commercial activity directed toward application in the

field of automation, in order to tie together analog and digital computers and instruments. This is the subject of Chapter 9.

(6) Several steps have been taken toward automating the insertion of a problem into an analog computer. Examples of this are the digital keyboard, used in conjunction with a servomechanism and clutches, for establishing the potentiometer settings, and the punch-tape input, also used to perform this function as well as to establish the connections between components.

(7) Some of the analog computers and simulators designed more recently, for example the submarine simulator described in Section 6.1, include automatic check systems for performing static checks and, to a limited extent, dynamic checks. The static check permits the verification of all component connections other than the inputs to integrators; the dynamic check verifies the rates of integration.

6.7 REFERENCES

1. G. A. Korn and T. M. Korn, *Electronic Analog Computers* (D-c Analog Computers), McGraw-Hill Book Co., Inc., New York, 1952.
2. W. W. Soroka, *Analog Methods in Computation and Simulation,* McGraw-Hill Book Co., Inc., New York, 1954.
3. E. L. Harder and G. D. McCann, A Large Scale General-Purpose Electric Analog Computer, *Trans. AIEE, 67:* 664–673, 1948.
4. G. D. McCann, H. Wilts, and B. N. Locanthi, Electronic Techniques Applied to Analogue Methods of Computation, *Proc. IRE,* 954–961, Aug. 1949.
5. G. A. Philbrick and H. M. Paynter, The Electronic Analog Computer as a Laboratory Tool, *Ind. Lab., 3,* No. 5, May 1952.
6. V. Bush and S. H. Caldwell, A New Type of Differential Analyzer, *J. Franklin Inst., 240:* 255–326, Oct. 1945.
7. V. Bush, The Differential Analyzer, A New Machine for Solving Differential Equations, *J. Franklin Inst., 212,* 447, Oct. 1931.
8. E. Janssen and D. Lebell, Applications of a Mechanical Differential Analyzer to Electrical Engineering, *Elec. Eng., 70:* 432–435, 1951.
9. F. J. Murray, *The Theory of Mathematical Machines,* King's Crown Press, New York, 1948.
10. R. E. Scott, Network Synthesis by the Use of Potential Analogs, *Proc. IRE, 40:* 970–973, Aug. 1952.
11. H. E. Burke, A Survey of Analog-to-Digital Converters, *Proc. IRE,* 1455–1462, Oct. 1953.
12. E. A. Goldberg, Stabilization of D-c Amplifiers, *RCA Rev., 11:* 296, 1950.
13. C. D. Morrill and R. V. Baum, Stabilized Time-Division Multiplier, *Electronics, 193,* Dec. 1952.
14. C. M. Edwards, Survey of Analog Multiplication Schemes, *J. Assoc. Comp. Mach., 1,* No. 1: 27–35, Jan. 1954.
15. W. A. McCool, An AM-FM Electronic Analog Multiplier, *Proc. IRE, 41:* 1470–1477, Oct. 1953.

16. E. A. Goldberg and G. W. Brown, An Electronic Simultaneous Equation Solver, *J. Applied Phys., 19:* 339–345, 1948.

17. J. H. Lanning, Jr., and R. H. Battin, *An Application of Analogue Computers to Problems of Statistical Analysis Project Cyclone, Symposium II,* Pt. 2, 79–87, Reeves Instrument Corp., New York, 1952.

18. R. R. Bennett, Analog Computing Applied to Noise Studies, *Proc. IRE,* 1509–1513, Oct. 1953.

7. Digital Computers

WILLIS H. WARE

The RAND Corporation
Los Angeles, California

7.1 INTRODUCTION

The digital computer, or as it is becoming known, the data-processing or data-handling machine, is another tool for use in automation applications. It is the second broad type of equipment which is available for performing the computation necessary in many automation or control processes. Compared with analog computing, the digital technique is relatively recent. Originally conceived for general engineering and scientific computation, machines constructed for this purpose were, in general, too expensive and too powerful mathematically for control or automation. It is only within the last few years that smaller, lower-priced machines have become available and have made economic the application of digital techniques to the automation field.

7.2 ANALOG VERSUS DIGITAL (1)[1]

In the analog technique we obtain a solution to a problem by making measurements on a physical system. The simplest type of analog system is the model, in which a scale version of the problem of in-

---

[1] Numbers refer to the references listed at the end of this chapter.

terest is constructed. From measurements made upon it the solution to the full-scale problem is achieved by extrapolating these measurements. Another example of the analog computing system is the direct analog, in which some alternate system is found whose equations of behavior are the same as the equations of behavior of the stated problem. Measurements made on the alternate system are then directly transferable to the problem of interest. The last type of analog system is the differential analyzer, which may be either mechanical or electronic. In this case the machine actually solves the mathematical equations which describe the problem. In any type of analog system it follows that there is a direct correspondence between parts of the problem and parts of the computer. In the case of the differential analyzer, or electronic analog computer as it is sometimes known, the correspondence is between parts of the equations which describe the problem and parts of the computer.

In contrast, the digital system does not deal with a physical system per se, but instead it deals only with numbers representing the characteristics of the problem of interest. The principal characteristic of a digital computing system is that it counts or manipulates numerical expressions which represent physical behavior but which are dissociated from a physical system.

The problem of deciding whether a given system is digital or analog is sometimes difficult. There are fundamental characteristics of each type, however, and these indicate how a given system should be classified. The analog system is one that involves the process of measurement on a physical system. It deals with variables that are continuous, and the results that may be obtained from an analog system are limited in precision since measurements are made with a physical device. On the other hand, the digital computing device is concerned with the counting or manipulation of numerical expressions. In general it deals with information that is in discrete form; and it yields results that may be of unlimited precision since, in general, numerical processes may be carried to as precise a degree as desired. Questions that might be asked to indicate whether a given system is digital or analog are:

Can precision be traded for time?

Must a better measuring device be used to get a better result?

Is this device counting something or is it measuring something?

Does this device deal with a physical system or does it deal with numbers that are independent of a physical system?

The classical example of an analog computing device is the slide rule, which measures physical lengths, which in turn logarithmically represent numbers. The classical example of the digital computer is the common desk calculator, which by an arrangement of linkages and wheels performs arithmetic operations. Many systems are mixed in nature. For example, a watch might be thought of as an analog balance wheel, plus an analog-to-digital converter in the escapement mechanism, plus a digital system which counts the number of oscillations of the balance wheel.

7.3 THE TYPICAL DIGITAL COMPUTER (5, 7, 8, 33, 42, 43)

A digital computer, data-handling system, or information-processing system, may be defined as follows: It is a system or device that produces output information derived from accepted input information by some process of logic. The process of logic must be demonstrably free from self-contradiction.[1] The crux of this definition is the "some process of logic." If there is included in the "process of logic" only the operations of arithmetic, then the digital computer is merely a device for carrying out numerical computations. If, on the other hand, there is also included in the "process of logic" the operations of formal mathematical logic, then the digital computer becomes adapted to general information-handling problems and its power on the mathematical problem becomes considerably enhanced.

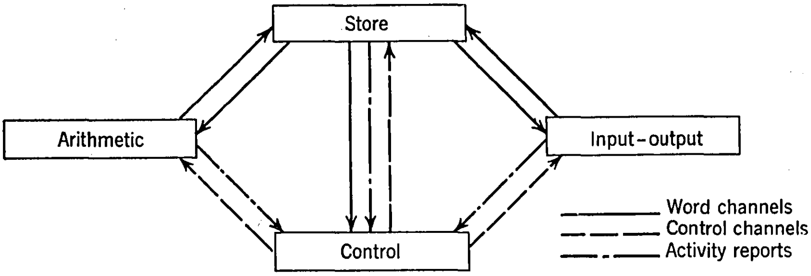


Fig. 7.1 Block diagram of a basic digital computing machine

A block diagram of a typical computing machine system is shown in Fig. 7.1. The function of the *arithmetic section* is to perform all of the arithmetic operations of which the system is capable. These may include addition, subtraction, multiplication, division, sometimes square rooting, and any other special features demanded by the particular class of problem for which the machine is intended. The

[1] Definition from the Standards on Electronic Computers, *Proc. IRE,* March 1951.

arithmetic section may also include certain logical operations which are required in the manipulation of information.

The *input-output section* is the transducer through which the machine system communicates with the external world. In general, it processes some physical medium such as paper tape, magnetic tape, punched cards, etc., which have been prepared by the user and which on input state the problem to the machine in a language intelligible to the machine. In the other direction it must provide, in a form that is intelligible to the external world, information which has been prepared by the machine system. The input-output section is the means for entering information into the machine and for withdrawing information from it. In the automation application the input and output may communicate with other machinery or with instruments, rather than with people.

Storage is the section in which initial data, intermediate results, final results, and the statement of the problem are stored. The statement of the problem to the machine system appears as a list of instructions which direct the system to perform the appropriate manipulations of the data it has been given. The storage acts as an erasable and reusable scratch pad. It provides both data and instructions at the rapid rate required by the balance of the machine, and it accepts results that have been computed by the arithmetic section.

Control is the section that interprets the instructions which have been given to the machine, and which then causes all other parts to do the appropriate function at the appropriate time, so as to achieve the desired operation. It is the mastermind of the system.

An analogy may be drawn with the equipment assembled to perform a longhand calculation—desk calculator, operator, scratch pad, and pencil. The keyboard of the desk machine is the input, its dials, the output; the arithmetic ability of the desk calculator is akin to the arithmetic section. The scratch pad is the storage and the operator is the control section.

7.3.1 Terminology (2, 3)

The unit in which information is handled in a digital system is the *word.* This term has been chosen because a word may be either a piece of data, or it may be an *instruction* which indicates to the machine system what function is to be performed. A word may even be an abstract collection of digits without physical or machine meaning in the whole, but having meaning only in selected parts. The plan for solution of a problem on a digital machine is known as a *program.* The collection of instructions in an assortment and sequence that describes this program to the machine is known as the *routine.* Both the program and the routine are the result of an analysis of the problem to be solved by a mathematician. Finally, the statement of the routine in machine language is known as the *code.*

The format of an instruction is indicated in Fig. 7.2. An instruction consists of an *operation part* plus one or more *addresses.* The operation part indicates what function the machine is to perform as a result of the instruction, and the address part indicates the location in storage where the information to be operated on is located. If there is but one address per instruction, this type of organization is known as a *single address* machine. The system may be organized as two address, three address, etc., in which case the additional addresses specify additional locations of data, instructions, and control information. If the number of digits, or characters, in a word is constant, then the system is said to be a *fixed word length* system; if the word length is variable, the system is said to be a *variable word length* system. The operation part and the address part of an instruction must be expressed in some language that is intelligible to the machine. This is generally in some numeric code, and since data are represented to the system in this same numeric code, instructions and data look alike to the machine. One function of the control section is to keep track of which parts of the storage contain data and which parts contain the instructions.

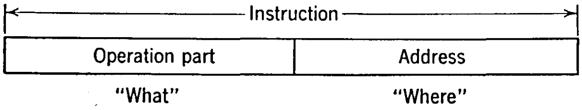


Fig. 7.2 Instruction format

7.4 SAMPLE PROGRAM (4, 5, 6, 8)

The address that appears in each instruction may be thought of as a tag indicating a location at which the desired information is held. The storage may be regarded as an array of pigeonholes, where each of the pigeonholes is identified and located by one of these tags, i.e., its address. The address bears no relation to the data contained in that location but only identifies the location. The address in an instruction is analogous to a street address locating only a particular house but indicating nothing about the inhabitants of that house.

Typical instructions that might appear in a machine are: add, multiply, stop, get a word from the input, give a word to the output, transfer a word from the storage to the arithmetic section, transfer a word from the arithmetic section to the storage, stop, etc.

As a simple example of a routine, consider the problem of computing the algebraic function $N_1N_2/(N_1 - N_2)$ for particular values of N_1 and N_2. Assume that N_1 is stored in location 100, that N_2 is stored in location 101, and that the desired value of this function is to be stored in location 102 after it is computed. Reserve locations 103 and 104 for storage of intermediate results. The following list of operations would then accomplish the solution to this problem for a single-address machine:

200	Get the number in 100	Forms N_1N_2
201	Multiply it by the number in 101	and
202	Put the answer in 103	stores it in 103
203	Get the number in 100	Forms $(N_1 - N_2)$
204	Subtract from it the number in 101	and
205	Put the answer in 104	stores it in 104
206	Get the number in 103	Forms $N_1N_2/(N_1 - N_2)$
207	Divide it by the number in 104	and
208	Put the answer in 102	stores it in 102
209	Print 100	Prints out
210	Print 101	N_1, N_2 and
211	Print 102	$N_1N_2/(N_1 - N_2)$
212	Stop	

Particular care must be taken to distinguish between the address *of* an instruction and the address *in* an instruction. The routine or list of instructions is held in the storage, and, therefore, each instruction occupies a storage location that has an address. These are the addresses 200 through 212 in the example above. These addresses *of* the instructions have no connection with the execution of the instruction in a particular location but only serve to indicate to the control where instructions may be found. In the given example the control section will have been told to seek the first instruction at location 200, and it will thereafter proceed sequentially down the list. On the other hand, as each instruction is obtained, it contains with*in* itself an address that specifies (in the example) the location of data or the location for results.

7.5 CONTRAST BETWEEN THE ANALOG AND DIGITAL SYSTEM (36)

To adapt the analog system to a new problem requires that its equipment be reconnected in a different fashion. In a digital system, however, the equipment is always connected in the same fashion and is always able to execute the same basic set of instructions. A new problem for the digital system means merely a new sequence of instructions—that is, a new routine. In each case this routine is available to the machine system from its position in the storage.

The analog system does some arithmetic operations (add and multiply) and also, many mathematical operations (differentiation, integration, other mathematical operators). It performs these operations on real physical variables. In contrast, the digital system does only arithmetic operations and the operations of formal logic; it does these on numerical representations of information. The process of performing the mathematical operations consists of replacing the infinitesimal calculus by the calculus of finite differences. This process, plus mathematical methods for solving systems of equations, handling matrices, general manipulation of data, etc., constitutes the subject matter known as "numerical analysis."

7.6 MACHINE DECISIONS

Machines have generally been given the ability to make decisions for themselves. This usually consists of detecting whether a number is of positive or negative sign, or, of detecting whether a number is zero or not. However, from a sufficiently repeated use of some sequence of such simple yes-no decisions, rather involved criteria may be established on the basis of which the machine may judge events. Since an instruction and a piece of data look alike in that each of them appear in the same numeric code, it is possible for a machine system to perform arithmetic on its own instructions and thereby alter them. Further, this alteration of instructions may be performed on the basis of decisions that it has itself made. The machine can, so to speak, determine its own future course of progress. Note that this is not "thinking," since all of the alternatives available to the machine are contained in the original program and, hence, had to be thought through *a priori* by the person who wrote this program. It may be said that a machine thinks, or learns, only to the extent that every alternative contained in such thinking or learning is at the outset already in the program (34, 35).

Such decision instructions are typically implemented as follows:

(1) If the test for the decision is satisfied, e.g., the number is positive, do not obtain the next instruction from the following storage location (i.e., the storage location next after the one containing the test

instruction itself), but rather go to the address contained *within* the test instruction and there obtain the next instruction to be executed.

(2) If the test for the decision is not satisfied, e.g., the number is not positive, proceed directly to the next storage location in sequence for the next instruction to be executed.

7.6.1 Instruction Classes

Typical instruction classes are:

(1) *Arithmetic,* which contains add, subtract, multiply, divide, occasionally square root, shift right, shift left, store a word, load a register, etc.

(2) *Input-output,* which contains the instructions necessary to govern both mechanically and electrically the behavior of all input-output equipment.

(3) *Logical,* which performs the formal logical manipulations of *and* or *or* on words or parts of words; or which combines parts of words to form new words (*extraction*).

(4) *Control,* which causes the control section to jump from one part of the routine to another, either under any circumstance (*unconditionally*) or in response to some decision (or test) made by the system (*conditionally*); also contains the stop.

For the automation application, the instruction list will contain these usual ones, except that there may be many instructions which are particularly intended for receiving information from, or transmitting information to, a large variety of other kinds of equipment (a set of specialized input-output instructions). In addition, there may be specialized instructions for convenient manual control and intervention into machine activity by an operator.

7.6.2 Distinguishing Features of the Digital System

The number of kinds of instructions may be large, and they may be used in any quantity of each kind; hence, the routine is essentially infinitely flexible. This is why digital systems are adaptable to such an enormously wide class of problems. This feature should be contrasted to the analog system which has certain kinds of equipment whose interconnections are changed to suit the problem. The digital system remains unchanged physically, as the problem changes. It adapts to the problem by performing a different sequence of the instructions built into it at the time of construction. Typical machine systems have from twenty to upward of a hundred distinct instructions, and large routines may contain tens of thousands of instructions.

In addition to its flexibility, the digital machine also exhibits two other distinguishing characteristics. First, it proceeds at enormously great speeds of operation. In the smaller, slower machines, individual operations may occur at a 1 to 15 millisecond pace, whereas in the largest and fastest scientific machines, individual operations may proceed as rapidly as a 10 to 20 microsecond pace. After making some allowance for internal operations that do not produce useful results, it follows that a digital computing system may accomplish work at a rate as great as 10,000 to 50,000 operations per second. Secondly, the digital machine system exhibits a relatively large freedom from error. Because of the high operating speeds and with the currently attainable intervals of error-free performance, it can be shown that the digital system in general makes an error only once in perhaps 10^8 or 10^9 operations. This is to be contrasted with the corresponding statement for manual operations in which a mistake occurs every few hundred operations.

7.6.3 Subroutines

It is possible for a routine that performs a fairly well-defined and often repeated operation to be so arranged that it is available to the machine from storage on demand, under the control of some other master routine. Such a routine is called a *subroutine*. Commonly, a routine will contain many subroutines that are prepared only once and are reused thereafter. Examples of subroutines might be ones for computing the sine, the cosine, or the exponential from the respective series expansions. Alternately, a subroutine might be arranged to perform linear interpolation, to fit an approximation to a function that is possibly empirical, or to fit a curve to a prescribed set of points. Subroutines that perform operations of the latter kind are especially important in the automation application.

In some applications the digital system may be so fast that it is not busy full time with a single control problem but may actually time-share its availability with other control problems. In this case the input and output are not continuous data but are handled on a sampled basis. The judicious use of subroutines makes such a time-sharing arrangement readily possible. In the analog control system, on the other hand, the computer must usually remain on the job full time.

7.7 CLASSIFICATION FEATURES

A digital system may be characterized according to the following features:

General purpose or special purpose.
Serial or parallel.
Type of arithmetic; e.g., binary or decimal.
Type and quantity of storage.
Word length.
Number of addresses per instruction.
Type and quantity of input-output equipment.
Number of levels in the storage hierarchy (see Section 7.13).
Operating speeds.

7.8 SIZE AND RELIABILITY (44)

A digital data-handling system may contain exceedingly large amounts of equipment. The largest systems intended for scientific or engineering work may contain in excess of 5000 vacuum tubes, with correspondingly large quantities of other components. The smaller machines intended for slower solutions of scientific or engineering problems, or for the control applications, may contain fifty to several hundred vacuum tubes with corresponding amounts of other components. Hence, in order to achieve reasonable behavior of this amount of equipment, it is necessary to maintain an extremely conservative attitude toward the equipment design. Furthermore, a digital system is so complex in nature, and possesses so many different possible configurations of action, that it is not possible to check every situation which might arise; this is another reason for a conservative design philosophy.

It is absolutely necessary to determine a set of ground rules for the design procedure that treats all components as conservatively as possible and forces large safety factors into all design criteria. In general, components are heavily derated in addition to being carefully tested and screened before use. As a result of such an approach, for large scientific systems, performances of tens of hours between errors on the average have been achieved; and for the smaller systems, there have been performances of up to hundreds of hours between errors.

In the automation or control application in which a breakdown might be serious, it may be desirable to provide duplicate equipment. It is possible to minimize unexpected breakdowns by a program of systematic preventive maintenance when the equipment is not in use. This generally consists in varying supply voltages, frequencies, and other parts of the operational environment in an effort to cause the incipient failure to occur. It is possible to formulate detection and trouble location problems that indicate by the nature of the error in their solution what part or parts of the machine system are at fault.

There may be three distinct sources of errors. The operator who is manipulating the machine or the person who wrote the program may have committed an error. The machine in its internal operation may be malfunctioning; but, just as important, the mathematical procedure may be failing. Since the continuous operations of analysis must be replaced by the finite difference operations, it is possible for the mathematical procedures that have been adopted for a particular problem to fail to converge or to fail to behave properly in some other way. Hence, Man, Machine, and Mathematics—these are the three broad classes of trouble.

It is possible in the design of a digital system to protect any parts of the system that are troublesome by nature. This may be done by incorporating error-detecting features into the numeric code used to represent information or, possibly, by even incorporating error-correcting features (11). This procedure amounts to adding redundancy to the numeric code in which the information is represented. The penalty for such protection is the extra equipment, or the extra time required for operation.

In this connection it is interesting to note that the analog computing system is full of feedback paths which tend not only to keep individual components in line but also to keep the whole system behaving properly. The digital system has no such logical feedbacks to help its cause. One small error in a digital system can spell disaster. It can mean that an instruction has become changed to a different instruction; or it can mean that the machine has gotten a piece of data and thinks it is an instruction, or the converse. This distinction between the two types of computing equipment has strong bearing on what is meant by "reliability." The digital system must be absolutely reliable and this must be by means of a conservative design approach. The analog machine, on the other hand, need only be reliable on the average and can depend, at least in part, on feedback paths to help overcome component difficulties.

It is the nature of a digital system to be made up of a very few basic building blocks which are repeated over and over again but are interconnected in different ways. The design of a digital computer in this sense is markedly different from much electronic equipment in which there is little repetition of individual circuit configurations. Hence, it is common to find that a relatively long period may be expended on the design and perfection of just one of the fundamental circuits that are required.

7.9 NUMBER SYSTEMS (7, 8)

Traditionally, the decimal number system has been used in recent times for arithmetic procedures. The decimal system makes use of digits 0 through 9 and hence contains ten permissible symbols. There is no fundamental requirement that there be any such restriction on the number of permitted symbols; and in particular, number systems that use only the digits 0 and 1 (binary), digits 0 through 7 (octal), and digits 0 through 9 plus six additional symbols (hexadecimal or sexadecimal) are popular and in common use in the organization of a digital computing system. It is obvious that a pile of objects could be counted in any of these number systems; this implies that a number whose numerical representation is given in a specified number system must have an equivalent numerical expression in any other number system. In other words, numbers expressed in the decimal system can be converted to an equivalent representation in octal, in binary, in hexadecimal, or in a system of arbitrary choice. Conversely, numbers

Decimal	Binary	Octal	Hexadecimal
0	0	0	0
1	1	1	1
2	10	2	2
3	11	3	3
4	100	4	4
5	101	5	5
6	110	6	6
7	111	7	7
8	1000	10	8
9	1001	11	9
10	1010	12	A
11	1011	13	B
12	1100	14	C
13	1101	15	D
14	1110	16	E
15	1111	17	F
16	10000	20	10
17	10001	21	11
.	.	.	.
.	.	.	.
.	.	.	.
29	11101	35	1D
30	11110	36	1E
31	11111	37	1F
32	100000	40	20

Fig. 7.3 Comparative table of various number systems

expressed in any of these systems can be converted back to a decimal representation or to a representation in still another system. Such conversions go on all the time in digital systems.

Figure 7.3 tabulates the first few entries of the binary, octal, decimal, and hexadecimal systems. It is apparent from this table that the expression "1 0" (*not* to be pronounced "ten") appears in each number system as the entry that represents the number of permitted symbols in that system. In any system the number of permitted symbols is called the *base* or *radix* of that system.

Examples of binary arithmetic with their equivalent decimal operations follow.

$$
\begin{array}[t]{r}
9.\\
+\ 5.\\
\hline
14
\end{array}
\qquad
\begin{array}[t]{r}
1001.\\
+0101.\\
\hline
1110
\end{array}
\qquad
\begin{array}[t]{r}
9\\
\times\ 5\\
\hline
45
\end{array}
\qquad
\begin{array}[t]{r}
1001.\\
\times 0101.\\
\hline
1001\\
0000\\
1001\\
0000\\
\hline
101101
\end{array}
$$

$$
\begin{array}[t]{r}
13.\\
-\ 7.\\
\hline
6
\end{array}
\qquad
\begin{array}[t]{r}
1101.\\
-0111.\\
\hline
0110
\end{array}
\qquad
\begin{array}[t]{r}
\frac{1}{4}\\
+\frac{3}{16}\\
\hline
\frac{7}{16}
\end{array}
\qquad
\begin{array}[t]{r}
0.0100\\
0.0011\\
\hline
0.0111
\end{array}
$$

These can be readily checked by noting the meaning of a numerical expression:

$$827.39 = 8\times 10^2 + 2\times 10^1 + 7\times 10^0 + 3\times 10^{-1} + 9\times 10^{-2}$$

Hence a binary expression 1101 actually implies

$$
\begin{aligned}
1101 &= 1\times 2^3 + 1\times 2^2 + 0\times 2^1 + 1\times 2^0\\
&= 8 + 4 + 0 + 1\\
&= 13
\end{aligned}
$$

Similarly, binary fractions are

$$
\begin{aligned}
0.01111 &= 0\times 2^{-1} + 1\times 2^{-2} + 1\times 2^{-3} + 1\times 2^{-4}\\
&= 0 + \tfrac{1}{4} + \tfrac{1}{8} + \tfrac{1}{16} = \tfrac{7}{16}
\end{aligned}
$$

From consideration of equivalences like this, rules for converting from one number system to another can be readily derived.

The importance of the binary number system lies in the fundamental nature of the electronic techniques now available. For the most part, the circuits and the physical phenomena suitable for use in a digital system are bistable in character; hence, it is natural to organize such devices into an arrangement for performing binary manipulations. If it then becomes necessary to organize a machine system which must deal with data expressed either in decimal notation or in mixed decimal and alphabetic notation, it is necessary to resort to a trick (40). Decimal information may be represented by associating 4 binary digits (bits) together as a character. There are 16 possible combinations of this tetrad; of these some 10 may be selected to represent the decimal digits and the balance discarded. Such an organization is known as a *binary-coded decimal machine.* If binary digits are associated in 6-bit groups, then of these 64 combinations 10 may be assigned to represent the decimal digits, 26 may be assigned to represent the alphabet, and as many others may be assigned to represent punctuation and special characters as are required; the balance is discarded. A machine organized to deal in such 6-bit groups is commonly known as an *alphanumeric* or *alphameric* machine.

7.10 LOGICAL ALGEBRA (8, 9, 10, 12, 13)

In the design of certain parts of a digital system, it is convenient to make use of a special form of algebra known as *Boolean* or *logical algebra.* In it the variables are not permitted to have a continuum of values but are permitted only the discrete values of 0 and 1. A value of 1 is usually associated with a statement's being true and 0, with its being false. It is then possible, with this convention, to describe completely each and every operation occurring within the digital system in the form of algebraic expressions, which may then be manipulated according to a set of theorems.

Because of the restriction placed upon the permitted value of any variable, the theorems that appear in this algebra are, in some cases, unlike those of more familiar algebras. For instance, since a given variable may at most be 1, it follows that $X + X = X$. Further, since a variable may at most be 1 or true, it is nonsense for this variable to be (true)^2; hence, $X^2 = X \cdot X = X$.

In formal logic it is convenient to have the connectives *or, and,* and *not* (negation), as in the following examples:

(1) It will hail if rain clouds are present *and* if there are strong updrafts of wind.

(2) The ground will get wet if it rains *or* if it snows.

(3) The ground will *not* get wet if it is sunny *or* if it is *not* cloudy.

It can be shown that two operations exist in logical algebra which represent the connectives *and* and *or*, and that these operations in many ways behave the same as the addition and multiplication operations of conventional arithmetic. It must be remembered that this algebra contains unusual theorems, so that there is not complete equivalence between the operation of logical algebra which is often designated by the plus sign, and the addition process of ordinary arithmetic. Similarly, there is not complete equivalence between the operation of logical algebra often denoted by the multiplication sign and the multiplication process of ordinary arithmetic. The following examples indicate how a plus sign in logical algebra is the equivalent of the *or* connective, and how the multiplication sign in logical algebra is equivalent to the *and* connective.

$P = A + B$ P is 1 (true) if A *or* B is 1 (true)

$P = A \cdot B$ P is 1 (true) if A *and* B are both 1 (true)

Logical algebra may also be stated in terms of the language of set theory. In this event, the *or* operation becomes the *union* and the *and* operation, the *intersection.* The importance of this technique is that it is possible to completely describe the behavior of a machine system in a fashion which may be conveniently manipulated according to well-defined rules; it is therefore possible to present equivalent designs with a minimum of effort.

7.11 BASIC BUILDING BLOCKS (5, 7, 8)

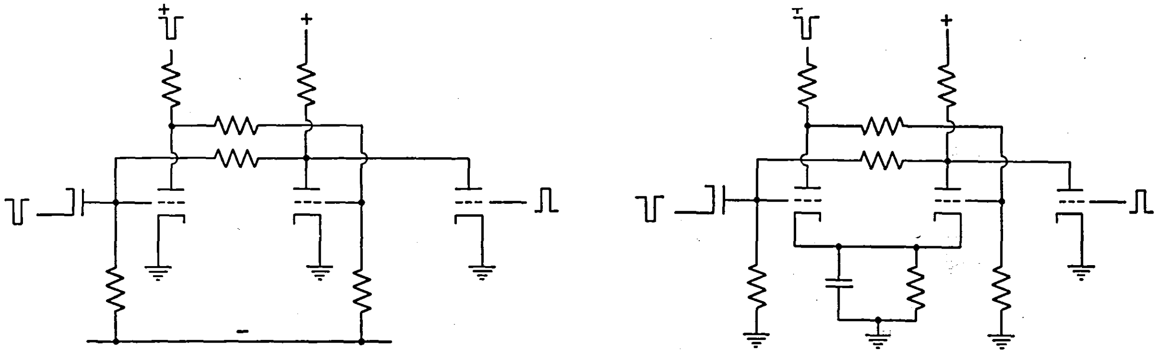
As previously indicated, the design of a digital computing system is essentially the problem of designing a few basic building blocks which are then used repeatedly. It can be shown that if circuits which perform the three operations of formal logic—*and, or* and *negate*—are available and if there is some means for storage of information, then no other fundamental building blocks are required. There are a variety of circuits that will perform the function of storage and also a large number of choices for the function of decision making, i.e., "gate circuits." In a particular case the choice of a circuit for a given job is made on the basis of many involved engineering considerations and perhaps, in addition, on the basis of the field of application and the environmental conditions to which the equipment may be subjected.

7.11.1 The Flip-flop (24, 25, 26)

A circuit that is applicable as a unit storage device and that can remember a simple *yes-no* decision, i.e., one bit of information, is any two-stable-state device. Such a circuit is commonly called a *flip-flop*. It appears in many forms; Fig. 7.4 indicates three such forms: the Eccles-Jordan or symmetrical flip-flop, the cathode-biased flip-flop, and the Schmidt or asymmetrical flip-flop. In each case the two stable states correspond to either left-tube conducting and right-tube nonconducting, or vice versa. One stable state may be assigned to represent binary 0 and the other, binary 1. This figure also indicates a number of ways of changing the state, or of flipping the flip-flop. In the examples illustrated in this figure, it will be noted that a signal applied to a given input places the flip-flop unambiguously in a specified state; i.e., a particular tube is always conducting after the signal is complete. In general, a flip-flop circuit has two symmetrically placed inputs which are in one-to-one correspondence with the two stable states; alternatively, there is one input on which two different kinds of signal appear (the asymmetrical flip-flop), these two kinds of signals being in one-to-one correspondence with the two stable states.

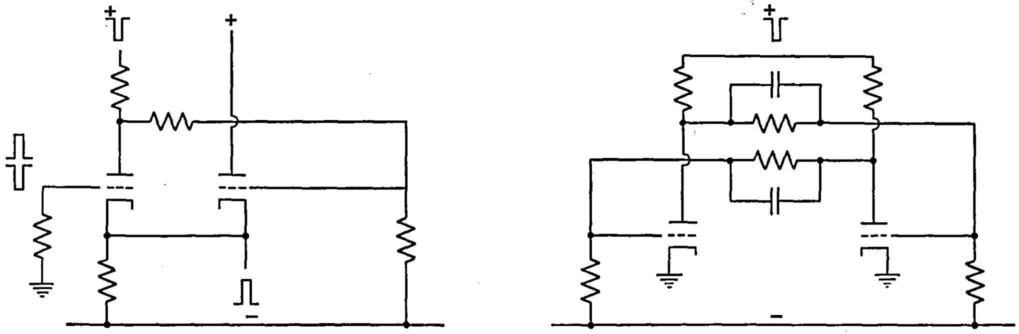
Another type of operation of the flip-flop circuit is known as *complementing*. In this case there is one input terminal and the flip-flop is expected to change state for each signal at this terminal (Fig. 7.4*d*). One such input terminal of the symmetrical flip-flop is the common plate supply bus or the common cathode connection. It should be noted that under this mode of operation, capacitors must be added to the circuit to provide a "memory" facility. The flip-flop circuit in this case must have some recollection of its last state in order to know which new state to enter. The charge upon the capacitors plays this role.

The flip-flop circuit will appear many, many times in a complete digital system, and the design must be such that perhaps hundreds of them will continue to operate satisfactorily, even though the resistors, the supply voltages, the tube parameters, the environmental conditions, etc., drift in a statistical fashion. The design of a flip-flop circuit resolves itself into two distinct phases. First, there is the requirement that the circuit have two stable states and that these stable states continue to exist, even though the statistical perturbations noted above occur. This may be thought of as the d-c design of a flip-flop. Second, there is the dynamic behavior of the circuit, which relates to



(a) Symmetrical flip-flop

(b) Cathode-biased flip-flop



(c) Asymmetrical flip-flop

(d) Complementing operation of the symmetrical flip-flop

Fig. 7.4 Flip-flop circuits

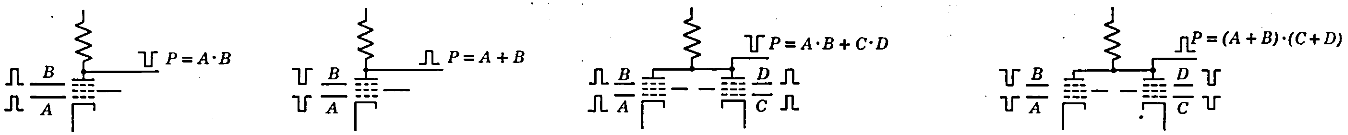
the speed with which, under the influence of the driving circuit, it switches from state to state. Approximation methods have been derived for the d-c design of flip-flop circuits (14, 15), but the complexity of the circuit is sufficient that a good design method for prescribed dynamic characteristics has not yet been evolved.

7.11.2 Gates (9, 10, 22, 23)

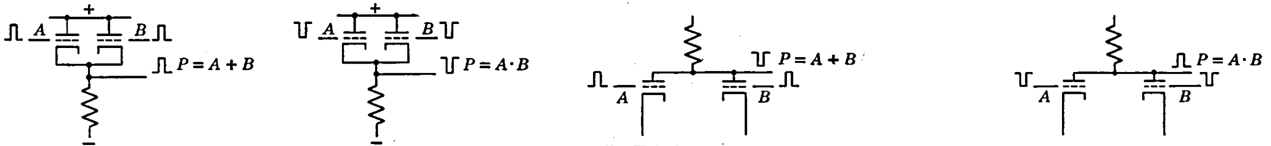
A second principal building block is the *gate* or electronic switch. The function of this circuit is to perform the *and* or *or* connectives of formal logic. A number of examples of circuits that will behave as gate circuits are indicated in Fig. 7.5. In the discussion of any gate circuit it is essential to define carefully the polarity of the signal and to indicate what this polarity signifies with respect to its meaning in formal logic. In the example of the pentode, which is shown with positive signals applied to the first and third grids, it is assumed that the lower or quiescent level of the pulse represents the logical *false,* and that the upper or plateau level of the pulse represents the logical *true.* Hence, it follows that plate current will flow in the pentode if, and only if, grid A is positive *and* grid B is simultaneously positive. If we assume that in the plate circuit a low plate potential defines true, then the pentode circuit is an *"and"* gate for positive grid signals. For negative grid signals it immediately follows that the plate is positive if, and only if, grid A *or* grid B is negative. And, hence, for negative inputs the pentode circuit is an *"or"* gate. If, however, the convention in the plate circuit is that a high potential represents true, then the pentode circuit with positive inputs is the *and not* function, or with negative input signals, the *or not* function.

This example illustrates a duality that exists throughout gating circuits. A circuit that is *and* for a given polarity of input is automatically an *or* circuit for the opposite polarity of signal, and vice versa. As indicated by the figure, relays, triodes, and diodes may also be used to implement the logical connectives of *and* and *or.* When triodes share a common cathode resistor, it is obvious that the cathode will rise if either grid A *or* grid B rises and, hence, this circuit is an *or* circuit for positive signals. When diodes share a common cathode resistor, it is evident that the cathode potential will fall if, and only if, plate A *and* simultaneously plate B are relatively at a low potential; hence, this circuit is an *and* circuit for negative signals.

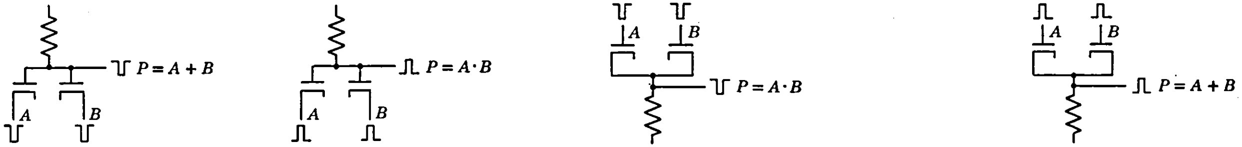
It should be noted that all of the *or* circuits indicated here are the "inclusive *or*" function; i.e., the output is true if input A *or* input B *or* both are true. The exclusive *or*—the output is true if input A *or*



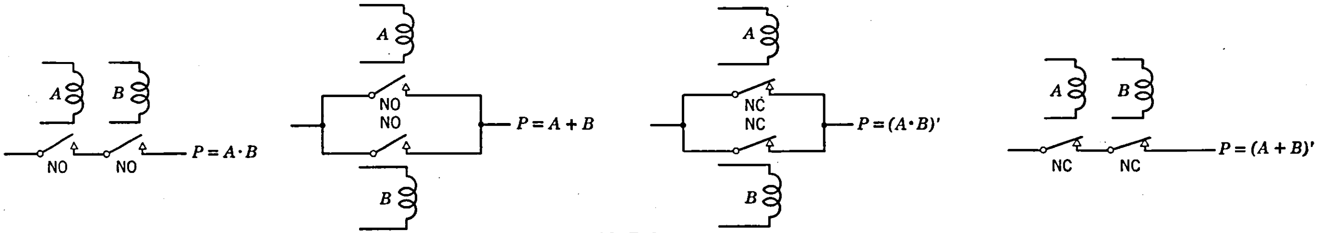
(a) Pentode gates



(b) Triode gates



(c) Diode gates



(d) Relay gates

Fig. 7.5 Gate circuits

input B is true but *not* if both are true—is considerably harder to implement. It may be expressed in Boolean fashion as:

$$(A + B)(AB)'$$

or the equivalent form:

$$AB' + A'B$$

in which the primed characters represent the complement or negation of the variable. In either form it is evident that the expression is 1 (true) if A *or* B equals 1 (true) but the other equals 0 (false), but the expression is 0 if both A and B are 1 (since then A' and B' are each 0).

The third logical connective of *not* is easily realized with any circuit that is inverting, i.e., a single-stage amplifier, a transformer, etc. If A represents a Boolean variable, then its negation is expressed as A' or $\bar{A}$; i.e., if $A = 1$, then $A' = 0$ and vice versa.

Here, again, the design of gates must be so conservative that many hundreds or even thousands of these circuits, when combined into a digital system, continue to work even though the electrical environment of the circuit (voltages, component values, frequencies, rise times, temperatures, etc.) change in an unpredictable statistical fashion.

7.12 ARITHMETIC SECTION (5, 7, 8, 37)

The arithmetic section of a machine may operate either on the successive digits of a number, one after the other time-wise, utilizing the same equipment for each digit; or it may operate simultaneously on all digits of a number, using duplicate equipment to handle all digits. In the first case, the machine is said to be a *serial machine* and it performs arithmetic operations as people are accustomed to do. In the second case, the machine is said to be a *parallel machine* and it obviously achieves a higher speed of operation at the expense of additional equipment. It is generally necessary for the arithmetic section to store three numbers within itself: the two factors entering an arithmetic operation and the result. A device that has the ability to store one word of information is called a *register,* and hence an arithmetic section generally contains three registers. A register may be made from flip-flop circuits or from other kinds of storage devices. It must be capable of accepting a number, of holding it for as long as required, and of then emitting this number upon request. Frequently a register may also need to be able to shift its contents, either left or right, one place at a time. Such shifting is equivalent to moving the point and can therefore be used to change the relative values of the number in the register, i.e., to scale the number. Such shifting features are also used in the general manipulation of data.

In addition to the three registers of the arithmetic section, there will also be contained an *adder* which can form the sum of two numbers presented to it. The adder may be either serial or parallel and, as before, the operating rate may be increased at the expense of additional equipment. As an example of the design technique making use of the logical algebra approach, consider the binary adder which must obey the following addition table. Note that a stage of an adder has really three inputs: one digit from each of the two numbers that are to be added, and a possible carry from the previous place.

Digit A	Digit B	Carry-in C	Sum-out	Carry-out
0	0	0	0	0
0	1	0	1	0
1	0	0	1	0
1	1	0	0	1
0	0	1	1	0
0	1	1	0	1
1	0	1	0	1
1	1	1	1	1

It can easily be verified that the following propositions (logical statements) completely define the behavior of this binary adder.

$$\text{Sum-out} = AB'C' + A'BC' + A'B'C + ABC$$

$$\text{Carry-out} = AB'C + A'BC + ABC' + ABC = AB + BC + AC$$

Here the primed letters are negations of the unprimed variable; that is, if A has value 1, then A' has value 0, and vice versa. For example, the sum-out is to be 1 if $AB'C'$ is 1 *or* $A'BC'$ is 1 *or* $A'B'C$ is 1 *or* ABC is 1. $AB'C'$ is 1 if $A = B' = C' = 1$ which in turn implies $A = 1$ but $B = C = 0$; this is the third entry in the addition table above. From this logical description of the adder it is a straightforward transition, by way of any of the previously described gate structures, to physical hardware which has adding ability.

Figure 7.6 illustrates a typical block diagram of an arithmetic section. The combination of register I and the adder is generally called the *accumulator*. Hence, one number that is presented to the adder is from register I. The second number that is to be presented to the adder for addition to the first comes from storage via register III and via the *complementer*. The function of the complementer is to represent numbers in what is known as "complement form," and thereby to effect the process of subtraction although actually performing an addition. Hence, this relieves the complexity of the design problem with respect to the adder, which otherwise might have to have both adding and subtracting features. After the new sum is formed, it replaces the previous information in register I.

The end-to-end connections between register I and register II indicate that it is possible to shift information from one register into the other, one character or digit at a time. Such shifting properties are sometimes convenient in the logical manipulation of data. Paths are shown from storage into registers II and III, and return paths to storage are shown from register II and the accumulator.

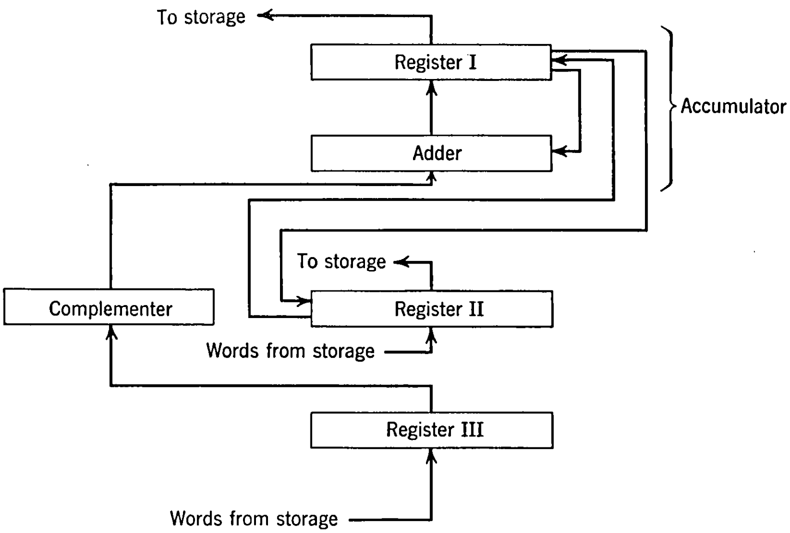


Fig. 7.6 Block diagram of a typical arithmetic section

7.13 THE STORAGE (5, 7, 8, 28, 38)

The storage must retain data and instructions for as long as required and must be able to locate any information within itself in response to a request accompanied by an address. Many physical phenomena may be used as a basis for a storage device. One of these is magnetism, in which the two states of magnetic flux represent the two values of a binary digit. Magnetic devices appear as drums, tapes, and the recently introduced magnetic core.

In each of these magnetic devices the two states of the binary variable (0 and 1) are made to correspond to either two physical orientations of elementary magnets (the drum or tape) or to the two states of remanent flux in a closed magnetic circuit (the core). Figure 7.7 indicates the salient features of the magnetic-core storage technique (27, 30). The annuli which are used to store one bit each are typically 0.1 inch outside diameter by 0.06 inch inside diameter by $\frac{1}{16}$ inch thick. A usual construction is in the form of a rectangular grid, at each intersection of which a magnetic core is threaded by three or more wires. The material from which such cores are made exhibits an essentially square hysteresis loop; therefore, it can be seen from Fig. 7.7 that a current of value I will not cause the core to transit from one state of saturation to the other, but that a current of $2I$ will. Hence, by selecting a particular wire in the X coordinate and a particular wire in the Y coordinate, the core lying at the intersection can be subjected to a driving current of $2I$, but all other cores will be subjected to a driving current of only I. Thus one core from the entire array is selected for reading.

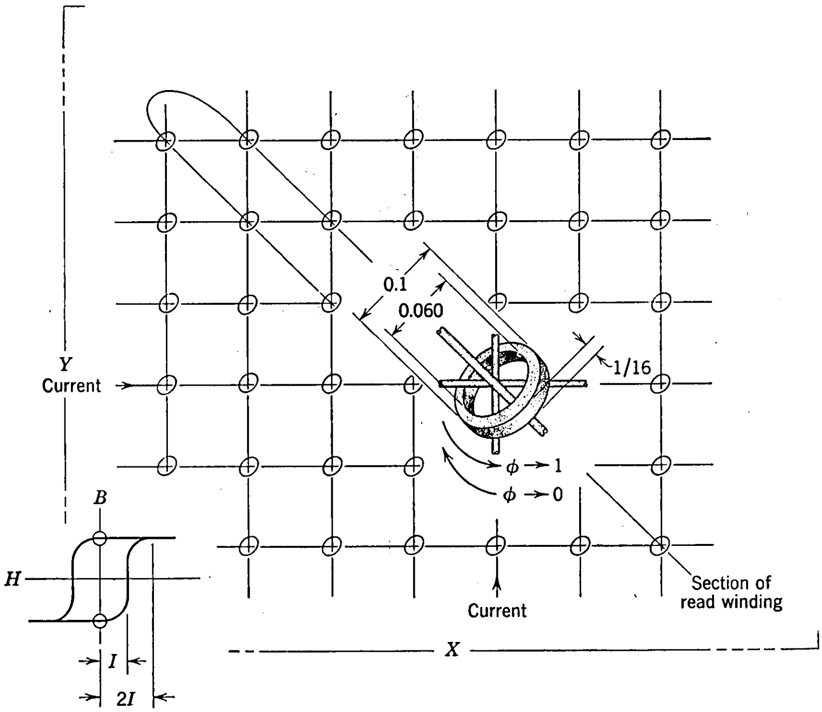


Fig. 7.7 Basic principles of a coincident current magnetic-core store

If the core is driven by the interrogating current toward the upper saturation level, then there will be or will not be a large flux change within the core, depending on whether it was previously at the lower saturation point or at the upper saturation point. Hence, an inspection for the presence or absence of a voltage in the read wire which threads all cores will indicate the nature of the information previously stored.

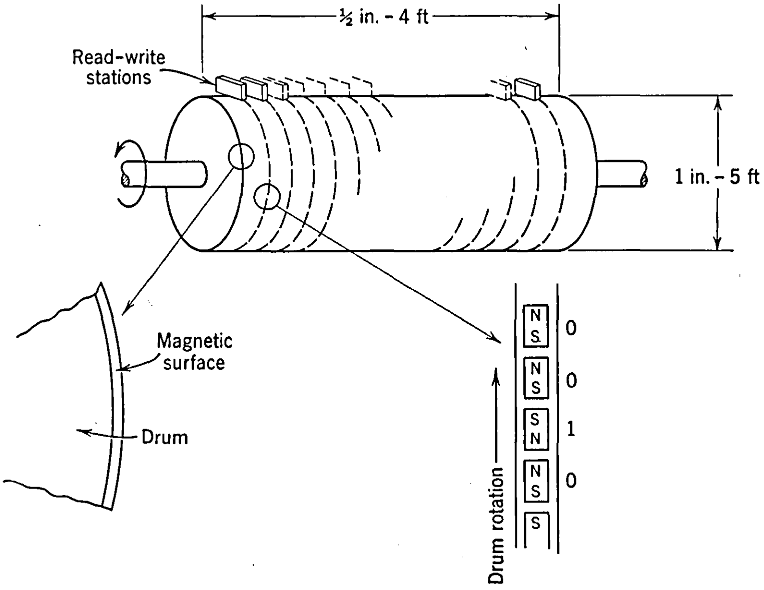


Fig. 7.8 Basic principles of a magnetic-drum store

Figure 7.8 indicates the principal features of the magnetic drum (39); and Fig. 7.9, of the magnetic tape. In each case, there is a magnetic surface on which elementary magnets can be produced by suitably constructed read-write stations. These are usually transformers with a small amount of leakage flux in an air gap which can then influence the magnetic surface beneath. By a suitable control of the current in the writing head, a succession of elementary magnets may be laid down on the magnetic surface. In Figs. 7.8 and 7.9, the convention has been taken that a magnet presenting its south pole first to the reading station represents a binary 1, whereas a magnet presenting its north pole first represents a binary 0. The leakage flux of the elementary magnets threading the air gap and magnetic structure of the read heads induces a voltage in the coils of the read head; hence the magnetic disposition of the surface can be obtained and so the stored information recovered. Typical geometrical configurations and dimensions are indicated in the respective figures.

Electric charge, either through presence versus absence, or as distribution (31), may also be used as a basis for a storage device. In the Williams-type store (32), a conventional cathode-ray tube is used, whereas in the selectron (29) or in the MIT storage tube, a special form of cathode-ray tube is used. Any physical device capable of producing a delay (in the same sense as in an electrical delay line) may also be used as a store. In this case the information is retained by causing it to propagate through the delay medium; and as it emerges from one end, it is retimed and amplified and reinserted at the other end. Hence, storage is accomplished by causing the information to circulate continuously. Electrical delay lines, acoustic delay lines of quartz or of mercury, magnetostrictive lines of nickel, or suitably used magnetic drums have served as the delay element.

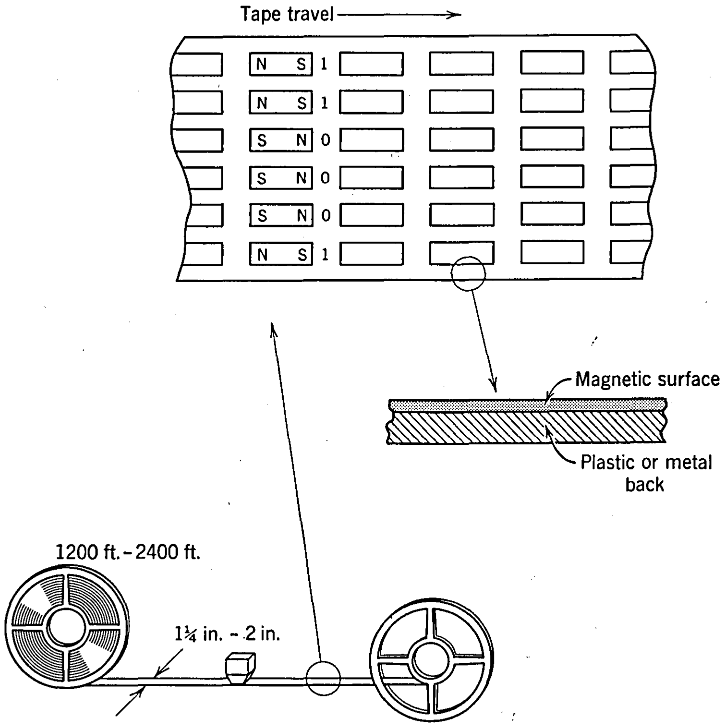


Fig. 7.9 Basic principles of a magnetic-tape store

In the interest of economics and conservation of equipment, the principal storage is frequently organized in a hierarchy. Because of the high cost of the fastest type storage (magnetic core or electrostatic for a parallel machine), not all the required storage facilities are provided of this type. Rather, the main storage is backed up by a larger, slower, and therefore less expensive storage (typically a drum) which in turn is backed up by a yet larger, yet slower, and yet more inexpensive storage (typically a large number of tapes), etc., as required.

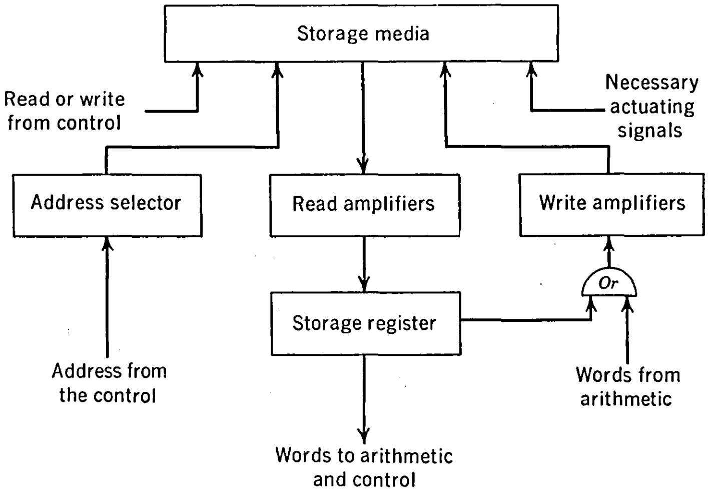


Fig. 7.10 Block diagram of a typical store

In addition to instructions for transferring data from storage to and from the arithmetic, input-output, and control sections, there may also exist special instructions for transferring data among the several levels of the storage hierarchy.

A block diagram of a typical storage system is illustrated in Fig. 7.10. Generally the output of the storage media is of sufficiently low signal level that read amplifiers are necessary. Very commonly the act of reading the information from the storage media also destroys this information ("destructive reading"); hence, in order to provide a "nondestructive read" facility, the information from the read amplifiers is inserted into the storage register; and subsequently, via the write amplifiers, is reinserted into the same location in the store. Alternatively, information from the arithmetic or input-output section may be routed via the write amplifiers into specified locations of the storage media.

In addition to being presented with information specifying the word, the store must also receive information from the control specifying the address at which this data is to be stored. Note that the address is not stored in the storage. The address merely serves to select one of the locations within the storage media; and at that location, the information from the write amplifiers is actually stored. In addition, the store will need to know—again from the control—whether the cycle it is performing is to be a read or write function; the store will also require a necessary sequence of actuating signals.

7.14 THE CONTROL SECTION (5, 7, 8)

The control section must withdraw instructions one by one from the storage, decide what each instruction means (decode it), and then cause the balance of the machine system to perform the appropriate operations in the proper sequence to achieve the result indicated by that instruction. The control is chiefly a collection of flip-flop circuits to remember that certain steps of a process have been passed, of gates to make the necessary decisions, and of pulse signals to effect the operation desired in the balance of the machine. The control, since it paces the rest of the machine, effectively supplies the actuating signals to the balance of the machine. In general, there are two ways of accomplishing this. The control may contain an oscillator (a *clock*), in which case the machine is said to be *synchronous* (or *clocked*) and no event happens within the machine except at clock intervals. Alternatively, the control may be organized such that the entire system is allowed to operate at a rate that is determined solely by its internal time constants, in which case events occur at a rate determined only by the physical configuration, the design, and the stray capacitances of the circuit. In this case the machine is called *asynchronous* or *self-timed*. In the asynchronous machine it is possible for the rate of operation to change radically without any ill effects; the clocked machine requires that all events occur on a regular schedule if proper operation is to be achieved.

Figure 7.11 illustrates a block diagram of a typical control. In this figure the solid lines indicate the flow of actual information, whereas the dotted lines indicate only control channels. The instruction word removed from storage is kept in an *instruction register* for the duration of the execution of that instruction. A *decoder* inspects the operation part of the instruction and decides what parts of the machine must be operated, and in what sequence. The output of the decoder, when combined with a clock or other signal generators, constitutes the signals that are routed to all other parts of the machine at appropriate times and in appropriate sequence.

The address part of the instruction may be directed to the storage when the instruction requires a piece of information from the storage; or it may be directed to the instruction counter, if the particular instruction being executed requires that the next instruction be brought not from the immediately following storage location but from an entirely different part of the storage (a *branch* or *jump instruction*). The instruction counter is normally advanced by one step at the close of execution of each instruction and, hence, in the normal course of events, instructions are withdrawn from the storage sequentially; only when there is a conditional or unconditional jump to another part of the routine is the action of the instruction counter interrupted and the contents replaced by a new value. When the instruction is a shift instruction, in which the address indicates not a location in storage but the number of places by which a given register is to be shifted, the address part of the instruction must be routed to the multiply-

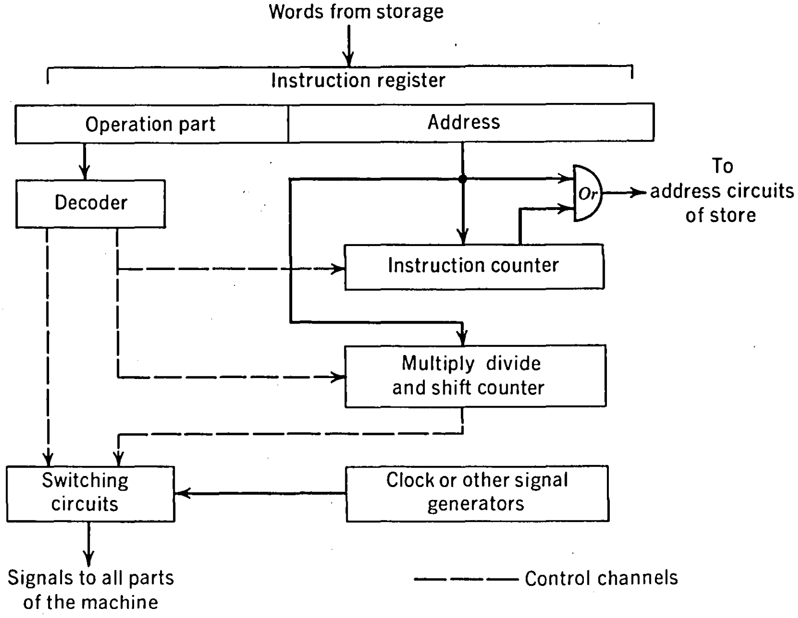


Fig. 7.11 Block diagram of a typical control section

divide and shift counter. This counter also must receive control information from the decoder in the event that a multiplication or division instruction is indicated.

7.15 THE INPUT-OUTPUT SECTION (5, 7, 8, 20)

In the automation application, the input-output section assumes great importance because of the necessity for communicating between the computer proper and possibly large numbers of many kinds of other equipment and instruments. In the scientific, engineering, or business application of a computer, the machine system commonly communicates only with man, whereas in the control or automation application, the machine system may more commonly communicate with other machines or instruments. The input-output may be thought of as a transducer between the machine and the external world. Alternatively, it may be thought of as a dictionary that accommodates machine language on the one hand and world language on the other hand. The variety of input-output possibilities is so vast that details of each kind cannot be given. A number of the more common input-output devices in each of the several categories is tabulated:

(1) Devices for communicating from man to machine.
  (*a*) Keyboards.
  (*b*) Punched-card readers.
  (*c*) Punched-tape readers, mechanical and photoelectric.
(2) Devices for communicating from machine to man.
  (*a*) Printers.
  (*b*) Typewriters.
  (*c*) Card punches.
  (*d*) Paper tape punches.
  (*e*) Plotters.
(3) Devices for communicating from digital machine to other machines.
  (*a*) Language translators.
  (*b*) Analog-to-digital converters (18, 41).
  (*c*) Digital-to-analog converters.
  (*d*) Magnetic tape devices.

In each case the particular input-output device must be surrounded with sufficient electronic gear so that it can provide two distinct kinds of information on the two sides, world language on the one side, machine language on the other. It may also be necessary to provide additional equipment on each input-output device to accommodate a mismatch in speed of operation between that device and the faster computing system proper.

7.16 THE DIGITAL DIFFERENTIAL ANALYZER (16)

An example of a special-purpose digital computing device is the digital differential analyzer. This machine performs essentially the same operations as the electronic analog computer, except that it accomplishes the process of mathematical integration by means of a numerical approximation. This machine is a very highly organized machine internally and does not have an instruction list in the same sense as the general-purpose machine previously described. It has the advantage over the mechanical differential analyzer in that it can perform integration with respect to any variable, whereas the electronic analog computer must integrate with respect to time. In addition, the digital differential analyzer can exhibit unlimited precision where in general the precision of the result can be traded against the speed of solution. It has the difficulty of all digital systems, that an error may arise from the replacement of the infinitesimal calculus by finite difference approximations. An additional error may also arise, again as in all digital systems, as a result of the truncation of numbers to the length that the machine system can accommodate. Generally, the digital differential analyzer accomplishes the process of integration by approximating the value of the integral as the sum of the areas of a sequence of rectangles. If required because of greater precision, such a machine could also be constructed to perform parabolic or higher-order numerical integration techniques.

Information within the digital differential analyzer is represented partly in a true number system and partly as a rate of a train of pulses. The initial value of the integrand is retained in a register whose contents may be altered by incremental values as required by changes in the value of the integrand. The contents of this register is then inserted into an adder each time the independent variable or variable of integration is considered to advance one increment. The sum attained by this adder over the period of integration then represents the value of the desired integral. A convenient internal arrangement is provided in the digital differential analyzer to permit communication between the various integrators which the machine contains. Commonly, digital differential analyzers may contain as many as eighty to a hundred integrators.

7.17 APPLICATIONS

Because the digital data-handling system is a relatively recent technique, the number of applications for which it has been used is relatively small. What can be described as an application depends considerably on how automation itself is defined. If it is assumed that a legitimate application is any one which contributes to the notion of an automatic factory, then inventory control, production scheduling, scheduling of shop loading, scheduling of machine loading, accounting, distribution of costs, payroll, sales quotas, and similar problems generally classed as business applications are legitimate applications of the digital system to automation. To most of these business problems the scientific and engineering machine may be applied, although it is sometimes inconvenient to use a machine organized for scientific work. For the application noted above, notice a common feature: The input is all numeric or alphanumeric in nature and for the most part does not come from measurements on physical systems. These applications are data processing or information handling and represent for the digital system the largest field of application—other than scientific computation—to date.

Another level of application is the direct control of manufacturing processes. At least one paper has been published describing a projected application in oil refinery control (17); here the digital machine receives input from measuring devices in the refinery, computes the refinery performance, and returns control information to valves, heaters, etc., in the refinery. A similar application which has been described publicly is the machine that receives flight plans of aircraft and predicts their progress and times of appearance at frequent check points, searches for collision courses, and automatically alarms when planes are overdue or fail to report.

Typical factors suggesting that digital equipment is the appropriate choice for a particular application are these:

(1) Is an answer of great precision demanded?

(2) Is the data in numeric form, or alphanumeric form, and must the result also be in this form?

(3) Is something more than simple control of a variable required? Must the statistical distribution of the variable or the trend of the variable be controlled; or must the future behavior of the variable be predicted?

(4) Must relatively involved decisions be made by the machine system?

(5) Is the time of solution or time of response required to be exceedingly short?

(6) Can money, equipment, or labor be saved by using one digital machine to handle several problems of control on a time-sharing basis?

(7) Is uncanny freedom from error required?

7.18 THE FUTURE

It appears that for the future the digital machine will share with the analog machine system the control applications which are directly at process level. However, above direct process control there is a level of management activity which requires handling large masses of data, performing extensive computations, and making complex decisions. In this area of general information processing, the digital system has a clear field. After all, any information whatsoever can be represented in a suitable numeric or alphanumeric code, and the digital system can deal with such information to any extent required.

Examples of new fields of application currently being proposed or possible in principle are these:

(1) The control of street traffic flow.

(2) Completely automatic bank systems, including sorting and handling of checks and documents (19).

(3) New high-speed chemical processes with digital control.

(4) Digitally controlled machine tools (21).

(5) Airport controller assistance.

(6) Personnel training.

This last represents perhaps the ultimate in sophistication. Here the digital machine contains within itself a mathematical model of the given situation—say, an economy. The trainees, perhaps fledgling executives, are effectively playing a game—here the game of "business"—against the machine. Each move of the opponent is weighed by the machine in terms of its model and the effects of the move on the model computed. Hence, the effects of a given business decision is promptly known; and, in this example, no company funds, contracts, or position are risked while exposing the trainee to a greater-than-a-lifetime's breadth of experience.

7.19 REFERENCES

1. J. M. Salzer, *Proceedings of the National Electronics Conference,* Vol. 8, 621, 1952; *Radio Electronic Eng.,* 13, Feb. 1953.

2. Standards on Electronic Computers, *Proc. IRE, 39:* 271, March 1951.

3. First Glossary of Programming Terminology, Association for Computing Machinery, June 1954.

4. W. H. Thomas, *Proc. IRE, 41:* 1245, Oct. 1953. G. M. Hopper and J. W. Mauchly, *Proc. IRE, 41:* 1250, Oct. 1953.

5. A. D. Booth and K. H. V. Booth, *Automatic Digital Calculators,* Academic Press, Inc., New York, 1953.

6. Maurice V. Wilkes *et al., Preparation of Programs for an Electronic Digital Computer,* Addison-Wesley Publishing Co., Inc., Cambridge, Mass., 1951.

7. *High Speed Computing Devices,* Engineering Research Associates, McGraw-Hill Book Co., Inc., New York, 1950.

8. R. K. Richards, *Arithmetic Operations in Digital Computers,* D. Van Nostrand Co., Inc., New York, 1955.

9. *Synthesis of Electronic Computing and Control Circuits,* Harvard Computation Laboratory, Annals 27, Harvard University Press, Cambridge, Mass., 1951.

10. William Keister *et al., Design of Switching Circuits,* D. Van Nostrand Co., Inc., New York, 1951.

11. R. Hamming, *Bell System Tech. J.,* April 1950.

12. Robert Serrell, *Proc. IRE, 41:* 1366, Oct. 1953. A. W. Burks and J. B. Wright, *Proc. IRE, 41:* 1357, Oct. 1953.

13. C. L. Shannon, *Bell System Tech. J.,* Jan. 1949.

14. J. O. Paivinen and I. L. Auerbach, *Trans. IRE Professional Group on Electronic Computers,* June 1953.

15. M. Rubinoff, *Elec. Eng.,* 905, Oct. 1952.

16. M. J. Mendelson, *Aeronaut. Eng. Rev., 13,* Feb. 1954.

17. Sibyl Rock, *Computers and Automation, 3,* March 1954.

18. Chapter 9 of this book.

19. G. Brown and L. N. Ridenour, *Proceedings of the Western Joint Computer Conference, February 1953,* IRE, June 1953.

20. Review of Input-Output Equipment Used in Computing Systems, *Proceedings of the Eastern Joint Computer Conference, December 1952,* AIEE, March 1953.

21. Chapter 16 of this book.

22. S. E. Gluck *et al., Proc. IRE, 41:* 1388, Oct. 1953.

23. D. Brown and N. Rochester, *Proc. IRE, 37:* 139, Feb. 1949.

24. C. L. Wanlass, *Trans. IRE Professional Group on Electronic Computers,* Aug. 1952.

25. J. C. Chu, *Proceedings of Toronto ACM Meeting,* Sept. 1952.

26. *Proc. IRE, 41:* 1287 (Ross), 1300 (Greenwald *et al.*), 1380 (Elbourn *et al.*), Oct. 1953.

27. J. Rajchman, *Proc. IRE, 41:* 1407, Oct. 1953.

28. I. L. Auerbach, *Elec. Mfg.,* 100, Oct. 1953, and 136, Nov. 1953.

29. J. Rajchman, *RCA Rev.,* 53, March 1951.

30. W. Papian, *Proc. IRE, 40:* 475, April 1952.

31. J. P. Eckert *et al., Proc. IRE, 38:* 475, May 1950.

32. F. C. Williams and T. Kilborn, *Proc. IEE* (London), *96,* Part II: 183; Part III: 77, 1949; *97,* Part IV: 455, 1950.

33. A. L. Samuel, *Proc. IRE, 41:* 1223, Oct. 1953.

34. M. V. Wilkes, *Proc. IRE, 41:* 1231, Oct. 1953.

35. C. L. Shannon, *Proc. IRE, 41:* 1234, Oct. 1953.

36. M. Rubinoff, *Proc. IRE, 41:* 1254, Oct. 1953.

37. H. D. Ross, *Proc. IRE, 41:* 1287, Oct. 1953.
38. J. P. Eckert, *Proc. IRE, 41:* 1393, Oct. 1953.
39. J. H. McGuigan, *Proc. IRE, 41:* 1438, Oct. 1953.
40. G. S. White, *Proc. IRE, 41:* 1450, Oct. 1953.
41. H. J. Burke, *Proc. IRE, 41:* 1455, Oct. 1953.
42. *Proc. IRE, 41:* 1262 (Buchholz), 1275 (Frizzell), 1294 (Huskey *et al.*), 1313 (Sheretz), 1325 (Erickson), 1332 (Crossman), Oct. 1953.
43. Review of Electronic Digital Computers, *Proceedings of the Eastern Joint Computer Conference, December 1951,* AIEE, Feb. 1952.
44. *Proceedings of the Eastern Joint Computer Conference, December 1953,* IRE, 1953.

8. Data Processing

JOHN W. MAUCHLY

Remington Rand UNIVAC
Philadelphia, Pennsylvania

8.1 THE ECONOMIC JUSTIFICATION FOR DATA-PROCESSING EQUIPMENT

8.1.1 The Economic Basis for Data Processing

Although the term "data processing" is relatively new, the process which this term describes has been an essential part of our business life for a long time. Data processing has always required a great deal of manual and clerical work, but some of the steps in any data-processing job have been recognized as sufficiently repetitious and routine to be dealt with by special equipment. All sorts of adding machines and desk calculators have thus found their way into our offices and our industrial laboratories. Although in recent years punched-card equipment has been used in many engineering and scientific calculations, the largest use of such machines has been in accounting, bookkeeping, and related work.

The basic fact behind almost every use of mechanized equipment, whether in the office or elsewhere, is that the *cost* of data processing is thereby reduced. The equipment is justified upon economic grounds.

8.1.2 A Handy Yardstick

When the application being studied is already being carried out by some method, economic justification may be obtained by comparison of the cost of proposed methods with that of the one presently in use. Let us take as a simple first example the case in which the present method involves no capital investment in equipment. The only cost associated with that method is the salary of a clerical worker. If that worker receives $3000 a year, then the substitution of equipment to perform the same work can be justified if such equipment costs less than $3000 per year. However, this rough rule-of-thumb is bound to be conservative, since there is always an overhead or burden cost associated with the direct salary or wage cost.

In most cases, the clerical worker is still needed for the operation of the equipment and an alternative statement of the rule-of-thumb is to say that a piece of equipment which doubles the production obtained from one person is worth at least $3000 per year.

When equipment is to be purchased rather than rented, the consideration of interest rates, depreciation, amortization, maintenance, and other factors which may vary depending upon the particular kind of equipment involved are used to assess the capital investment which is reasonable to obtain double productivity from the clerical worker. A device that depreciates quickly and has high maintenance costs might be worth only $10,000. For typical equipment, a reasonable figure is higher, perhaps $40,000. Bearing in mind the overhead on clerical work, this is still likely to be on the conservative side.

Another factor to be considered when making such a comparison is the fraction of the total machine time likely to be usable. Service and maintenance may slightly reduce the total, but even on the very largest electronic general-purpose computers no more than a few per cent of the otherwise available time is lost in this way. It must be recognized that sometimes a device cannot be fully loaded by a particular application under study. We must then decide whether to find other applications for the remaining available time, or whether the particular application will already provide an economic justification, even though the machine may be idle for a considerable fraction of its available time. In general, companies that can use a large general-purpose computer will easily find additional tasks which will occupy all of its available hours.

Smaller office equipment, more specialized in application, will often be found to pay off even though idle during a large part of the day.

8.1.3 Examples for Small Machines

Let us consider a bookkeeping machine costing $2500. Assume that it is capable of doubling the production of a clerical worker. Since we have seen that we can certainly afford to spend more than $10,000 for a device which will do this, it is clear that the bookkeeping machine will still be a valuable investment, even if it is used only 25 per cent of the available time.

As a second example, let us consider an adding machine which costs $250. This is only 2.5 per cent of our very conservative $10,000 investment, but if productivity is increased by a mere 20 per cent, while it is used, and it is used only 2 hours out of every 8, we might well figure its value as 5 per cent of $10,000, or $500.

8.1.4 Semiautomation by Punched Cards

Punched cards have made data processing more automatic and have reduced costs by increasing productivity in accounting and record keeping. The area of profitable application of punched-card equipment overlaps that of adding machines, desk calculators, and bookkeeping machines but includes a great deal more. Card machines are good at sorting, merging, tabulating, and various extensions of these operations. These are not handled at all by the simple adding, calculating, and bookkeeping machines which must receive almost all of their data from the keyboard for each new operation they undertake. Although key-punching machines are usually required to record data on punched cards, such data need be recorded only once. Thereafter, the rate of handling such data is no longer limited by the "recording rate." Standard punched cards accommodate either 80 or 90 digits or characters of information (depending on the card type), and a full punched card may often represent 100 key strokes. Various card machines process these cards at rates from 100 to 1000 cards per minute.

Before the advent of electronic computers, card machines were very slow on purely computational work, although they were, and still are, well adapted and fast for sorting. A typical nonelectronic multiplier in a punched-card system used to require several seconds to produce the product of two six-digit numbers, whereas less than a second is required to read the factors and punch the card. A girl, using a desk calculator costing $1000, can turn out at least three such products per minute. If we assume thirty products per minute from the nonelectronic card multiplier, we find that the card machine rate is thus ten times greater. A rough economic comparison will then indicate that such a card multiplier might be worth $2000 per month. The rental for the older equipment was actually much less than this. Hence, such a nonelectronic card calculator could be justified in a system requiring enough multiplication to keep such a special-purpose machine occupied for a reasonably good fraction of its time.

With electronic card calculators available today at rates of less than $2000 per month, dozens of multiplications or other arithmetic operations may be carried out in less time than it takes to read a card and punch the results.

Moreover, the pre-electronic card machines could do only one or two operations on each card pass. The same cards had to be transferred from one machine to another to accomplish different steps of a sequential process, and all such card transfers were done by hand. The present-day electronic card calculators not only speed up the rate of processing because of their fast internal operation, but they decrease the human labor involved in card handling and contribute to a much higher degree of automatic operation as a result.

8.1.5 Transition to Automation through Large Computers

Before the entry of electronics into the computing and data-processing field, several types of large calculating instruments were in use by engineers and scientists. These were, for the most part, merely larger and more flexible versions of the small special-purpose analog devices which had long been used by engineering designers for the solution of special problems. Network analyzers, which employ a model electric network to study the behavior of large electrical power distribution systems, became popular tools for the solution of nonelectrical problems as well. A variety of linear partial differential equations, with many kinds of boundary conditions, can be readily solved on a network analyzer. However, nonlinear equations can present extreme difficulty, since nonlinear elements are then required in the network.

A quite different analog device, originally conceived by Lord Kelvin and introduced as an engineering tool at MIT, is the mechanical differential analyzer. The chief use of the differential analyzer is in the solution of sets of ordinary differential equations. Methods have been described for solving certain partial differential equations on a differential analyzer, but this does not represent the routine bread-and-butter application of such equipment. Although there are still a number of mechanical differential analyzers in use, several computing centers at which modern electronic digital computing equipment is now available have stopped using the differential analyzers on which they once depended. Electronic differential analyzers become handy tools for engineering design of control systems. The cost of an electronic analyzer, equivalent in complexity to one of the older mechanical analyzers, is considerably lower, *and* the speed at which it produces solutions is substantially greater. Also, the electronic type is easier to set up, and it is much easier to vary parameters.

All analog computers have inherent limitations upon the accuracy of the solutions that they are able to provide. Digital devices, by contrast, can be constructed or programmed to satisfy any desired accuracy requirement.

It is interesting to note that there is a type of digital device which functions somewhat like the analog differential analyzer. Such machines are known as digital differential analyzers, although they are actually "finite" difference calculators rather than true differential devices. Their chief application is again in the solution of ordinary differential equations. Being digital, they can work to a greater accuracy. Certain variants of the digital analyzer which can give drift-free calculations appear to be useful computers for incorporation into complex servo control loops.

Some of the big analog computers cost as much as the largest of the new electronic digital computers. An investment of this magnitude in an analog computer is justified only if the class of problems that this computer can solve are *certain* to remain at high volume for a number of years. In the light of current advances in automatic coding for general-purpose digital equipment, it is far from obvious that the investment in a large analog device is wise, even supposing that enough problems within its scope can be supplied to keep it busy full time.

8.1.6 The First Large Automatic Digital Computer

The first large automatic digital computer was a nonelectronic machine built at Harvard by Dr. Howard Aiken and IBM. This electromechanical computer, known as Mark I, is a general-purpose device; hence it suffers none of the limitations of scope or accuracy characteristic of analog computers. Under the control of a punched paper tape, Mark I carries out long and complex numerical calculations without human intervention. As Dr. Aiken has often pointed out, Mark I followed closely the principles expounded by Charles Babbage in the last century. Through the use of modern electromechanical components, Dr. Aiken was able to bring to practical and operational reality a dream that Babbage was unable to translate into equipment because the necessary technology was lacking.

Mark I became the first true example of automation in computation since it carries through an entire calculation, however long and complex, without human attention. Because Mark I is a general-purpose digital computer, it can handle types of problems entirely outside the scope of any of the analog computers. Like all such digital computers, it can produce results of any desired accuracy. Its cost has been estimated as $1,000,000. Dr. Aiken went on to design further computers, and an electronic Mark IV now stands facing Mark I at Harvard. Modifications have been made in Mark I to enable convenient use of subroutines in its program. Mark I is still operating day in and day out at the Harvard Computation Laboratory.

8.1.7 Automatic Computation by Electronics—the ENIAC

An entirely different approach toward the automation of digital computing was taken by a group at the University of Pennsylvania during the war years when the ENIAC, the first large-scale electronic digital computer, was designed and built for the Aberdeen Research Laboratory. Begun in June 1943, and completed in December 1945, the ENIAC demonstrated the practical possibility of using the high speed of electronic circuitry to reduce the cost of complex computations. Today, when stock models of electronic computers for business and scientific applications are available from a number of different manufacturers, it may seem incredible that the very first electronic computer was put in operation only ten years ago. With numerous more powerful computers around us today, it may still be of interest to note that the economic justification for the ENIAC was not difficult to show, once the workability of such equipment was demonstrated.

For example, the ENIAC is able to compute a trajectory of an artillery shell in 30 seconds—about half the time required for the shell itself to reach its target. On the differential analyzer of the mechanical type used prior to 1945 for such a trajectory the same computation would take 20 minutes. The same work done by hand with the aid of a desk calculator requires something like 20 hours. We can, therefore, say that the ENIAC is, on this problem, equivalent to 40 mechanical differential analyzers or to 2400 girls equipped with desk calculators. The cost of the ENIAC is estimated at less than half a million dollars. We see that, if hand calculation is used as our basis of justification, ENIAC is easily valued at $10,000,000, even if used only *half* of the time and left idle the other half.

However, comparison with hand calculation is not proper when some better means is already available. The differential analyzer, which compares with hand computation by a margin that is not exciting, might be valued at $600,000, whereas its actual cost was perhaps half of this. The ENIAC, on the other hand, having a productive rate equal to forty mechanical differential analyzers, is then worth about $12,000,000 when used full time for one shift.

It should also be taken into account that these machines can produce two, or even three, times as much output with only a small additional expense for operating staff on several shifts, but without increase of the total investment.

8.1.8 The Fundamental Economics of Electronic Computation

There need be no mystery concerning the advantage that an electronic general-purpose digital computer has over a nonelectronic or mechanical device for the automation of data processing. Signals can be switched electronically at rates of over 1,000,000 per second, while a switching rate of 1000 per second is fast for reliable electromechanical relays or other nonelectronic devices. Yet when the cost of the electron tubes and circuitry is compared with the cost of the electromechanical relays which might perform a similar function, there is not a great deal of difference. It is true that a single relay may carry a number of independent contacts, but increasing the contacts on a relay cannot be carried far without sacrifice of speed or reliability. Hence vacuum tubes or other electronic elements should be able to turn out somewhere between 100 and 1000 times as much work for the same capital investment. Of course this is an extremely crude argument; in practice, ratios approaching 1000 have not been achieved. There are many problems in the handling of the data *before* it gets to the electronic processing stage, and still other problems have to do with the recording of the output results. In these input and output categories, the economies deriving from electronics are not nearly as great as those in the pure arithmetic and logical functions which are performed in the computer proper.

Even such rough considerations make it clear that, in seeking better and cheaper methods for the processing of data, much may be gained by exploring what electronics has to offer, whereas possible improvements in the performance of mechanical systems are likely to be much less spectacular.

8.1.9 Fields of Application

Every data-processing device has some proper field of utility. Lest the general statement in the preceding section be taken without reservation, let it be pointed out that every kind of data-processing instrument has some field in which its application is economically justified. No matter how good electronic machines become, there will still be valid applications for punched cards (or for some successor to punched cards which involves a unit record that can be handled, mailed, and read by the general public). Likewise, there will always be applications for desk computers and adding machines. To carry this to the extreme, it is unlikely that pencil and paper will soon disappear.

Pursuing the same idea in a different direction, we must recognize that not all computation should be done digitally—there are innumerable places where special analog computers will always prove to have better economic justification. There is, of course, no general rule whereby we can draw an exact boundary line and say that on *this* side of the line the digital machine is always the better one and on *that* side the analog machine is always to be preferred.

Just so, we cannot lay down any strict rule for when a special-purpose computer is better than a general-purpose one. If such a rule were worked out today, it would be outmoded tomorrow with the advance in computer techniques, both analog and digital. More important than any rule is the fact that there is not today a sufficient familiarity with the ways in which general-purpose digital computers can be made to perform (in an economic fashion) jobs which have heretofore been done in other ways, sometimes manually, sometimes by analog computer, and sometimes by special-purpose equipment. The development of automatic coding techniques, which will be discussed later in this chapter, are expected to have a considerable influence on this situation.

There is a human tendency to subdivide fields of application categorically and say that one type of machine takes care of all work in one classification, and that some other type of machine must be used when attacking problems in another class. With respect to general-purpose electronic digital computers, such statements are often heard. In the next major section of this chapter, devoted to the requirements of data-processing machines, this question is discussed.

8.2 BUSINESS AND SCIENTIFIC COMPUTER REQUIREMENTS

8.2.1 Economic Basis Common to Business and Scientific Applications

In almost any discussion of electronic digital computers, a distinction is made between scientific applications and commercial and business applications. The implication is that a type of computer which is *good* for scientific work is *not* suited to commercial and business problems, and that one which is *well suited* to the latter will be *ill-adapted* to scientific work. There are circumstances and particular examples of application which make such a distinction seem reasonable, but to the present author this dichotomy is misleading. In what follows the applications will be distinguished rather by the demands they make on the input-output equipment relative to the internal processing.

First, however, it should be apparent that this very question of the suitability of one or another computer for one or another application implies a common basis for judging the applicability of various computers to such divergent application. That common basis is, in general, the economic one. Whether a computer is to be used mainly in the solution of partial differential equations or principally in keeping accounting records, the ultimate criterion which should govern its selection is usually the same—the computer is a tool for producing results, and the proper choice is that one which gets the desired results at the lowest cost.

Generally, but not invariably, the *larger* automatic data-processing systems are able to show the *lower* cost if kept busy. A more or less constant work load at a high level can usually serve to justify a large data-processing system. In scheduling and managing such a system, the general aim is to avoid idle time and peak loads by spreading the work as evenly as possible.

Any system which deserves the name "general-purpose" is sufficiently flexible to handle a variety of problems so large that there should be little difficulty in finding low-priority problems which can be inserted into the computer production line at times when the immediate demands for high-priority results do not exist. A computer designed primarily for accounting problems may nevertheless be used on scientific problems, even though it may not be as efficient as a scientific computing system expressly designed for a given type of problem. Likewise, a computer normally loaded with scientific work may be used for accounting or other business purposes when its scientific problem load is low. Clearly, if one and the same type of computer can be easily used for *all* sorts of problems, the chances of getting an even flow of work and a balanced load are thereby improved.

8.2.2 The Computer-Limited Problem

Some problems make very small demands upon the input and output facilities of a data-processing system. The computation of firing tables and trajectories is of this type. The angle of elevation, muzzle velocity, and a drag function which specifies air resistance as a function of speed, plus certain constants, are specified at the start of the computation. From then on, thousands of steps, each involving many multiplications and other arithmetic operations, must be carried out in order to arrive at a comparatively few results. The solution time is therefore dependent entirely upon the rate at which the arithmetic work can be done. Even rather low-speed and primitive facilities for loading the original data and taking out the final results will not make an appreciable difference in the economy of the operation.

Such a problem is called a "computer-limited" problem. (The opposite extreme is termed "input-output limited.") Computer-limited problems are, more often than not, scientific problems. Some are computer-limited in the extreme. Any computing system which has an extremely fast arithmetic processing unit, but very mediocre input-output gear, is likely to be kept busy, because engineers and scientists can find many problems which require no more than this. However, problems of the computer-limited type are rare in business operations, and such a machine would be of little use in accounting or in any other business record-keeping activity.

The fact that scientists can use a machine having poor input-output facilities, whereas businessmen cannot, certainly does *not* mean that scientific applications are always handled better on a nonbusiness computer. As technical applications of computers grow, the engineering and scientific applications that demand more of the input-output system are coming to the fore.

8.2.3 Business Problems Require Fast Input-Output Facility, Input-Output-Limited Applications

There is an entirely different type of application which may result in limitations associated with the input and output facilities of an electronic computer. Take as a possible example the maintenance of a record file for something like a million customers, each account running to at least 100 digits. Periodically, this file must be brought up to date, even though not every account will be changed in the process. We might assume, for sake of concreteness, that about 10 per cent of the accounts have shown activity during the period since the last updating of the file. Eventually, some system of random-access files will be developed for such purposes, but at present the economics are *against* keeping a file of 100 million digits in any fast-access storage or even a moderately fast-access storage such as a magnetic drum. It *is* economically feasible to keep such a file on magnetic tapes, and in fact this kind of problem is well within the economic range for current magnetic tape machines. Somewhere between 50 and 100 tapes might be required, and the scanning time would be about 5 hours for existing computers. In recently announced systems soon to be available, the scanning time for this file should be less than 3 hours.

On the average, only every tenth item would need any manipulation whatsoever, and the items acted upon might require only a simple addition to one member or transfer to a different tape in the output listing. This then is an example of a tape-limited or input-output-limited problem. The arithmetic or other operations performed on the data by the computer are exceedingly trivial, and the computing system which has such a problem will almost certainly be limited in its capacity to perform by the rate at which the input-output system can feed the data through.

This is what is generally termed a "business problem," i.e., this is a problem typical of business applications. Whether such a problem is found in scientific work depends on the field of scientific application being considered.

8.2.4 The Complete Spectrum of Applications

It is a logical fallacy to assume that the scientific calculations with which our computers have been busy in the past represent a typical cross section of scientific problems in general. First of all, we may note that the computers which have been somewhat less capable with respect to input-output have for the most part been provided with larger internal storage capacity ranging from 4000 to 16,000 words. On the other hand, computers with fast magnetic tapes intended to serve well in handling business problems, and with internal storage of only 1000 or 2000 words, have nevertheless been able to handle much larger scientific problems than could be efficiently solved on the scientific machines.

It is a fact to which many men of scientific and engineering experience can attest, that their field abounds with problems that are fully capable of taxing the best input-output facilities and storage capacities of any proposed data-processing system.

For instance, matrices of order 200 by 200 have 40,000 elements, and at the present time there is no electronic computer with sufficient internal storage capacity to accommodate a matrix of this size without recourse to tapes as an auxiliary storage medium. Nevertheless, matrices of this order *have* been inverted. Such jobs have been done on tape computers commonly described as business rather than scientific computers.

Again, if the solution of partial differential equations is attempted by the usual finite difference techniques, there are already severe limitations as to what can be done in a two-dimensional problem. A 20-by-20 net will have 400 points, and for each of these points not one but a number of values may have to be stored. As soon as we attempt to carry out similar solutions in *three* dimensions, even this coarse lattice requires that the storage capacity be multiplied by 20. The *systematic* nature of the process whereby such a lattice of points is repeatedly scanned in solving a partial differential equation makes tapes an effective auxiliary storage for such purposes. Only a small portion of the total data need be brought into the high-speed or rapid-access internal storage of the computer at any one time.

To demonstrate the feasibility of this systematic use of tapes, a moderate-sized three-dimensional Poisson problem was run on UNIVAC I several years ago. A complete iteration over 2000 net points was carried out every 40 seconds. This was done with tapes since UNIVAC I has only 1000 words of internal storage. Both IBM and Remington Rand now have computing systems with larger internal storage, on which this 2000-point problem would run faster. These machines will have to rely on auxiliary storage for handling anything over 30,000 points.

Another field of application, scientific in nature and capable of using all that can be supplied in the way of input-output facility, includes certain types of large-system simulation problems. For instance, for flood control and flood prediction, any mathematical model that can adequately represent a large river system including many tributaries is complex and exceedingly nonlinear. Presumably many of the properties of the system must be determined empirically by analysis of previous observations. If we attempt to solve this problem using a scale model of the actual system, difficulties arise regarding the hydraulic analogy. In scaling down the dimensions of the system, we should also scale the viscosity of the fluid used. However, it is not practical to use in the model anything other than water; hence, the analogy is by no means exact. In order that difficulties on this score will not be too drastic, the model must therefore be rather large; it then becomes an expensive matter to maintain such a model, which extends over vast areas. In fact, the maintenance cost of such a model may well exceed the cost of a digital computer which could solve the same problems. The computer code could have its constants altered to represent a change in the river bed, or an entirely new channel of flow, much more readily than can the physical scale model.

As a final example of a scientific problem which can easily make strong demands upon the input-output equipment of a computer, consider the field of "linear programming." Here linear inequalities are used to describe the restrictions which must be observed in a scheduling problem, such as the allocation of manufacturing machines in a large plant for the most efficient production of certain specified products in accordance with a specified delivery schedule. That is, we seek to minimize the cost of production or to maximize the amount produced, while at the same time keeping within certain imposed restraints, which of course includes the fact that the manufactured quantities and the goods dealt with must always be nonnegative.

Such problems are characterized by rectangular matrices which in practical cases may be of order 500 by 5000. Even though many of the elements of such a matrix may be zero, it is nevertheless impractical to store all of the pertinent information within the internal storage of existing large-scale computers. Auxiliary storage, such as magnetic tapes, must be used for this data, yet such data must be repeatedly available during the course of solution.

These do not exhaust the examples which could be cited, but it should be clear that there are large classes of problems, usually described as scientific in nature, which are *not* well handled on a computer lacking good input-output facilities.

8.2.5 How Scientific and Business Problems Converge

There is still another reason to be extremely wary of any attempt to distinguish sharply between the "scientific" and the "business" category of computers. There is a growing tendency for the applications themselves to come closer together. For instance, the linear scheduling problem discussed in the preceding section was classed as a scientific application. It may well be considered also as a business problem. The same mathematical techniques are being used at this time to enable an operator of an oil refinery to determine what proportions of different end-products he should make from his crude petroleum, in order to maximize the profit to be obtained through the operation of his refinery. From time to time, as demands fluctuate and the market prices at which he can sell consequently vary, he needs to solve the same type of problem repeatedly in order to readjust his output for the maximum profit.

Is this a scientific problem or a business problem?

The businessman may also use a related mathematical technique for minimizing the cost of transportation between numerous factories or warehouses and the market destinations of his products. Similarly, utility companies which distribute electric power over a large area may have a number of generating stations, each of which operates at a different efficiency which varies with the load. Such companies are now seeking mathematical solutions which optimize their profits, in delivering the required power to consumers, through proper regulation of the various generators to minimize the cost of the *delivered* power. This is different from minimizing the cost of the *generated* power, since the cost of distributing or "transporting" the power from the generator to the customer is different for different patterns of generation and consumption.

The examples just cited certainly have their technical aspects, as well as their economic side. Let us therefore pick as a final illustration something that comes much closer to the normal accounting and business operations. Almost every large company has a great deal of money tied up in inventory. Significant savings can be made if it is possible to determine accurately the proper operating level at which such inventory should be maintained. Human judgment based on past experience is the usual basis for the rules governing the minimum point at which an order should be initiated, and the quantity to be ordered, both of which of course differ from item to item. An electronic inventory control system can follow such rules and imitate exactly the system which has been worked out on the basis of such experience.

However, even if the rules established originally were ideal, as time goes on the situation is bound to change, and the established rules become outdated. Why not let the computer work out the rules? The electronic data-processing system is handling all of the data that constitute the experience pertinent to the formulation of such rules. Give the computer a program of the right kind, and it will statistically reevaluate the inventory operation on each item, taking into account the most recent experience as well as earlier data. It appears possible, in other words, to have the computer carry out a continuous reevaluation of the system, properly adjusting the action it takes in maintaining the inventory on each item so that the capital tied up in inventory is minimized while at the same time the risks attendant upon too low an inventory position are controlled. Such an application for an electronic business machine certainly lies within the field of business operations, but any approach that attempts to do this in a rational way must at the same time lean heavily upon mathematical and scientific techniques.

Indeed, the last-mentioned example is only one of an inexhaustible variety in which the methodology of science is applied to business problems. This is sometimes known as "Operations Research." If a distinction between the business and the scientific application exists, it is certainly being strained by such developments, and it is the author's opinion that attempts to draw such distinctions do more harm than good.

In general, we can say that as business organizations learn more about the capabilities of the data-processing machines they are acquiring, they will want to do more sophisticated things with them, and many of the new techniques which will be developed to aid the businessman in the management field will have a very strong scientific flavor to them. Hence the firm that acquires a digital computer should favor equipment which is likely to be flexible in application, even if present requirements do not seem to require all of the abilities that make a general-purpose data processor efficient on a wide range of jobs.

8.3 THE BASIC REQUIREMENTS ON EQUIPMENT FOR AUTOMATIC DATA PROCESSING IN BUSINESS AND INDUSTRY

8.3.1 Fast-Access, Reusable Storage

The speed with which electronic components can perform arithmetic and logical operations cannot be utilized without storage devices that are capable of supplying the data to be processed at an adequate sustained rate and of storing the results so obtained at the same rate. Much of the economics of computer design and computer application revolves around the balance of computer storage capacity and access times. Since faster access devices are generally higher in cost, a mixed strategy turns out to have advantages. A limited amount of storage can be made extremely fast. This is then supplemented by some other form of storage which, though slower in access, effectively has an unlimited capacity. Some processing equipment also uses an intermediate level of medium-access-time storage with a capacity well beyond that of the fast storage. Until recently, fast storage was usually provided either by mercury delay lines or by electrostatic storage on the face of a cathode-ray tube. UNIVAC I uses 1000 words of delay line storage with an average access time of about 200 microseconds per word. Sperry Rand, IBM, and many others have built computers using electrostatic storage for much faster access, but magnetic-core devices now appear most attractive for reliable large-capacity, fast-access storage.

Computers with magnetic-core storage have already been put into service, and this type of memory device seems likely to be dominant for some time to come. Up to 10,000 words of core memory will soon be obtainable in the Remington Rand UNIVAC II, and IBM offers large computers with core memories ranging up to 30,000 words in the 704. As with the electrostatic memory, special designs could be used to give extremely short access times, ranging down to about 1 microsecond, but the commercial models being offered are in the 10 to 25 microsecond range. There is every indication that core storage requires very little maintenance and is highly reliable.

The medium used to provide unlimited storage capacity is usually magnetic tape, although under some conditions, punched cards are preferred to facilitate communication with other existing pieces of auxiliary equipment. Some scientific calculators continue to use Teletype tape as the principal and sometimes sole input-output medium, but this severely limits the flexibility and utility of such equipment for general processing.

The fact that magnetic-tape records can be transcribed faster than card or paper tape records is important, but the reusability of the magnetic tape also contributes directly to making the processing operations more automatic. As will be seen in Section 8.5, where sorting and merging will be discussed, the magnetic tape is used to extend the effective capacity for "temporary storage." Data must be shuttled back and forth from tape to tape. If fresh tapes had to be provided every few minutes because the medium was not reusable, much of the automatic character of such an operation would be lost. It is for this reason, rather than the obvious material cost saving, that a reusable medium is essential.

Many large computer systems also have a magnetic drum as an intermediate storage facility. Typical drums range in capacity from 4000 to 16,000 words. The drum may be loaded from tape or cards, as the case may be, and results placed on this storage medium can be read out to tape or cards. Some computers have instructions for selecting a single word from the storage drum, but it is customary to rely mainly upon transfers which move a block of words between the high-speed storage and the drum. Access times comparable to those of the mercury delay line memory can be secured by extremely high-speed drums, but drums in current use are slower and give correspondingly longer access times. Thus, with a drum turning at 3600 revolutions per minute, sometimes a maximum wait of one-sixtieth of a second might occur in locating a desired word. For such a drum, the average access time would be approximately 10 milliseconds.

8.3.2 Common Storage of Data and Instructions

A second basic requirement for any general-purpose electronic data-processing machine is that the instructions which direct the operations to be performed should be stored in the same facilities used for the data, so that they may be operated on and transformed or modified by the same kind of instructions as those used to process the data. This immediately provides an enormous degree of generalization in the processing that can be set up by coding. It means that programs can be written so that one sequence of instructions modifies or even generates another instruction sequence. It is clear that there is no theoretical limit to this creative and self-modifying ability. A first set of instructions may operate upon a second, and these in turn upon a third, and these in turn upon a fourth, etc. The practical limit depends upon the ability of the human personnel who must think through the consequences of such a pyramid in order that the end result will be understandable and useful.

The fact that a computer program can modify its own instructions or create new instructions to be used in a later processing operation has several important implications. When the processing to be done is essentially an almost repetitive one with only minor changes in the calculation from step to step, this facility makes it unnecessary for the human programmer to write out in detail every one of these steps and carefully check that each one has exactly the proper variation from the last. Programming becomes more like the formula writing of mathematics, in that the same program can be applied again and again to different pieces of data. Sometimes the only change is the identification of the data to be used as the program is repeated. Of course, the data itself may supply some cue which dictates a modification in the program, and thus a single set of instructions may allow for a very large number of different types of treatments of the item, depending upon the item characteristic.

The facility to use instruction to modify instructions also has an important bearing upon the computer design and the balance of computer operation speeds, fast-access storage, and input-output rates. Since instructions may be modified, it is possible to provide a small program which, through modification, can do a large number of jobs, and this replaces an extremely lengthy set of alternative programs that would otherwise have to be written out explicitly. Consequently, a great economy in storage of programs is possible, but at a sacrifice in effective processing speed. This will be illustrated in Section 8.4.

8.4 EXAMPLES OF CODING FOR DATA PROCESSING

8.4.1 A Simplified Single-Address Code

It is not necessary to become an expert at programming and coding to understand some of the problems facing those who wish to use large digital computers. Through a study of a few typical examples in coding, we can make certain generalizations that will be helpful in considering the larger strategy of data processing.

For present purposes we need only a few instructions. In the following list the first letter of the instruction indicates the operation to be performed, and the succeeding digits *"xyz"* are the code for a storage location. The "accumulator" is a one-word register separate from the main storage and having no location code *"xyz."* Its use will be evident as the instructions are described. Parentheses are used to indicate the contents of the register named within them. Thus, (*xyz*) means the contents of the register *xyz*.

Sample instructions for coding examples:

A	*xyz*	*Add* (*xyz*) to accumulator, result in accumulator.
S	*xyz*	*Subtract* (*xyz*) from accumulator, result in accumulator.
C	*xyz*	*Copy* accumulator into location *xyz*, leaving accumulator clear.
H	*xyz*	Same as C *xyz*, except *hold*, do not clear, accumulator.
J	*xyz*	*Jump* control to *xyz*.
N	*xyz*	Jump control to *xyz* and clear accumulator, only if accumulator is *negative*.
Z	*xyz*	Stop computer, sound buzzer. Digits *"xyz"* have no effect.

A typical general-purpose digital computer would, of course, have numerous other instructions and in particular should have instructions for multiplication and input-output facilities. The following additional instructions, not used in the examples here, have their counterpart in most general-purpose data-processing devices. A second special register, here designated as the "factor" register, is convenient to facilitate multiplication.

F	*xyz*	Reproduce (*xyz*) in *factor* register.
M	*xyz*	*Multiply* (*xyz*) by factor; product, rounded off, is left in accumulator (all decimal points at left).
R	*xyz*	*Read* next group of 100 words from tape number *"z,"* placing these words in consecutive locations beginning with *xy*0.
W	*xyz*	*Write* a group of 100 words on tape number *"z,"* said group being consecutive words beginning with location *xy*0.
Y	*xyz*	Rewind tape number *"z."* Digits *"xy"* have no effect.

8.4.2 Straight-Line Coding

Suppose it is desired to sum 100 different numerical values, which have already been stored in locations 100 through 199. The sum of these 100 values is to be placed in location 200. Clearly, this can be done by a straightforward sequence of instructions, such as the following:

A100, A101, A102, ···, A198, A199, C200.

Simple as this program may be, there are still two ways in which it could fail. In the first place, we have not made certain that the accumulator was cleared to zero before starting the accumulation. This can be rectified by preceding the instructions above by one instruction, "C200." Secondly, the sum of 100 values might turn out to exceed the digit capacity of the accumulator and "overflow" would take place. Coding methods to handle overflow, or to prevent its occurrence, will not be discussed here. It is sufficient to note that such problems do come up, and that it is a necessary part of the job of programming and coding to foresee and provide for such matters.

8.4.3 Modification of Instructions to Form Iterative Loops

The coding just given for the summation of a set of numbers is simple, but such "straight-line coding" is often inexpedient for several reasons. The instructions themselves may occupy a great deal of storage, particularly if the number of values to be summed is large. Furthermore, if all of these instructions must first be written by the human coder, time is wasted and there is opportunity for error, both on the part of the coder and on the part of the clerk transcribing such coding to tapes or other machine media. All and all, it is more reasonable to have the computer itself create the required sequence of instructions.

If the computer is to generate such a sequence, then two principal courses of action are open. It can be arranged that all of the required instructions are generated sequentially before any of them are executed. At the opposite extreme is the strategy of executing each instruction just after it is created. Both methods will be illustrated.

First consider the coding indicated as Code *A* in Table 8.1. This has no other purpose than to increase by one unit the location digits "*xyz*" of an *add* instruction. The serial numbers at the left of the instructions indicate the register numbers in which such instructions are to be stored. Thus, the instruction A100 is in location 300. This information must be specified because the instructions themselves are to be operated upon; therefore these location numbers will occur in the instructions we write. (Note that the straight-line coding, as written above, would have operated correctly, no matter where it was stored, provided it was *not* in the locations 100 through 200, inclusive.)

Table 8.1 Some Examples of Coding—Instructions Modified during Program Execution Are in Brackets

```
Code A (7 *)          Code B (8 *)          Code C (8 *)
300  [A 100] *        300  [A 100] *        300   A 313
 01   A 200 *          01   A 200 *          01   C 303
 02   C 200 *          02   C 200 *          02   C 200
 03   A 300 *          03   A 300 *          03  [A 100] *
 04   A 307 *          04   A 309 *          04   A 200 *
305   C 300 *         305   H 300 *         305   C 200 *
 06   J 300 *          06   S 310 *          06   A 303 *
 07   0 001            07   N 300 *          07   A 312 *
 08                    08   Z 311            08   H 303 *
 09                    09   0 001            09   S 314 *
310                   310   A 200           310   N 303 *
                                             11   Z 315
                                             12   0 001
                                             13   A 100
                                             14   A 200
                                            315
```

```
                      Code D (10 *)
400   C 300           410   N 401 *         420   0 001
 01   A 421 *          11   A 402            21  [A 100]
 02  [C 300] *         12   H 416            22   A 200
 03   A 402 *          13   A 420            23   C 200
 04   A 420 *          14   C 418            24   Z 425
405   C 402 *         415   A 423           425
 06   A 421 *          16  [      ]
 07   A 420 *          17   A 424
 08   H 421 *          18  [      ]
 09   S 422 *          19   J 300
```

```
              Code E (11 *)
300   A 318          310   A 317 *
301   H 305           11   H 305 *
 02   A 317           12   A 317 *
 03   C 306           13   H 306 *
 04   C 200           14   S 319 *
305  [A 100] *       315   N 305 *
 06  [A 101] *        16   Z 320
 07   A 200 *         17   0 001
 08   C 200 *         18   A 100
 09   A 306 *         19   A 200
```

* Instructions within the iteration loop. Parenthesis after code letter gives total number.

In Code *A*, the instruction in 303 brings the contents of 300, which is A100, into the accumulator. One unit is added to "*xyz*" of this instruction by the command in 304, since the contents of 307 is one. This is followed by C300 in 305, which sends the augmented instruction to the original location 300, after which the jump control in 306 closes the loop by selecting the new instruction in 300 as the next one to be executed. Each time this loop (involving the seven instructions in locations 300 to 306 inclusive) is traversed, the instruction in 300 appears with an operand "*xyz*" one unit greater. The contents of the location so designated add to the accumulated subtotal stored in location 200, after which the new subtotal is returned to 200.

It is clear that Code *A* accomplishes the purpose of generating and executing new instructions for a continued accumulation, but there is no control for stopping this iterative process, once it has been started. To avoid the endless cycling in this loop we must add instructions that test for the completion condition. In Code *B*, two more instructions have been added. Also, a "constant" has been stored in 310. Note that (305) records the modified A *xyz* in 300 using an H rather than a C instruction, following which the constant A200 is subtracted by the instruction S310 occurring in 306. As long as the modified instruction is less than A200, the subtraction of (310) will produce a negative result in the accumulator, and the negative jump instruction in 307 will turn the program back to 300 and continue recycling around the loop. However, when the modification of the contents of 300 has proceeded to a point where this modified instruction becomes A200, the result in the accumulator will be zero after subtraction of (310), and the "negative jump" will not take place. Instead, the next instruction for the computer will be taken from location 308, which contains a stop instruction.

Code *B* accomplishes the task for which it was intended but still leaves a few loose ends which are taken care of in Code *C*. Neither Code *A* nor Code *B* guarantees that the accumulator was cleared before starting the summation. Furthermore, location 200, which acts as an accumulator, is likewise able to spoil the result if not initially cleared. Finally, it should be noted that Code *B* is not "reusable." It is often desirable to code routines in such a way that they may be used any number of times for the same purpose, and this requires that those instructions which have been modified in the course of execution of the routine be reset or restored to their original values, when the routine is again entered for subsequent use.

In Code *C*, these matters have been taken care of by storage of the extra constant A100 in location 313 and by the provision of three instructions at the outset, which reset the modified instruction using this constant and also clear location 200. What initial assumption is made in Code *C*?

8.4.4 Generation of Straight-Line Coding

Code *D* illustrates the use of an iterative routine for setting up the straight-line coding originally suggested in Section 8.4.2. When the full set of 100 *add* instructions has been produced, Code *D* transfers control to the generated instructions. Some of the extra instructions found in Code *D* are necessary to transfer the stored terminal instructions C200 and Z425, which must be appended to the generated set of *add* instructions. Code *D* is not an example of elegant refinement in coding technique; the reader should readily be able to devise improvements. It is certainly inefficient to bring the same contents of 421 into the accumulator on two different occasions (note the instructions stored in 401 and 406). However, the purpose of these examples is not to exhibit polished coding but to show the variety of methods of attack that can be used on even the simplest sort of problem and the wide implications of such a diversity of techniques.

8.4.5 Storage Requirements versus Execution Time

Although the task which the preceding codes have been designed to accomplish is an exceedingly trivial one—the summation of 100 numbers—nevertheless the coding given illustrates several extremes that are typical of many more complicated, but essentially repetitious, data-processing jobs. When the entire set of operations are in straight-line form without resorting to modification of instructions, the storage required is often prohibitive. By using the computer to modify its own instructions, we may, in effect, create from a smaller set of more general instructions the set that actually operates upon the data. This reduction in storage requirements is bought at the price of extra computer operations. Sometimes these extra duties which the computer must perform are referred to as "bookkeeping" or "housekeeping" operations.

From Code *B* or Code *C* above, it is seen that in the iterative loop there are eight instructions, only one of which directly operates on a data word. It is a reasonable assumption that the same execution rate applies to all of the instructions used in the example. Thus, seven-eighths of the time used by routine *C* is spent on "bookkeeping." A time estimate based only on useful operations which directly process the data must be multiplied by a "bookkeeping factor," which in this instance is 8.

It should be clear that the bookkeeping factor will vary from problem to problem, and since most processing involves more than a single operation on each word of data, there are usually ways to cut the bookkeeping factor well below 8. Also, some automatic computers have automatic address modification facilities designed into their hardware, and this helps a great deal in reducing the bookkeeping factor.

As newer types of large digital computers are appearing with increased high-speed storage capacity, it is worth noting that such additional storage facilities can do more than merely increase the size of the data-handling problems which can be run through them. The extra storage can often be effective in increasing the processing rate in those cases where it becomes feasible to use more straight-line coding and spend less computer time on the generation and modification of instructions.

8.4.6 Compromise Coding

A final example will illustrate that it is not necessary to adopt either of the two extremes which have been illustrated, and that, in some cases, a compromise may be made between the storage-saving techniques and the time-saving techniques. In Code *E* above for the same problem, advantage is taken of the fact that once an instruction word has been brought to the accumulator for modification, it can be used to create two modified instruction words, before the accumulator is released for the execution of these instructions. The iterative cycle of Code *E* contains 11 instructions, but this cycle includes 2 *add* instructions and therefore the bookkeeping factor is 5.5. The increase from 14 instructions in Code *C* to 19 in Code *E* indicates that the space requirement has gone up in about the same proportion as the time requirement has been reduced.

The fact that coding methods able to effect a compromise between the competing limitations of storage and execution time can often be found has one further implication with regard to the basic computer requirements. The implication is that there is no precise point at which the storage of a general-purpose computer can be said to be "adequate" or "inadequate." Not only are the requirements bound to differ from one application to another, but it should now be clear that, even for the same application, there are enough different strategies in the coding process for it to be extremely unlikely that any user could justly claim a fast-storage capacity of 2500 words as absolutely necessary for efficient handling of this job, whereas anything above that would be of little value to him. To be sure, there may, for certain special applications, be sharp cutoff points peculiar to that application, but for the broad class of data-processing jobs normally occurring in business and industry, it is difficult to determine where the optimal point might be.

So far, we have discussed only the internal problems connected with coding. A discussion of the input-output facilities will throw further light on the variety of strategies available in data processing by automatic computers.

8.5 INPUT-OUTPUT CONSIDERATIONS—TAPE STRATEGY

8.5.1 Cards and Tapes for Sorting and Merging

The related processes, variously described as sorting, merging, collating, ordering, sequencing, and alphabetizing, occur frequently in many data-processing applications. The volume of data to be so treated is usually rather large and may involve many thousands of items. In some cases an item may run to hundreds of digits. There seems to be no practical way of providing enough fast-access storage to hold at one time the entire body of data that would normally enter into such a process. Medium-access storage on magnetic drums can provide enough capacity for some moderate-sized problems, but, in general, large volumes of data to be sequenced must be recorded either on punched cards or on magnetic tapes.

At first glance, it might appear that we should not attempt to sort on tape equipment. Cards have certain obvious advantages for sorting, and card-sorting machines are relatively inexpensive. Card sorters are also faster than most other card equipment; some present-day sorters are capable of feeding 600 or more cards per minute.

A closer look, however, reveals that tape equipment can be made to sequence data more automatically, and there are also other considerations that make card sorting less attractive. At 600 cards per minute, about 1000 characters per second are being passed through the sorter. Magnetic tape units operate at ten to twenty times this rate. To keep ten or twenty sorting machines supplied with cards at the 600 per minute rate, a number of attendants are required (ranging from eight to over twenty). Although personnel are needed to change tapes on the automatic digital machine, the requirements are appreciably lower. Moreover, in card sorting, human errors in card handling can be costly, whereas the computer operation can be set up to guard against mix-ups in tape handling.

Furthermore, the standard card sorter looks at but one digit, or column, per card during one passage of the data through it,[1] whereas the general-purpose digital computer can examine whole words or combinations of digits from different words during a single data pass, thus allowing more sophisticated sequencing methods to be employed. Merging or collating, an operation that is relatively slow on card equipment, is actually a preferred method of sequencing for many applications of tape machines.

Finally, automatic computers can often perform such jobs as summary tabulations *without* sequencing the data being summarized; much of the sorting done on card machines is required solely because this is the only expedient way to summarize and tabulate on such equipment.

8.5.2 Tape Requirements for Sequencing Data

In general, for an automatic computer to be efficient at sequencing, it should have a number of associated tape units. How large this number should be depends on many interrelated computer characteristics. However, it is easy to set a minimum.

A digital computer should have, at the very least, four tape units if it is to carry out properly the sorting of one or more reels of randomly arranged items. It is easy to see why this must be so. In the first pass of data through the computer, information read from one reel is subjected to test by some criterion and, depending on the outcome of such a test, each item is recorded on one or the other or two output reels. This two-way, or binary, sorting of the data is but the *first* of a series of binary sorting operations which must be carried out before the proper final sequence is attained. In the second step of this sorting process, the two tapes that were used just previously to record the *output* from the first sort now become the *input* to the second step, and this means that two more tapes must be available for the output of this second sort. The original data tape can be used for one of these outputs, if erasure of the original data is acceptable. If the original data is to be saved, that reel is removed from the tape unit and another substituted. At the end of step two, the output from step one will have been read back through the computer and redispersed on the other pair of reels in accordance with whatever sorting criterion is being used. Succeeding steps in this process shuttle the

---

[1] There are special machines, such as the IBM 101 statistical machine, which have the ability to examine many different columns and discriminate among various combinations of the information so read.

information back and forth between the two pairs of reels until, in the final step, the desired sequence has been secured.

It is clear that for a very large number of items many passages of the information from one tape to another to the computer will be required. If we use only the binary or two-way process just described, then an extra tape pass will be required each time the number of items to be sequenced is doubled. The obvious way to reduce the number of data passes through the computer is to use more tape units, distributing the items on three or more tapes during each pass.

However, to make such a process continuous and automatic, the computer should have at least six tape units for a three-way process, eight units for a four-way process, and so on. Thus, to duplicate the ten-way sorting process customary on card-sorting equipment, at least twenty tape units would be necessary. Although this is not an impossible requirement, the cost of such an extended input-output system is not justified in most processing systems. Fast tape units are expensive, and slow ones would fail to give the gains we are here seeking. Actually, in machines presently available, the internal rates for comparisons and transfers ultimately limit the sorting speed, and sequencing methods using only eight or ten tape units can be computer-limited. When this is the case, there is little benefit from additional tape units which must stand idle part of the time because the bottleneck is in the internal processing speed.

Several other considerations make it unwise to assume that a large number of tape units for sorting can be economically justified. For one thing, the column-by-column or card-sorting methods usually require more passes than the collating and merging methods, which are easily carried out by electronic processing equipment. Also, in an automatic computing system we can frequently combine any necessary sequencing with other processing operations during the same passage of data through the computer. In card systems, sorting and other processing operations are of necessity done by different pieces of equipment.

Finally, we must not presume that all the sorting done in a card system must also take place when large data-processing computers are used. Sequencing of data can often be partly or completely avoided on electronic equipment. This is because many characteristics of a data file can be summarized and tabulations built up in the internal storage during a single scan of a data tape. To prepare the same summary tabulations with cards, sorting is essential. This is one example of the general principle that all data-processing methods need basic re-examination when new equipment is to be used.

8.5.3 Overlapping Input-Output, Computation, and Rewind Times

Previous discussion of computer-limited and tape-limited problems has already assumed that facilities exist for overlapping some of the time required for these different operations, but it may be helpful to present some hypothetical figures (for the "sample computer" of Section 8.4.1) to illustrate the desirability of such strategy. Let the tape transcription rate of the sample computer be 1000 words per second. Assume an item length of 10 words. Then each block of 100 words contains 10 items. For present purposes, assume that instructions such as A, S, and N, required for transferring and testing data, are executed in 100 microseconds each.

Suppose that the sequencing operation is now at the stage where a block of 10 items from one input tape is to be merged with another block of 10 items from a second input tape, thus creating 2 blocks of output data consisting of 20 items in proper sequence. The typical instruction sequence, Add, Subtract, and Negative jump (A, S, N), for a comparison of pairs of items from the 2 input blocks, will have to be repeated something like 19 or 20 times. All other operations required to keep track of addresses and the number of items dealt with can be considered as included in the "bookkeeping factor" which we will here take to be 5. Then, a rough time estimate for the internal operations is 1.5 milliseconds per test, and thus 30 milliseconds for each pair of blocks merged.

At 1000 words per second, 200 milliseconds are used in reading 2 blocks from tape and an equal time is required for writing the merged sequence of 20 items.

If reading, writing, and processing cannot be overlapped at all, then 430 milliseconds will be consumed for each 20 items so processed. Merely overlapping the tape-read and tape-write operations cuts the time requirement to 230 milliseconds. A further reduction is possible through overlapping the computing or processing time with the input-output operations. In the particular example used, this saving does not seem large; only 30 milliseconds have been cut from the 230 milliseconds by overlapping the reading of new data, the processing of current data, and the writing of already processed data. The merging process is here a tape-limited one, and the computer has 170 milliseconds to spare while waiting for the tape operations.

There are several possibilities for improving the efficiency of such an operation. If more tape units are available, it may be quite profitable to merge items from 4 tapes and thus convert sequences that are only 10 items long into sequences that are 40 items long in 1 tape pass.

Since the total number of items to be processed is assumed fixed, the total time for each complete passage of all the data through the computer will not change. Hence the final sequencing will be accomplished in half the time, because only half as many passes through the machine are needed when four-way merging is substituted for two-way merging. The more complicated comparisons required to accomplish four-way merging require considerably more than four times the computer time—in this case, more than 120 milliseconds—and might turn out to be over 200 milliseconds. Such a change in procedure can easily result in converting what was originally a tape-limited process for a simple merging scheme into a computer-limited process for a more complex merging scheme. Such a change in strategy is of no advantage if the computer time cannot be overlapped with the input-output time.

A second strategy to overcome the inefficiency of the merging process used in this example is to incorporate other data-processing operations so as to make effective use of the extra computer time not used in the actual merging operation.

Another kind of time overlap that may contribute to efficiency is obtained when the rewinding of certain reels of processed data can occur while other tapes are recording newly processed data. This too has limitations. The nature of the process may be such that the data just recorded is immediately necessary for the next step in the sequencing process; overlap of rewind time is then impossible. In such instances, a computer capable of reading data back into the computer while tapes are running in reverse, can operate more efficiently than one that must rewind the tapes before they can be read again.

8.5.4 An Example of Tape Strategy in Merging

There are many ways in which the efficiency of utilization of a data-processing computer depends upon the strategy used in handling tapes. An interesting example which can be explained without getting into complicated details has to do with the merging of many tapes (already internally sequenced) to form a single sequence. For a particular computer and for a particular kind of item, some particular merging process will often be most efficient. In some cases, a four-way merge might be the most efficient type, whereas under different circumstances a three-way merge or a five-way merge would turn out to be most efficient. For this example, it will be assumed that the characteristics of the computer and of the items being handled are such that a three-way merge is the most efficient.

The tapes entering into this merging process have already been internally sequenced, and the proper strategy for merging three or nine or twenty-seven such reels into one continuous sequence is then clear, since this can be done through the use of three-way merging at every step.

Suppose, however, that the amount of data occupies exactly eleven reels, so that we must employ two-way merging in some cases. We might enthusiastically embark on a process that uses three-way merging as fully as possible at the outset and resorts to two-way merging only when there are no longer any opportunities to use the three-way merge. This procedure is shown graphically in Fig. 8.1.

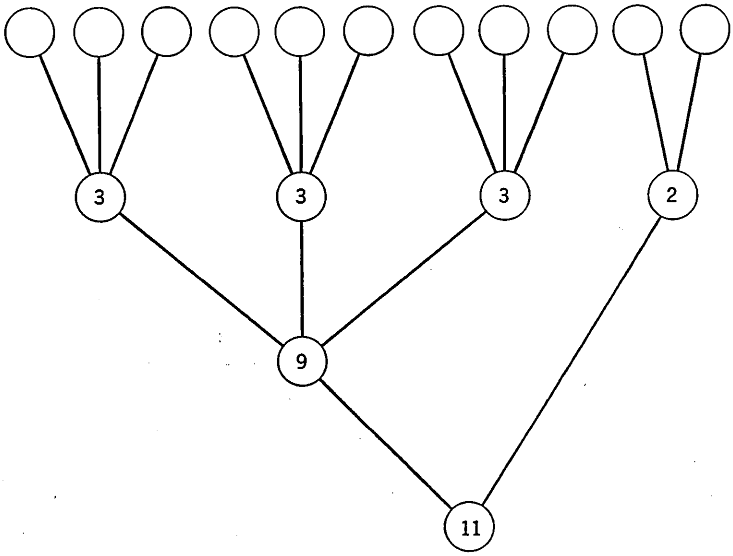


Fig. 8.1 Merging eleven tapes: first scheme, thirty-one tape passes

By three-way merging of the first three sets, we obtain three sequences of three reels each, which are then merged into a single sequence of nine reels. The two remaining reels are subjected to a two-way merge, and the two-reel sequence so obtained is then merged with the nine-reel sequence by a final two-way merge, so as to obtain the desired eleven-reel sequence.

Consideration of Fig. 8.1 may lead us to try a different scheme, such as that shown in Fig. 8.2, where the need for a two-way merge has been eliminated and nothing but three-way merges are used. Is this a better strategy?

If, for the sake of comparison, it is agreed that all mergers, whether two-way or three-way, are tape-limited, then the efficiency of these strategies may be judged by the number of tape passes required to accomplish the final sequence. For the scheme in Fig. 8.1, thirty-one tape passes are required, whereas for the scheme in Fig. 8.2, twenty-nine suffice. This change in strategy therefore gives a 6 per cent improvement.

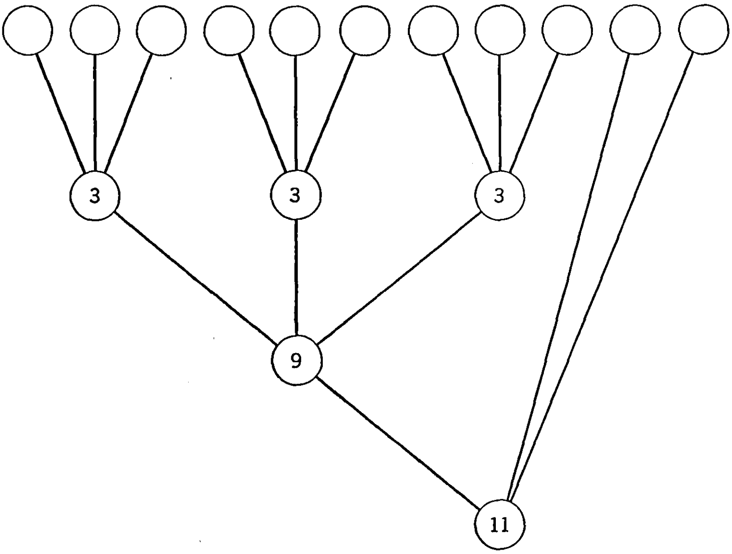


Fig. 8.2 Merging eleven tapes: second scheme, twenty-nine tape passes

Since the second method uses only three-way merging and never resorts to two-way merging, there is a temptation to assume that Fig. 8.2 represents the optimum strategy.

This is an unwarranted assumption, as can be seen from Fig. 8.3, which achieves the same eleven-reel sequence by a different combination of three-way merging. The third scheme requires only twenty-five tape passes, in contrast to the twenty-nine required for Fig. 8.2. Thus a further improvement of about 16 per cent has been made. In a typical operation where this kind of sequencing job has to be done every week throughout the year, many thousands, or even tens of thousands, of dollars of processing cost can be saved by this simple change of strategy.

If we no longer assume that the two-way merging process is just as time consuming as the three-way process, then other strategies, involving a combination of two-way and three-way merging, may be found that are better than the one shown in Fig. 8.3. This condition may arise when the item size is small, thus requiring more comparisons for each batch of data read in from tapes. For smaller item sizes, more complicated merging methods sometimes turn out to be computer-limited rather than tape-limited.

Although this discussion of tape strategy has been in terms of the sequencing process, there are many computer applications that do not involve sequencing at all; but for these the computer efficiency is strongly affected by the way in which the tapes are used. Tape strategy is important whenever large volumes of data must be read in and out. As mentioned earlier, the tapes form an important auxiliary memory for a variety of large scientific computations, and the proper use of such tape storage can make a large difference in the efficiency with which a scientific problem is solved. Matrix multiplications and inversions, for a matrix of the order of 200 by 200 (thus having 40,000 elements), are but one example of a type of problem which exceeds the practical capacity of an internal high-speed storage system. As mentioned earlier, partial differential equations in three dimensions quite often require the systematic handling of tapes for auxiliary storage for the best efficiency to be achieved.

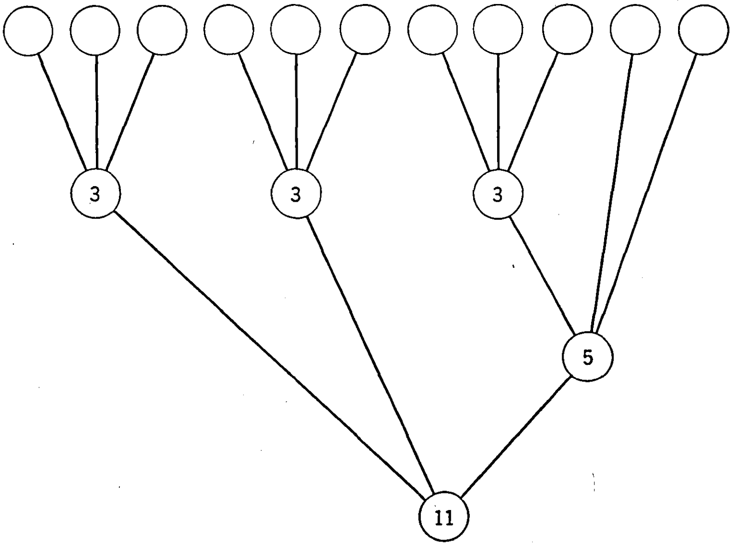


Fig. 8.3 Merging eleven tapes: third scheme, twenty-five tape passes

8.6 RELIABILITY AND ERROR CONTROL ARE BASIC TO SYSTEM DESIGN

8.6.1 Machine Faults and Human Mistakes

Error control procedures are a necessary part of every system design. Under the name of "quality control," a great deal of theory and practice has been developed to deal with the variations that can occur in a tangible manufactured product. The amount of human and machine time spent in controlling the errors that occur in data processing must be enormous, but such controls are usually devised without adequate theory and without reference to the factory quality control methods. Accountants have required their books to balance to the penny. Engineers and scientists have often had their calculations repeated in detail by a different group of people, just to make sure that no serious errors have been made.

Since the control of errors is not a new problem, past methods and experience are likely to be relevant but should not be relied upon exclusively when new equipment and new processing procedures are being considered. We can usually attribute any deviation from correct results to one of two general sources—the human mistake and the machine fault. Automatic electronic data-processing equipment has in fact established new standards of reliability with respect to machine faults. Automation has also reduced the opportunity for certain human mistakes and provided some additional methods for control. Several aspects of these questions will be discussed briefly.

8.6.2 Machine Reliability and Checking

Digital computing systems are, by definition, insensitive to small changes in the dimensions, electrical characteristics, or other properties of their elementary physical components. One of the fundamental responsibilities of the design engineer for a digital system is to examine the tolerances of the components which he uses and to use this knowledge to create a system in which there can be no error in the computed result arising from changes that remain within chosen tolerance limits. Furthermore, these tolerance limits must be realistically set on the basis of both experience and theory, so that the probabilities for error are limited to extremely low values.

However, it is manifestly outside the bounds of economic possibility to design circuits that never fail. The practical compromise often involves duplicating certain portions of the equipment. Also, by including error-detecting devices, most failures can be recognized and the otherwise damaging effects on the computed output can then be prevented. Currently available computers differ greatly in the kind of automatic checking provided and the degree to which reliance is placed on such checking.

Although different applications may have quite different requirements for the reliability and accuracy of the processed data, there is probably no application in which the control of error can be dispensed with entirely. The man on the street may be perplexed when he hears that it was considered necessary to proofread a recently published table of random numbers. However, even random numbers must pass a quality control or inspection procedure.

When, through malfunction, an error does occur in a digital system, it may be trivial or it may be devastating. The designer of the hardware cannot control the importance of the digits being processed, or the damage a fault can create. The computer engineer provides the basic reliability that makes the system function properly almost all the time, and he can also provide checking circuits that serve to detect almost all the faults which do occur. Those who apply the computer and specify the procedures have the ultimate responsibility for error control. When automatic checks are not included, then the programming personnel must decide which possible faults should be checked, how they should be checked, and then work out the added coding that performs the check.

8.6.3 Controlling Human Mistakes

There are many developments in the data-processing field which reduce the extent by which human mistakes can affect the input to a computer. In scientific-data reduction, instruments may be read automatically and the digital values equivalent to those readings sent to the computer without any human transcription intervening. As devices for reading printed numbers from bank checks and other documents become available, the opportunity to introduce wrong data through human mistakes is decreased. Redundant typing or punching may sometimes be the only practical method for cutting down the errors that enter from human sources.

Human mistakes can also occur in providing a computer with its instructions. This might be just as serious as the faulty data, but it need not be. The errors in instructions can be found by systematic testing before actual use in production runs. Once eliminated they will not recur. A human clerk may be inconsistent in following instructions, but a computer is monotonously consistent.

There are also ways in which a computer can forestall and nullify the effect of certain types of human mistakes. An interesting example is in the handling of tapes in large data-processing applications. We might take, for example, the merging operation discussed in Section 8.5. A number of tapes must be mounted and removed from the machine as the merging process continues. The personnel who handle these tapes can occasionally make a mistake.

Since the entire operation can be spoiled by one wrong tape, it is very important to do everything possible to check that the proper tapes are supplied for each step of the operation. Every tape should have a visible label to guide the operator. To allow the computer to check, each tape must also carry a label which the computer can read. Therefore, the first information to be found on a tape should be that needed by the computer program to determine whether this tape is the one that should be where it is.

Tapes that contain primary data going to the computer for the first time should have such information recorded on them as a part of the original data preparation and recording job. For all tapes that are themselves the output of the computer and later become input, the labels should be automatically provided by the computer program. Tape-labeling procedures should be deliberately incorporated into the programming at the outset, not added as an afterthought. Each program must provide for the proper identification record on the tapes it produces and check such identification data on the tapes it uses.

8.7 KIND OF SAVINGS POSSIBLE THROUGH USE OF AUTOMATIC ELECTRONIC DATA-PROCESSING SYSTEMS

8.7.1 Economic Advantages Do Not Come from Speed Alone

The speed with which electronic data-processing systems operate is often so impressive that it obscures the major business reason for use of such equipment. Fundamentally, the principal applications for electronic data processing are justified by the greater useful output per dollar invested. To be sure, the high speed of the elementary arithmetic operations contributes to the low-cost factor. However, a significant part of the cost reduction comes from the elimination of human participation in the overall task—for instance, the need for preparing and transferring paper records from one desk to another, or the conveying of punched-card records from one machine to another. In some cases, several different files that had to be maintained separately in older methods can now be combined into a single file which is useful for a multiplicity of purposes. Rate tables or other auxiliary bodies of information that were not easily made accessible to systems employing smaller single-purpose machines are more readily introduced into the automatic systems which have large auxiliary storage facilities as a part of their integrated equipment.

8.7.2 Savings Come from High Reliability and Freedom from Human Mistakes

No system is ever completely free of error, and automatic electronic computers are certainly not trouble-free. However, the standards of reliability that have been achieved in the field of electronic computers far surpass those that have been customary in previous equipment. (The error detection circuits, which are standard in some computer systems, make it possible to secure better estimates of reliability for electronic equipment than any available for earlier devices.)

For example, in transcribing data to and from magnetic tape, we may find, under certain conditions of maintenance and adjustment, that a transcription error will occur on the average once every 3 hours. Since 10,000 to 20,000 characters per second are characteristic of the tape transcription rates, this can be translated into the statement that an error occurs about once in every 100 million characters. A similar performance for a card machine, running at 100 cards per minute, would mean that there would be only 1 character in error during a week of operation in which the machine operated 24 hours a day for 7 days. Furthermore, in the instance of the electronic computer, such errors are automatically detected and are thus prevented from propagating to the point where their correction becomes a major task.

Moreover, the computer is unfailingly consistent in its application of the rules of procedure. In nonautomatic systems, some cost must always be charged to the failure of the human being to carry out in a consistent manner the discriminations and judgments he is expected to exercise with regard to the handling of exceptional cases. In fact, the more monotonous and computer-like the tasks performed by the human being, the more likely it is that the human clerk will occasionally slip up.

8.7.3 Savings Come from Improved Systems and Procedures

Before any task can be turned over to an automatic computer, that task must be meticulously and unambiguously defined and the correct program written to direct the computer. In those applications where punched-card systems have been used, the care with which procedures must be analyzed and specified is already well known. However, electronic data-processing equipment can, because of its versatility, handle many applications that have never been previously subjected to punched-card methods. In attempting to apply the electronic equipment to these broader fields of application, the necessary systems and procedures analysis discloses weaknesses in the older procedures, which, when corrected, lead to appreciable savings and increased efficiency. Some of the indicated improvements, once discovered, can lead to savings without the introduction of electronic equipment. However, it does seem to be true that the extremely thoroughgoing analysis which must precede the application of electronic data processing to any task is often more successful in uncovering the areas in which improvement can be made. Some part of the savings so effected may well be attributed to the careful and logical methods developed out of the necessity of satisfying the exacting demands of the electronic devices.

8.7.4 Savings by Automatic Coding

Out of the systems analysis that has just been discussed must come a program for the computer to follow. The program must ultimately be reduced to machine code, such as that discussed in Section 8.4. Furthermore, this final code must be thorougly checked to make certain that the data-processing equipment is controlled exactly as it was meant to be, and does, in fact, carry out in every detail steps that were intended. Actually, a great deal of the labor of reducing the original system requirements to the machine code is *routine* labor, which can be executed in accordance with precise rules. Hence, such work is in itself an example of the kind of labor that should be relegated to the automatic computer.

"Automatic coding" is the term now used to designate the field of computer application that aims to turn over to the computer itself all the systematic and routine operations involved in translating the general plan of operations to accomplish any specific task, as envisaged by the systems and procedures analyst, into the final detailed machine coding which constitutes the "running code."

A great deal of development work is going on right now in this area of computer application. Some of the methods being devised are intended primarily for the translation of algebraic and other mathematical equations directly into codes to be used in solving scientific and engineering problems. Others are aimed at providing faster and easier ways to set up codes for commercial processing operations. The idea is to provide a library of subroutines which are sufficiently general, so that each will have many possible applications as a part of a larger system of operations. Putting together a large program from a set of such library routines is called "compiling."

Through the use of "pseudo codes" and compiling systems, it is possible for a programmer to specify exactly what he wants the computer to do, without ever having to concern himself over the detailed coding of the parts already worked out once and for all in the subroutines contained in the library. In fact, the notation in which he describes the process to be carried out may be one deliberately chosen to be most convenient to him and need not resemble the elementary instructions of the computer in any way. The notation he has used is then translated by a computer program into instructions directing the computer to assemble from its library the required running program, which will then appear as an output of the translating and compiling program.

The advantage of having such a pseudo code for the programmer to use is, in part, that such a pseudo code may be a more natural language for the programmer to use. A second advantage is that the information he then has to write is considerably condensed; consequently he is spared the labor and drudgery of the extended writing which would otherwise be necessary. A further aspect of the use of pseudo codes is that the programmer may more easily check to see that what he has written is indeed what he intended to write, and such a pseudo code therefore reduces considerably the chance of human mistakes in the programming process. The compiling operation which follows is free of human error. Consequently, if the final running program does not carry out the operations as intended, hunting down the source of the discrepancy is easier because the trouble must lie in the information supplied by the programmer.

The importance of the savings anticipated from the developments in automatic coding can easily be appreciated when we find that in many of the large computer installations the budget for programming and coding has run to hundreds of thousands of dollars per year. There still are many installations that make no appreciable use of the existing automatic coding and compiling systems. But sufficient experience has been gained at some installations to estimate that at least 50 per cent of these annual costs can be saved through efficient application of the automatic coding techniques. As the development of better automatic coding methods continues, and as the use of such methods becomes more widespread, it is to be expected that even greater savings will accrue.

Additional savings, over and above the mere reduction in dollars spent to produce a particular program, may sometimes be realized through the reduction in elapsed time from the statement of a problem to its solution. For instance, in the case of an engineering problem, the elapsed time from the moment a problem is proposed to the time its solution is obtained may be more important than the precise amount of money needed to effect that solution. By automatic coding, problems that would have required something like three weeks for the programming time, using nonautomatic coding methods, have been formulated and solved within two days.

In commercial data processing, the same situation may obtain. From time to time, management finds it necessary to require special reports which cannot well be anticipated. If several weeks were required to produce the code that would turn out such a report, this would seriously impair the value of the data-processing system for management purposes. Any technique for dealing with situations of this sort is almost sure to depend upon automatic coding.

8.7.5 Major Savings in New Applications

The greatest savings from the application of automatic electronic data-processing equipment are likely to come from the solution of problems which have heretofore gone unsolved because there was no reasonable and practical way of coping with them.

Many of the tasks now being done by punched cards or by manual methods can be done with greater economy and dispatch with automatic data-processing equipment, but the savings in converting to these more automatic tools will frequently be modest rather than spectacular. This is particularly the case if we seek to do with the automatic equipment no more than has been done by the older methods.

However, if imagination and ingenuity are applied to seek out those tasks which become feasible for the first time with these new tools of office automation, it is almost certain that every large organization can find applications leading to tremendous savings. For instance, large electronic computers make possible a new approach to the job of factory scheduling, by methods entirely out of the question with older equipment. Many management problems are now reducible to variations of what is mathematically known as "linear programming." The many variables that enter into a complex process (such as are found in chemical and petroleum industries) have never been adequately studied to determine the proper combination of conditions for maximum yield or maximum profit. Entirely apart from the internal operations of a manufacturing plant, savings can be made in the cost of transportation from plant to warehouses, and from warehouses to other distribution points. Studies have shown that in some cases millions of dollars per year can be saved through proper mathematical solution of the transportation problem.

8.8 REFERENCES

1. A. D. Booth and K. H. V. Booth, *Automatic Digital Computers,* Academic Press, Inc., New York, 1953.

2. B. V. Bowden (Ed.), *Faster than Thought—A Symposium on Digital Computing Machines,* Sir Isaac Pitman & Sons, Ltd., London, 1953.

3. R. E. Canning, *Electronic Data Processing for Business and Industry,* John Wiley & Sons, Inc., New York, 1956.

4. N. Chapin, *An Introduction to Automatic Computers: A Systems Approach for Business,* Technology Center, Chicago, 1955.

5. R. F. Clippinger, B. Dimsdale, and J. H. Levin, Automatic Digital Computers in Industrial Research, *J. Soc. Ind. Applied Math., 1, 2,* Philadelphia, 1953–1954.

6. M. J. Dooher (Ed.), *Electronic Data Processing in Industry,* American Management Association, New York, 1955.

7. G. Kozmetsky and P. Kircher, *Electronic Computers and Management Control,* McGraw-Hill Book Co., Inc., New York, 1956.

8. Proceedings of a Symposium of Industrial Applications of Automatic Computing Equipment, Midwest Research Institute, Kansas City, Mo., 1953.

9. Analog-to-Digital Conversion Units

BERNARD S. BENSON

Benson-Lehner Corporation
Los Angeles, California

GEORGE G. BOWER

Motorola Laboratories
Corona, California

9.1 THE NEED FOR CONVERSION

It is our purpose here to show (1) how analog-to-digital converters fit into the picture of automation in business and industry; (2) some fundamental characteristics of these devices; (3) the basic types, how they are built, and where they may be obtained; and (4) some specific examples of their use in industry.

An analog-to-digital converter is a device that is connected to conventional measuring instruments or is an integral part of their design; it permits measurements to be expressed automatically as numbers, i.e., digitally. The converter replaces the person who reads the scale of the instruments and writes the results on a data pad. It is much faster and more reliable than the person replaced, and the data is in a convenient form for further use.

The world today is humming with activity. Millions of people are cutting things, shaping things, mixing things, weighing them, measuring them, and selling them. The purpose of all this activity is to produce returns for the individual in goods, in money, and in knowledge. This tremendous creative activity must continually be monitored. Stock must be taken of what is going on so that each individual can be rewarded according to his fair share of the contribution. The human memory is obviously not sufficient to keep track of all of this activity, and accurate, permanent records must be maintained on a continuous basis.

Furthermore, in the field of scientific research, millions of measurements must be made to find out what is going on, and a structure of complex computations must be built on this foundation in order to reach forward and provide progress based on experience and understanding. This is our particular method of operation as compared with creatures, for example, which evolve apparently without understanding.

If the accounting and the calculating are to keep track of the building and experimenting, it is evident that a tremendous wealth of information has to flow between these two camps, and this means a busy thoroughfare of conversion between the language of one (digital) and the language of the other (analog).

Size, weight, temperature, pressure, lapse of time, all these things in their original form are in the analog language. The monitoring, the recording, the accounting, the computing, and recently the controlling, or at least a part of it, use their own digital language. The two camps have existed for a long while. It is the increasing activity in each one and the tremendous increase in traffic between the two that are new.

In the old general store there were gunny sacks of coffee and there were ledgers, and every once in a while a pound of coffee would give way and its equivalent number of cents would appear on the ledger. The storekeeper, when he was not playing the role of analog-digital converter, had plenty of time to gossip by the stove.

The stoker in the boiler room, when he was not shoveling coal, would take an occasional reading of the pressure gage and jot it down for the benefit of the young boss who was all for this new "scientific" approach. Contrast this with the tempo today. A thousand thermocouples on an atomic pile are being quizzed dozens of times a minute by a machine with a mania for consuming numbers.

A supersonic fighter streaks across the sky, with every part of its stubby wings fighting against the mounting pressures and temperatures. Engineers on the ground are clamoring for more and more measurements to greater and greater accuracy in less and less time.

Tens of thousands of people are dashing madly in and out of stores whisking up premixed, prepackaged, oftentimes dehydrated or reconstituted foods, while the accounting department must maintain complete control from the raw material to the final inventory.

And partly here, partly on the drafting board, and partly in the designers' imagination is the automatic factory. Sparked by new concepts of factory organization, by special product designs, and by the expanded use of automatic control, this scheme of production promises a new order of magnitude in output. It also very humanly promises to relieve workers of their responsibility for making and executing irksome and tedious routine decisions. This requires tremendous amounts of data. The parameters and performance of the processes must be monitored and the data transmitted remotely to central control stations. Some must also be transmitted to the office for use in inventory control, payroll computation, maintenance forecasting, cost accounting, and the like.

Data of itself is of no use unless it can be resolved quickly so that appropriate decisions can be made as a result of it. The digital computer and its various counterparts in the business field allow such decisions to be made quickly. For this reason, it is advantageous to convert measurements to the digital form. There are other reasons too, accurate electrical transmission and convenient and compact storage with rapid access being foremost.

Given the tremendous expansion in the need for data, the vastly increased tempo of modern commercial and industrial activity, and the advantages of dealing with numbers in the digital form, it is necessary to have devices for reading instruments automatically at high speeds and for providing the results as numbers. The analog-to-digital converter satisfies this need.

9.2 SOME FUNDAMENTALS OF ANALOG-TO-DIGITAL CONVERTERS

We have stated that an analog-to-digital converter is essentially a device for presenting the result of a measurement as a number. Measuring instruments are built around a convenient physical effect which will produce a response analogous to the parameter we wish to measure. For example, the fact that a wire carrying current can be moved by a magnetic field is the basis of our electrical instruments. This effect combined with clever linkages moves a pointer across a graduated scale. The movement of the pointer is an analog—the observer, in reading the scale and recording the number, performs an analog-to-digital conversion. The converters we are discussing do the same thing automatically and they do it faster, cheaper, and more reliably. It is possible to handle a large variety of parameters—temperature, pressure, *p*H, flow, voltage, current, color, weight, time, frequency, and almost any others of interest. Not that converters have been or will be developed for all of these, but it is possible to interpose transducers that will transpose any of them into, say, angular position, voltage, or time or frequency for which converters have been designed.

The designs of these converters are patterned after the processes an observer might follow in making a measurement, i.e., they count, they compare, or they read. It can be argued that the latter two are versions of the first with conveniences added. This is true, but the classification is still convenient. As an example, the conventional meter stick requires both reading and counting. The centimeter divisions are all marked and are read off, but the millimeter divisions between the centimeter must be counted. The best example of the comparing process is in bridge measurements where a galvanometer is made to read zero and thus indicate equality by the addition of known values to one arm of the bridge.

The converter must also do one thing that a human reader does almost unconsciously. It must make a decision about the nearest division on a scale.

9.3 SPECIFICATION OF CONVERSION UNITS

It is worth while listing and describing the characteristics of converters which are of primary interest to the user. Among these are the analog input, the range of input, the sampling rate, the number of channels, the type of read-out, the base of the number system, and the number of digits.

9.3.1 Analog Input

The analog input refers specifically to the parameter a converter is designed to measure. The most convenient parameters from a design point of view have been voltage, angular position, and time or frequency. Fortunately, these parameters can form the basis for complete systems whereby any of the parameters of a manufacturing process can be transformed readily to these.

There are many instruments whose final indication is produced by a voltage moving the pointer on a meter. This is true of the thermocouple for measuring temperature, for the photosensitive cells for measuring radiation, for the *p*H meters which measure acidity, for the strain gages used extensively in material testing and conceivably forming part of pressure measuring instruments, and many others. Also, a large number of instruments rely for final indication on a turning shaft deflecting a pointer. Therefore, it is not too big a jump to fit conventional instruments into converters already designed.

9.3.2 Range

Since the converters are basically measuring instruments, they share the problem of having proper ranges. As in the conventional measuring instruments, this problem is handled by providing multiple ranges, usually through analog arithmetic operations on the input. For example, on electrical devices, shunts and pads are provided; and on mechanical devices, different sets of gears.

9.3.3 Sampling Rate

One feature of handling data in the digital form is that we deal with a finite set of discrete values as numbers rather than with continuous values as on a curve of a graph. The parameters of primary interest to us vary with time. We recognize intuitively that we do not need to know the value of the function at every instant of time in order to make intelligent decisions about it. For example, when we drive a car, we look at instruments at intervals rather than continuously. Also, when we plot graphs, we are satisfied in connecting discrete points with smooth curves. The faster the parameter varies, the more frequently we sample its value in order to feel that the data represent the function. These intuitive feelings can be dealt with mathematically and a sampling rate assigned according to the bandwidth (1) occupied by the parameter or according to the accuracy demanded of the approximation (2). In general, the faster the parameter varies and the more accurately it must be approximated, the more frequently it must be sampled. Most decisions on this can be made empirically by experienced plant engineers, without resorting to more elegant methods.

9.3.4 Number of Channels

It is not always necessary to devote a converter to a specific instrument. Instead, many instruments may be fed into one converter and the output of each sampled in turn in a cyclical fashion. This is nothing more than a mechanization of the process whereby a human operator with a data pad made a tour of a plant writing down the indications of the instruments. The automatic process is more efficient and permits current and complete knowledge.

The number of channels refers, of course, to the number of instruments feeding into the converter. For most applications, a simple commutator of many segments is all that is needed for this time-sharing idea. Should rapid sampling rates be needed, electronic commutators are available or can be designed to handle almost any need. In still other cases, the sampling sequence and rate can be controlled at will by a program stored in an associated digital computer.

9.3.5 Type of Read-out

It is in the read-out that the big advantage in the use of analog-to-digital converters over manual methods is realized. The output in numbers can be in a form that is usable directly for transmission or computation, or for typing on prepared forms. Common methods include the card and tape punches for the slower speeds and magnetic recording for higher speeds. These methods also provide for compact storage with fairly rapid automatic access. The numbers could also be fed directly into a digital computer and used immediately, or they could be stored in magnetic-core registers which give very fast random access.

9.3.6 Number System

Automatic reading of instruments and automatic handling of data bring up the question of what number system to use. This rarely comes up specifically when manual methods are used. It is convenient from an equipment standpoint to use the binary system in many data-handling systems, particularly in the digital computer. It saves equipment and the equipment is simpler. The binary system, however, is difficult to read and is not as compact as the decimal system. Converters have been built to read out in both systems, and also in modifications of the systems.

One of these modifications is the binary-coded decimal. This consists essentially of coding each digit of the decimal number in the binary form. For example, the number 1956 is written 0001 1001 0101 0110. This retains, to some extent, the easy-reading feature of the decimal scheme and yet permits arithmetic to be performed very nearly as simply as in the straight binary scheme.

Another modification of the binary system which has none of these advantages but does have an advantage in design of converters is the reflected binary code (also called modified code or cyclical code). This is just a permutation of the binary code which has the property that only one digit changes from zero to one or one to zero in going from one number to an adjacent number. Further details and applications can be found in the literature (see Table 9.1 and references 3 and 4).

Table 9.1 Comparison of the Conventional Binary and Modified Binary Codes

Decimal Number	Conventional Binary	Modified Binary
0	0000	0000
1	0001	0001
2	0010	0011
3	0011	0010
4	0100	0110
5	0101	0111
6	0110	0101
7	0111	0100
8	1000	1100

9.3.7 Number of Digits

The number of digits affects the precision with which a measurement is expressed. A high degree of precision is not required unless the accuracy of the measurement is consistent with it. Much of the instrumentation in manufacturing processes is accurate to, say, no better than 2 per cent of full scale. Two decimal digits (0–99) or six binary digits (0–63) represent sufficient precision for this accuracy. If greater accuracy is needed, the number of digits can be increased to match it. In certain counters for measurement of time, as many as five decimal digits are used. An advantage of dealing with numbers is that the accuracy of measurement can be retained through transmission and data processing.

9.4 SOME TYPICAL EXAMPLES OF ANALOG-TO-DIGITAL CONVERTERS

So far we have been talking about converters very generally. It is worth while now to show some specific examples, so that the reader can get a clearer idea of these devices and also see the tremendous range of characteristics obtainable from quite different techniques. The converters have already been applied to such diverse fields as graphic-record analysis where relatively slow sampling rates suffice, to the communication applications where sampling rates are very, very high. The former field can use relatively simple mechanical methods, but the latter must rely on the fastest electronic methods. It is quite possible that automation may be able to borrow most from those techniques with slow sampling speeds. However, it may be more advantageous in many cases to use fewer converters of many channels on a time division basis. This would require sampling rates that could be satisfied only by electronic means. For this reason, we are illustrating some schemes which may not now be particularly appropriate but may become so in the future. In any case, they help to show more clearly the workings of these converters and may stimulate newer and better ideas. It is convenient again to break them down into three classes: those that count, those that compare, and those that read.

9.4.1 Converters That Count

A familiar device which qualifies as a primitive analog-to-digital converter is the Veeder-Root counter. This is essentially a device which counts revolutions of a shaft and indicates the answer on a decimal digital scale. With simple linkages it can be applied to other parameters, and we find them used on clocks for time, on gasoline pumps for volume measure, and on kilowatt-hour meters for electrical energy. Although the result must be read by someone, it is easy to see that electrical closures could be included which would permit reading and recording on cards or type.

The Veeder-Root counter is strictly a mechanical device and is limited to a relatively slow counting rate. The electronic counter, which is quite similar in principle, can count at extremely high rates on the order of 1,000,000 counts per second. When it is combined with an accurate crystal oscillator, it is very useful for measuring time intervals. This is shown in Fig. 9.1*a*. A start pulse and a stop pulse defining the beginning and the end of a time interval are fed into an electronic gate. The start pulse opens the gate and allows pulses generated by the crystal oscillator to be fed to the electronic counter. The answer appears in glowing neon lights on the counter, or on other read-out equipment. Also, by supplying an accurate time interval, an unknown frequency can be measured in place of the crystal oscillator frequency. These devices are being made as production items by many companies. The oldest in the field are the Hewlett-Packard Company, the Potter Instrument Company, and Berkeley Scientific Corporation.

It is possible to convert an analog voltage to a time interval, and thus the scheme just described can be used for a wide variety of parameters that can be transformed to voltages. One method for doing this is shown in Fig. 9.1*b*. A timing generator produces a start pulse and also initiates a ramp voltage which is compared with the unknown voltage. When the two are equal, a stop pulse is produced.

Other devices in this class are used for measurement of shaft rotation. They make use of index marks on a disc attached to the shaft. These marks may be holes in the disc, or serrations on the rim, or locally magnetized spots on the face. Optical or magnetic means are used for detecting the passage of the marks. Fairly elaborate control arrangements are required in these schemes, however. These are marketed by Engineering Research Associates, Telecomputing Corporation, and others.

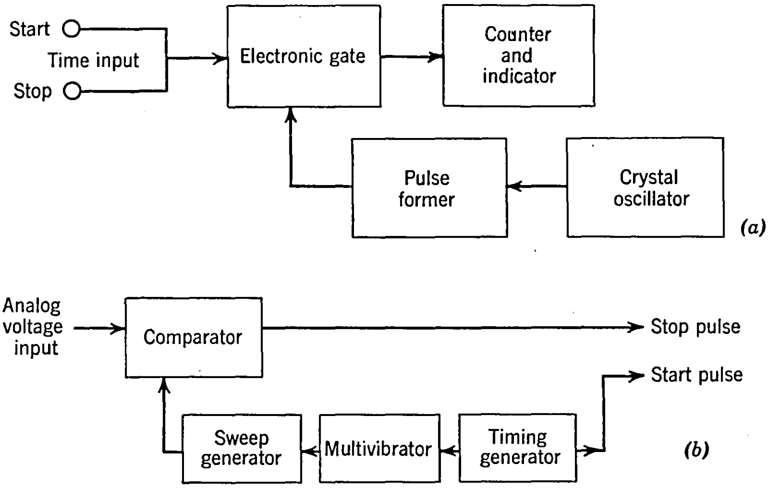


Fig. 9.1 (*a*) Counter-type converter for time intervals; (*b*) arrangement for transforming an analog voltage input to a time interval

9.4.2 Converters That Compare

Converters that compare resemble self-balancing-type potentiometers and other servomechanism-type devices. They contain an error-sensing device which tells when the unknown value is equal to the known value.

One multichannel device made by Consolidated Electrodynamics Corporation and used in the data reduction field is a self-balancing potentiometer. The conventional slide wire is replaced by three groups of calibrated resistors with ten resistors in each group. These groups represent the digits of a three-digit decimal number. Three servo-controlled stepping switches sweep through these groups of resistors. The switches act until a voltage equal to the unknown input voltage is produced, as indicated by an error-sensing device. Digital contact closures are provided to actuate conventional read-out equipment.

Another device, which uses electronic switches rather than relays, is shown in Fig. 9.2. This is a servomechanism-type arrangement in which the reading of the counter is compared to the unknown input. If there is a difference between the two, the error signal opens a gate, allowing pulses from a generator to feed into the electronic counter. A digital-to-analog converter transforms the numbers in the counter into a negative feedback voltage proportional to the number. When this voltage agrees with the input voltage, the gate closes and the counter stops. The counter must be able to count down as well as up. The direction is determined by the sign of the error. It should be noted that the comparison is obtained by introducing a count of one unit at a time.

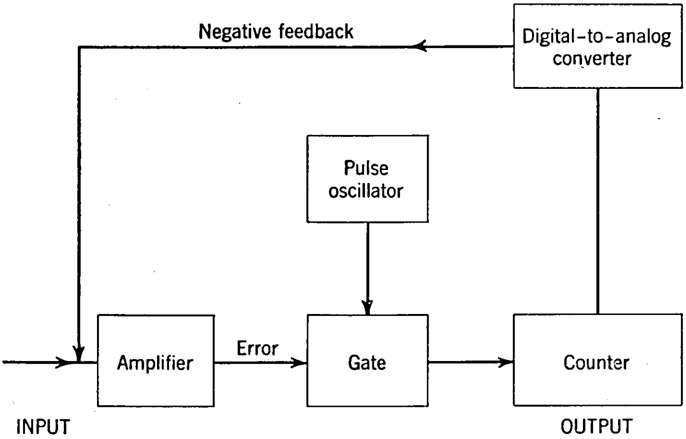


Fig. 9.2 Comparison-type converter for voltage inputs; comparison is obtained by advancing counter one unit at a time until no error is indicated

Another comparing-type converter is shown in Fig. 9.3. It is electronic but differs from the preceding one in that the comparison is brought on by starting with the highest-value digit rather than the units digit. The input voltage is sampled and the sample is stored on a capacitor. Assuming the result is to be expressed as a seven-digit binary number or as 1 part in 128, the sequence of operations is as follows. The voltage on the capacitor is first compared with a known voltage of 64 units. If it is larger, an output pulse representing the first digit of the number is generated, and at the same time 64 units are subtracted from the capacitor. If it is smaller, no pulse is generated and no subtraction takes place. The capacitor voltage is then compared with known voltages of 32 units, 16 units, 8 units, 4 units, 2 units, and 1 unit in turn, and the same sequence is followed. This method of comparison is more economical in time than the previous one. A converter of this type was used on one of the first pulse-code modulation systems (6) and provided sampling rates of 8000 a second.

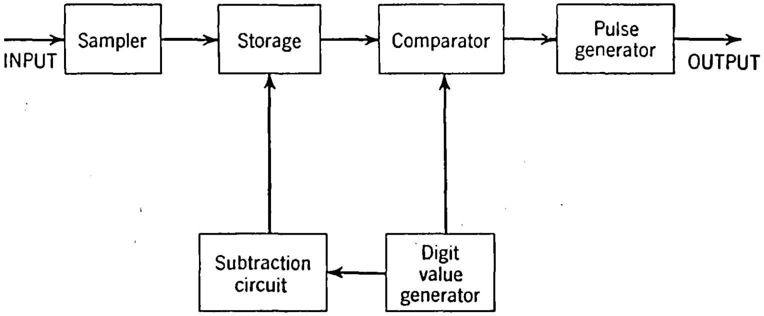


Fig. 9.3 Comparison-type converter for voltage inputs in which unknown is compared with successions of digit values starting with the highest

9.4.3 Converters That Read

It is convenient in many cases to transform the parameter to be measured to an indication that will permit a number to be directly associated with the indication.

Consider, for example, the commutation arrangement shown in Fig. 9.4. Here, not much transforming is required because the input is an angular position. The input drives a brush on its own shaft and also drives two other brushes on shafts geared down in ratios of 10:1 and 100:1, as shown. Each brush is in contact with a commutator of 10 segments; each commutator represents a digit of a three-digit decimal number. The segments can be connected to the usual machine read-out equipment. More commutators could be geared together to give more digits. Although simple in concept, this scheme is more difficult to realize in practice. The brushes and the segments, and the insulation between the segments, are of finite size. This makes it possible for the brush to straddle more than two bars if it is large, or to lie in between two bars if it is small. Either one is a source of ambiguity in the result. A number of different schemes have been used to eliminate this difficulty. One scheme interposes a relay control (7) system which favors the lower number until the brush has moved completely off the lowest-order bar representing that number. Other devices make use of star wheels or other mechanical arrangements to position the brushes properly without ambiguity either continually or during read-out. Some companies making mechanical-type converters are Cox and Stevens Aircraft Corporation, Streeter Amet Corporation, Taller and Cooper, Benson-Lehner Corporation, and Genisco, Incorporated.

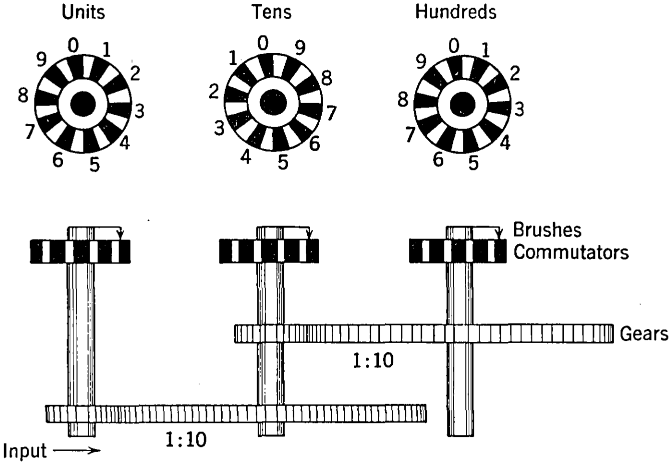
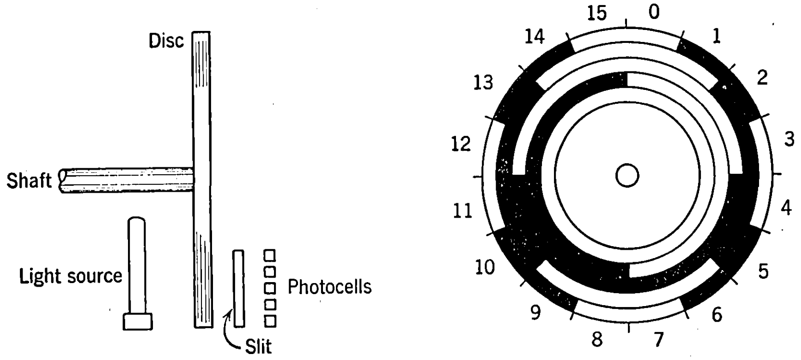


Fig. 9.4 Direct-reading converter for angular shaft position; decimal commutator

Another ingenious example (8) is shown in Fig. 9.5. This is also meant for an input of angular position but its output is read optically. It consists of a transparent disc mounted on the shaft which is to be monitored. The disc is divided into a number of sections which give the desired precision. Each sector is then coded radially in opaque and clear areas corresponding to the reflected binary code. Facing the disc is a narrow slit. Behind the slit is a row of photocells, one for each digit of the number. The cells are excited by a light source on the other side of the disc, providing an opaque area is not interposed. Here a straddling of two numbers is also possible, but the "reflected" code can be in error by at most only one unit distance, i.e., the proper number can be mistaken only for one adjacent to it. This device and the preceding one could be coupled directly to the servomotor of conventional bridge-balancing-type instruments. This means that a whole realm of process parameters could be very conveniently converted to digital form. The G. M. Giannini and Company, Inc., has announced the manufacture of this type of device.

The third example (9, 10) of this group is shown in Fig. 9.6. Here a preliminary transformation is required because the input is voltage. This is converted to a deflection in a modified cathode-ray tube. The unknown voltage on one pair of plates deflects the beam vertically, according to the voltage present. A sweep voltage then moves the beam across a mask which contains openings corresponding to the binary number for each position. The reflected code can be used to eliminate ambiguity, or a system of quantizing grid wires can be built into the tube to accomplish the same result for the straight binary code. A



Modified, reflected, or cyclical binary code

Fig. 9.5 Direct-reading converter for angular shaft position; binary-coded disc

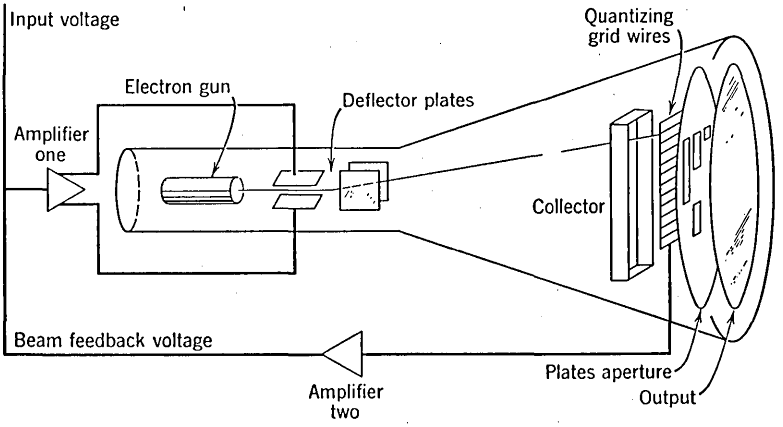


Fig. 9.6 Direct-reading converter for voltage inputs; modified cathode-ray tube and binary-coded mask for electron beam

serial output is obtained and the results can be fed to shift registers or to tape. Parallel output can also be obtained by using a fan-shaped beam and breaking the target plate up into elements corresponding to the number of digits in the number. This is a very fast converter and has been used successfully in coding the video signal of television.

The examples described should give a rough idea of the principles used in the design of these devices and may suggest others. The most common input to these devices are angular position, voltage, frequency, and time. These were good choices because transducers are available to transform most of the interesting parameters to these. A wide range of sampling rates has been obtained. The slowest are those requiring manual resetting, i.e., those that put out a number when someone presses a button. Others operate at the relatively slow speeds of 10 through 100 samples a second. Extreme rates of 8000 samples per second and above are secured in those devices meant for pulse-code modulation in telephone and television systems.

Multichannel operation is quite common, with the number of channels varying from 10 to 100. Many companies are now engaged in supplying converters. In Section 9.8, Appendix, they are listed with their addresses, so that the reader will be able to refer requests for more detailed information to the particular companies.

9.5 SOME SPECIFIC EXAMPLES OF THE USE OF ANALOG-TO-DIGITAL CONVERTERS

9.5.1 Problem: Time Recording

One of the oldest examples of analog-digital conversion is that of converting time—the length of time a person has worked, for example, converted into units of pay. Time is strictly an analog value whereas dollar bills are measured by counting. The problem is one of establishing a time standard, making measurements in relation to the standard, and at any required instant recording the time, in the digital language of the accounting department.

Manual solution. In the manual solution the escapement of a spring-driven clock or the frequency standard of a power station in the case of the electric clock provides the standard. The hands on the face of the clock present an analog display of the accumulated total, and the dial provides a scale against which a human being can compare and read the total. In reading the time to write down on his time card, the man first notes in digital form the major divisions on the dial nearest to the hands of the clock; then he interpolates the position of the hand between these two numerical marking posts. He stores this information momentarily while he hunts for his pencil and writes the number down on his time card.

Automated solution. In an industrial time clock, a man slips his card into a slot and the time is printed automatically. In operation, the clock is really a pulse accumulator. It receives impulses every minute and accumulates them through a gear train system. A mechanical detent takes the responsibility for the final alignment of the print wheels which are driven by the gears. This system is essentially a Veeder-Root counter.

9.5.2 Problem: Monitoring Oil Storage Tanks

Oil storage tanks are dotted all over a group of hills in the most inconvenient places. It is important to know at least once a day and preferably twice a day what the oil level is so that the people who are concerned with productivity can keep track of what is going on.

Manual solution. In most oil fields the human analog-digital converter is still used today. He wears rubber boots in wet weather and a sun helmet in hot weather and trudges from tank to tank on his daily rounds, writing down the height against a scale of a ball hanging on a piece of cord which is attached, by means of a pulley, to a float on the surface of the oil.

Automated solution. Here we have a mechanical displacement in the form of a float-operated cable. This can readily be passed over a drum, and on the drum can be attached a commutator-type converter like the one shown in Fig. 9.4 or the disc type shown in Fig. 9.5 A similar unit can be placed on each of the other reservoirs in the field, and then each unit can be connected electrically to a central tabulating unit in the office. The converters can be read out on demand by the operator pressing a button, or they can be triggered by a clock to read automatically at certain times. There are, of course, many variations of the details. For example, a two-wire telephone system could be used with a scanning device reading the numbers sequentially, or a multiple-wire system can be used with no scanner. At the office end, a typewriter can do the logging or the oil-level information can be fed directly to punch cards or punch tape.

9.5.3 Problem: Seismographic Work

Seismographs are busy in the field collecting information with a geophysical exploration crew. There is, of course, quite a high motivation to find out where the oil deposits lie. However, the data will need a lot of processing before the information it contains can be broken loose.

Manual solution. In the majority of cases the information is still recorded on oscillographs. Later the record is studied manually, measurements are made on it, and decisions are made on the basis of the measurements.

Automated solution. More and more of the work in connection with seismographic exploration is being processed on large digital computers. This involves converting the seismographic information into the digital computer's language. An entirely different form of analog-digital conversion is involved as compared with the previous examples. The information as it is received from the geophones is in electrical form. If the subsequent processing is to be entirely automatic, it has to be recorded in a form that is easy and convenient to play back. For this reason it is being converted into a frequency-modulated form and recorded on magnetic tape. A converter like the one shown in Fig. 9.1*a* can pick up the information from there and, by counting out the cycles in precise short time spans, convert it to digital form and load it into a computer.

9.5.4 Problem: Analysis of Graphic Records

Tens of thousands of feet of records are daily pouring out of oscillographs, pens, not to mention high-speed motion cameras and ballistic cameras. In the majority of cases these data are in the raw form. The displacements have to be measured, often simple computations such as additions and multiplications have to be applied, and the data have to be tabulated or fed into a computer.

Manual solution. Although this job has been extensively mechanized, there still remain installations where girls hover over endless rolls of oscillographic records, looking at the mysteries of scores of crisscrossing lines and judging their positions in relation to hand scales. Usually each girl has a companion who busily writes down the numbers as she calls them out. This job involves some interpretations, a great deal of analog-digital conversion, and extensive recording.

Automated solution. Inasmuch as the majority of records are complex and the human being is more economical in the actual interpretation than is an automatic machine, the problem involves placing a convenient pointer or reference line in the operator's hands and then automatically converting the displacement into digital form. We can handle this two ways, depending on the details of the problem. A mechanical converter such as that of Fig. 9.4 or Fig. 9.5 can be attached to the reading reference line by a cable and drum arrangement. The linear motion is now converted to a rotary motion and is available for the commutator to convert to digits. If we want to do a certain amount of simple analog computation on the data which does not require the complexity of a computer, we can use an analog transducer, a potentiometer, to convert the mechanical displacement to a resistance value. With a simple bridge circuit we can add and subtract, multiply and divide, and now we can take our modified information in electrical form, ending up as a voltage, and convert it to a number with devices like those shown in Figs. 9.1, 9.2, 9.3, and 9.6.

9.5.5 Problem: Process Control Logging

Thermocouples, pressure transducers, and other instruments are spotted around the crucial points of a chemical processing plant. There is a need for continual logging of the overall operation to provide general information on performance and also to indicate deviations from predetermined limits. There is also a separate growing need which stems from the fact that engineers are forging ahead with plans for automatic plant control. They would like to be able to connect a digital computer to the whole system. This requires a very fluent conversation from the instruments to the computer and from the computer back to the implementers, with a large part of this conversation in digital form.

Manual solution. To consider just the logging part of the problem, the information is normally displayed by manometers, on dials, strip chart recorders, and other instruments. Here the human being must perform analog-digital conversion. He walks down the line taking readings from all the relevant instruments as accurately as he can and with as short a time delay as possible between the readings. However, the sampling rate has a fundamental speed limitation when the information is being routed through a human being in this way.

Automated solution. The majority of our quantities to be measured are in electrical form, and those that are mechanical displacement can quite cheaply be converted to electrical values by potentiometers. If we were starting from scratch we could route all of this information through a converter line (Figs. 9.1, 9.2, 9.3, and 9.6), sharing it in turn with all instruments, just as the human analog-digital converter shares himself in turn with all instruments. However, for the sake of description, let us assume that we already have many self-balancing potentiometers such as the Bristol, Brown, and Leeds & Northrup instruments. We can attach angular-position converters like those of Figs. 9.4 and 9.5 onto the servomotor to form an electromechanical system. This has the possible advantage of making use of the existing instruments with which the crew are already familiar and of maintaining the customary dials. This approach also provides for simultaneous parallel output from several channels, although these benefits are lost if the information is then fed to a serial recorder such as a typewriter. These are but examples of factors that come into the choice of systems. There are, of course, many others.

9.6 CONTROL APPLICATIONS

Similar examples of the use of converters can be expected to multiply rapidly in the future. Beyond this, however, are other exciting prospects for their use, particularly in those control applications where the digital computer is made an integral part of a process. The computer will make many routine decisions based on the incoming data and a programmed criterion of performance. These applications are, for the most part, still largely in the future, but already two very interesting illustrative problems have been explored.

The first (11) deals with the flight control of an airplane responding to fire control signals. The second (12) is concerned with control of the airplane for navigation from loran-type signals. Both represent examples of fairly difficult process control, because a number of parameters must be monitored, and fairly complicated equations must be solved. Also, the final equipment must be consistent in size and weight with aircraft use. The analog-to-digital converters in the first example were similar to those of Fig. 9.5. The original parameters transferred to a shaft rotation were pitch, pitch rate, altitude, and other aerodynamic terms, plus others like mass, Mach number, and control surface positions. The computer converted the reflected binary code into conventional binary before performing computations; the computer also smoothed much of the data statistically.

The parameters converted to digital form in the second example, the Digitac system, are altitude, heading, air speed, and time. The time is expressed digitally to a precision of 1 part in 30,000. These are all manipulated in the computer and an output is fed to the autopilot. The aircraft could be successfully navigated by these means to go through several check points on a given route.

These two examples show very vividly the possibilities for a manufacturing process, because not many processes would require such close control, and none would be subject to such severe limitation on size and weight.

9.7 DOES THIS FACET OF AUTOMATION APPLY TO ME?

The decision of whether to automate or not is purely an economic one. How long will the automatic system and its ramifications take to pay for themselves in comparison with manual methods? This much is simple, but what is more difficult is establishing the criteria for evaluating savings and increased productivity. There are a very large number of factors depending on the particular situation, a few of which are as follows:

(1) Saving in labor costs.

(2) Saving in input and output equipment connected with human beings.

(3) Increased security in the operation from greater speed and from greater accuracy.

(4) Increased productivity in the operation from the same factors, i.e., speed and accuracy.

(5) The opportunity to undertake new types of operations based on the capabilities of the automatic system.

However, we must have a representative manual and an equivalent automatic system in order to make these sorts of comparisons. To arrive at the automatic system, it is very important not to start off with any particular converters or units in mind but to lay out what fundamentally has to go into one end of the system and what is basically required at the other end of the system. The trick is then to connect these two with the very least number of elements in between.

We have to consider the number of different inputs into the system, what type they will be, and where the sources of information have to be located. We have to consider the required accuracy and required speed. At the other end of the system we have to exercise great care in considering storage, recording, and display elements to make sure that they are designed to do as complete a job as possible in helping either the automatic machine or the human being to make a decision. Only when the requirements of the system have been laid out can we begin to consider components.

When the chain of elements has been built up, it is a good idea to ask ourselves, "If this whole system were built, what would be the weakest link?" This can, of course, create a problem of how to bypass the weak element, but it is a lot better to meet and solve these problems at the thinking stage rather than at the doing stage.

Finally, it is a very good thing to sit back and make sure that we have not evolved a very good solution of a problem that did not really exist in the first place. In laying out a process it is very easy to design a measuring system to feed a converter to supply information to a computer which will instruct implementers to meet requirements—which could have been avoided in the first place!

9.8 APPENDIX: SOME MANUFACTURERS OF ANALOG-TO-DIGITAL CONVERTERS

Atomic Instrument Company
  Cambridge, Massachusetts
The Austin Company
  New York, New York
Beckman Instruments, Inc.
  Fullerton, California
Bell Telephone Laboratories
  Murray Hill, New Jersey
Bendix Aviation Corporation
  Pacific Division
  North Hollywood, California
Benson-Lehner Corporation
  West Los Angeles, California
Berkeley Division, Beckman Instruments, Inc.
  Richmond, California
Bristol Company
  Waterbury, Connecticut
Brown Instruments Division
  Minneapolis-Honeywell
  Philadelphia, Pennsylvania
Coleman Engineering Company
  Los Angeles, California
Consolidated Electrodynamics Corporation
  Pasadena, California
Dayton Instruments, Inc.
  Dayton, Ohio
Electro-Instruments
  San Diego, California
Engineering Research Association
  St. Paul, Minnesota
Epsco, Inc.
  Boston, Massachusetts
Fischer & Porter Company
  Hatboro, Pennsylvania
Genisco, Inc.
  Los Angeles, California
G. M. Giannini and Company, Inc.
  Pasadena, California
Hanson-Gorrill-Brian, Inc.
  Glen Cove, New York
Helipot Corporation
  Division of Beckman Instruments, Inc.
  Pasadena, California
Hewlett-Packard Company
  Palo Alto, California
Kearfott Company, Inc.
  New York, New York
Kellogg Switchboard & Supply Company
  Chicago, Illinois
Leeds & Northrup Company
  Philadelphia, Pennsylvania
Librascope, Inc.
  Glendale, California
Arthur D. Little, Inc.
  Cambridge, Massachusetts
Minnesota Electronics Corporation
  St. Paul, Minnesota
F. L. Moseley Company
  Pasadena, California
Non-Linear Systems, Inc.
  Del Mar, California

Oerlikon Tool & Arms Corporation
Asheville, North Carolina
Potter Instrument Company
Great Neck, New York
J. B. Rea Company
West Los Angeles, California
Streeter-Amet Corporation
Chicago, Illinois
Taller & Cooper
Brooklyn, New York
Telecomputing Corporation
North Hollywood, California
Victor Adding Machine Company
Chicago, Illinois

9.9 REFERENCES

Referred to in Text

1. B. M. Oliver, J. R. Pierce, and C. E. Shannon, The Philosophy of PCM, *Proc. IRE, 36,* No. 11: 1324–1331, Nov. 1948.

2. Staff of Engineering Research Associates, *High Speed Computing Devices,* McGraw-Hill Book Co., Inc., New York, 1950, Chapter 7.

3. George G. Bower, Analog to Digital Converters, *J. Instrument Soc., 1,* No. 12: 15–19, Dec. 1954.

4. R. K. Richards, *Arithmetic Operations in Digital Computers,* D. Van Nostrand Co., Inc., New York, 1955, Chapter 6.

5. Kay Howard Barney, The Binary Quantizer, *Elec. Eng., 68,* No. 11: 962–967, Nov. 1949.

6. W. W. Goodall, Telephoning by Pulse Code Modulation, *Bell System Tech. J., 26,* No. 3: 395–409, July 1947.

7. Arthur L. Klein and Kenneth P. Gow, A Dial-Reading Translator for Digital Machine Inputs, *Trans. AIEE, 71:* 1–8, 1952.

8. B. Lippel, A High Precision Analog-to-Digital Converter, *Proceedings of the National Electronics Conference,* Vol. 7, 206–215, 1951.

9. R. W. Sears, Electron Beam Deflection Tube for Pulse Code Modulation, *Bell System Tech. J., 27,* No. 1: 44–57, Jan. 1948.

10. W. M. Goodall, Television by Pulse Code Modulation, *Bell System Tech. J., 30,* No. 1: 38–40, Jan. 1951.

11. Maier Margolis and Eric Weiss, An Experimental Digital Flight Control System, *Proceedings of the Western Computer Conference, AIEE,* 23–37, April 1954.

12. D. W. Burbeck, E. E. Bolles, W. E. Frady, and E. M. Grabbe, The Digitac Airborne Control System, *Proceedings of the Western Computer Conference,* AIEE, 38–44, April 1954.

General

1. Ralph B. Bowersox and Chester G. Hylkema, High-Speed Recording Potentiometer, Memo No. 20–69, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, Jan. 16, 1952.

2. J. F. Brinster and E. Donath, Data Reduction Methods for Radio Telemetering Systems, Applied Science Foundation of Princeton, presented at the Western Electronic Show and Convention, Long Beach, Calif., Aug. 27, 1952.

3. J. H. Brown, Automatic Measurement of Lineal Motion Using Optical Gratings, *DIC 6873 Eng. Rept., 3,* 1951.

4. D. W. Burbeck and W. E. Frady, Precision Automatic Time Measurement Equipment, Hughes Aircraft Co., Culver City, Calif., presented at IRE National Convention, New York, March 5, 1952.

5. H. E. Burke, Jr., A Survey of Analog-to-Digital Converters, *Proc. IRE, 41:* 1455–1462, 1953.

6. Bert A. Coss, D. R. Daykins, Leonard Jaffe, and Elmer M. Sharp, A Digital Automatic Multiple Pressure Recorder, *NACA Tech. Note 2880,* Jan. 1953.

H. G. Follingstad, J. N. Shive, and R. E. Yaeger, An Optical Position Encoder and Digit Register, *Proc. IRE, 40:* 1573–1583, Nov. 1952.

7. G. E. Foster, The Metrotype System of Recording and Telemetering, *Elec. Eng., 69:* 427–430, May 1950.

8. B. M. Gordon, M. A. Meyer, and R. N. Nicola, A Shaft to Digital Encoder, *Proceedings of the Western Computer Conference,* AIEE, 128–133, April 1954.

9. H. J. Gray, Jr., P. V. Levonian, and M. Rubinoff, An Analog-to-Digital Converter for Serial Computing Machines, *Proc. IRE, 41:* 1462–1465, Oct. 1953.

10. Gerhard L. Hollander, Criteria for the Selection of Analog-to-Digital Converters, *Proceedings of the National Electronics Conference,* Vol. 9, 670–683, 1953.

11. Donald H. Jacobs, A Device for Converting a Shaft Rotation to an Input for a Digital Computer, Jacobs Instrument Co., Bethesda, Maryland, presented at Annual Meeting of American Physical Society, New York, Feb. 1, 1952.

12. Arthur L. Klein and Kenneth P. Gow, A Dial-Reading Translator for Digital Machine Inputs, *Trans. AIEE, 71:* 1–8, 1952.

13. M. L. Kuder, Anodige, An Electronic Analogue-to-Digital Converter, *Nat. Bur. Standards Rept. 1204-34-1202,* Aug. 24, 1951.

14. M. L. MacKnight and P. A. Adamson, Multichannel Analog Input-Output Conversion System for Digital Computer, *Convention Record of the IRE* (Electronic Computers), Part 7, 2–6, 1953.

15. J. O. McDonough and A. K. Susskind, A Numerically Controlled Milling Machine, Joint AIEE-IRE-ACM Computer Conference, AIEE Special Publication S-53, 1952.

16. L. A. Meacham and E. Peterson, An Experimental Multichannel Pulse Code Modulation System of Toll Quality, *Bell System Tech. J., 27:* 1–43, Jan. 1948.

17. H. W. Mergler, A Digital-Analog Machine Tool Control System, National Advisory Committee for Aeronautics, Lewis Flight Propulsion Laboratory, Cleveland, Ohio.

18. James M. Mitchell, A High-Speed Multichannel Analog Digital Converter, *Proceedings of the Western Computer Conference,* AIEE, 118–127, April 1954.

19. A. D. Scarbrough, An Analog-to-Digital Converter, Rept. M-58, Hughes Research and Development Laboratories, Hughes Aircraft Co., Culver City, Calif., Aug. 1953.

20. W. S. Shockency, Multi-Channel Analog-Digital Conversion System for D. C. Voltages, *Proceedings of the Western Computer Conference,* AIEE, 113–117, April 1954.

21. Sink and Slocomb, The SADIC, A Precision Analog-Digital Converter, Consolidated Electrodynamics Corporation, presented at the IRE Data Handling Symposium, Long Beach, Calif., Aug. 1951.

22. R. L. Sisson and A. K. Susskind, Devices for Conversion between Analog Quantities and Binary Pulse-Coded Numbers, A.M. thesis, MIT (1950).

23. D. W. Slaughter, An Analog-to-Digital Converter with an Improved Linear Sweep Generator, *Convention Record of the IRE* (Electronic Computers), Part 7, 1953.

24. Richard L. Smith, A High-Speed Potentiometer for Recording on Punched Paper Tape, NACA, Lewis Flight Propulsion Laboratory, Cleveland, Ohio.

25. A. K. Susskind, Approaches to the Design Problems in Conversion Equipment, *Proceedings of the Western Computer Conference,* AIEE, 105–112, April 1954.

26. Richard L. Trent, A Transistor Reversible Binary Counter, *Proc. IRE, 40:* 1562–1572, Nov. 1952.

10. Input-Output Equipment

WALTER F. BAUER

The Ramo-Wooldridge Corporation
Los Angeles, California

10.1 INTRODUCTION AND HISTORY

Input-output equipment for the modern digital computer provides the means of communication between the machine and its users. This equipment, together with the control panel, forms the "sensory organs" of the computer through which information entering or leaving the computer must pass. The importance of speed and reliability cannot be overemphasized. The situation with regard to speed can be likened to the jet aircraft which becomes ineffective as a means of transportation because the average airport runways are too short for landing. As for reliability, it is easy to imagine the reduction in effectiveness of a scientist with poor eyesight who, although intelligent and quick-witted, can read at the rate of only one word in ten seconds and with a high frequency of mistaking various words for others.

Input-output equipment is defined as the equipment used to get information into and out of the internal components of the computer, usually the high-speed memory. A category of equipments very closely associated with input-output equipment, but strictly not of that type, is the category known as "off-line" equipment. This means that the equipment is not directly connected to the computer, for example a printer that operates from information on a magnetic tape, the tape having served previously as an output medium for the computer. This equipment is alternatively called "auxiliary" or "peripheral" and is used to transfer information from one medium to another external to the computer. In some cases equipment may be used in either role—input-output (on-line) or off-line operation.

Historically, the development of input-output equipment, usually of an electromechanical character, lagged behind the development of the electronic computer proper. In the period 1946–1950, when most of the first stored-program computer systems were planned, it was difficult to foresee the need for scientific computer output at the present-day rates of 2000 words per second and higher. Also, extensive commercial applications with their great demands on input-output systems seemed more than ten years away. Hence, although the objectives in the development of input-output equipment have been speed and reliability, only in recent years has sizable effort been expended in those directions.

In Chapter 8 the terms "computer-limited" and "input-output-limited" are defined, with reference to the speed at which a computer system solves problems or performs data processing. When a system is computer-limited with respect to a problem, the computer internal speed is the limiting factor, and no substantial savings in time would be obtained by speeding up the input-output process. When the system is input-output limited, the situation is reversed. Generally, scientific problems with their heavy demand on arithmetic speeds are computer-limited, whereas the business data-processing problems are input-output limited.

There is a rapidly increasing awareness, however, that the input-output requirements for the scientific computer are not as easily met as had been first estimated. Activity in two areas is the cause: automatic programming (or coding) and the extensive use of the computer in data reduction. The development of automatic programming implies systematizing certain clerical procedures so that they can be performed quickly by the computer rather than slowly by the programmer, with economy favoring the computer. In the light of the shortage of computer applications personnel, automatic programming looms as an extremely important development. Automatic program compilation and code checking in automatic programming places heavy demands on input-output. With most computers used for scientific purposes, upward of 35 per cent of the time is used in code checking, which involves the input and output of vast amounts of data. The other activity, data reduction, involves the handling of the large amount of engineering data from test runs in missile, aircraft, and other industrial development work. Radio telemeter data from the craft itself, ground radars, and ground optical instruments amount to millions of recorded points for one missile or aircraft test. A large fraction of this data must be read into a computer for calibration and analysis.

From humble beginnings in 1950–1951, when computer input-output consisted of only Teletype equipment operating at about 10 characters per second or less, the modern computer has emerged with many different devices operating at speeds up to 15,000 characters per second. The IBM-704 is such a computer (Fig. 10.1). This computer has the following input-output devices: punched-card and magnetic-tape input; punched-card, line printer, cathode-ray tube, magnetic-tape output; and card-to-tape-to-card equipments. All of these devices will be discussed in this chapter.

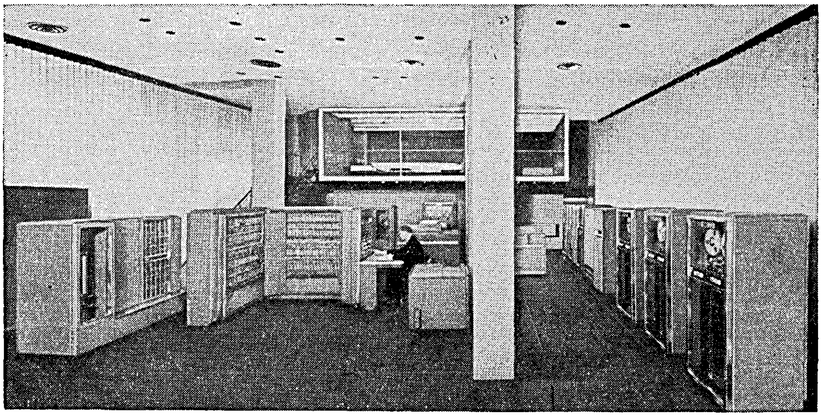


Fig. 10.1 IBM-704 computer system

The SEAC (Standards Eastern Automatic Computer) is generally considered to be the first operating stored-program computer. Approximately 20 minutes was required to fill SEAC's 512-word memory with the Teletype equipment first used. During this 20 minutes the computer was inhibited from performing any of the million or more arithmetic operations which could have been performed. An important step in the development of input-output equipment was that made at the Bureau of Standards when a magnetic-wire input-output dumper system was added to that computer's facilities. Although the speeds are not high by today's standards, the ability to transfer data in and out of the computer at a rate of about 50 words per second was a significant achievement in 1951. In this development a commercially available office-dictating unit was adapted for use. A length of wire was contained in a cartridge, and mounting and dismounting the wire during operations amounted to a simple snap-in and release to establish a level of convenience seldom seen in modern equipments. Because multichannel recording on magnetic wire was impossible and because the wire was vulnerable to breakage, the system was not adopted by other computing installations.

Another important step in the development of input-output devices was the use of magnetic tapes with the UNIVAC computer, the first of which was delivered in 1951. The 60-word input-output buffers of the UNIVAC, the first buffer system to appear, allowed a highly efficient use of the magnetic tapes as input-output devices and auxiliary storage. The concept of a buffer is explained below.

Reliability has always been a problem. With most computing systems, over half the down time is caused by input-output failures. Paralleling the increase in speed, reliability has increased greatly over the years. Probably the greatest motivation for increased speed and reliability has been the use of computers for business applications. Potential customers for business data processing have insisted that reliability be kept at a high level and have given as very compelling reasons the irritated employee who receives an incorrect salary check and the irate customer who receives an incorrect billing.

The lack of redundance in the form of information handled in automatic systems often implies that an entire set of calculations is invalidated when one character is incorrectly read. The message in a telegram is not rendered incomprehensible by such an error, although thousands of dollars of computer time can easily be lost by one incorrect character. Thus reliability of the input device is important in all computer applications.

The fact that input-output equipment constituted approximately 1 per cent of the computer system cost in the system of 1950–1951, whereas in current systems this equipment constitutes 15 per cent or more of the total computer system cost, demonstrates the increasing awareness of the importance of input-output equipment and the great strides that have been made in its development.

The first equipments used extensively with computers were the punched-paper-tape readers, punches, and page printers made by the Teletype Corporation. Their many years of use and proven reliability in communications made them attractive, and they were used on the Harvard-IBM MARK II and Bell Telephone Laboratories computers. Somewhat later, punched cards were used as input-output on IBM's

SSEC (Selective Sequence Electronic Calculator). Later, Teletype equipment was used with the first input-output developed for the SEAC and EDVAC (Electronic Discrete Variable Automatic Computer) computers, although it has since been replaced by faster equipment. The six-character-per-second input-output speed of the five-hole Teletype tape was soon found too slow, and computers appeared with the somewhat faster card readers, punches, and printers made by IBM. The proven reliability of punched-card equipment in accounting and the increase of speed over paper tape equipments were attractive. At about the same time higher-speed equipments for reading and punching paper tape came into use. Not long after came the extensive use of magnetic-tape equipments with designs that allowed dismounting of the tape reels for off-line use. The use of magnetic tapes thus allowed an extremely rapid transfer of data out of the computer. The slower processes of printing information from the tape and transferring information from printed documents or cards to the tape could be done outside the computer, without interfering with the central processing effort. The next developments for greater speed were devices that can read a printed page and nonmechanical printing devices utilizing cathode-ray tubes.

There are many devices that can be regarded as the input-output type which are not discussed in this chapter. These are special-purpose devices, such as those used in department store sales-data-gathering systems and input to automatic reservation equipment used in the railroad and airplane transportation industry. The analog-digital converter is in this same class, since it must be designed and used in close conjunction with the equipment generating the analog information. Also, many of these instruments generate data which are read into one of the equipments specifically mentioned here. Information from the analog-digital converter, for example, is frequently read into a conventional card punch or magnetic-tape unit and the resulting cards or tape used in the next step of the process.

10.2 RECORDING MEDIA

As information is transferred from written documents into the computer, a usual intermediate step consists of recording the information on a medium such as punched paper tape or punched cards. The cards or tape are used in the input process. In addition, this medium then forms an external high-volume storage for the computer. This type of storage can indeed be regarded as another stage in the memory hierarchy of the computer (see Chapter 7). In the case of punched cards, which are stored in file cabinets, the information is relatively inaccessible as compared with information stored on magnetic drums and magnetic cores. However, this type of storage external to the computer has been and will continue to be important for computers. In designing and choosing a recording medium, there are a number of objectives: low cost, reliability, permanency, erasability, durability, and handling ease. Depending on the design goals of the equipment, these objectives vary in importance. Reliability of the medium means that the medium lends itself to use with equipment that has a high degree of reliability. Erasability is considered by some to be a disadvantage, since the information may be inadvertently erased.

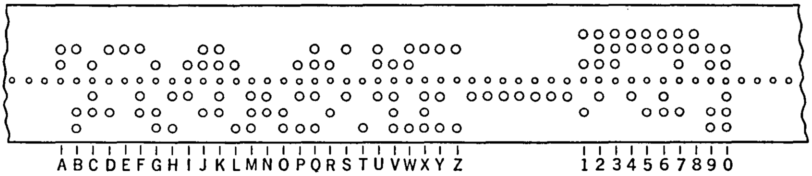


Fig. 10.2 Seven-level Flexowriter punched paper tape showing alphanumeric codes

Four types of media are in extensive use today: paper tape, punched cards, magnetic tape, and printed page. Paper tape and punched cards make up the type which is often referred to as "perforated media." Magnetic tape is the more modern form of recording medium with important advantages of compactness and higher-speed operation. The printed page as a form of output data is of the most importance, but there is the disadvantage that it cannot be read directly back into the computer. Equipment for direct reading of the printed page is under development and will find extensive use.

Punched paper tape has been used with computers since the first automatic equipments were developed. Teletype equipment uses five-level tape eleven-sixteenths of an inch wide; Flexowriter equipment uses seven-level tapes seven-eighths of an inch wide. Figure 10.2 is an example of seven-level punched paper tape with the Flexowriter code indicated for the alphabetic and decimal characters. In this code the seventh level is not used. Each column of hole positions is called a "frame." The tape is drawn past the reading station with sprocket drives, making use of the small sprocket holes or feed holes punched at the same time information is recorded. In some applications the tape is drawn past the reading station by friction drives, and no direct use is made of the feed holes, except for timing purposes. An advantage of paper tape is that the information cannot get out of sequence, since the information is recorded on one continuous piece of tape, and codes can be read visually. The disadvantages of paper tape are that it requires frequent manual windup, and that high information rates are difficult to achieve.

The modern punched card and the first equipment to use it were developed about 1900 by Hollerith, an engineer with the Bureau of Census. In Fig. 10.3 there is shown the standard 80-column, 12-row punched card showing the Hollerith code. This card is in very extensive use with accounting and tabulating machines, especially those built by IBM. In most applications, the cards are removed from a deck placed in the hopper of the reading or punching mechanism by picker knives. They are moved by rollers to the various stations for reading and punching and then stacked in a pile for removal. Higher input-output speeds are possible with punched cards than with punched paper tape, and many kinds of punched-card equipment are available. Cards are easy to handle and require no manual operation, such as winding in the case of the paper tape.

With the Flexowriter equipment it is possible to use upper-case characters, since one of the "codes" is the upper-case shift. This is not possible with Teletype equipment. Occasionally the holes in paper tape are not punched entirely through, thus allowing printing to appear on the tape "over" the holes. In the case of the punched card, the symbols corresponding to the hole combinations punched are printed in the appropriate column positions. This printing is shown in Fig. 10.3 on the card prepared with the IBM Type 026 Printing Punch.

Both paper tape and punched card are convenient for recording information in binary form. In a computer using a 36-bit word, for example, six punched-tape frames using six levels could describe a word. A total of 960 binary digits can be recorded on each punched card. For the 36-bit word, for example, 72 columns could be used to record 24 words. The first word could be recorded in the first 36 positions of the "9's row," the second in the second 36 positions of the "9's row," and so on.

In practice there is an important distinction between paper tape and punched cards. Cards frequently use the column position of the information to identify an item of information. Certain groups of card rows, called card fields, are usually reserved for certain types of data. This is necessary when data is processed by certain equipment such as sorters. An item of information recorded on paper

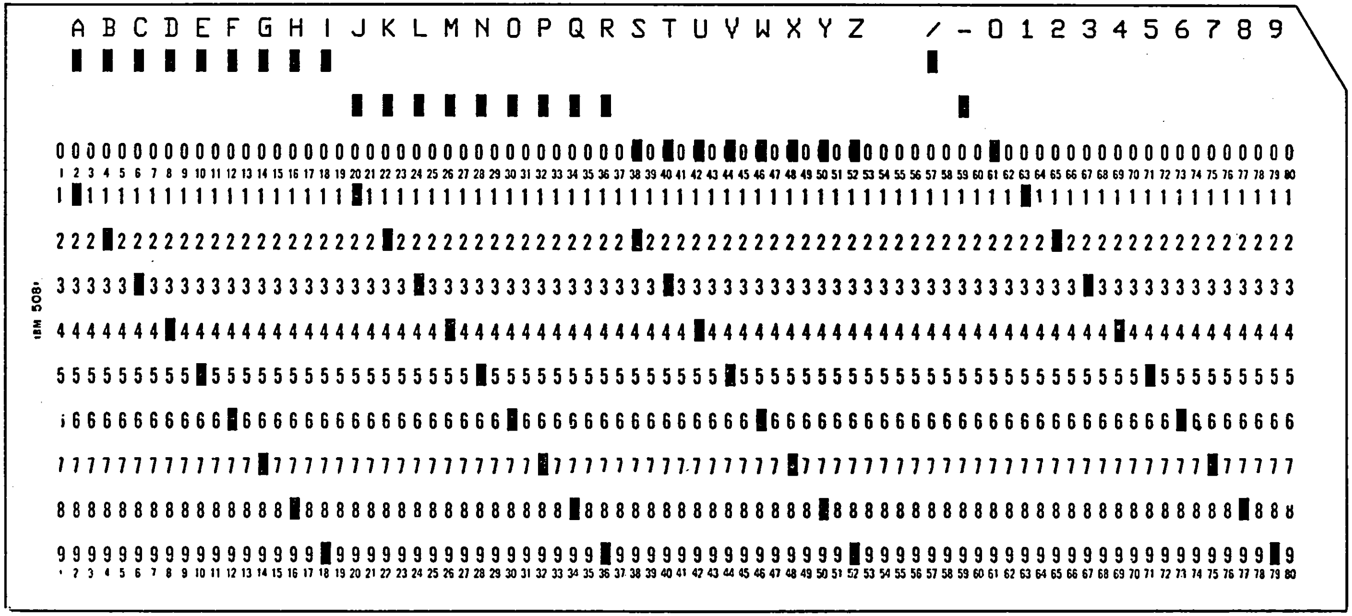


Fig. 10.3 Standard 80-column punched card with Hollerith code

tape is usually identified by its serial position. This implies that data on paper tape is compressed when certain information items are not present. This distinction is not a result of inherent differences between the two media but rather one resulting from practice and the characteristics of the equipment.

Magnetic tape is rapidly gaining popularity as a recording medium. Magnetic tape is manufactured by placing a magnetic oxide coating on a plastic or metallic base. Binary digits are recorded on the tape by magnetizing the surface in a pattern that can be recognized during the reading operation (see Chapter 7). The first use of magnetic tape was as an auxiliary memory device. It provided an inexpensive medium for the storage of bulk data during problem solution and allowed transfers to and from high-speed memory devices at acceptable rates. More recently, equipments have been designed for easy dismounting of tape reels. This feature combined with the availability of tape-to-card and card-to-tape converters, keyboard-to-tape recorders, and magnetic-tape printers make more convenient the use of magnetic tape as an input-output medium.

Two types of magnetic-tape base are in use, plastic and metallic. The advantages of the plastic tape are that it is cheaper and lighter than metallic tape. The advantages of metallic tape are that it has a great resistance to destruction by fire, does not tend to become brittle, does not stretch in use, and is not affected by humidity. A newcomer is Mylar plastic tape; it is stronger, is less affected by humidity, and is thinner than conventional plastic tape; it holds promise for replacing plastic and metal tapes now in use.

Figure 10.4 shows a schematic of eight-channel recording on magnetic tape. In the code used in storing the alphabetic and decimal characters, six channels are used to record the data, and a seventh channel is used to record a "parity bit." The parity bit is selected as a one or a zero to make an even number of binary ones across each "frame" of the tape for checking purposes, so that an error in one binary digit or an incorrect recording can be detected. The eighth channel is a sprocket channel or timing channel for recording pulses to time the reading of frames of information as they pass the reading heads. This eight-channel recording is very frequently used with magnetic tapes half an inch wide. Wider tapes using many more channels are used for recording more information per frame. The magnetic tapes in use with the ORACLE (Oak Ridge Automatic Computer and Logical Engine) computer at Oak Ridge National Laboratory, for example, employs a much wider tape with 40 information channels. Information is generally recorded in the density range of

100–200 bits per inch in each channel. However, the special tape units of the NORC computer employ a pulse density of 510 bits per inch.

Information is recorded on tapes in "blocks" or "unit records." These records may be a standard length or they may be of variable length. Interblock or interrecord gaps of 1 to 2 inches are provided to allow stopping and starting of the tape between records. This is necessary since the tape must be traveling past the read and write heads at a standard speed of 50 to 100 inches per second for successful operation. An innovation used in IBM Type 727 tape units involves placing "longitudinal parity bits" for each channel at the end of a block or unit record, providing additional checking. In many magnetic-tape systems, block addresses are recorded in a separate channel to allow automatic tape search independent of the computer operation. In case block addresses are not recorded, units of information are identified simply by their position by counting or by identifying the recorded information itself. The advantages of magnetic tape over paper tape and punched cards are the high-speed transfer rates attainable and the high density of recorded information. The speed and storage density are compared with other media in Table 10.1. The information density on magnetic tape is such that the information stored on one punched card of size 3¼ by 7⅜ inches can be recorded on a ⅖-inch length of magnetic tape. Over a million characters can be stored on a 1500-foot reel of tape about 8 inches in diameter. Approximately 15,000 cards requiring ten trays

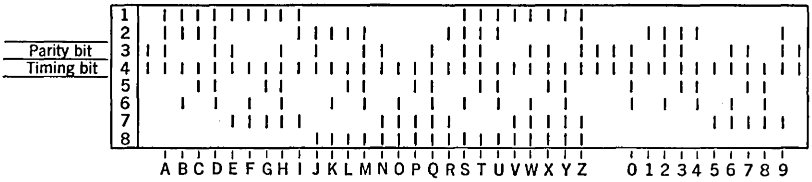


Fig. 10.4 Code for magnetic tape recording

Table 10.1 Recording Media Characteristics

Medium	Bits per Square Inch	Input Speed, Characters per Second	Output Speed, Characters per Second
Paper tape	70	200–400	60–200
Punched cards	40	300–600	130–260
Printed page	. . .	. . .	2,000
Magnetic tape	1,000–3,000	13,000–70,000	13,000–70,000
Photographic	3,000,000	160,000	. . .

of a standard card filing cabinet would be necessary for the equivalent amount of information recorded in punched-card form.

Although photographic film is an important medium for input-output, it has not come into extensive use. Photographic techniques have a wide range of use in recording. Film is used to record printed matter from the face of cathode-ray tubes and is used as a high-density storage medium in recording binary information. It is in its latter use that the high-resolution properties of the film are exploited. The problems of both original recording and readback of the data are difficult. One of the disadvantages is the nonerasable character of the medium. For this reason, photographic techniques fall into the category of high-volume auxiliary memory rather than input-output.

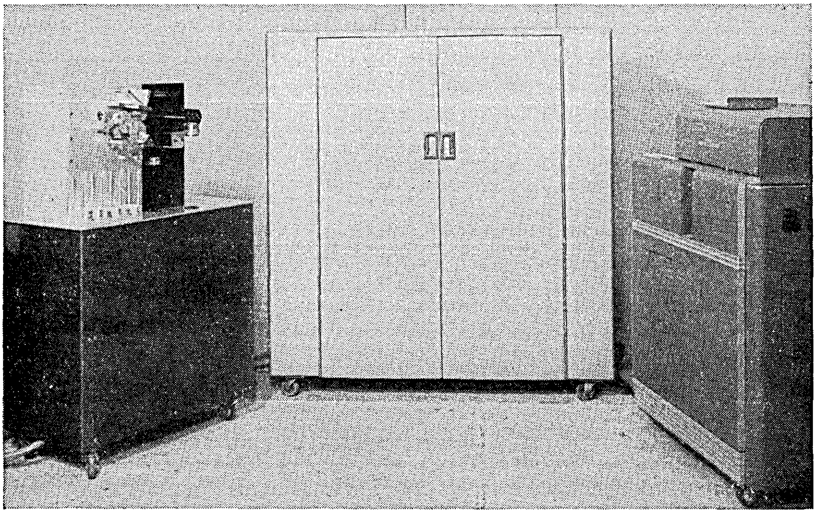


Fig. 10.5 Intelligent Machines Corporation print reader

The printed page is frequently used as a medium in output—the reverse process of reading the printed page is receiving much attention. Figure 10.5 shows experimental equipment developed by Intelligent Machines Company. Number reading is important for checks in banking operations. Numbers can be read directly, or a code printed in magnetic ink can be used.

Many modern computers employ the binary number system internally and, therefore, require conversion from decimal to binary somewhere between input and problem solution. Although this was formerly done by hand external to the computer, this conversion is now accomplished in most cases by the internal stored program of the computer. Also, in reading data it is common practice to form sum checks. As each word or unit of information is read into the computer, a sum is accumulated and checked with a sum obtained external to the computer or by a preceding data input operation. If the two sums thus formed agree, this is taken as evidence of the correct operation of the input device. Another operation on input occurs when full alphanumeric keyboard information is read into a binary computer. Although early computers accepted fewer than 20 different characters (the SEAC accepted 16 characters of the "hexadecimal" number system), the modern computer accepts the full range of keyboard information. Frequently this information must be rearranged within the computer.

In Table 10.1, where a range of figures is given for recording media, the lower figures are obtained easily and used frequently. The higher figures are obtained by an unconventional arrangement of information in the medium or in experimental equipment. Where no figure appears, it is considered that the particular information is nonapplicable.

10.3 BUFFERING AND COMPUTER CONTROL

The first stored-program computers were inhibited from performing any arithmetic operations while input-output was proceeding. During input and output, using the Teletype equipment, the effective megacycle rate of the SEAC computer was slowed to the input rate of 24 bits per second. In most cases the mechanical input-output devices will be slower than the internal speed of the computer. The problem is to reduce the rate of information transfer on output to match the output device, and on input to speed up the data rate to make it acceptable to the central computer. The problem of changing the speed and creating signals to make the two equipments electrically compatible was certainly challenging enough at the beginning. A greater challenge was to increase the efficiency of computer operation during extensive input-output operations.

To meet this challenge, most computer designers developed small memory devices, such that information could be shifted in and out at different rates, to act as a buffer between the computer and the input-output mechanism. The modern buffering system does more than simply slow down the information flow during output and provide electric compatibility between the computer and the input-output devices. In addition, it allows the computer to communicate with the buffer at computer speed and, after such communication, to continue with its normal operation, meanwhile allowing the input-output device to communicate with the buffer at its own lower rate. With the modern computer, for example, fewer than 50 microseconds are necessary to transfer a word into buffer storage, whereas close to 500 microseconds are necessary for the modern magnetic-tape unit to transfer one word out of the buffer to the tape. The buffer makes the 450 microseconds between successive word transfers available to the computer for arithmetic and logical operations.

Following the development of the UNIVAC I in 1951 with its 60-word buffers for input and output, the IBM-701 computer appeared with a buffering-type system which allowed the usage of a great percentage of the input-output time cycle for arithmetic operations. In the 701, for example, 292 milliseconds of the card cycle of 400 milliseconds are available for useful calculating. The UNIVAC Scientific Model 1103 computer was the first commercially available computer to be designed with a general-purpose buffering system. In this computer two buffer registers are provided through which communication can be accomplished with any device that will give and accept electric signals compatible with it.

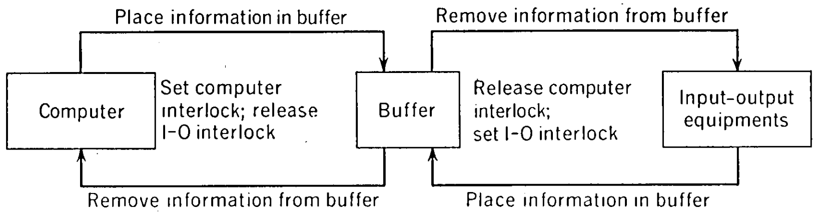


Fig. 10.6 Buffering system schematic

Because the input-output equipment works asynchronously with the computer and because the programmer must allow periodic communication with the buffer, a system of interlocks is necessary to allow for the possibilities of poorly timed operations. Figure 10.6 illustrates an idealized buffering system. When the computer communicates with the buffer, the computer is locked out and the I-O interlock is released, allowing communication of the buffer with the input-output device. When communication of the I-O with the buffer occurs, the computer interlock is released and the I-O interlock is set. In most cases the input-output equipment is "free-running," which implies that, during output, the output equipment continually and periodically samples the buffer. Information must be there or the interlock causes an alarm. By the same token, if the computer attempts to deposit a number in the buffer and the computer interlock is not released, the computer will idle until the release occurs. Similar statements can be made about input.

Besides the transfer of the information to the buffer, the input-output equipment at some stage must be selected and put into motion. The speed of selection then is important, and it is highly desirable that this selection process take place at internal computer rates. The size of the buffer is likewise an important item. The larger the buffer is, the less frequently information transfers to it are necessary; and the less frequent the transfers, the greater the efficiency becomes. For example, very few operations could be accomplished between transfers which are 500 microseconds apart in a computer which has an add time on the order of 50 microseconds. In this instance probably no more than 50–70 per cent of the time between transfers could be utilized because the apportionment of the computing between transfers must be such that information is transferred to the buffer at the correct rate. Compare this with the buffer 20 times as large, thus allowing transfers 10 milliseconds apart. Here the apportionment of computing between transfers is a much simpler job and would probably allow as great an efficiency as 90–95 per cent. The increased efficiency provided by effective buffers is shown by comparing the IBM-701 with the IBM-705. In the IBM-701, card half-rows had to be transferred at proper intervals, but the IBM-705 has 80-character buffer storage available which allows information for an entire card to be transferred at once.

In lieu of large buffer registers, it is possible to design an external control of the input-output equipment so that transfers and selection proceed completely independently of the main computing activity. With this system design, the size of the buffer is unimportant, since successive transfers to the buffer are accomplished automatically without explicit instruction. In this arrangement, the whole memory essentially becomes the buffer for the input-output equipment. Interlocks are important in this instance, as in the other, since it is possible that the input-output equipment and the computer itself might conflict in attempting to communicate with the same memory cells in an improper order. In this instance of the "advanced buffer," a highly complex piece of equipment that controls the input-output, it is possible for the mechanism itself to be programmed to permit not only a simple transfer but to provide for processing of the information as it leaves the central computer. This idea is being exploited in the LARC (Livermore Atomic Research Computer) being built by Sperry Rand for the Livermore Radiation Laboratory of the University of California.

Buffering systems with considerable flexibility have been designed. For example, consider the buffer storage that allows the storage of a number of words up to some large number. In this system the item drawn out by the input-output device is the one that has been in the buffer longest. The buffer is designed to provide signals in the event that it has been emptied or that it is full and cannot accept another item. Such a buffer has been designed by Telemeter Magnetics in Los Angeles. This buffer, of modular design utilizing magnetic cores and transistors, forms the heart of a number of systems which allow transfer between various output media, such as between paper tape and magnetic tape.

10.4 READING AND RECORDING EQUIPMENT

Teletype equipment was first used for input-output with automatic computers. Principally because of its greater speed of ten characters per second, Flexowriter equipment, manufactured by Commercial Controls Corporation, replaced Teletype for punched-paper-tape input and output. The Flexowriter is essentially an electric typewriter especially adapted to prepare printed page matter from paper tape and to punch paper tape. The paper tape used is seven-level tape rather than the five-level tape used with the Teletype equipment. Flexowriter equipment is still in extensive use today, specially for medium-sized magnetic-drum computers. It is also used for monitoring the output of large-scale calculators. In order to obtain higher-speed input than that obtained by means of adapting the Flexowriter, many computer systems use the photoelectric reader made by Ferranti Limited, an English company. This reader, shown in Fig. 10.7, is a well-engineered device which reads seven-level paper tape at 200 characters per second and has a high-speed clutch which can stop on a character, that is, read a character, receive a signal from the computer, and halt before traversing another frame. The "Ferranti Reader" is widely used in this country and in England. The Teletype Company makes a 25-character per second punch as well as a 60-character per second punch, both of which punch seven-level paper tape. Of the large-scale systems, only the UNIVAC Scientific Model 1103 uses paper tape as an input-output medium (in addition to punched cards).

There are many other photoelectric readers of various designs. The operation philosophy for some of these units is that control of the reader is accomplished through information stored on the tape. In the case of the Ferranti Reader, the start and stop of the reader are under computer control exclusively. This system allows greater flexibility of operation but requires the high-speed start and stop mechanism referred to. Also, a number of paper tape punches have been developed, some with considerably higher speed than the 60-character per second Teletype punch mentioned.

As a natural consequence of its accounting machine business, International Business Machines Corporation makes a full range of punched-card readers and punches. The readers generally operate at 150–250 cards per minute, and the punches operate at 100 cards per minute. Many of these units are adaptations of machines that have been used over the years in tabulating work. For example, the conventional reproducer can be adapted rather easily as a card reader and card puncher for a computer system. The serial ten-character per second electric typewriter operation has largely given way to the "line printer," occasionally referred to as a "gang printer," which allows the simultaneous printing of the characters in an entire line. Again, most of these printers used as computer outputs are modifications of conventional accounting machines. As an example, the IBM Type 716 printer is a modified Type 407 accounting machine.

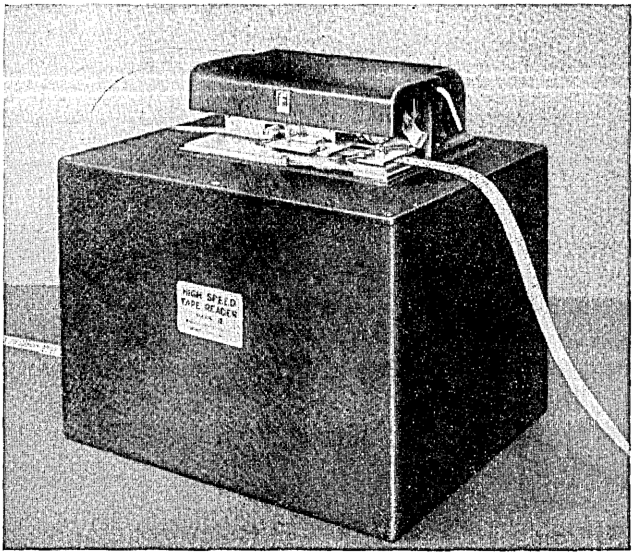


Fig. 10.7 Ferranti photoelectric reader

Although IBM makes most of the punched-card equipment and line printers used with electronic computers in this country, there is one exception. The Compagnie des Machines Bull, of Paris, makes card-reading and card-punching equipment that handles the conventional 80-column card and is used extensively with the UNIVAC Scientific line of equipment. The outlet for this Bull equipment in the United States is Sperry Rand. Although the equipment has characteristics similar to those of the IBM equipment, there are certain limitations on the number of holes which can be punched in any single card or in any one row of any card. The Bull company also makes a line printer that operates at 150 lines per minute and is available as an on-line printer with the UNIVAC Scientific computers.

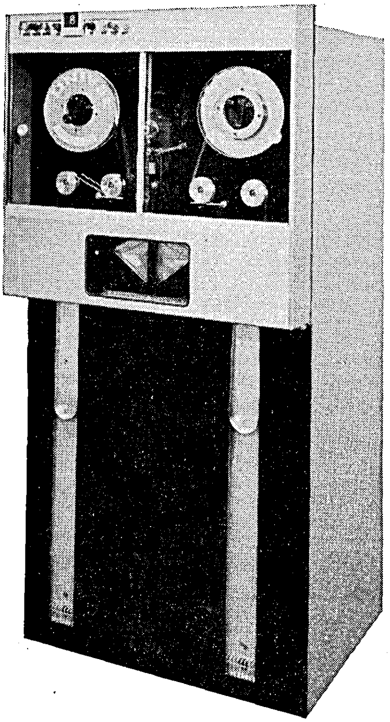


Fig. 10.8 Uniservo II tape reader and recorder

The highest-speed input-output devices are magnetic tapes. Magnetic-tape input-output has been available with the UNIVAC computer and with IBM computers, starting with the IBM-702 computer. For successful reading and recording, the tape must be moving past the magnetic read-write heads at a standard rate in the range of 50–100 inches per second. The time wasted in waiting for the transfer and the length of unused tape traversed during acceleration should be minimized. Hence, acceleration to speed is accomplished as quickly as possible, usually in 5–10 milliseconds. Accelerating the tape is much easier than accelerating the reels holding the tape. Although it is possible to do without reels altogether and simply let the tape fall into a bin, most units are designed with reels and a servomechanism to control them. When tape motion past the read-write heads starts, the tape loop between the heads and the reels shortens (or lengthens), and a signal is sent to turn the tape reels to keep a loop of correct size. The length of the tape loop is measured by spring tension, by vacuum switches, or by photoelectric devices. Because of the servomechanism aspect of the equipment and the precision manufacturing necessary, tape mechanisms are expensive devices.

The design feature important in making the tape unit usable as an input-output device is the ease in dismounting and mounting tape reels. With early tape devices it was necessary to perform a painstaking job in properly threading the tape through a maze of capstans and rollers. With the modern device the tape reel can be mounted and the tape unit made ready for use in about half a minute.

Table 10.2 On-Line Input-Output Devices

Equipment	Use	Operating Speed	Computer Applications
Teletype	Five-level paper tape read and punch	6 characters/second	Early computers
Teletype	Seven-level paper tape punch	60 characters/second	Numerous computer systems
Flexowriter	Seven-level reader and punch	10 characters/second	Numerous computer systems
Ferranti	Seven-level paper tape reader	200 characters/second	Numerous computer systems
Bull	Read and punch 80-column cards	120 cards/minute	UNIVAC Scientific Model 1103
IBM Type 711, Model 1	Card reader	150 cards/minute	IBM-701
IBM Type 711, Model 2	Card reader	250 cards/minute	IBM-704
IBM Type 714	Card reader	250 cards/minute	IBM-705
IBM Type 721 (722)	Card punch	100 cards/minute	IBM-701, 704 (705)
IBM Type 716 (717)	Printer	150 lines/minute	IBM-701, 704 (705)
Bull	Printer	150 lines/minute	UNIVAC Scientific Model 1103 (1103A)
UNIVAC	Printer	600 lines/minute	UNIVAC Scientific Model 1103A
Shepard	Printer	600 lines/minute	NCR-107, and others
ANalex	Printer	900 lines/minute	Johnniac, Maniac, and others
Charactron	Cathode-ray tube printer	10,000 characters/second	UNIVAC Scientific Model 1103, UNIVAC LARC
IBM Type 727	Magnetic tape	15,000 characters/second	IBM-701, 702, 704, 705
Sperry Rand Uniservo	Magnetic tape	12,800 characters/second	UNIVAC, UNIVAC Scientific Model 1103A

A picture of Sperry Rand's Uniservo tape unit is shown in Fig. 10.8. A summary of the characteristics and use of a number of on-line input-output devices is presented in Table 10.2. The Charactron, included in this table, is discussed below.

10.5 OFF-LINE EQUIPMENT

The equipments commanding the most attention (and highest rental fees) are the off-line units used in transferring data at high speeds to and from magnetic tape. Machines for transferring data from cards to printed page, from paper tape to cards, and from paper tape to printed page have been available in many forms for many years. The magnetic-tape equipments are more recent developments. Certain of these devices can be used either on-line or off-line.

The Eckert-Mauchly Corporation, now part of Sperry Rand, was the first to develop equipment to place information directly on to magnetic tape from the keyboard and to transfer tape information directly to the printed page. The Unityper, available since about 1951, has the direct keyboard-to-magnetic-tape recording. This device, still in use with UNIVAC systems, stores characters on tape in 120-character "blockettes" for subsequent read-in to the computer. The Uniprinter, also available when the first UNIVAC was completed, is essentially an electric typewriter that prints information stored on magnetic tape at a rate of ten characters per second. To replace or supplement the Unityper, Sperry Rand has developed the Tape Verifier, which will allow automatic tape verification as well as provide more convenient initial recording of data. The Tape Verifier is pictured in Fig. 10.9.

Both IBM and Sperry Rand have machines to transfer information from magnetic tapes to punched cards and vice versa. The IBM Type 714 card-to-tape converter is pictured in Fig. 10.10. The magnetic-tape-to-punched-card converters take information from tapes to the conventional 80-column, 12-row cards at a rate of about 120 cards per minute. The punched-card-to-magnetic-tape converters transfer information stored on the conventional cards to magnetic tape at rates of about 250 cards per minute. The Sperry Rand version is somewhat unusual in the sense that a suction mechanism, rather than the "picker-knife" mechanism usually seen, picks cards from the hopper pile. In addition IBM makes the Type 774 Tape Data Selector which performs a selection of magnetic-tape data for simultaneous output to two or more devices such as card punches or accounting machines.

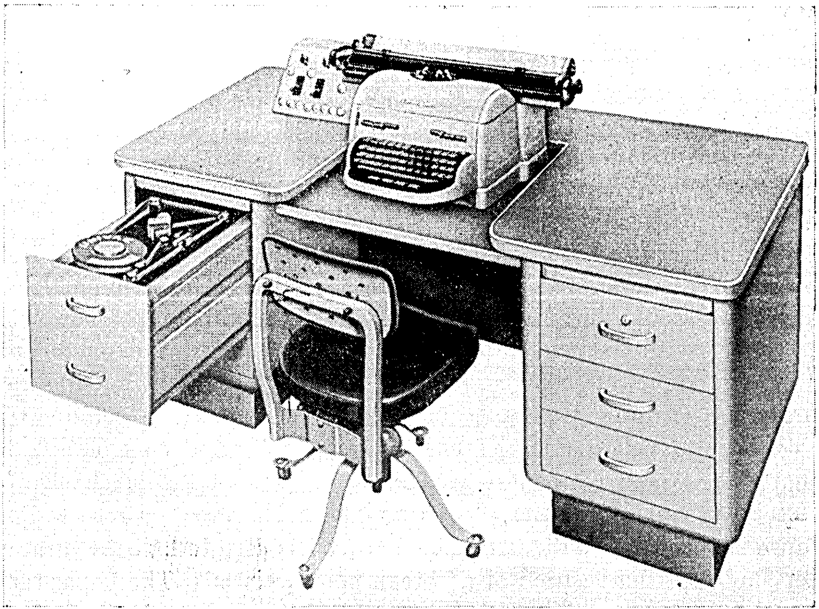


Fig. 10.9 UNIVAC tape verifier

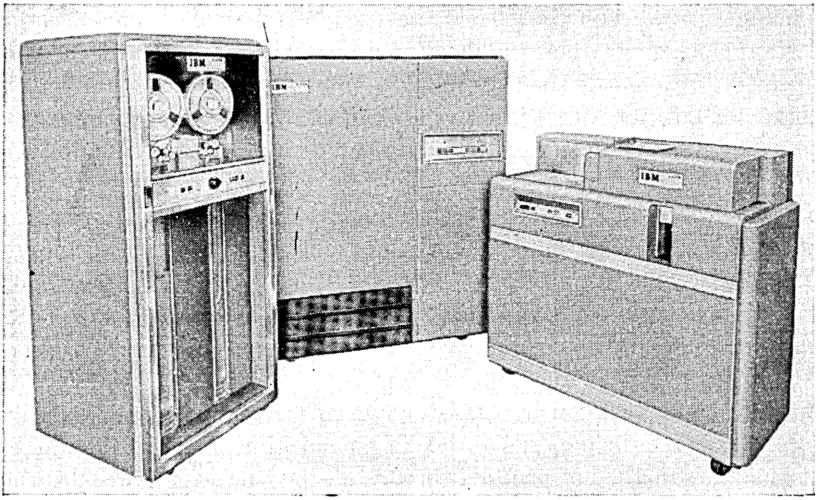


Fig. 10.10 IBM Type 714 card-to-tape converter

Magnetic-tape-controlled line printers are a most important new development. These machines, with printing rates of 500–1000 lines per minute and upward, are as expensive to buy or rent as many small magnetic-drum computers and, in many instances, take up more space. When such printers are used, information is transferred at high rates from the computer to magnetic tape in blocks of about 120 characters, the amount of information to be printed on one line. Tape reels are then dismounted and taken to the tape drive mechanism in use with the line printer where the information is read from the magnetic tape and transferred to the printed page at magnetic-tape rate.

Control of the printer is accomplished by four means: manual, information stored within the block recorded, a paper tape loop such as is used with the conventional accounting machine, and plugboard. Most of the manual controls involve switches to control the stop and start and positioning of tape and are sometimes used in conjunction with the information stored on the tape. In tape control certain characters stored in the first character position of the block signify certain special operations. For example, a character is reserved to cause the printer to halt. Another character is used to halt the printer, depending on the position of a "break point switch." The paper tape loop is used to provide fast feed or line skipping in certain operations, especially in dividing the printed information into pages. The plugboard is used in much the same way as with conventional tabulating machinery. The plugboard can be used for the following: selecting certain characters from the block of information recorded on tape for selected placement in the printing position; multiple printing of certain character positions; multiline printing of characters; line spacing; and zero suppression. In most cases the plugboard is not necessary to perform editing of the data, since it can be done by the computer before output. However, the use of the plugboard removes from the programmer much of the onus for editing output and reduces the computer time necessary.

Two types of printing mechanisms are used in high-speed line printers, and the differences between them are noteworthy. In one type of printing, characters are formed by selection of a pattern or matrix of wires chosen by a five-by-seven array. The wire ends are forced to strike the paper, and the dot pattern thus forms the character. In the other method of printing there is a type wheel for every printing position. Each type wheel turns, and a hammer strikes the paper in each printing position at the precise time the wheel is in position to produce the character chosen. Although an assessment of the relative merits of each method of printing is difficult, some observations can be made. Printing quality is probably somewhat better with the type wheel method. On the other hand, vertical registry of characters in their line position is not as good with the type wheel method, and the number of carbon copies possible is smaller.

The IBM magnetic-tape-controlled printers use the five-by-seven wire array technique for printing in its Type 719, Type 720, and Type 730 line printers, as is used with the Type 026 Printing Punch, a printing sample of which is shown in Fig. 10.3. The Type 719 operates at 1000 lines per minute with 60 printing positions, the Type 720 at 500 lines per minute with 120 printing positions, and the Type 730 at 1000 lines per minute with 120 printing positions. The UNIVAC High-Speed Printer, made by Sperry Rand and shown in Fig. 10.11, utilizes the type wheel design. This printer operates at 600 lines per minute and has 130 printing positions 120 of which are chosen arbitrarily from the tape; the remaining 10 are selected as duplicates of the 120 tape-recorded characters. The ANalex Corporation and Shepard Laboratories both make line printers that operate on the type wheel principle.

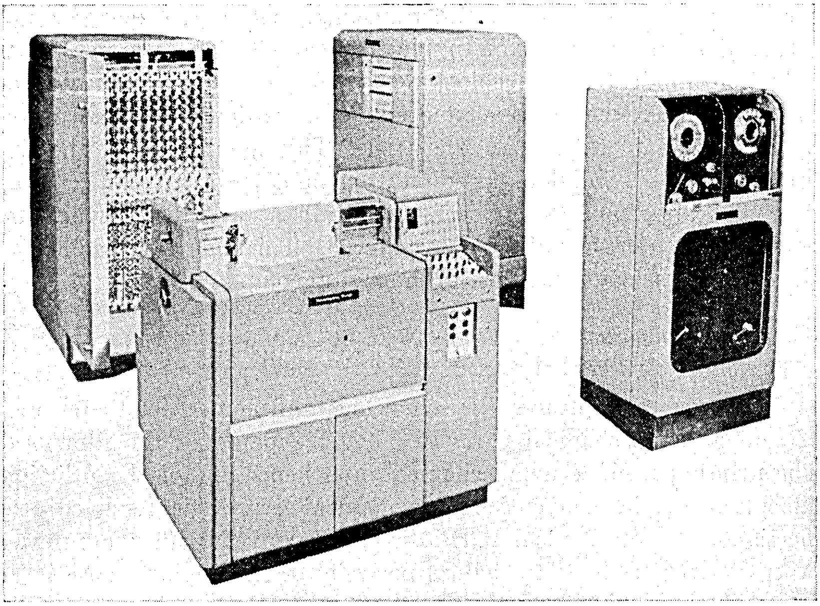


Fig. 10.11 UNIVAC high-speed printer

An interesting adaptation of the line printer makes it suitable to print data in the form of analog-type plot. In this adaptation, type wheels, or type bars, of the line printer are changed to remove certain characters and replace them by vertical bars and marking symbols. The vertical bars allow the marking of scale marks during output, and the special marking symbols allow for each character position a subdivision into four parts. The line printer with 120 character positions could then handle quantized graph plots with information good to 1 part in 480. The adaptation also features changing the gears on the carriage control to allow a smaller spacing, such as ten lines per inch rather than the usual six. In plotting samples of data in a data reduction application, when looking at the paper as it is printed on the line printer, the positive time axis extends downward and at the positive ordinate extends to the right. This adaptation of the line printer makes available, as digital computer output, an analog-type plot important to engineers to obtain at a glance certain important characteristics of the data reduced. This innovation adds considerably to the convenient use of the digital computer.

10.6 NONMECHANICAL PRINTERS

"Nonmechanical printers" is somewhat a misnomer, for there must, of course, be mechanical parts to move the paper or the film used. The printing itself is nonmechanical, and hence one of the principal difficulties of the conventional mechanical printer is removed—the inertia of the mechanical parts directly involved with the printing process. At speeds of 1000 lines per minute and higher, there is an extremely heavy burden placed on the mechanical design in making parts move as fast as necessary, for with these devices the ink must be transferred from a ribbon to the printed page by means of a hammer strike. In the nonmechanical printing devices to be discussed here, the part that must move at high rates of speed is metallic powders, an electric spark, or, in the case of the cathode-ray tube, the electron beam.

The Atomic Instrument Company has produced a high-speed printer called the Dataprinter, capable of printing characters at rates up to 500 per second in a column. The manner by which the characters are printed is similar to that of the IBM line printers in the sense that numbers are formed by dots from an array. In this case, however, a three-by-five array is used to represent the digits, and the dots are formed by sparking wires through Teledeltos paper, a Western Union product. In the printing process a black layer beneath the reading surface is exposed by the spark. The paper runs from one reel to another, passing the stationary spark wires. Early models recorded only three-digit decimal numbers, but by duplicating the printing mechanisms numbers of larger sizes can be printed.

As another example of a nonmechanical high-speed printer, the General Electric Company has examined the possibility of using a process called ferromagnetography. In this process, finely divided charged particles are placed in contact with a surface containing a magnetized latent image. The particles adhere to the magnetized areas and make the image visible. The surface is then brought into contact with another paper to which the particles adhere to provide a permanent copy. The design illustrates the importance of reducing the inertia of the elements in the basic printing process. In this case the elements are the small ferromagnetic particles.

Perhaps the most important type of nonmechanical printers are those using the cathode-ray tube. One of the first such devices was the Magnetic Numeroscope Printer, developed by Engineering Research Associates in 1950 and delivered to the National Bureau of Standards. With this device, two magnetic drums were used to make permanent recordings of the deflection voltages to be impressed on the cathode-ray tube to cause the tracing of appropriate characters. Access to different parts of the drum and therefore to the different voltages corresponding to the various characters was made in accordance with the character to be printed. A 16-millimeter microfilming unit was used for permanent recording. The design of this system in connection with cathode-ray tube output has not been used widely.

Another early application of the cathode-ray tube for output was that used with the Whirlwind I Computer at MIT. This type output received considerable impetus when the Rand Corporation began using a tube and a camera to record information from their IBM-701 computer in a military application. The general design has been embodied in an output system which IBM makes available for the 704 computer. Under computer control, the beam can be directed to one of 1024-by-1024 positions in a square array. Characters must be formed as combinations of those spots. An automatic X- and Y-axis generating feature is available with this device for analog-type plotting of functions and outputting graphic information like charts or figures. Characters must be formed out of spots directed to the 1024-by-1024 position array under complicated program control. The cathode-ray tube output device for the Whirlwind I Computer operates in much the same fashion.

In employing these devices in which a single spot must be manipulated to form characters, considerable programmer inconvenience is incurred and extra computer time is consumed. The Charactron tube, a development of Consolidated Vultee Aircraft Corporation, is designed to present alphanumeric and quantized analog plot information on the face of the tube with considerable operational convenience. With this tube the electron beam is directed toward a metal plate into which characters have been cut. When a diffused electron beam passes through the matrix, the characters formed are on the tube face, and secondary focusing directs it to any position in a 1024-by-1024 square. The advantage in the use of this tube is that the characters are formed automatically in the tube rather than by directing the electron beam in a prescribed pattern under complicated and time-consuming computer control.

The Charactron tube with associated camera equipment provides an output device which is available for the UNIVAC Scientific Model 1103 computer. In this system a 35-millimeter camera is used to photograph the tube one frame at a time, develop the picture, and then project it onto a built-in screen for viewing. The film can be removed for later viewing or for making prints. Under control of the computer, characters are formed on the face of the tube, and when the tube face is filled as desired, another film frame is brought into position for exposure.

A nominal rate for the tube is 100,000 characters per second. However, after necessary computer information, editing, and camera control, an output of about 50 words (600 characters) per second is achieved in a program in operation at the Convair Corporation in San Diego. An exposed film frame can be developed and projected for viewing in 2 or 3 seconds. The device is pictured in Fig. 10.12.

A Charactron tube high-speed output system is planned for the LARC computer. A different model Charactron tube will be used in this application and will operate at the rate of character output of 200,000 per second. A number of other cathode-ray tube output systems have been proposed, designed, and built. One of these is built by RCA-Victor Corporation and is included in the complement of output equipment for the RCA-BIZMAC computer.

All of these cathode-ray tube output systems require a means for permanent printing of the information. Photographic methods can be utilized to record the data permanently. When high speeds are particularly desirable, the dry process called Xerography developed by the Haloid Company affords an interesting possibility for recording.

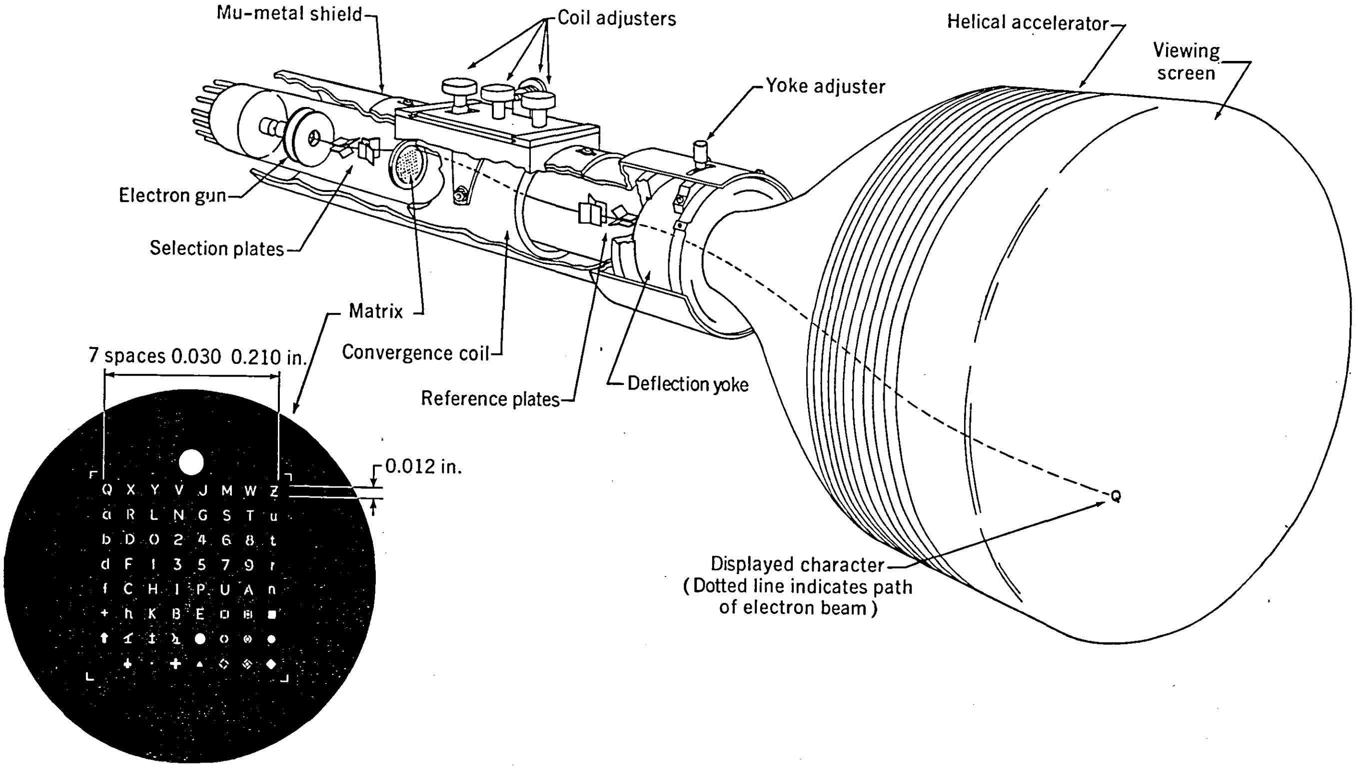


Fig. 10.12 Charactron shaped beam tube (Courtesy of Stromberg-Carlson, a division of General Dynamics)

10.7 CONCLUSIONS

A number of interesting trends in the development of input and output equipment for computers are evident. The first involves the transition from sequential operation to simultaneous operation for greater speed. Just as in the early days of stored program computers when the transition was made from the serial computer to the parallel computer, so also is this trend in evidence in input-output equipment. The development of the line printer to replace the single-action serial electric typewriter is an example. Since internal computer speed far surpasses even the fastest magnetic-tape unit, there is a trend to make the computer operate two or more units simultaneously to allow the input-output to keep pace with the computer. Another trend is the higher speed of selection of input-output mechanisms and the greater efficiency allowed by buffering and control of the input-output mechanism apart from the computer proper. When carried to the logical extreme, the selection of the input-output mechanism and the transmission to the input-output mechanism control unit will take place at the highest internal speed of the computer. The transfers to the input-output units will then take place completely independently of the program and its central processing efforts.

There is also a steady trend toward faster electromechanical devices and toward the increased use and development of nonmechanical devices, especially those utilizing the cathode-ray tube. Another interesting trend marks the confluence of system design of the scientific calculator and the data processor. More and more those operating scientific computing installations are seeing the need for facile and rapid input-output equipment. Undoubtedly in many cases this will result in a single line of computers by each manufacturer, allowing various types of input-output mechanisms to be attached as desired.

A number of challenging problems exist in printing mechanisms, in addition to greater speed and reliability. One of these is printing of the nonalphabetic, nonnumerical data. For example, a future computer might output a complete prognosticated weather map with its isobars and symbols for direction of wind, weather, and other items of weather information. As another example, a battlefield map might be printed from a computer on the basis of information received for a hypothetical or conjectural move or placement of friendly or enemy forces. The printing problem here is one of placing many symbols in a large number of positions, frequently in a continuous form, to delineate accurately the battlefield situation.

The ultimate goal in computer input is to read printed or written matter from the original document. An especially challenging problem is to read successfully the printed or written matter, even if the printing or writing is imperfect. The fact that the conventional alphanumeric printing is done in a form which, information-theory-wise, is redundant makes this seem possible. The speed of this reading must be at least comparable to the speed obtained by the highest-speed paper tape and punched-card devices available today.

10.8 REFERENCES

1. *Proceedings of a Symposium on Large-Scale Digital Calculating Machinery,* Harvard University Press, 1948.
2. *Proceedings of Electronic Computer Symposium, April 30, May 1 and 2, 1952,* Los Angeles.
3. Review of Input and Output Equipment Used in Computing Systems, Joint AIEE-IRE-ACM Computer Conference, December 10–12, 1952, New York.
4. Data Handling Systems Symposium, *Trans. IRE Professional Group on Instrumentation,* PGI-2, June 1953.
5. *Proc. IRE, 41,* Oct. 1953.
6. *Proceedings of a Symposium on Large-Scale Digital Computing Machines,* Argonne National Laboratory, August 3–5, 1953, ANL-5181.
7. Trends in Computers: Automatic Control and Data Processing, *Proceedings of the Western Joint Computer Conference, February 11–12, 1954* (IRE-AIEE-ACM), Los Angeles.
8. *Proceedings of the Eastern Joint Computer Conference, December 8–10, 1953* (IRE-AIEE-ACM), Washington, D. C.
9. *Proceedings of the ACM, May 2–3, 1952,* Pittsburgh.
10. *Proceedings of the ACM, September 8–10, 1952,* Toronto.
11. *Convention Record of the IRE,* 1954 National Convention, Pt. 4, Electronic Computers and Information Theory.
12. G. L. Hollander, Bibliography on Data Storage and Recording, *Trans. AIEE, 73,* Pt. 1 (Communication and Electronics), No. 11: 49–58, March 1954.
13. R. Bird, Computing Machines Input and Output, *Elec. Eng., 25:* 407–410, Oct. 1953.
14. C. W. Fritze, Tape Recorder Stores Computer Output, *Electronics, 27,* 7: 166–169, July 1954.
15. FOSDIC, a Film Optical Sensing Device for Input to Computers, *Tech. Bull., NBS, 38:* 24–27, Feb. 1954.
16. G. W. King, A New Approach to Information Storage, *Control Eng.,* 48–53, Aug. 1955.
17. Digital Recording Techniques and Recorders, Naval Ordnance Test Station, Inyokern, Microcard: U:13380.
18. W. B. Groth, Principles of Tape-to-Card Conversion in the AMA System, *Commun. Electronics, 5:* 42–52, March 1953.

19. J. Hartmann, Monoscope Tube for Computer and Other Applications, *Elec. Eng., 73,* 3: 208–212, March 1954.

20. H. L. Daniels, Boundary Displacement Recording, *Electronics, 25:* 116–120, April 1952.

21. A. D. Booth, On Two Problems in Potential Theory and Their Application to the Design of Magnetic Recording Heads for Digital Computers, *Brit. J. Applied Phys., 3:* 307–308, Oct. 1952.

22. J. T. Potter and P. C. Michel, High Density Digital Recording System, *Trans. IRE Professional Group on Electronic Computers,* 60–72, Dec. 1952.

23. Letter-Printing Cathode-Ray Tube, *Electronics, 22,* 6: 160–162, June 1949.

24. J. M. Carroll, Electronic Addressing Aids Publishers, *Electronics, 26,* 2: 98–100, Feb. 1953.

25. L. D. Wilson and S. Meyer, Model II Unityper, *Trans. IRE Professional Group on Electronic Computers,* EC-2 2: 19–27, Dec. 1953.

26. John J. Wild, High-Speed Printer for Computers and Communications, *Electronics, 25,* 116–121, May 1952.

27. Leroy Packer and William J. Wray, Germanium Photodiodes Read Computer Tapes, *Electronics, 25:* 150-151, Nov. 1952.

28. B. G. Welby, Intermittent-Feed Computer-Tape Reader, *Electronics, 26:* 115–117, Feb. 1953.

29. W. H. Blashfield, A Subscriber Toll Dialing Tape Reader, *Commun. Electronics, 5:* 17–21, March 1953.

11. Applications of Electronic Data-Processing Machines

CUTHBERT C. HURD[1]

International Business Machines
New York, New York

11.1 INTRODUCTION

We shall consider the application of a large-scale data-processing system to the solution of the information-handling problem in life insurance policy operations, automobile spare-parts stock control, and public-utility billing and cash accounting. We shall assume familiarity with the characteristics of large-scale electronic data-processing machines—including those of high-speed magnetic tape and stored programming.

11.1.1 Kinds of Data-Processing Applications

Data processing refers to processing information expressed in numerical or alphabetical form and has applications arising in three non-mutually exclusive fields: engineering and scientific research, research and development in management science, and the control of a business. Applications in engineering and scientific research, which led to the development of the large-scale electronic machine and hence have

---

[1] The author is indebted to the Prudential Insurance Company of America, the Chrysler Corporation, and the Commonwealth Edison Company, and to a number of associates in IBM for valuable assistance in preparing this chapter.

the longest history, have been described in the literature and will not be referred to again in this chapter.

Applications of data processing in research and development in management science have a fascination in their own right but, in the main, involve a mathematical formulation and hence processing characteristics which resemble scientific computation. For example, formulations in terms of linear equations, differential equations, and simulations are common. Consequently, even though management science data processing may differ sharply from scientific computation in the ratio of input to internal processing, it will not be discussed here.

Therefore, we shall consider only applications arising in *commercial* data processing, also known as *accounting* processing, *business* processing, the processing necessary to control a business.

11.1.2 Classes of Data-Processing Machines

A machine that processes numerical or alphabetical information in an automatic fashion and whose operations include at least that of comparing will be called a data-processing machine. Here, by automatic, we shall mean that at least two operations are carried out in sequence without the intervention of a human operator.

Within this rough definition we shall then say that data-processing machines can be classified as small, medium, and large.

Of these three classes, we shall consider only the latter for the reason that the unit cost of processing operation is by far the lowest in this machine. That is to say, a medium-sized machine renting for $3000 per month can execute of the order of 100 internal operations per second. Its larger brother, renting for $30,000 per month, can execute of the order of 10,000 internal operations per second. This apparent relative efficiency of ten to one of the large machine can never be completely realized in practice because of time out for preventive maintenance, problem testing, and scheduling. Nevertheless, it is so large that it is important to examine those instances in which this efficiency can be utilized, and this we propose to do. Finally, then, of the nine combinations of problem and machine classes, we shall consider only one: business problems and large machines.

11.2 ELECTRONIC DATA-PROCESSING MACHINES IN BUSINESS

For the past several years, the managements of virtually every sizeable organization in America and in many other parts of the world have attached great significance to the application of electronic data processing, automatic data transmission, and office automation in general. The urgency with which these leaders are studying developments in this field arises from the astounding fact that in a single generation we have filled this world with more new products than in all the previous generations of man combined. In reaching this tremendously accelerated pace of technological development, we have freely utilized recorded information of our predecessors in science and industry. Thus, in science, we have used the accumulated data concerning the operating characteristics of a chemical process in order to design a new plant. In business we have used historical information concerning investment policy and credit risk in order to make business judgments. It is obvious that as the total amount of information available becomes more gigantic, the problems relating to its organization and communication become more acute. We have reached the point today where, without great facility and speed in recording, classifying, summarizing, and extracting facts, we shall not be able to take full advantage of our cumulative information.

Moreover, as a business grows beyond the size controllable by one man, its management must depend upon reports and records for the news of the events of the business. The process of selecting, interrelating, and communicating useful facts from the total amount of operating data takes time; often, the manual methods are slow enough that by the time the information reaches its destination and is analyzed, the decisions for which the data were collected in the first place were of necessity made without the information.

For example, a principal executive of a large and highly successful chemical organization recently remarked that without the availability of high-speed electronic data-processing machines, the monthly statements of cost by product were available to operating personnel only during the last third of the month following the closing date. Under such a time schedule, an operating foreman had the tendency to say, if confronted by a poor cost report, "Oh, well, the end of the month is almost here. Let's wait and see what happens." Close and effective control of costs will now be much more effective to enforce from a psychological point of view since, with modern data-processing equipment, the reports will be available in the first third of the month.

Characteristics of business machines available up until the recent past have brought about the recording in multiple form of the most minute events of a business. The objective was to cross-index this data a sufficient number of times so that the time required to isolate and extract wanted facts was in some closer relation to the time allowable for its assimilation.

Without implying that any final solution to these problems has been found, we must recognize that important gains have been made in the last several years. Today's latest large-scale electronic data-processing machines offer to us the most promising means yet to maintain control over information handling in large enterprises.

11.2.1 Fundamental Requirements for Automation in Data Processing

Like all advances in automation, modern automatic data processing depends upon two characteristics which were first present in factory automation seventy or eighty years ago. These characteristics are those of complete definition and of repetition. That is to say, the engineer can build a tool provided the functional requirements of the tool and the specifications can be written. Whether the tool will be built depends upon the possibility of economy and therefore probably upon the amount of repetitive use to which a tool is to be put.

Exactly similar analysis can be made concerning the application of automatic data-processing machines. The problem must be defined completely and unambiguously before the machine can be programmed. Whether the program for the problem will in fact be written depends upon the number of repetitions of the problem. This number, of course, will decrease as more completely automatic programming methods are developed.

A second similarity between factory machines and office machines is the following: in factories, at first, all machines were of the vertical type in function. That is, they accomplished only one function, drilling, for example. Later, horizontal-type or multiple-function machines appeared. Also material was transported from one operation to the next by such means as endless belts. At the present time, factory machines now combine functions and transport to such an extent that a large-scale transfer machine may perform in sequence several hundred combinations of a few dozen operations with all transport being performed automatically.

11.2.2 The Growth of Office Automation

In the office, we find a similar evolution. The first office machines were all of the vertical type, for example, the typewriter which writes or records and the adding machine which adds. Later, machines performing several operations appeared, such as the typewriter combined with the adding machine, generally called a bookkeeping machine. The modern punched-card accounting machine is an example of the multiple-function machine. It can match one record against another, add, subtract, cross-foot, post, position its own forms, etc. Likewise, in the beginning, virtually all business data were communicated or transported entirely in paper-recorded form. Later, after the transmission of information by wire became widely accepted, we witnessed the gradual development of wire transmission of business data, first by key depressions, then by magnetic facsimile and wireless.

The existence of electronic data-processing machines has altered the whole economy of data transmission significantly because of this tremendous speed and capacity. One notable characteristic is that they have substituted an internal means of transmitting data from one function or operation to another instead of by the conventional external transmission of paper-recorded information.

We shall now proceed to discuss in order: data recording and handling, characteristics of an efficient data-processing system, and characteristics of the business problem.

11.2.3 Data Recording

There are three principal methods of recording data from source documents: by hand, by sound, and automatically. Hand and sound recording are quite familiar; automatic, in this case, refers to the automatic by-products of previous processing such as balance forward punched cards, balance forward magnetic tape, plus any new information available such as the automatic output from production recorders or from automatic transmission devices. Here "balance forward" refers to a file of information that has been updated by a file maintenance procedure, i.e., all changes in the file have been incorporated.

In both hand and sound recording, a conversion to machine language must be made. There has recently been considerable progress toward making this conversion an automatic rather than a manual process. The punched-paper-tape or punched-card outputs of radio, mechanical, and telephone transmission systems are examples, as are the combined recording systems which produce simultaneously both card and tape. Practically every major business machine company is introducing some kind of device in this field.

But whatever the automatic data-processing machine used, machine language, on a practical basis today, is either punched cards, punched tape, or magnetic tape; and in this connection, we note that virtually every large-scale electronic data-processing machine depends upon the punched card either for direct input or for the most practical way to create, verify, and correct magnetic tape.

Once having converted source data to machine language, however, the subsequent arranging, computing, summarizing, and transportation of data between these functions become an automatic internal process, conducted at extremely high speed. If the output is printed, no further reconversion is required. If output is in some form of machine language, usually some additional processing step is needed before a human decision or analysis can be made from the results.

11.2.4 Characteristics of an Efficient Data-Processing System

Keeping in mind the fast input, large storage, and high-speed operation of the modern large-scale data-processing system, we shall now state the facilities essential to an efficiently designed application. Because the business problem is characterized by multiple inputs, multiple classes, and multiple outputs, the data-processing system must provide for (1) consolidation of several files into a single magnetic-tape file; (2) processing of several classes of information simultaneously; (3) production of all operating reports simultaneously and elimination of intermediate results with no operating application.

A typical data-processing application generally involves the collection of large volumes of data from multiple input sources, selection of a class of data, operation on this class of data, and the generation of information for output reports and documents. Associative or associating data from each of multiple sources are contained within a unit record which includes descriptive and identifying information. Unit records from a given source are ordered on their control or identifying information and matched with records from other sources. Data from one source are then consolidated with those from another source, and a class of data is identified and used as the basis for calculation of output information. Processing then consists of repetitive cycles of operation on associated units of information. The large volume arises from the number of times the cycle must be repeated—i.e., the volume of original input information.

11.2.5 Characteristics of the Business Problem

We shall now discuss the business problem in general and indicate how the characteristics of the efficient data-processing system are important.

Classes of data. A file of unit records from a given source usually contains several classes of data. For any single class, the calculations to be performed during a single operating cycle may be simple but a large number of logical decisions may be required to determine what class is to be treated. In the past, the selection of class has been simplified by segregating the different classes of data from a given source. For example, a payroll involving salaried, hourly, and piecework employees would be divided into three separate processing operations for the calculation of gross earnings. The three classes would receive somewhat different treatment for the gross-to-net calculation. Because the subclassifications such as overtime, sickness and accident benefits, and training pay must also be considered, the segregation of classes has in the past involved complex document-handling operations and very often multiple passes of the same data in several operations.

With a large memory and the flexible control of stored program, all classes are treated in a single run in the modern data-processing machine. Instructions stored in memory are used both to determine the class and to carry out the appropriate processing on this class.

Multiple inputs and outputs. The consolidation of data files is possible in the large-scale data-processing machine because the storage density of magnetic tape permits a record length which is, for practical purposes, unlimited. This very consolidation, however, in a long sequential file containing a maximum of associated information concerning a given account, emphasizes the need to pass this information as infrequently as possible. This is particularly true in view of the fact that very often the active items in the file for a given processing cycle are a small percentage of the whole file. Multiple tape units provided on a large-scale data-processing machine permit the matching of records from several sources in a single run. At the same time, multiple output files may be generated. By associating unit records from several input sources, all the data for a particular operation cycle can be brought within the large memory of the machine at one time. The operations necessary to generate various forms of output information can then be performed in one operation cycle. For example, data from an attendance record and a voucher record may be consolidated with data from a matching payroll record to produce an earnings statement, a payroll register, and a check record. For a given employee number, the data from these three sources can be consolidated and used to calculate gross earnings and apply deductions to obtain net pay. In the same operation cycle, the employee master record can be updated in year-to-date totals, bond balances, etc.

In order to accomplish the consolidation of data from multiple sources containing a maximum of associated information in a given record, it is, as implied in the previous paragraph, necessary within an operation cycle to recognize variations in the form and content of input data. The logical choices possible in the stored program machine with large storage facilitates the distinguishing of these variations.

The processing cycle should be designed so that operations are performed on data from input sources to produce output information required for implementing operating procedures. Consequently, all presently existing reports must be examined critically. They may have been required as justification for an intermediate processing operation on a segregated class of data. Simultaneous processing of the varied classes is likely to remove such justification.

Detailed listings which may have been necessary in a system involving multiple files are, in many cases, unnecessary because of the consolidated master-file concept and the facility with which reference information can be obtained within the processing cycle. The ability of the large-scale machine, through this logical capacity, to assume routine clerical functions lessens the need for detailed information for manual reference. In designing a data-processing application, the engineer should therefore not necessarily attempt to parallel or duplicate existing procedures in order to generate the same set of reports and documents.

11.3 EXAMPLES OF APPLICATIONS OF LARGE-SCALE DATA-PROCESSING MACHINES

To illustrate the efficiencies of a large-scale data-processing machine, we shall use three case studies. Detailed planning and programming of these systems have already been accomplished and the results evaluated.

11.3.1 Company A: Life Insurance Policy Operations

When this insurance company opened its doors for business in 1875, its bookkeeping was hardly a problem. Its first application was filed in a cardboard box and placed on a wooden shelf. Desks were pine boards set on saw horses, and one clerk kept all the records and prepared all the policy statements.

Quick reference to one file folder enabled the company to prepare a premium notice, approve a change in the contract, calculate a reserve or a dividend, and accomplish nearly any other data-processing task required in the fledgling organization.

Within ten years, however, the policy file had grown to 400,000 folders, and there were over $1,000,000 in company assets to keep track of. Because all the postings and statement and check writing, as well as all calculations, were manual jobs entirely, it was becoming obvious by this time that the single-policy folder file accounting system was no longer meeting the company's requirements.

Perhaps it was some now obscure clerk with a fine Spencerian hand, or perhaps it was a "mathematical genius" who could mentally calculate dividends quickly who first demonstrated the efficiency of the specialization of function. At any rate, it became more and more efficient to subdivide the office clerical functions. Proficiency was achieved through this system of specialization, and the transmission of data from one function to another was handled with the latest available recording tools.

When punched-card accounting machines became available, the company was quick to take advantage of the multiple operations they offered. By this time, however, the volume of their operations was such that, although the punched-card machines provided a way of consolidating, record-keeping functions were more effectively handled as separate products.

Thus, the basic data about each policy, for example, was recorded in several different files, and this basic system was maintained until the ability to transmit and handle information electronically at extremely high speed made practicable a fundamental change in their data-processing method.

The problem. The first major data-processing area studied and programmed for large-scale electronic system application was the ordinary policy operations. Already highly mechanized through conventional punched-card procedures, these operations are centered around eight major files in as many separate locations (Fig. 11.1).

(1) *Policy Master File.* This is the basic file for all premium-paying policies. It is maintained in district and policy number sequence. Further, it is subdivided by cycle billing periods so that processing can proceed on a relatively level-volume basis. The principal output of this file is the premium notice, when combined with the policy holder's name and address file.

The limitation of the punched-card record length makes expedient the separation of the accounting record and its respective name and address record. In addition to its primary output, the file is used frequently for general policy reference and contains the most up-to-date policy detail. It might well be considered the "master file."

(2) *Policy Address File.* This is essentially a supplement to the policy file and contains the current billing address for each premium-paying policy.

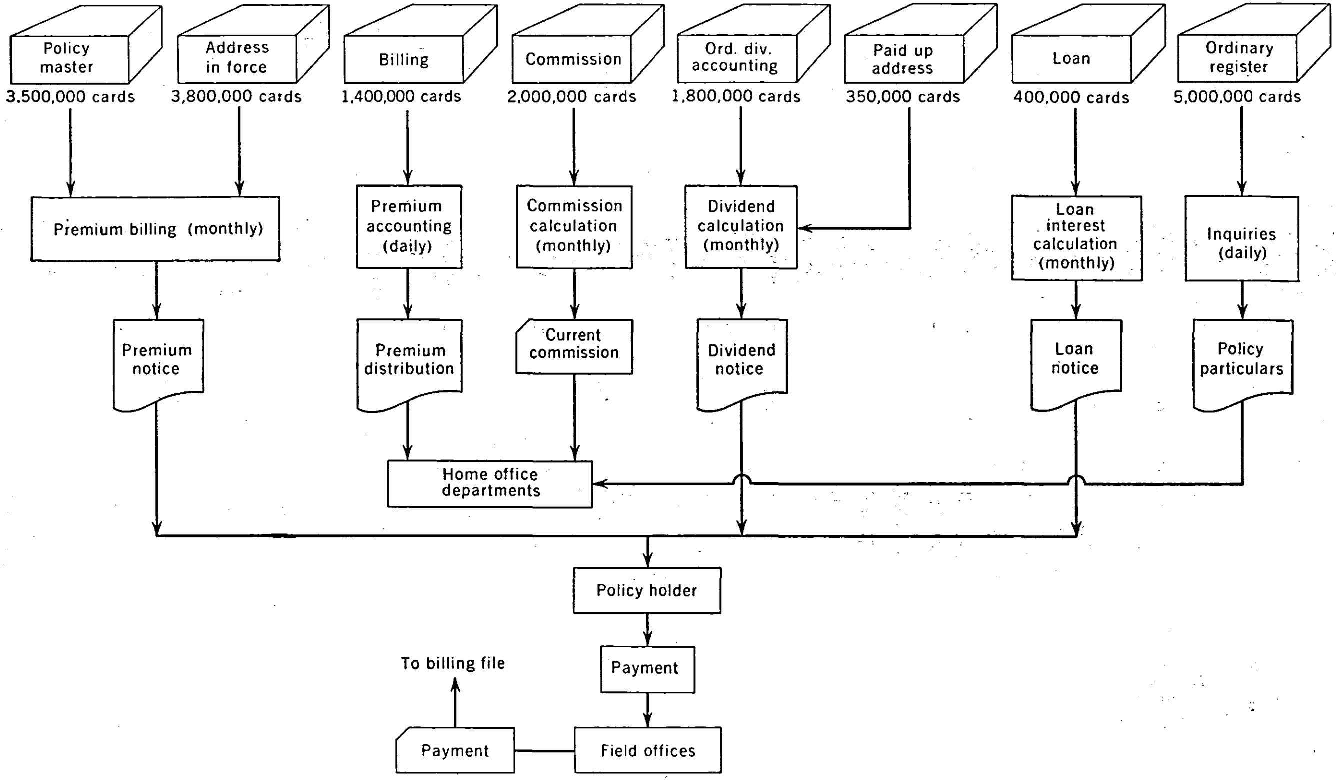


Fig. 11.1 Ordinary policy billing and accounting, punched-card accounting system (simplified)

(3) *Billing File* (*due and unpaid*). This file consists of one punched card for each premium billed and unpaid. It is relieved as payments are received and maintained in the same sequence as in the policy file. Its primary outputs are statements of premiums collected, premiums due, and production reports by agent and office.

(4) *Commission File.* This file consists of a punched-card record for each premium-paying policy on which a commission is paid to company agents on a level-contract insurance plan. Its basic product is a periodic current-commission statement.

(5) *Ordinary Dividend File.* This file consists of one punched-card record for each policy whose holder elected to account for his dividends due in any one of six basic options not including automatic reduction of premium. The main output is the dividend notice and associated calculations and accounting registers.

(6) *Paid-up Address File.* This file consists of punched-card name and address records for all paid-up or nonpremium-paying policies. It is associated with the dividend record file for the preparation of dividend notices.

(7) *Loan Files.* Divided into two sections, this file consists of a ledger record for every current policy loan, and a punched-card accounting record used to calculate and bill interest payments. The loan notice and its associated accounting registers are the basic output.

(8) *Ordinary Register File.* This file contains one manual card record, i.e., not a punched-card record, for each policy. It is essentially a chronological record of all policy transactions. It serves as the basic random-access file for policy history.

Objectives. The application of electronic data-processing systems to commercial data-handling problems can be particularly well illustrated by the following fundamental objectives established for the handling of the inquiry, maintenance, input, and product of these eight files:

(1) Consolidate multiple files into a single magnetic-tape file.

(2) Minimize the number of processing steps, or passes, required for particular data.

(3) Process several classes of information simultaneously.

(4) Eliminate intermediate results that have no operating application.

In this example, the consolidation of the different punched-card files, their associated record output, controls, and operations is the key to the electronic system. This plan is entirely practicable except for the ordinary register file which requires random access and which can be handled more economically on a manual-card basis. Because the company delegates much of the current policy status to its field offices, where duplicate policy history is maintained, the frequency of reference to the ordinary register file is low. The manual record, therefore, is relatively highly efficient.

Procedure. After having set the primé objective as consolidation, it was found that all the file records, with the one exception mentioned above, could be merged into one magnetic tape. Against this master policy record, all types of activity can be applied and the corresponding changes effected. The basic procedure required to accomplish this follows (Fig. 11.2).

(1) The master policy record is read into the memory of the electronic data-processing machine, one record at a time.

(2) All policy changes and new business, recorded on tape in policy number order, are read into machine memory from a second tape-reading unit.

(3) All payments, recorded on tape from payment cards received from the branch offices, are read into memory as the third tape input.

(4) If a change is indicated in record *A*, the change is effected and the updated record transferred to the output policy master file. Where no change is indicated, the entire input record passes intact into the output tape. Accounting, machine, and systems controls assure the accuracy of all tape processing.

(5) As the policy record is stored in the memory of the machine, it is interrogated to determine whether any premium, dividend, or loan interest is due. If so, a notice record, containing all identifying and quantitative data, is recorded in a consolidated output notice record tape. Simultaneously, an accounting record, reflecting the condition, is produced as the third output tape.

(6) Whenever these changes, new business, premium billing, and payment, affect commissions, an individual commission record is reflected on a fourth output tape. This tape also contains payment status control of the policy master file and records of overdue premiums; it reflects any discrepancy in input data. In addition, it contains summary information such as premium distribution by state by class of business, and accounting and systems control totals.

(7) The output tapes are either converted into punched cards for subsequent subsidiary machine processing or converted to printed results through a direct tape-to-printer operation.

Advantages. The effect of this consolidation of records is threefold: simplification, speed, and a higher degree of accuracy.

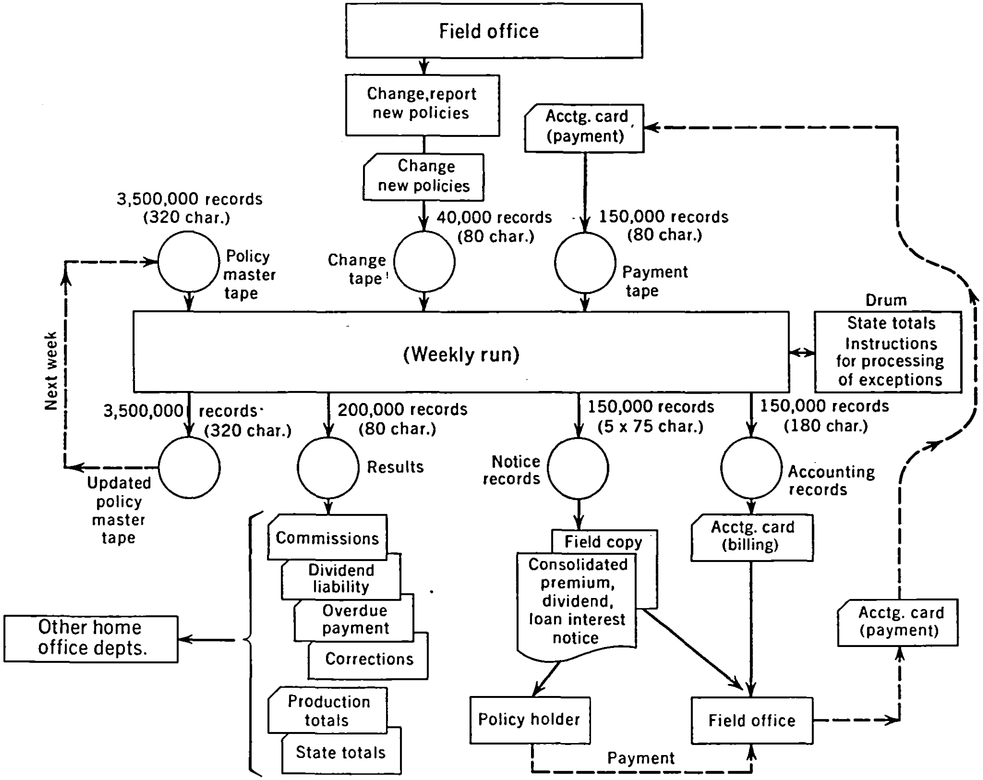


Fig. 11.2 Ordinary policy billing and accounting, electronic data-processing system

First, *simplification:* In addition to the obvious elimination of duplication of maintenance and processing brought about by the consolidation, the policy holders also benefit directly. A consolidated and simplified notice replaces separate statements for premium due, dividends, and loan interest. Also, the branch offices benefit equally from this comprehensive, simplified statement. Further simplification is made possible by the logical decisions developed within the system. For example, where contracts are sold that involve automatic premium changes, this condition is conventionally determined by routine and periodic examination of the actuarial records. The electronic data-processing machine recognizes and handles these changes automatically as a regular part of the routine processing. Redundant identifying information needed to tie in multiple records is eliminated, and file maintenance is accomplished in a single operation. Finally, the several hundred punched-card machine and related manual processing steps are reduced to approximately twelve under the electronic data-processing system.

Second, *speed:* Because of the high-speed processing, the time required for both inquiry and maintenance of file records is tremendously reduced. For example, the machine will process a policy and record all required output data, including all logical decisions and calculations, in about 400 milliseconds. In addition, the lead time for the preparation of all kinds of policy holder statements is greatly reduced. The volume of statements and conventional-system time requirements combine to impose an average of 10 weeks' lead time for statement preparation. Lead time is cut to an average of 4 weeks. This significantly reduces the number of manual changes, such as addresses, that have to be effected, with a corresponding reduction in the number of errors so incurred. This time and error elimination is also reflected in the corresponding branch office operations. A reduction of 10,000 manually prepared statements per month is made possible by the speed of statement preparation and the resultant reduction of lead time.

Third, *accuracy:* The application of record maintenance to one consolidated file in place of multiple files greatly reduces the exposure to error. Because decentralized files are necessarily functionalized, the data they contain are accurate, to a large degree, in direct relation to the frequency of their use. For example, in one file, agent number is used every month and thus its degree of accuracy is relatively high. In another file, the same data, used only once per year, may contain errors that are undetected for long periods and thus may result in costly error location and correction or in equally costly file maintenance. Because there is only one master file to maintain under electronic data processing, this problem of error correction is greatly eased.

In addition, as records are consolidated, the systems and accounting controls are simplified. These controls are accumulated and applied while processing takes place. The system is designed to detect automatically data input errors. All policy identification on current transactions must match the data on the master tape record before processing proceeds. Such errors are written out for subsequent manual analysis and processing is not inhibited.

Any occasional machine error is detected and automatically corrected by the machine before any recording takes place, and many clerical or operator errors are prevented or detected automatically before they enter the system. For example, the prevention of such a simple error as loading the wrong input tape is handled in the following manner. The first record on each input tape is a "tape label" containing details such as file name, serial number of tape reel, and date. Instructions, previously stored in machine memory, test each tape label. Any incorrect reel is detected immediately, the machine stops, and the console typewriter writes a message referring to the error.

Finally, management expects to improve the factual basis for the definition of standards in many operations. The large-scale electronic machine can so speed the collection, classification, and display of data that new techniques are offered for the incorporation of formulas to indicate deviations from the norm or standard.

Operating characteristics. The operating characteristics of the data-processing machine on premium accounting using a single-address instruction system, variable field length and variable record length are as follows:

Instructions		
Total Instructions		5000
Input Instructions	150	
Output Instructions	200	
Processing Instructions		
To pass inactive records	90	
To process 81% of active records (Regular transactions)	1260	
To process remainder of active records (involving exceptions)	3300	5000

Collating Ratio

One transaction to four master records

Master Record Length: 315 characters

11.3.2 Company B: Automotive Spare-Parts Stock Control

This inventory encompasses 60,000 service parts for passenger cars, trucks, and marine and industrial engines. The company sells these parts to independent distributors, automobile dealers, and retail customers. Approximately 40 per cent of the parts are manufactured by the company in its car-producing plants, and the remainder are procured from vendors. In some cases several vendors supply the same items.

The distribution of these parts emanates from three master warehouses. In addition, there are four outlying warehouses that stock 20,000 fast-moving items. Each warehouse supplies customers in its respective area. All material is distributed to the four outlying warehouses from a central freight-forwarding station located at one of the master warehouses. The daily volume is about 40,000 line-item transactions per day, of which 32,000 are sales.

The problem. Each warehouse performs its own billing by punched-card methods. The punched cards are forwarded to a central point where the stock control records are maintained.

Each month's sales, receipts, and transfers are summarized by part number. The summary is completed two weeks after the close of the month's business. Individual stock ledger cards are posted and manually reviewed for the required action. All ledgers are filed in part and warehouse sequence, which prohibits specialization of review by type or classification of part.

When the method of manual review is used, several weeks elapse before needed action can be recognized. After determination of a needed action, a number of separate data files provide the additional information required.

Objectives. The following objectives were established:

(1) Consolidation of all information required for any stock control action into one master file.

(2) Consistent action for procurement and transfer (relatively inconsistent under manual review methods).

(3) Automatic and early incorporation of engineering changes.

(4) Automatic and early determination of procurement, transfer, follow-up, and surplus actions.

Procedure. *Input Data.* All information, including sales history, usage, costs, specification, procurement data, contained in the several reference files is consolidated into a master stock status ledger, magnetic-tape file (Fig. 11.3). In addition, the file includes the current

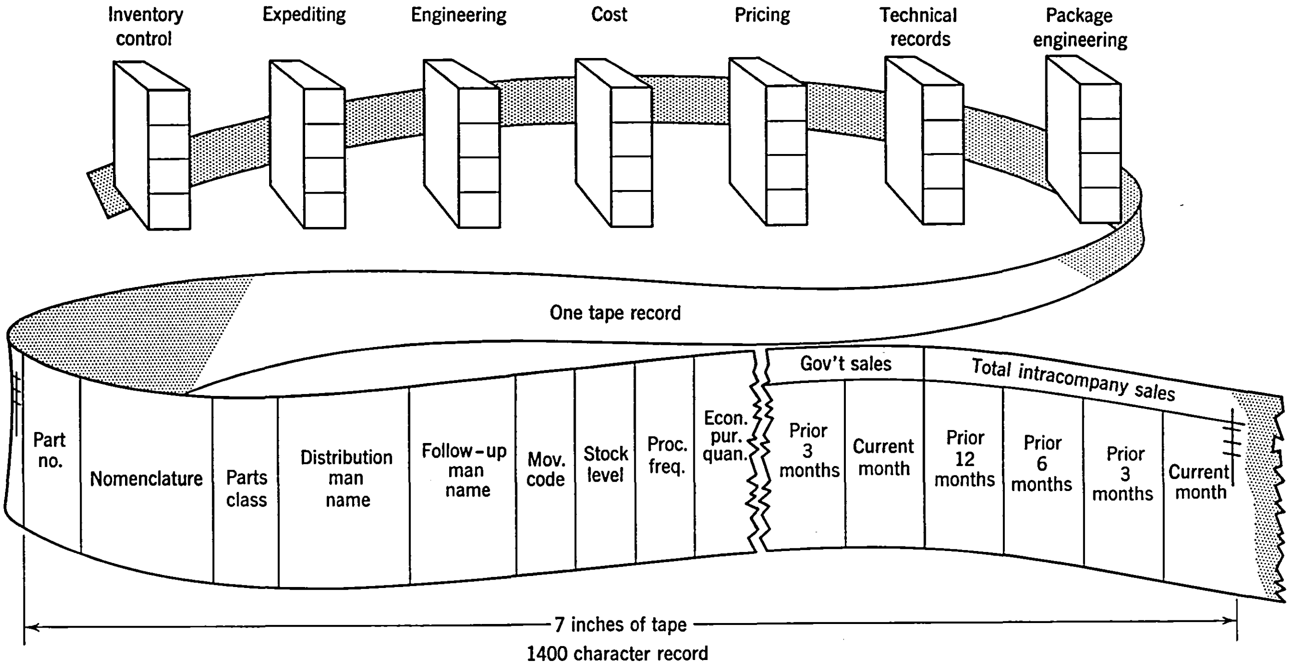


Fig. 11.3 Master stock ledger

GENERAL MANUFACTURING COMPANY
STOCK CONTROL
ACTION REPORT

ITEM NO.	NOMENCLATURE	ITEM CLASSIFICATION	PROCUREMENT MAN:	DISTRIBUTION MAN:	FOLLOW-UP MAN:	REASON FOR REPORT				MOVEMENT	STOCK LEVEL	PROC. FREQ.	ECONOMIC PUR. QTY.	MAX. ON HAND	INV. VALUATION ON HAND + ON ORDER	DATE OF REPORT
						P	D	F	M							
5739162	PULLEY 8 IN. HI SPEED	14 F-3.	62 COLLINS	O'BRIEN	MCDUFF	2	4	5		FAST	C	MO	400	1200	1813.73	7-15

MONTHS OF SUPPLY	DEMAND TO STOCK		LEAD TIME	TYPE	SOURCE		BLANKET PURCHASE ORDER	PUR. ORD. EXPIRATION	VENDOR UNIT PRICE	PROCESSING SPEC. NO.	PACKAGING SPEC. NO.	MAT'L CODE	UNIT PKG. SIZE			DRAWING NUMBER	DRAWING DATE	COMPONENT OF PKG. OR ASM NO.	QTY. PER PKG.	AVG. QTY. PER ORD.
	PIECES	ORDER			CODE	PLANT/VENDOR							LENG.	WID.	HT.					
0	20	7	1.25	P	PF	MOORE	W-51481	12-31-5	1.6725	36-R17492	02-19	M	2	3	2 B	14372	7-15-37	215261	1	3

SUPERSEDES PART NO.	INTERCHANGEABLE WITH	SUPERSEDED BY:	DATE OF ORIG. REL.	TAX CODE	% FOR EXPORT	UNIT COST	PACKAGING COST	LIST PRICE	END-PRODUCTS USED ON					
									B	E	H	P	R	S
									4 7 8 9 10 12 17 19	9 12 15 18 21 27 31	1 3 4 8	2 3 4 5 9 10	4 5 7 9 11 14	6 21 24 36 44
1948399	5742151		2-16-44	NT	30	2.0175	.0400	4.00	X XX X	X X	X	X X	XXX	XXXX

WAREHOUSE	RESERVATION	BALANCE ON ORDER	QUANTITY IN TRANSIT	QUANTITY ON HAND	CURRENT AVAILABILITY	REORDER LEVEL	QUANTITY VARIANCE	SALES											REFERRED ORDERS				BACK ORDERS		
								4TH QTR.		3RD QTR.		2ND QTR.		1ST QTR.		LAST MO.		INCREASE	PRIOR QTR.		CUR. MO.		3 MONTHS AGO	CUR.	ORD.
								QUAN.	ORD.	QUAN.	ORD.	QUAN.	ORD.	QUAN.	ORD.	QUAN.	ORD.	%	QUAN.	ORD.	QUAN.	ORD.			
*1	100	100		93	93	189	96	217	74	209	69	195	65	219	74	71	26	.01							
2		100	100	14	214	145	69-	197	68	236	82	207	84	185	65	63	21	8.36-							
3		75	50	37	162	134	28-	195	70	184	71	198	69	140	49	53	191	1.30-							
*4		25	25	40	90	121	31	164	57	193	64	195	73	93	38	49	12	8.84-					21	0	0
5		50	50	62	162	130	32-	186	63	178	69	201	49	136	44	54	14	7.57-							
*6		50		28	18	110	32	141	46	165	51	173	55	108	37	49	9	.00							

7																	.	14	3	2	1			
8																	.	10	5	4	4			
NET TOTAL	100	400	225	274	799	829	30	1100		1165		1169		881		339		5.72-	24	8	6	5	21	0	0

TOTAL DOMESTIC SALES				TOTAL EXPORT SALES				TOTAL GOVERNMENT SALES				TOTAL INTRACOMPANY SALES			
PRIOR 12 MONTHS	PRIOR 6 MONTHS	PRIOR 3 MONTHS	LAST MONTH	PRIOR 12 MONTHS	PRIOR 6 MONTHS	PRIOR 3 MONTHS	LAST MONTH	PRIOR 12 MONTHS	PRIOR 6 MONTHS	PRIOR 3 MONTHS	LAST MONTH	PRIOR 12 MONTHS	PRIOR 6 MONTHS	PRIOR 3 MONTHS	LAST MONTH
3242	1473	577	237	873	492	267	93	157	64	29	7	43	21	8	2

CODE LEGEND

REASON FOR REPORT

1. PROCUREMENT ACTION — DIST. WHSE.
2. PROCUREMENT ACTION — NAT'L. REORDER
3. REDISTRIBUTION ACTION — SETUP WHSE.
4. REDISTRIBUTION ACTION — WHSE. REPLENISHMENT
5. FOLLOW-UP ACTION — LESS THAN 30 DAYS ON HAND
6. FOLLOW-UP ACTION — LESS THAN 60 DAYS ON HAND
7. SUPERSEDURE
8. EXCESS MAT'L. ON HAND
9. DECLINE IN SALES, REVIEW FOR CANCEL
10. REPORT REQUESTED

NOTE: ASTERISK (*) BESIDE WAREHOUSE NO. INDICATES LOCATION WHERE DISTRIBUTION ACTION REQUIRED.

MOVEMENT CODE

CODE	PIECES SOLD PER YEAR
VERY FAST	OVER-5000
FAST	OVER-3000
MEDIUM	OVER-500
SLOW	OVER-100
VERY SLOW	OVER-25
INACTIVE	LESS THAN-25

STOCK LEVEL CODE

C — COMPUTED
F — FIXED
V — VARIABLE
N — NONE REPORTED WEEKLY
S — SPECIAL
G — GOVERNMENT ONLY

PROCUREMENT FREQUENCY

MO — MONTHLY
QTR — QUARTERLY
SA — SEMIANNUALLY
ANN — ANNUALLY
AT — ALL TIME
LT — LONG TIME
AR — AS REQUIRED

TYPE OF ITEM

P — SUPPLY ITEM
C — COMPONENT ONLY NONSUPPLY
CP — PACKAGE COMPONENT
C-S — COMPONENT AND SUPPLY ITEM
EX — EXPORT ONLY
UP — UNIT PACKAGE ONLY

MATERIAL CODE

A — PAPER	M — METAL
C — CORIT	P — PLASTIC
F — FIBER	R — RUBBER
G — GLASS	W — WOOD

SOURCE CODE

PF — PURCHASED FINISHED
PIA — PURCHASED IN ASSEMBLY
PSF — PURCHASED SEMIFINISHED
SA — SHOP ASSEMBLED
MFG. — MANUFACTURED
VP — VENDOR PROCESS ONLY

REMARKS:

Fig. 11.4 Stock control form

stock status for each of the 60,000 items and warehouse totals. The individual records on the tape vary from 600 to 1400 characters.

A second input tape is daily sales, an automatic by-product of the punched-card billing procedure. The following sales data are contained: warehouse, type of sale, quantity, part identification, date, and invoice number.

The third input includes all transactions except sales, including receipts, transfers, and adjustments. All input is in part number sequence.

Output. The first output is a completely updated master stock status record tape, which becomes the master input record for the next day. Second is a distribution or sales summary tape which is converted to a printed report. This is used to allocate material due in at the central forwarding station for distribution to other warehouses, to release back orders, and to expedite on order material. The third output is a stock control action report. This record is produced only when some kind of human action is required, and it contains all the information necessary to complete any action without reference to any other record (Fig. 11.4).

Program and Logic. The stored program calls for the input tapes to be read one record at a time. After the first item is read from the master tape, the first record from the transaction tape is read and compared. If equal, the master record is increased or decreased in the respective quantitative balances after the necessary arithmetic such as price extension is performed. If unequal, all the data on the master record pass unchanged into the updated output tape, and the next stock item is read. Similarly, the data from the sales tape are posted.

The machine program is such as to make an average of twenty-four decisions for each part number handled. In an average day's business, 10,000 parts are active. Thus, 240,000 decisions are made. These may be classified into four groups: (1) procurement, (2) distribution, (3) expediting, (4) distribution of surplus.

A typical example of a procurement decision follows: If the quantity on order nationally plus the quantity on hand nationally is less than the computed average of the last six month's sales multiplied by a given factor, procurement action is required and a stock status report is produced. Another example: If the quantity available satisfies both national and local requirements, but the on-hand balance is less than 30 days' supply for both conditions, a stock status report is produced and directed to an expediter for follow-up.

Finally, the machine automatically allocates parts received in stock to back orders by date or age until the quantity allotted for back orders is exhausted. Where the allotted quantity is insufficient to cover all back orders, proportionate partial shipments are effected automatically.

Advantages. (1) It is estimated that the electronic data-processing machine can handle eight times as much of the total stock control record keeping as the previous system.

(2) It is estimated that the speed of operation makes possible a reduction of 3 per cent in the total inventory investment. The cost of inventory maintenance may be estimated at 25 per cent per year. For a $50,000,000 inventory, this represents a savings of $375,000 per year.

(3) The logical abilities of the electronic data-processing machine permit the vast majority of stock control decisions to be resolved in formulas and programmed for automatic handling. This ability is not judged so much in terms of the savings of clerical time as it is in the elimination of error and the achievement of better control.

Operating characteristics. The operating characteristics of the data-processing machine for the inventory control application are as follows:

Instructions		
Total Instructions		2400
Input Instructions	170	
Output Instructions	90	
Processing Instructions		
To pass inactive records	300	
To process 95% of active records (Regular transactions)	1490	
To process remainder of active records (involving exceptions *)	350	2400

Collating Ratio

One transaction to six master records

Master Record Length: 700 to 1450 characters

* In this case an exception is defined as a transaction which will not occur more than twice in one day.

11.3.3 Company C: Public-Utility Billing and Cash Accounting

We are all familiar with utility billing, either from our experience as students of data processing or as consumers who must pay the bills. Therefore, the aims, scope, and requirements of data processing in public utility are fairly well within our understanding. Simply stated, the principal task of the utility office is to account for every unit of energy produced and consumed, and to turn out bills and records reflecting this consumption in a form in which the energy units are converted to their dollars and cents value.

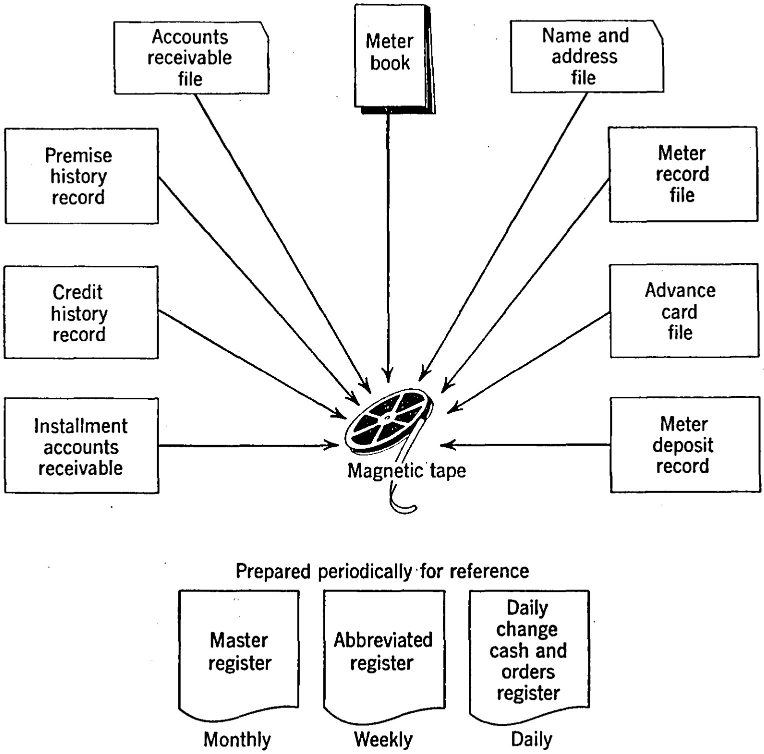


Fig. 11.5 Utility billing and accounting

Complexities arise in this basic process through the additional need to measure and bill consumption for different classes of customers at different rates, the sale of merchandise such as stoves and refrigerators, the adding and dropping of subscribers, and the many records and reports necessary to the conduct of modern business. The primary need is an accounting procedure that will be intelligible and swift to the customer and collectible and routine for the producer. The principal functions to be performed are those of billing, accounting for cash payments, and computing revenue and statistics (Fig. 11.5).

Billing. Several charges may be involved in the determination of the final amount of a bill. Although the customer's use (units consumed) is the primary item, demand charges based on the highest rate of use during the billing period is sometimes a factor. Local taxes, arrears, merchandise purchases, and fuel adjustment charges are frequently encountered.

The charge for usage is determined from a rate schedule in which the cost per unit decreases usually as a step function as the usage increases. In addition, schedules may be modified by special circumstances. An average company has eight rate schedules with five steps or blocks in each schedule.

Consumption is figured by subtracting previous from present meter readings. A so-called "hi-lo check" is made to determine whether the calculated use is reasonable when compared to previous period uses. Assuming the figure to be correct, the proper rate schedule is chosen and a table look-up executed. The proper block having been established, appropriate factors are applied to calculate the charge. Taxes, arrears, and other factors are added to this figure to arrive at the total bill. The bill amount and account number may be punched in an IBM postcard with a detachable stub. The charges are printed on one side of the card, and the name and address on the other. The bill is then mailed to the consumer.

Cash Payments. The stub is detached by the customer and forwarded with payment for posting accounts receivable. Open accounts must be reviewed periodically so that notices may be sent or special action initiated on delinquent accounts. Open items are also picked up during the following billing periods to provide the arrears charges.

Revenue and Statistics. Statistics are required at the time of billing. Amounts, rate, use, type of customer, location, transformer area, and similar information are recorded. These data are reflected in reports such as revenue by rate, use by transformer area, and revenue by type of customer. Revenue entries for the general ledger are also obtained at billing time.

This review of basic billing objectives illustrates the necessity for combining various types of information. A customer history is needed to indicate rate, previous meter reading, and previous use. Accounts receivable information is needed for arrears. Name and address information is required for mailing. Deposit and merchandise purchase information may also be required. Delinquent notices necessitate the merging of information from accounts receivable credit history and name and address files.

Under methods other than employing magnetic tape, it is necessary to maintain these various files separately and quite often in different departments, because unit records of sufficient capacity are impractical, and clerical decisions are often necessary to determine what actions are to be taken. The burdens imposed because of these restrictions hamper the handling of inquiries to an account either from inside or outside the company. The maze of intercommunication for obtaining the desired information is costly both from the point of view of personnel involved and the introduction of errors caused by manual handling of the files. The process of keeping duplicate files current creates another problem. Notification of changes must be routed to six or seven locations in many instances. The physical difficulties involved in maintaining parallel files cause delays in the posting of charges. When files are not up to date, subsequent corrections must be made. Gathering the required files together to perform the functions of customer accounting and the subsequent splitting apart of these files for inquiry and maintenance are costly operations.

Objective. Magnetic tape, with its high-speed input and output characteristics, places the following objectives within the grasp of companies in the utility field using data-processing machines:

(1) Elimination of duplicate files.

(2) Reduction of file maintenance effort to a single file, electronically updated.

(3) Keeping files more current.

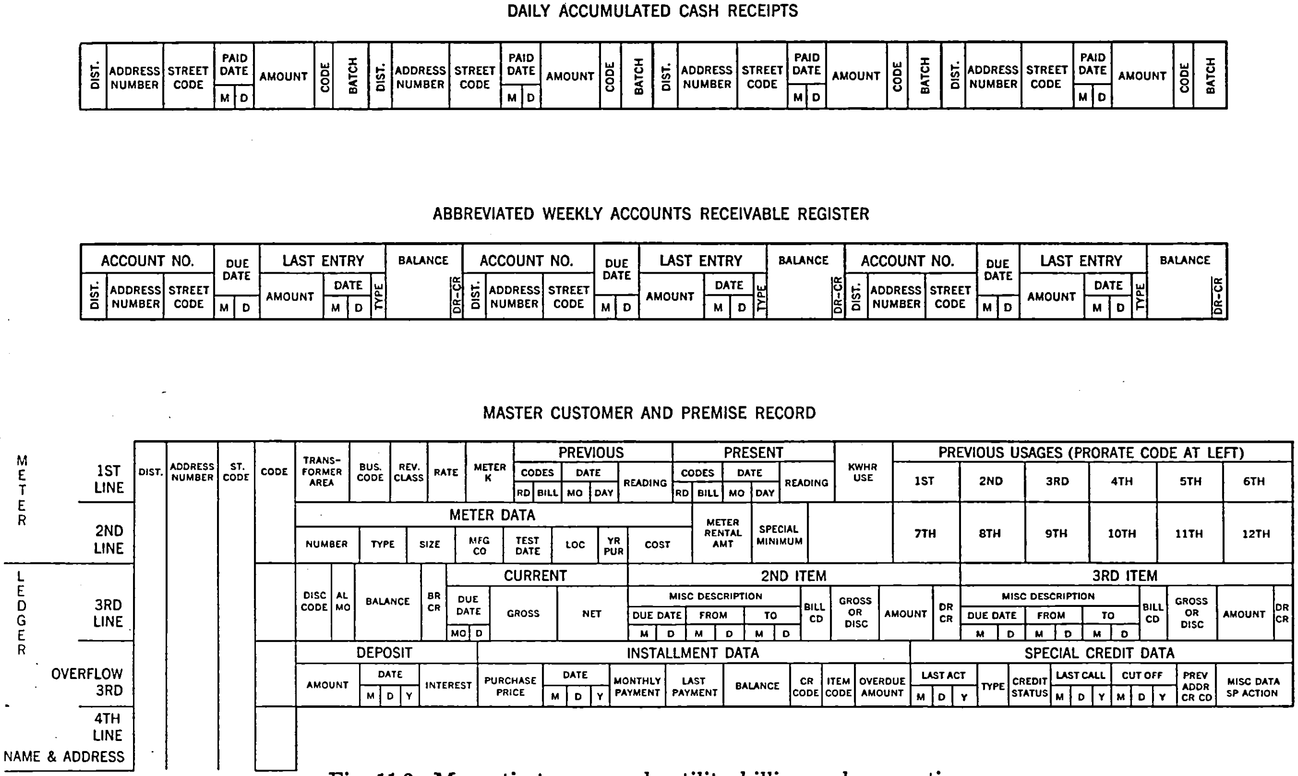
(4) Handling all exceptional billing matters without clerical intervention.

(5) Speeding up the preparation of bills.

(6) Economizing all office operations.

Coincident with obtaining these objectives, the capabilities of data-processing equipment allows consolidation of all identifying, historical, accounting, deposit, consumption, and merchandising information regarding a consumer on a few inches of magnetic tape. The high speed of tape reading in turn assumes that an average of thirty accounts can be processed per second. The high speed, extensive memory, and associated capacity for storing and executing a large number of instructions allow the equipment to provide more information sooner and more often with far less clerical control and decision and with greater accuracy at less cost (Fig. 11.6).

Procedure. To explain the flow of work, we shall use as an example a company with 1,800,000 electric accounts. This company bills each account every 40 working days. Approximately 50,000 bills and 60,000 payments are processed each day, and 2000 special and final bills are required daily. Changes to the master file are also encountered at the rate of 9000 per day. Delinquent accounts require 3000 collection notices daily.



DAILY ACCUMULATED CASH RECEIPTS

ABBREVIATED WEEKLY ACCOUNTS RECEIVABLE REGISTER

MASTER CUSTOMER AND PREMISE RECORD

Fig. 11.6 Magnetic tape records utility billing and accounting

The forty billing cycles are divided into five major groups for processing purposes. Each group consists of eight billing cycles with each cycle five reading days apart. For example, the first group consists of reading days 1, 6, 11, 16, 21, 26, 31, and 36 and the second group of reading days 2, 7, 12, 17, 22, 27, 32, and 37.

In file maintenance, all accounts in the same group as the day to be billed are updated. Thus, each account reflects current status every 5 days and can be selected for special or final billing. This 5-day cycle also makes it possible to check for delinquency and send collection notices at frequent intervals. A brief description of each of the ten daily runs follows:

Run 1: Batch-Balance and Sort. The tape is created from the cash stub portion of an IBM postcard bill which is returned with the customer's remittance. During bill preparation, these are punched with account number, type of payment, and the net amount of the bill. They are ready for processing unless the customer makes an irregular payment or does not take advantage of the discount.

The 50,000 stubs returned each day go through clerical batch control procedures and are then, along with cash correction cards from the previous day, transferred to the daily cash tape.

This tape, in turn, is sorted during run 1 to arrange cash and miscellaneous entries into account number order. Proof listings, produced in auxiliary tape-to-printer operations, meet the need for control figures and balancing.

Run 2: Merge and Select. The sorted cash tape contains entries for all reading days. Since only one group of days is scheduled for file maintenance, the cash payments for other days must be stored until called for by the schedule. Also, 5 days having elapsed since the selected group was previously processed, interim cash entries are picked up. The accumulated unposted cash file is merged with the sorted daily cash tape from run 1 to achieve this.

The reading day for each account from both tapes is interrogated upon entering the machine. If the day in question is on the current file maintenance schedule, it appears on the selected cash output. Otherwise, it is transferred to the accumulated unposted cash tape. Auxiliary tape-to-printer listing provides station clerks with reference information for inquiry handling. The tape from which it is written contains 200,000 items which are subject to the same electronic scrutiny of succeeding days until selected for output on the cash tape described in run 5.

Runs 3 and 4 perform the same functions for change orders as runs 1 and 2 perform for cash. These changes, approximately 1000 daily, include such things as final turn-offs, turn-ons, rate changes, and address changes. Run 3 sorts changes to account number sequence. Run 4 merges these changes with an accumulated file of unprocessed changes and selects those changes pertinent to the file maintenance for a particular day.

Run 5 is the file maintenance operation. The master records for the 8 reading days scheduled are on one input tape. The corresponding cash entries and changes selected on runs 2 and 4 are the other input files. While the master files are updated by the change orders and the balances are adjusted by the cash entries, several other operations are performed.

First, tapes are prepared for register-listing the cash entries and the details of open balances. Several types of delinquent notices can later be prepared, depending on such factors as extent of arrears, amount, and the customer's credit history. Rather than tie up a separate tape unit for each requirement such as merchandise statements, charge-offs, and collection action, advantage is taken of the machine's ability to consolidate all these activities on a single tape during the file maintenance routine and later, on run 6, to separate them into their proper categories. The other tape outputs are the main tapes of the file maintenance run—the updated master records for accounts to be billed and the updated records not in the current billing period.

Run 6 accomplishes the necessary separation and editing of various classes of data that were consolidated on one tape during the file maintenance operation. The combined records enter the machine on a single tape, and separate tapes are produced for meter testing, accounts receivable control, collection records, charge-offs, and requested copies of master records. These new tapes are used for further input or in auxiliary tape-to-printer and tape-to-card operations for production of the final papers involved.

Run 7 sorts the meter-reading records from the field into account number sequence. This is a simple, straightforward machine operation involving, at the first stage, prepunched cards pencil-marked with present meter readings. After the pencil marks have been automatically translated into punched-hole form, the cards to be billed for the day and any special reading cards for the other seven reading days in the group are combined for card-to-tape conversion and subsequent tape sorting into account order.

Run 8 combines the updated master records and the meter-reading records. All computations necessary for billing are performed. Consumptions are computed and hi-lo checked to insure accuracy of readings. Where discrepancies are encountered, estimations are made on the basis of both the readings for the past year and seasonal factors. Rates are selected and charges computed from internally stored schedules. All billing data are recorded in the new master records. Field calls and memos generated during the performance of this run are recorded on a tape which is subsequently converted to cards.

Run 9. On run 9, the information necessary to prepare the bills is extracted from the new master records. It is arranged for tape-to-card conversion. Once converted, the cards are processed on standard punched-card machines. The result is a postcard bill with a punched stub. The stub, punched with amount and account number, is returned with the customer's payment and is used in the posting of cash as described in run 1. A bill register tape is prepared which is subsequently printed in an auxiliary operation. Copies of the register are used as a reference document by the station clerks to answer inquiries. Another output from this run is the revenue and statistics tape, used to obtain accounts receivable entries and statistical reports. A summary tape is updated daily, and reports are produced for management on a monthly schedule.

Advantages. This description has been kept brief and general so that the overall results could be evaluated. The programs stored within the memory of the machine test the data for application of appropriate measures to such variables as multiple meters, credit history, mailing address different from meter address, present usage greater or less than previous usage, sliding rate schedules, meter changes, and similar departures from the norm. This decision-making ability—the so-called "logic" of the machine—ascertains which of many electronic paths the data should follow; what computations to include or exclude; and where to store or record the results. The machine program indicates in detail the discrete steps that must be taken in every case so that each transaction will be processed according to its needs.

Perhaps the most significant indication of what electronic data processing means to a utility is drawn from a recent newspaper story. Executives of a utility told the news reporter that electronic data processing is expected to speed up office operations so that bills will be in the mail within two days instead of the five now required. Furthermore, an estimated annual savings of three-quarters of a million dollars is expected.

Operating characteristics. The operating characteristics of the data-processing machine for the file maintenance application are as follows:

Instructions		
Total Instructions		4000
Input Instructions	400	
Output Instructions	680	
Processing Instructions		
To pass inactive records	25	
To process 61% of active records (Regular transactions)	200	
To process remainder of active records (involving exceptions *)	2695	4000

Collating Ratio

One transaction to seven master records

Master Record Length:		
	Minimum	202 characters
	Average	240 characters
	Practical Maximum	1000 characters

* The most complex exception required 804 program instructions to process on an automatic basis.

The operating characteristics of the data-processing machine for the bill calculation application are as follows:

Instructions		
Total Instructions		4500
Input Instructions	180	
Output Instructions	213	
Processing Instructions		
To pass inactive records	200	
To process 68% of active records (Regular transactions)	350	
To process remainder of active records (involving exceptions *)	3557	4500

Collating Ratio

Nine transactions to ten master records

Master Record Length:		
	Minimum	202 characters
	Average	240 characters
	Practical Maximum	1000 characters

* The most complex exception required 1200 instructions to process on an automatic basis.

11.4 CONCLUSION

In the preceding section, we have illustrated how the principles of data processing can be applied to problems which involve large volume and which represent the work of existing departments. The application of the large-scale machine became efficient through consolidation of files, the processing of several classes of information simultaneously, and the production of all operating reports simultaneously.

11.5 REFERENCES

1. *Business Applications of Electronic Machines,* an annotated bibliography, prepared and published by Controllership Foundation, Inc., One East 42nd Street, New York 17, N.Y.

2. *Data Processing Digest,* published monthly by Canning, Sisson and Associates, 914 South Robertson Boulevard, Los Angeles 35, Calif.

3. *Electronics Reference Handbook,* a compilation of twenty papers on commercial applications of electronic data processors, American Management Association, March 1, 1955, 330 West 42nd Street, New York 36, N. Y.

4. Papers presented at the American Management Association Conferences, October 20–22, 1954, New York, and February 28, March 1–2, 1955, New York.

5. Papers presented at the Joint Computer Conferences, March 1–3, 1955, Los Angeles, and December 8–10, 1954, Philadelphia, Pa.

6. Peter B. Laubach and Lawrence E. Thompson, Electronic Computers: A Progress Report, *Harvard Business Rev.,* 120–128, March–April 1955.

7. Richard F. Clippinger, Economics of the Digital Computer, *Harvard Business Rev.,* 77, Jan.–Feb. 1955.

8. Trends in Computers: Automatic Control and Data Processing, *Proceedings of the Western Computer Conference, February 11–12, 1954,* Los Angeles.

9. *Proceedings of the Western Computer Conference, February 4–6, 1953,* Los Angeles.

10. Herbert J. Groelinger, Inventory Control under the Program-Usage-Replenishment System Defined for Electronic Processing, *Nat. Bur. Standards Rept. 3651.*

11. E. F. Cooley, Electronic Equipment Applied to Periodic Billing, *Computers and Automation, 2,* No. 8, Nov. 1953.

12. Cuthbert C. Hurd, Automatic Digital Computing in Engineering, *Applied Mechanics Rev.,* 269, July 1955.

13. Cuthbert C. Hurd, Computing in Management Science, *Management Sci., 1,* No. 2: 103–114, Jan. 1955.

12. Automatic Control of Flight

DUANE T. McRUER

Control Specialists, Inc.
Inglewood, California

12.1 INTRODUCTION

12.1.1 General

From the earliest times, man has dreamed of flying. During the past fifty years this dream has become a reality; and as the size and speed of aircraft increased, the development of automatic flight control became a necessity. Elementary forms of automatic flight control date back to within ten years of the Wright brothers' first flight. Today, automatic pilots are standard on almost every type of commercial and military aircraft. The first automatic pilots were essentially devices to improve aircraft stability and thus relieve the pilot from the requirement of continuous concentration. The modern autopilot not only must provide short-term stability but must be adaptable to long-range navigations and automatic landing as well as handle stabilized turns, altitude control, and compensation for widely varying aerodynamic characteristics encountered in the changing conditions of modern flight.

The advent of the guided missile has given perhaps the greatest impetus to automatic flight. Although small drones or missiles were developed during the first World War, they were never used in combat until World War II. Several missiles with automatic control features were developed in this country during World War II. It was the successful operation of the German V-1 and V-2 missiles, however, that brought home a realization of the full potentialities of guided missiles.

A general problem of systems engineering is the synthesis of a mechanism to control automatically a physical element in a specified fashion while operating in an environment containing restrictive and disturbing influences which tend to upset the desired system performance. In the case of automatic flight control this problem can be resolved to one of building a system capable of achieving, in the presence of various aerodynamic and command disturbances, an actual airframe flight path consistent with some desired trajectory; this system must utilize equipment enclosed largely in the airframe itself. This is a formidable problem, basic to flight itself, as is attested by the fact that the major technological feat of the Wright brothers was one of attaining a craft and control configuration which man could control in the presence of disturbing wind forces. Aircraft and gliders flew fairly well before the Wrights' accomplishment, but men had enormous difficulty in controlling them. Today, after some fifty years of progress, the flight control art is in a situation similar to that confronting aviation's pioneers, for high-performance, manned-aircraft configurations are very difficult to control without automatic aids, and unmanned aircraft or guided missiles require a complete substitute for the pilot.

Fortunately, automatic flight control techniques have advanced, as part of aviation's general progress, to the point where automatic controls can be made available for augmenting or replacing the pilot's effort wherever required. Almost all modern high-performance manned aircraft have a full complement of automatic flight devices. These devices include: automatic pilots for navigation and general pilot relief; associated autopilot–command system couplers for automatic approach, tracking, guidance, etc.; and stability augmenters suitably coupled with an artificial "feel" system to help the pilot when he wants to fly under manual control. Even newer developments control the entire flight plan from ramp and take-off to landing, taxiing, and back to the ramp, with the pilot merely acting as a monitor to take over when something goes wrong. Just a few years ago the autopilot was installed as an accessory system for pilot relief. If used at all, the pilot turned it on cautiously, while staying alert and ever-ready for instant disengagement. A few years hence some particularly daring flyer may even more timidly turn off his automatic flight control system just to see what will happen.

The guided missile, by virtue of its name alone, has required an automatic flight control system from the outset. Here, as in manned aircraft, the overall flight control system usually has many subsystems. These are tied into the basic autopilot and provide coupling of the flight control system to the command or guidance device, thrust, and cruise controls, etc. As would be expected, the flight control system of a guided missile is strikingly similar to that of a manned aircraft from a general, overall standpoint, but it may be very dissimilar in detail. The major reason for both the similarity and differences is the application of the system concept, which requires a specifically designed control system for a given task, performance, and environment.

12.1.2 The Specific Flight Control Problem

In the automatic flight control case, the quantity to be controlled is normally the flight path of the aircraft, which is dependent upon motions of the craft. Control forces and moments are imposed upon the airframe by actuating the control surfaces, which conventionally consist of the elevator, rudder, and ailerons or composite control surfaces such as elevons, and special-purpose surfaces like dive brakes. In some instances thrust direction and magnitude changers are also used as elements to apply control on the craft. Various types of servomechanisms are used to actuate these controls; sensors are used to detect the motions of the airframe, and suitable equalizers and amplifiers are placed between the sensors and servo actuators to enable suitable overall system operation.

The logical approach to any control design problem stems from a consideration of the characteristics of the command inputs, the controlled system, the disturbances and environmental conditions, and the desired overall operation. (See Chapters 3 and 4.) The physical quantities to be controlled should be identified and the means established for imposing control upon the controlled system. The ability and means used to measure physically the controlled variables and to actuate the mechanisms which apply control to the controlled element should then be carefully considered.

The command input which sets the mode of operation is derived from a guidance or reference element. The difference between the input and the feedback signal calls for either maintenance of or a change in the flight path of the aircraft. When the difference is zero the craft is usually controlled in a level flight attitude. A difference other than zero calls for a turn, a climb, or a dive. The controlled system is, of course, the airframe, which also is a major contributor to the environment in which the control equipment must operate. The disturbances may be due to unwanted portions of the command input, noise, atmospheric forces acting upon the airframe, and miscellaneous unwanted inputs generated or imposed upon various items of the control equipment.

From the introductory statements above, it is apparent that a reasonable approach to take in describing the general aspects of automatic flight control is to discuss the following topics, which constitute the next five sections of this chapter: (1) characteristic motions of the airframe, or the dynamics of the controlled system; (2) environmental conditions imposed upon equipment by the airframe; (3) typical sensing elements; (4) typical actuating elements; (5) equalization and amplifying elements.

When the characteristics and environment of these "components" of automatic flight control systems are understood, it is then possible to discuss their general arrangement in the form of typical systems. This is done in the seventh section of the chapter.

12.2 CHARACTERISTIC MOTIONS OF THE AIRFRAME

An aircraft in flight may be thought of abstractly as a velocity vector in space, having both a direction and speed of travel. Considering the craft as a rigid body, the airframe then has 6 degrees of freedom, corresponding to rotations about and translations along its three axes.

To establish a reasonable reference frame for the description of the craft motions, imagine a rectangular coordinate system fixed in the airframe with the origin at the center of gravity. With the airframe in wings-level straight equilibrium flight the X axis will be directed in the direction of the relative wind; i.e., the equilibrium velocity vector of the craft will be along the X axis (see Fig. 12.1). While this axis is established along the relative wind for the equilibrium case, it is fixed to the airframe and will not be aligned with the relative wind (total airframe velocity vector), when the craft is disturbed from equilibrium. The Z axis is directed downward (aligned with the gravity vector if the craft is in wings-level horizontal straight equilibrium flight), and the Y axis sideways out the right wing. This axis system creates a reference frame, fixed in the craft for any one equilibrium flight condition, which is essentially that of the pilot if he were at the center of gravity. Hence, the forces, moments, and motions measured relative to it are essentially those observed by the pilot. With the reference frame established the components of the motion of the aircraft's velocity vector can be resolved into (1) yawing rotations, rotations about the yaw or Z axis; (2) pitching rotations, rotations about the pitch or Y axis; (3) rolling rotations, rotations about the roll or X axis; (4) forward velocity, linear velocity measured along the X axis; (5) side velocity, linear velocity measured along the Y axis; (6) downward velocity, linear velocity measured along the Z axis.

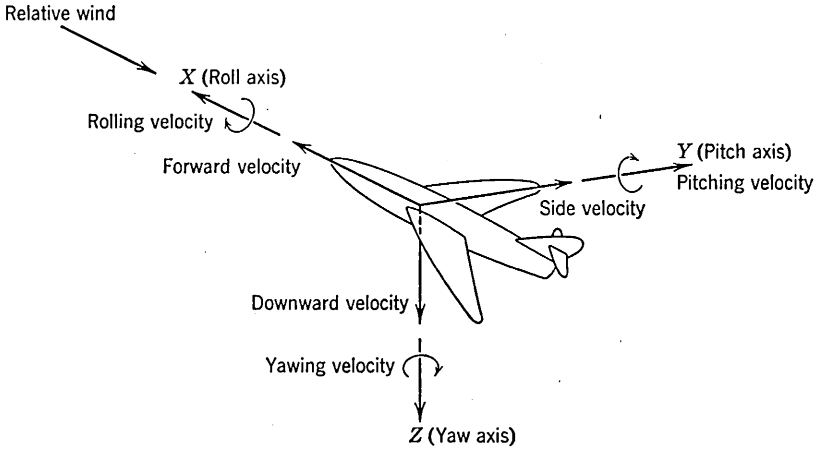


Fig. 12.1 Airframe axis system in equilibrium flight

Functions of all these quantities are capable of being measured in the aircraft itself by instruments properly aligned with the axis system fixed in the craft.

When the aircraft is disturbed from equilibrium flight its velocity vector will no longer be aligned along the X axis, and a pilot sitting in an open cockpit will feel the relative wind shift from directly head-on to some other angle. This requires the definition of two more angles to orient the aircraft velocity vector relative to the axis system fixed in the airframe.

These angles are shown in Fig. 12.2. In the YZ plane the angle between the relative wind and the velocity vector is called the angle of attack; in the XY plane the equivalent quantity is called the sideslip angle. These quantities cannot be measured directly by equipment within the airframe but rather require a sensor mounted externally on the craft, with the sensor's reference position fixed along the X axis. The angle of attack, as defined here, is dependent upon the downward velocity measured along the Z axis; the sideslip angle is similarly dependent upon the side velocity.

The forces acting upon the airframe are of several varieties; the inertial, centripetal, and gravitational forces are direct, but complicated, functions of the angular and linear velocities and accelerations measured about and along the X, Y, and Z axes. The aerodynamic forces are dependent upon the angles of attack and sideslip, as well as the various velocities and accelerations measured in the X, Y, and Z system. The detailed form of these forces and moments is quite complex, and discussion of them is beyond the scope of this chapter. The natural modes of motion of the aircraft arising from these forces and moments are of great importance, however, and require description, if not detailed technical explanation.

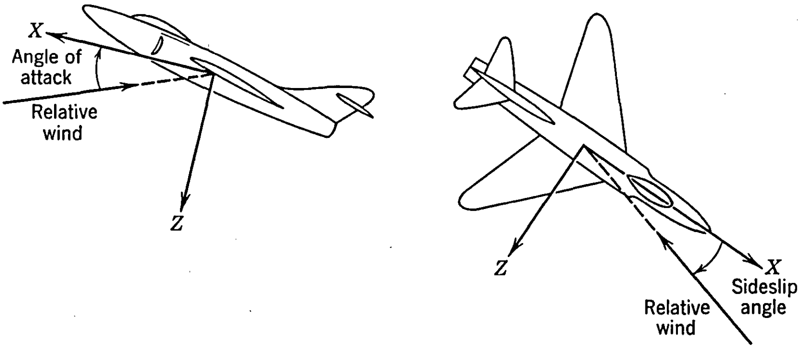


Fig. 12.2 The airframe in disturbed flight

When an airframe of a fairly conventional configuration is flown in wings-level straight flight, the fact that the aircraft is symmetrical about the XZ plane gives rise to an approximate separation of aircraft motions into those of the longitudinal variety and those of the lateral type. The longitudinal motions involve pitching-, forward-, and downward-velocity changes. The lateral motions involve rolling-, yawing-, and side-velocity changes.

In a conventional aircraft flying at subsonic speeds, disturbances in the longitudinal axis give rise to two oscillatory motions if the controls are fixed. The first of these is a very long-period (in seconds, about one-fifth the true air speed in miles per hour), lightly damped oscillation called the phugoid. Physically this motion is a cyclic change of potential to kinetic energy involving fairly large oscillatory changes in forward speed, pitch attitude and altitude, with essentially constant angle of attack. The second motion is a relatively fast, fairly well-damped oscillation called the short period. In a crude physical way, this oscillation is similar to one that would be observed if a weathervane were displaced and then released. The short-period motion has large changes in pitch angle and angle of attack, with essentially no change in forward speed. Typical time histories of these motions are shown in Fig. 12.3.

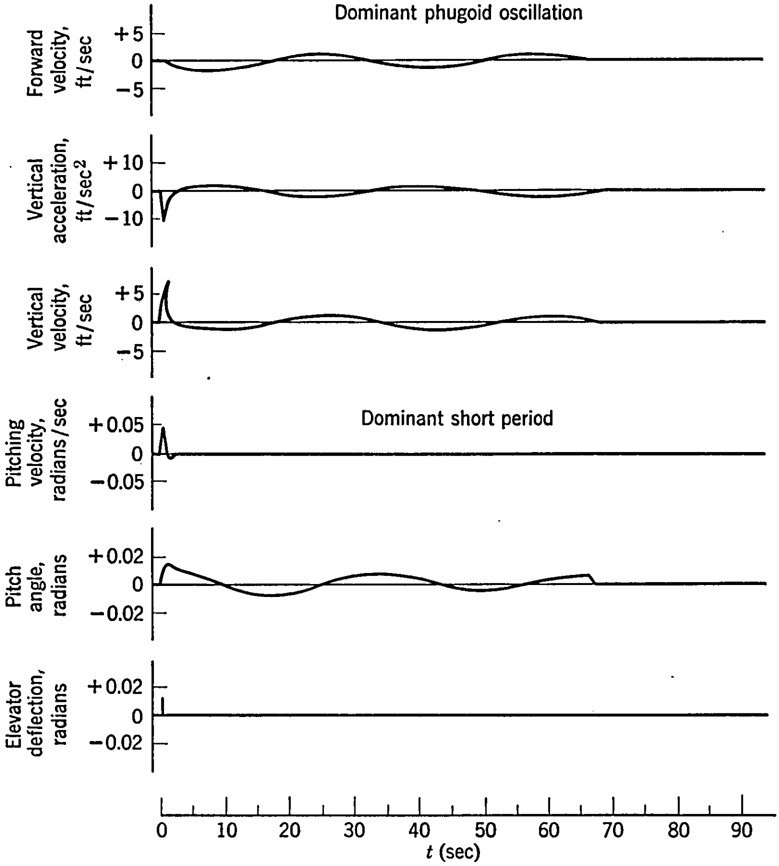


Fig. 12.3 Longitudinal time history

There are three more or less characteristic lateral motions in subsonic flight, the spiral divergence or convergence, the roll subsidence, and the Dutch roll oscillation. The divergent spiral mode is made up of a slowly diverging, ever-increasing yawing and banking motion, ultimately ending in a rolling dive. In some instances, this motion may be convergent, in which case the pilot would note a gradual directional subsidence to the previous equilibrium position after encountering a side gust. The roll subsidence is a mode associated primarily with the rolling behavior of the aircraft and is described largely by the time lag in attaining a nearly steady-state rolling velocity after a step application of the ailerons. This time lag is principally due to the large damping moment generated as the wing rotates and the roll moment of inertia. The Dutch roll mode is a lightly damped yawing, rolling, and sideslipping oscillation similar to the short-period mode in the longitudinal instance. Figure 12.4 illustrates these motions for a typical airframe. In addition to these lateral motions it should also be noted that the airframe is neutrally stable in heading.

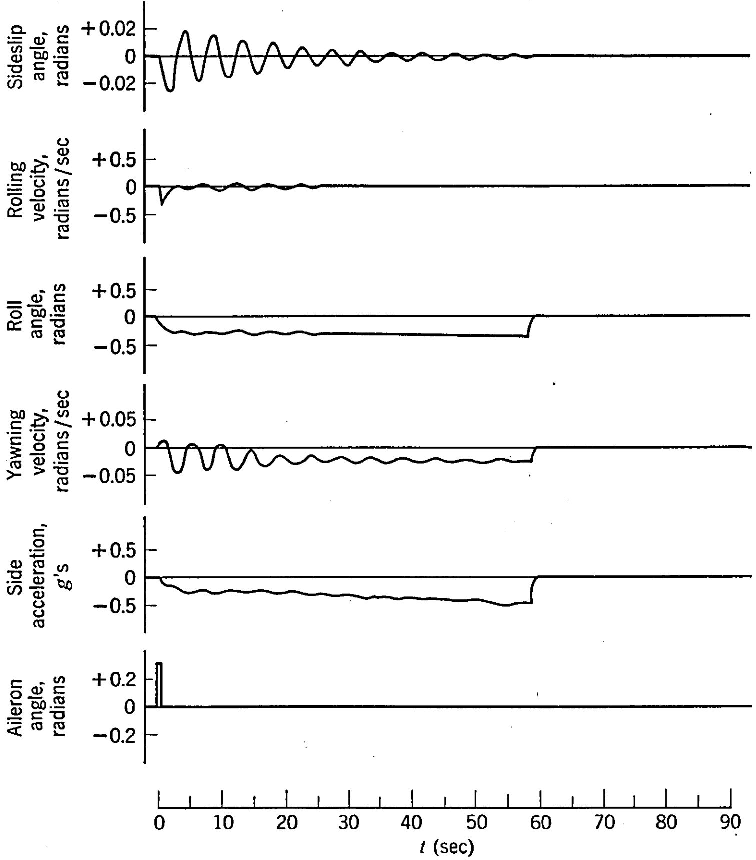


Fig. 12.4 Lateral time history

At transonic and supersonic speeds, and in unconventional aircraft, the "nominal" longitudinal and lateral modes of motion may be modified considerably from those discussed above. For example, longitudinally either the phugoid or short-period oscillations may be replaced by ever-increasing diverging motions coupled with subsiding ones. In other cases the short period may be very lightly damped or divergingly unstable. Laterally, the damping of the Dutch roll mode is often negative, resulting in a diverging oscillation. The spiral and roll subsidence motions sometimes couple and form a low-frequency oscillation which is occasionally divergent and is akin, in some ways, to the longitudinal phugoid motion. As a final example, there are often severe coupling effects between the lateral and longitudinal motions caused by gyroscopic-like phenomena occurring at large rolling velocities. Besides augmenting the control engineer's vocabulary with new descriptive words like "tuck under," "pitch-up," "wing drop," "supersonic yaw," "inertial coupling," etc., these transonic and supersonic effects have added a new dimension of difficulty to the flight control problem.

The simplified picture given above of the complicated dynamics of an uncontrolled airframe leads immediately to the conclusion that the characteristics of the craft alone may be highly deficient from the stability and control standpoint, and that a prime objective of any automatic flight control system is the modification of some of these characteristics.

The dynamic characteristics discussed above are basically those of a specially shaped and powered rigid body in space (though the low-frequency effects of elastic deformations usually alter the basic motions

Table 12.1

	(Basic) Output Quantities, Longitudinal	Actuating Quantities
u	forward velocity	δ_E elevator deflection δ_B fighter brake deflection
w	vertical velocity	
q	pitching velocity	
a_x	forward acceleration	
a_z	vertical acceleration	
α	angle of attack	
θ	pitch angle	

	Output Quantities, Lateral	Actuating Quantities
v	side velocity	δ_A aileron deflection δ_R rudder deflection
p	rolling velocity	
r	yawing velocity	
a_y	side acceleration	
ψ	yaw angle	
ϕ	bank angle	
β	sideslip angle	

slightly). The actual airframe is generally a light structure, deliberately designed to be relatively nonrigid, and hence has many more degrees of freedom than the six considered here. These additional degrees of freedom are the so-called aeroelastic modes, which are generally of much higher frequency than the rigid body characteristics. The aeroelastic modes are often of importance in a detail system design and should not be ignored in that phase of the systems engineering effort.

A summary of airframe output quantities capable of being used for control of certain modes and the actuating quantities suitable for imposing control is given in Table 12.1.

12.3 EQUIPMENT LIMITATIONS AND ENVIRONMENTS IMPOSED BY THE AIRFRAME

The components utilized in an automatic flight control system are quite similar to those used in many other control applications. However, the special environment encountered by virtue of the equipment being installed in an aircraft requires some attention.

The equipment environment created by the airframe is particularly troublesome because of the extremes of pressure, temperature, and vibration encountered in the craft, both in operation and when on standby. When the necessity of extraordinary system reliability is also realized (there being no maintenance men around when a missile or manned aircraft is airborne), it places the component design task at the very fringe of the state of the art. Add to these considerations the necessity of minimum weight and volume, ability to resist large accelerations or shocks during launching or landing, and suitability for servicing by technicians located anywhere in the world, and it becomes apparent that the designer is faced with a most challenging problem!

Besides the more general aircraft and missile equipment problems indicated above, the flight control component designer has several other difficult areas he must consider. A large number of these are unique to aircraft, though some are present in many other high-performance servosystems. A partial listing of these special problems is presented below.

(1) The frequency response of the sensing and actuating elements must be properly tailored from an overall system standpoint. For a high-performance flight control system the following points should be considered.

(*a*) From airframe rigid-body considerations, the actuating servomechanism and sensing elements should be fairly well damped and should possess response characteristics having natural frequencies several times that of the highest-frequency airframe mode to be controlled. The primary reason for providing such a large margin is to minimize the effects upon overall system performance of changes occurring in the airframe motion characteristics as the flight condition is changed. If these frequency responses are not tailored in this fashion, there may be severe and complex requirements placed upon elements that change the gains and equalizations of the flight control equipment as functions of flight condition.

(*b*) To avoid inadvertent coupling with the airframe motions caused by its lack of rigidity (the so-called aeroelastic modes), any resonant peaks in the positional servo or sensing element frequency responses should be well damped, well separated from the airframe nonrigid-body modes, and properly placed in frequency relative to these and other overall system high-frequency characteristics.

(2) On positional servomechanisms used directly to deflect an airframe control surface or reaction control, the following more or less steady-state airframe characteristics create important servomechanism requirements.

(*a*) *Maximum control deflections* for airframe control, which are determined by considering the full range of flight conditions and airframe configuration variables for all the required maneuvers.

(*b*) *Maximum control speeds* sufficiently great to allow rapid airframe response to command inputs or rapid recovery from the most severe disturbances.

(*c*) *Maximum actuator hinge moments,* which must be large enough to provide adequate dynamic and steady-state (trim) control of the craft.

(3) The nonlinear characteristics of the flight control equipment require special consideration. Nonlinearities of a threshold and hysteresis nature are of great importance from the standpoint of limit cycles and control sensitivity. Deliberate introduction of limiting-type nonlinearities, such as torque, velocity, or motion limiting on the servo actuator, are often required for safety reasons or to avoid excessive power supply weight and volume. As a matter of practical experience, it can possibly be stated that nonlinearities of one form or another are the major headaches of the system engineer in the development phase.

12.4 TYPICAL SENSING ELEMENTS

The automatic control of any process requires suitable sensing devices to measure either the output variables, or quantities which are functions of the output variables, and to convert these measurements into usable signals for control. In the flight control situation, the sensing elements of importance are instruments capable of measuring functions of the aircraft's motions. This section will be concerned with the problem of sensing the airframe output motions requiring control; it will include a brief general description of the principal characteristics and operations of typical sensing elements.

The airframe angles, velocities, and accelerations which it may be desirable to sense for particular applications are summarized in Table 12.1 (Sec. 12.2). The ability to pick up these quantities will be discussed for each of the sensors considered.

12.4.1 Rate Gyros

A rate gyro is a gyroscopic element restrained to move in only one degree of freedom (Fig. 12.5). The single gimbal houses a high-speed rotor and is restrained by a spring or equivalent device. When the gyro case is rotated about the sensitive axis (normal to the plane

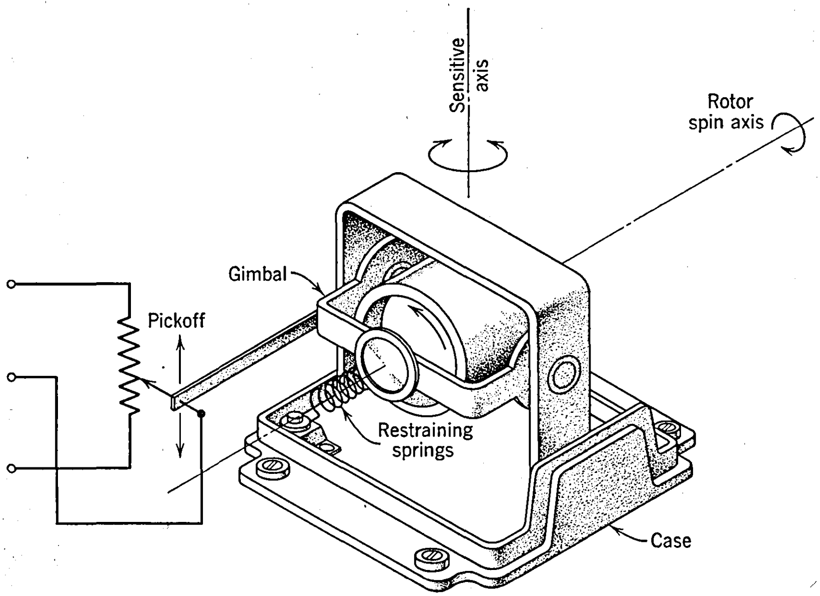


Fig. 12.5 Typical rate gyro

of the gimbal ring), the high-speed rotor acting as the gyroscopic element creates a precessional torque which is proportional to that portion of the case angular velocity normal to the plane of the gimbal. The gimbal ring then deflects until the spring-restraining torque exactly balances the gyroscopic torque. In the steady state this gimbal deflection is proportional to the angular velocity about the input sensitive axis. A suitable pickup may be placed between the gimbal and the case to measure this deflection. Since an applied angular velocity measured normal to the base of the gyro case is approximately equal to that normal to the plane of the gimbal (a small error is present due to angling of the gimbal ring), properly mounted rate gyros can be used for the direct measurement of angular velocity about any given airframe axis.

By adding a damping restraint on the gimbal ring, the rate gyro damping may be increased. Further, by modifying the form of the restraint, the gimbal deflection can be made a different function of the input angular velocity; for example, by using only a damper, the gimbal deflection becomes proportional to the integral of the input angular velocity.

When applying a rate gyro to the automatic control of aircraft, it should be remembered that friction and other nonlinearities give rise to a threshold sensitivity within which no rate is detected. Further, the rate measured by the instrument is that local one normal to its gimbal plane, so location in the airframe is an important factor, particularly in the less rigid high-performance craft of today and the future.

12.4.2 Amount Gyros

An amount gyro consists of a rotor spinning at high speed and mounted in a gimbal system intended to give complete angular freedom (Fig. 12.6). In a perfect gyro any motion of the case causes the gimbal system to move in a fashion which maintains the direction of the spin axis fixed in space. Suitable pickups arranged between the inner and outer gimbal and the outer gimbal and the case allow the measurement of particular airframe angles. For example, if the gyro shown in Fig. 12.6 were mounted in an airframe, in straight and level flight, so that the outer gimbal bearing axis and rotor spin axis were aligned with the airframe's X and Z axis, respectively, then pickup 1 would measure pitch angle and pickup 2 would measure bank angle. A similar setup could be used to measure yaw angle or to establish a heading reference.

In the practical gyro, inaccuracies in properly locating the center of gravity of the gimbal system, friction in the pivots, and the departure of the gimbal rings from a mutually perpendicular relationship cause difficulties which tend to give undesirable readings over a long period of time. Further, the datum supplied by the amount gyro (the spin axis) is one fixed in space, although an earth reference is more desirable. On the other hand, the gyro readings are highly accurate over a short period of time.

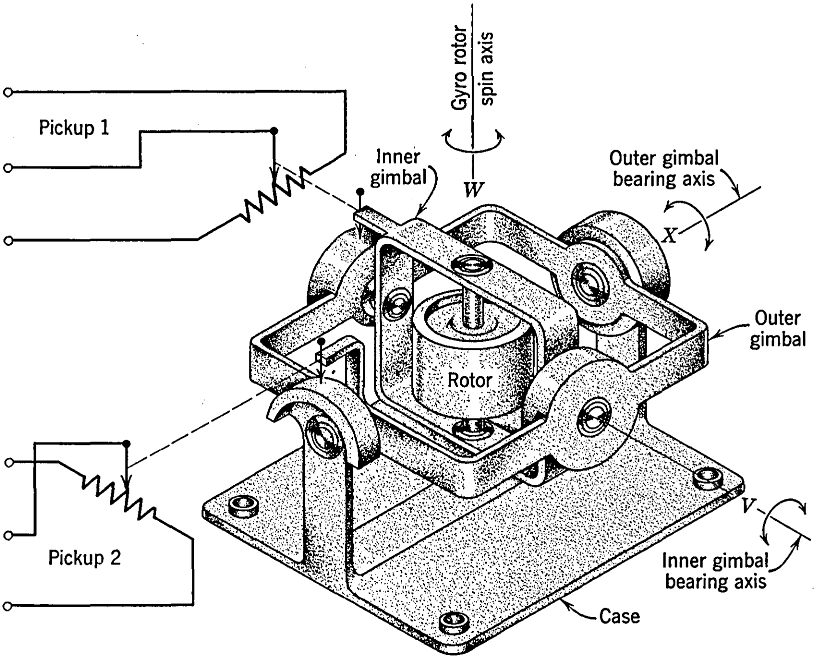


Fig. 12.6 Vertical amount gyro

To alleviate these difficulties and take advantage of the desirable features of amount gyros, a weak monitoring system may be used in conjunction with torque motors applying torques to the gimbal rings. (A torque applied to one gimbal causes precession of the other, similar to the action of a rate gyro.) To obtain a vertical gyro the usual method is to attach the gyro-torqueing (erection) system to a two-degrees-of-freedom pendulum or accelerometer; a directional reference can be established by utilizing a magnetic field detector for the monitor. The gyro systems then act as extremely low-frequency mechanical filters.

When the aircraft is in other than straight and level flight, the angles measured by the gyros are no longer proportional to simple integrals of the aircraft angular rates. This condition often requires complicated interaxis connections to allow the performance of even such simple maneuvers as a coordinated turn.

12.4.3 Accelerometers, or Force Pickups

An accelerometer is essentially a spring- and damper-restrained mass of one form or another, such as that shown in Fig. 12.7. Other typical examples used in aircraft control work are pendulums and spirit levels. If the natural frequency of the sensing device is much higher than the frequency content of the input, the displacement of the mass is proportional to the acceleration applied along the sensitive axis. The same device can be used as either a displacement or velocity pickup for inputs with different frequency contents. Probably the simplest description is that the device measures a function of the applied force. For flight control purposes a force pickup can be used for the measurement of side, forward, or vertical accelerations acting at various locations in the airframe. Since these quantities are proportional to functions of the applied air forces, in those conditions

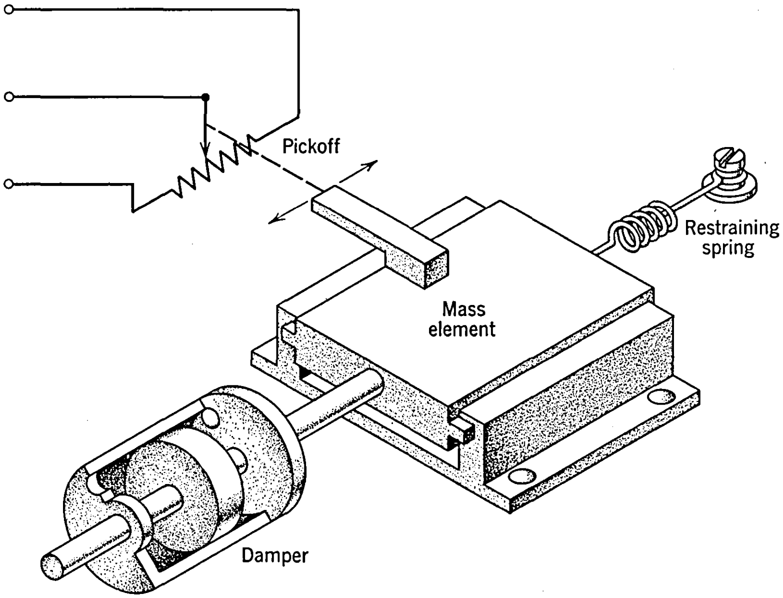


Fig. 12.7 Typical accelerometer

where the air forces are a dominant function of a particular airframe motion variable, the force pickup can give a reading which is an approximate function of that dominant airframe motion variable. For example, a side accelerometer mounted at the lateral center of percussion gives a reading which is a strong function of the side force on the aircraft due to its side velocity and, hence, is a fair indicator of sideslip angle. Indeed, the familiar "ball" of the bank and turn indicator has long been considered by pilots to be a measure of sideslip.

12.4.4 Local-Flow Direction Detectors

To measure directly such quantities as angle of attack and sideslip it is necessary to measure the direction of the relative wind. Since the presence of the airplane seriously affects the air flow, the best that can be done normally is to detect the direction of the local flow at some appropriate spot on the aircraft and then compensate the in-

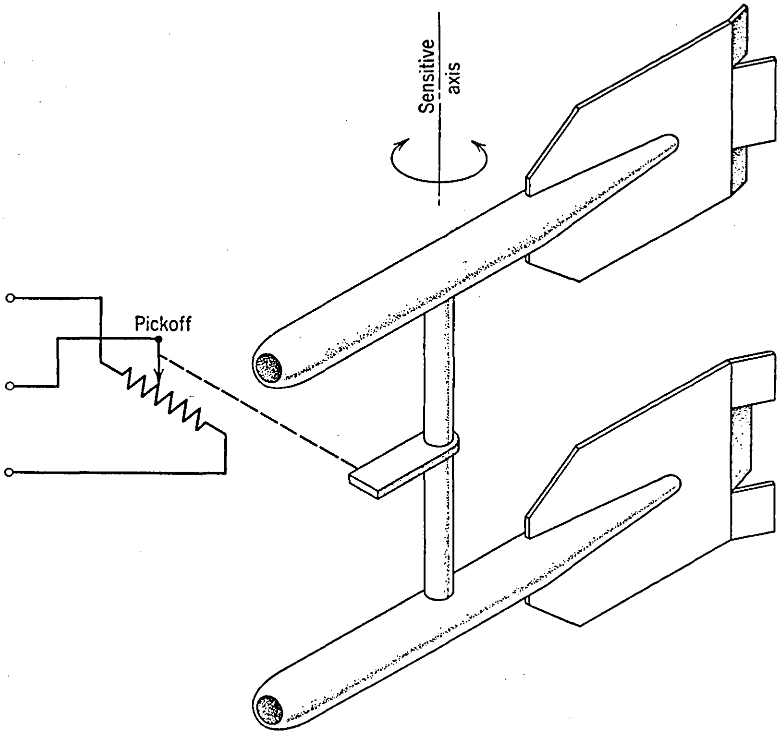


Fig. 12.8 Vane

strument reading to obtain an approximation to the desired quantity. These direct means, such as vanes (see Fig. 12.8), probes, dual-pressure pickups, etc., all suffer from two basic faults.

(1) The readings must be extensively calibrated, or averaged over a number of installed units, to obtain the desired results.

(2) The instruments must be mounted externally to the airframe, or at least require external connecting tubes.

Because of these disadvantages this type of device has been used only for special applications, though they are becoming more popular as system configurations depart more and more from the rather staid, relief, autopilot variety.

12.4.5 Local-Flow Magnitude Detectors

The types of instruments which may be used to measure the magnitude of the local air pressures, both ram and static, are very important in many flight control problems. From the magnitudes of the local pressures, properly compensated sensing elements can give indications of air speed, altitude, and Mach number. These indications are used both directly for control and occasionally as input elements for gain-compensating devices. This latter use is becoming more and more important as aircraft operate over a wider range of speeds and altitudes, thereby requiring, for an optimum airframe-controller system, automatic adjustment of the controller characteristics to compensate for changes occurring in the airframe motions as flight condition changes.

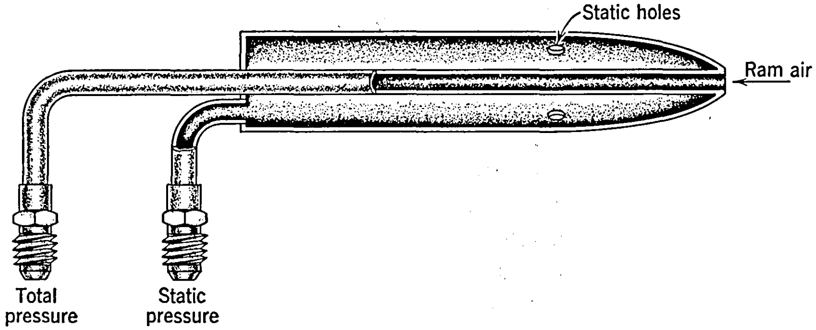


Fig. 12.9 Pitot tube

The usual type of local-pressure instrument consists of a pitot tube installed externally to the airframe, the sensing instrument usually consisting of a bellows-linkage combination and air lines connecting the source and the instrument. The pitot tube, shown typically in Fig. 12.9, is located at a favorable spot on the aircraft, is directed along the approximate X axis, and picks up both the local total and static pressures. Air lines connect the pitot head to the instrument. In a conventional altitude sensor, Fig. 12.10, only the static source is attached to the instrument. In a Machmeter, Fig. 12.11, both static and total pressures are required. Both instruments employ a bellows

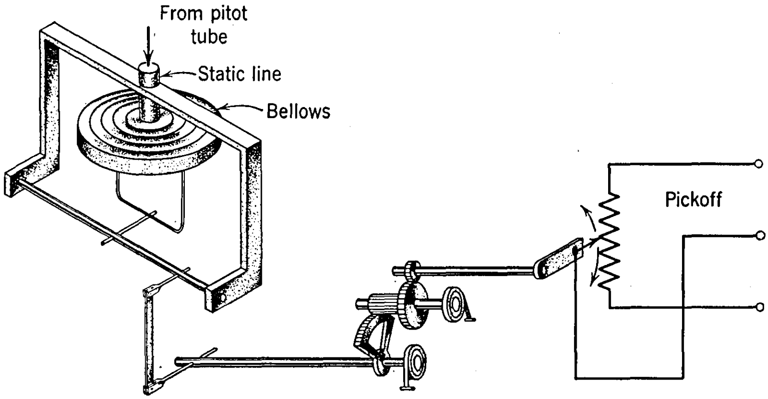


Fig. 12.10 Altitude control

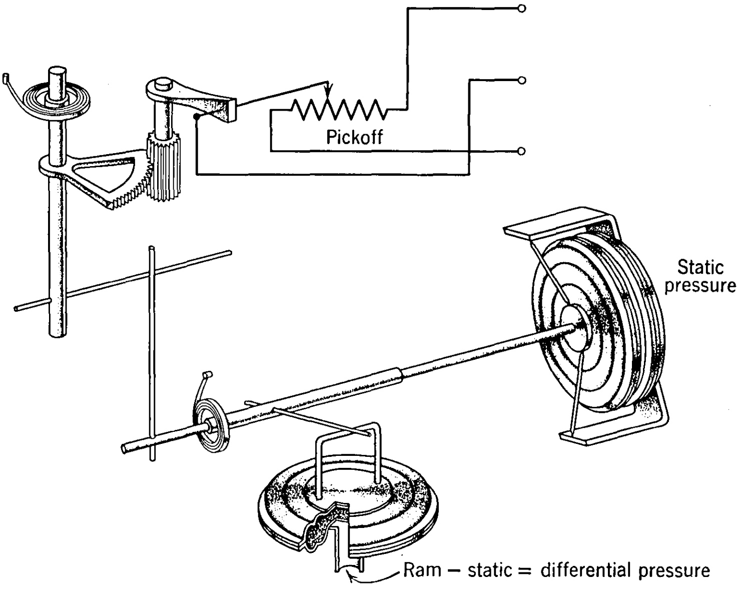


Fig. 12.11 Machmeter

system coupled with linkages to amplify the bellows displacement into a motion sufficiently large for convenient pickup devices.

Several other types of local-pressure devices are occasionally used, including indicated air speed and true air speed sensors. Some of these require nonlinear linkages and special compensation techniques to obtain the desired output.

Local-flow magnitude detectors have several sources of error which must be considered. A primary "error" is inherent in the positioning of the pitot head on the craft and the fact that only local instead of free-stream conditions are being sensed. If extreme accuracy is required in this regard, extensive flight calibration is necessary. The instruments themselves possess threshold and hysteresis nonlinearities which are sometimes troublesome, as are the transmission lags in the air lines connecting the pitot tube and the sensor.

12.4.6 Other Sensors Commonly Used

In the foregoing portions of this section the emphasis has been placed upon some of the more common sensors utilized in flight control work. A discussion of this nature would not be complete, however, without mentioning such devices as terrain clearance indicators, instrument landing approach systems, fire control and guidance computers, etc., which are often used to supply "command" signals into the flight control system. The major conceptual differences between "sensors" of this variety and the ones more thoroughly described above are that these devices are sensitive to other quantities, such as the motion of a target, as well as to the airframe's motions. Because of this, an additional outer loop is introduced which includes the automatic flight control system as an inner loop, or subsystem, of the overall system.

12.5 TYPICAL ACTUATING ELEMENTS

The surface-actuating systems on most flight control systems consist of high-performance positional servomechanisms. Electromechanical, hydraulic, and electrohydraulic actuators are all well represented in present systems, with the current trend being heavily toward the two latter varieties, since they more easily satisfy projected high-load, high-response requirements.

In many modern manned aircraft the flight control system actuation element consists of two separate devices, a surface actuator, usually of the fully powered, hydraulic type, and a controller actuator. The surface actuator is often a simple but high-performance valve-cylinder-

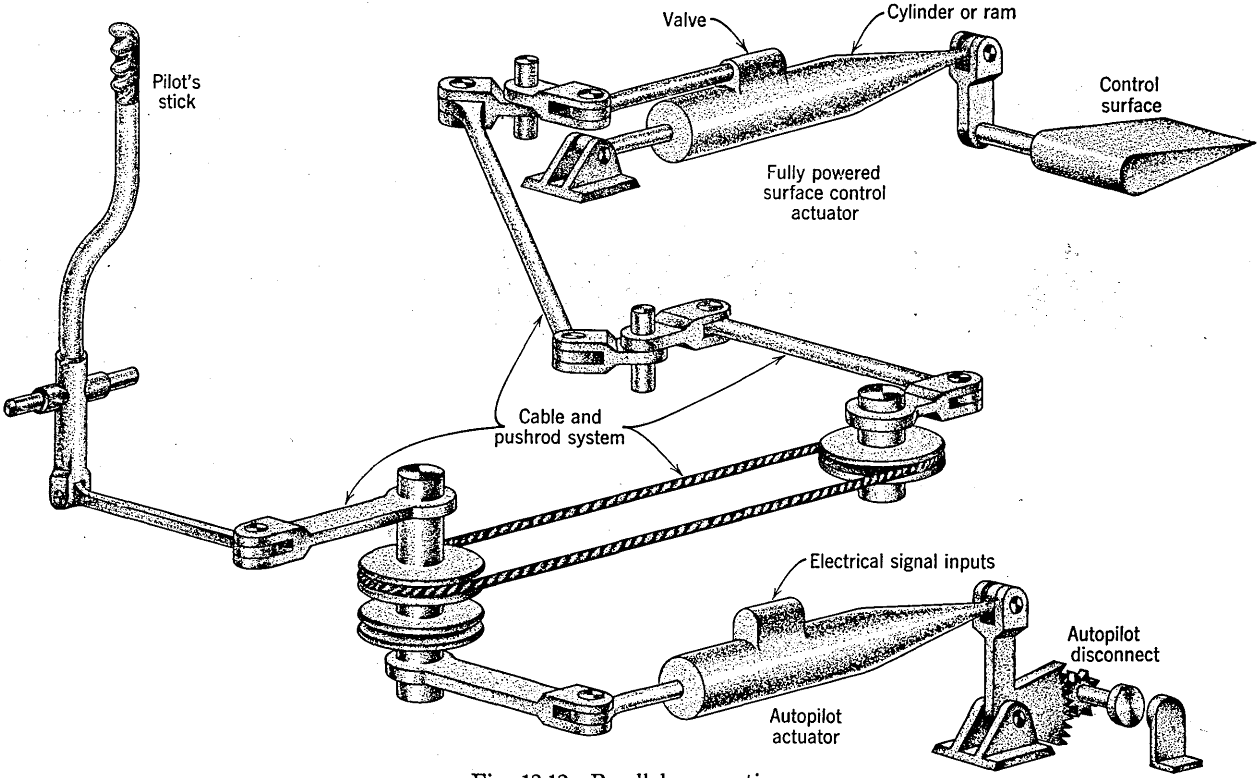


Fig. 12.12 Parallel connection

linkage combination, capable of amplifying pilot effort, or controller actuator effort, to the level required to control the surface position. The controller actuator is a much smaller servomechanism used to actuate a portion of the airframe's control system inboard of the valve of the surface actuator.

Controller actuators may be physically connected into a primary control system in two ways. The first, called a parallel connection, is the conventional installation for automatic pilot servomechanisms.

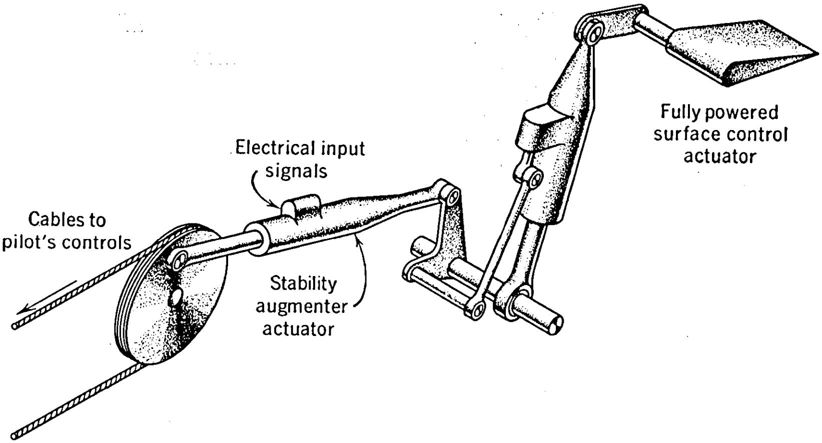


Fig. 12.13 Series connection

The second, called a series connection, is used intensively for stability augmenter applications. In the parallel connection, Fig. 12.12, the pilot and the control actuator operate against the same loads and, consequently, any movement of the actuator produces a corresponding movement in the pilot's cockpit controls. The series installation, Fig. 12.13, is so arranged that the fully powered, hydraulic surface actuator valve, and hence the surface, is moved without disturbing the pilot's controls.

The design of a manned-aircraft surface-actuating system is fairly complicated since it must be designed for both satisfactory airframe-pilot and airframe-automatic controller operation. Many elements, such as artificial feel devices, may be required for the pilot-airframe system whereas they may only complicate automatic operation. Certain other requirements, such as minimum backlash, may be less stringent for manual control than for automatic operation.

The rather obvious requirement for compatibility of the manual and automatic systems, and the unwanted duplication of system elements, such as actuators, has created a fertile field for the application of the systems engineering doctrine of integration. Considerable ingenuity has been and is continuing to be used in the detail design of integrated manual-automatic flight control systems, and in integrated components for these systems. Examples of equipment falling into the latter category are various multifunction actuators which combine into one structure the functions of the parallel (autopilot), series (stability augmenter), and surface actuators.

The actuating elements for missile applications are usually less complex than those for manned aircraft. In the majority of instances the controller actuator and the surface control actuator are the same unit, though on larger missiles the two may be separated to obtain sufficient output. The actuating system for a missile is often simpler to design than that of a manned aircraft, since no consideration need be given to a pilot. However, some missile applications, where a pilot would be unable to fly the missile, require extreme dynamic and load-carrying capabilities of the servomechanism.

12.6 EQUALIZATION AND AMPLIFYING ELEMENTS

The previous sections have considered the characteristics of various elements involved in automatic flight control systems. The components discussed have all been either relatively unalterable, or alterable only by choice between different items of the same general class. In particular, the airframe characteristics are essentially unalterable in any major way after a certain stage of the airframe's design. The sensors and actuators are alterable only by selection or development of another sensor or actuator. Consequently, although not as fixed as the airframe, the possible sensors and actuators are members of a narrow class of equipment.

The aim of any automatic control design is the integration of components into an optimum functional system. In the initial phases of the design, the designer must

(1) Compile and establish system requirements.

(2) Synthesize a system meeting the requirements; i.e., select and integrate components into the functional system. In this process, the designer must, to some extent at least,

(*a*) "Live with" the unalterable element.

(*b*) Select, or design, the best quasi-alterable elements for the particular job.

(*c*) Design equalization and amplifying elements to tie the unalterable and quasi-alterable elements into a well-integrated and functional system meeting the system requirements.

Nothing has been said, to this point, about the main alterable elements, the equalizing and amplifying devices. There is, of course, an excellent reason for this, since the characteristics of the equalizing and amplifying elements are almost completely determined by the system requirements and the characteristics of the unalterable and quasi-alterable elements. As explained in Chapter 3, a very large portion of the system design procedure is concerned with the problem of determining the required equalizer characteristics.

To determine the best estimate of the required equalizing and amplifying characteristics, a series of detailed studies are performed utilizing all of the more useful methods of experimenting with mathematical models, particularly root locus, frequency response, and analog computer techniques. Even in the earlier studies as much physical system information is used as possible, including all important nonlinearities. As physical equipment is built up, simulator techniques begin to carry a heavy load, ultimately leading to system tests with an analog computer being used to simulate the airframe and some sensors, with all other components being the actual physical parts. Fairly major changes can be expected in the detailed equalization and amplification requirements as these synthesis processes progress.

Equalizing and amplifying elements for automatic flight control systems have been made using a wide variety of techniques. The most common arrangements have involved vacuum tubes, although magnetic amplifiers have been used by some manufacturers.

Most manufacturers of automatic flight control systems have used vacuum-tube techniques involving 400-cycles-per-second signal circuits, with demodulators and passive networks and/or instrument servomechanisms to provide equalization. For applications requiring extensive and complex shaping, all d-c signal circuits are becoming more prevalent. In the latter case, the equalizing and amplifying elements are basically a permanently wired analog computer, with heavy emphasis being placed upon reducing the number of operational amplifiers by utilizing passive networks in both feedback and input positions. An example of an autopilot using this technique is shown in Fig. 12.14. Magnetic amplifiers have their major application in flight control systems as servo amplifiers, power supplies, and modulators.

Transistor replacements for vacuum tubes and magnetic amplifiers as well as special transistor designs are receiving a considerable amount of development effort at the present time and in the near future will undoubtedly be used a great deal. In addition to these developments, some efforts are being made to employ very different techniques, such as hydraulic equalizing and amplifying elements.

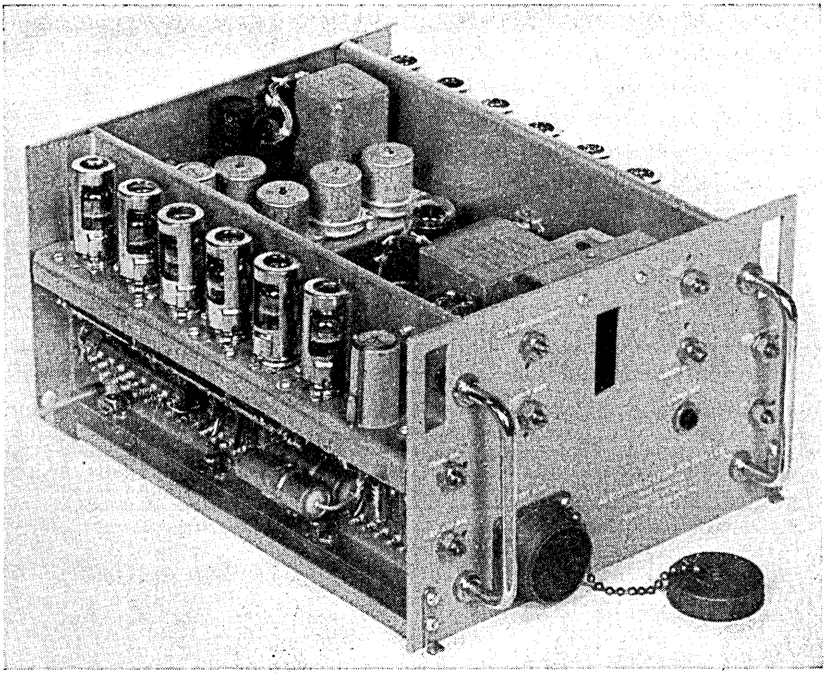


Fig. 12.14 Autopilot equipment (Courtesy Control Specialists, Inc.)

12.7 ILLUSTRATIVE FLIGHT CONTROL SYSTEMS

In this concluding section, the previously described components will be brought together to form two illustrative systems. The discussion on each will be confined to a presentation of the general situation confronting the system designer, a block diagram showing the general system configuration, and a discussion of its major features. In order, the systems to be discussed will be a sideslip stability augmenter for a manned aircraft, illustrating a single-axis controller, and a simple two-axis controller for a radio-controlled missile. Much more complicated missile and manned-aircraft automatic flight control systems are the rule today, but the systems shown are suitable for illustrative examples.

12.7.1 A Sideslip Stability Augmenter

In most high-speed jet aircraft the lateral oscillatory mode of motion (Dutch roll) is poorly damped. A small flight disturbance can excite this motion to the extent that the airframe oscillates almost continuously in a combination of rolling, yawing, and sideslipping motions.

Such undamped oscillations are undesirable from the standpoint of pilot comfort and safety and are to be avoided or eliminated wherever possible. If the oscillations occur only at fairly low frequencies, the pilot may be able to control the motion, but only by extreme concentration. If they occur at higher frequencies, his reactions will not be sufficiently fast to damp the oscillation, and if he attempts to do so he may actually cause the motion to become divergent.

In order to eliminate such deficiency in handling qualities, it is necessary to increase the Dutch roll damping. Theoretically this can be done for an airplane still in the design stage by increasing the fin size or some similar structural change. Actually such changes are generally impractical from the standpoint of their effect on airplane performance and efficiency, and some "artificial" means of improving the damping must be used.

The mechanization of most aircraft fire control computers assumes that the aircraft sideslip angle is zero. Therefore, when flying an attack course, the pilot must continuously divert his attention from the target and glance at the ball of his turn and bank indicator to make sure that he is coordinating all three surface controls during the maneuver.

A system providing aileron to rudder cross-feed, so that a perfectly coordinated two-control attack course can be flown, is therefore desirable to increase the effectiveness of the pilot-airframe combination.

A sideslip stability augmenter system, as shown in Fig. 12.15, automatically damps Dutch roll, provides two-control maneuverability, and eliminates sideslip. The sideslip stability augmenter system accomplishes this through components which perform the same function as does the pilot when he is centering the ball of his turn and bank indicator. The major difference between the augmenter and pilot is the greatly increased sensitivity and speed of response of the augmenter system.

An accelerometer senses the lateral acceleration of the airframe, as does the ball of the turn and bank indicator, and produces an elec-

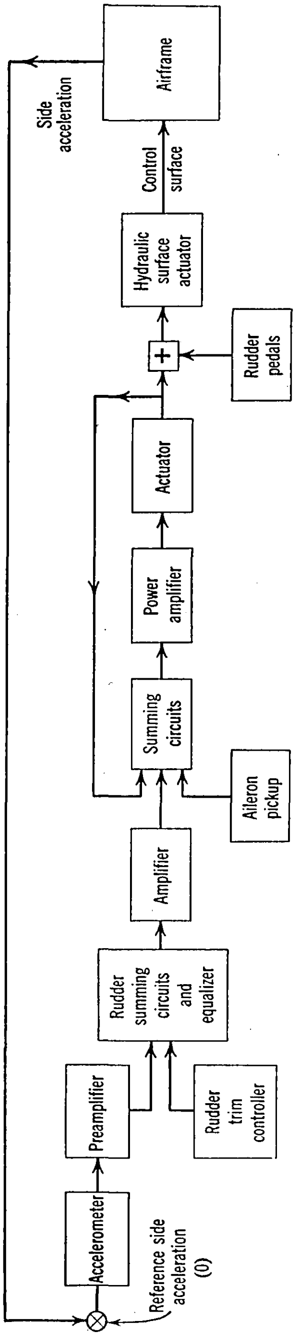


Fig. 12.15 A sideslip stability augmenter

trical signal proportioned to this acceleration. An aileron potentiometer senses the amount of aileron deflection and provides an electrical signal to the rudder servomechanism proportional to this aileron deflection. A compensation and equalization network provides the required judgment for the correct amount of surface displacement necessary for the given flight configuration. A servomechanism translates these signals into mechanical movement which causes the rudder to be deflected sufficiently to keep the ball of the turn and bank indicator centered.

Thus, sideslip, Dutch roll, and the problem of coordinating turning maneuvers are eliminated and the airplane becomes essentially a two-control (aileron and elevator) system.

12.7.2 A Two-Axis Control System for a Radio-Controlled Missile

In many missile applications exceptional airframe-controller dynamic performance is not required; the ability to launch the craft and then fly it in a straight line at constant altitude and to introduce climb and dive and left- and right-turn commands are the major requirements. In these instances an autopilot actuating only two axes is often acceptable. The selection of the elevator for longitudinal control is obvious. Since launch performance against asymmetrical winds is best with a tight roll control, the aileron is the selection for lateral control.

With a stable Dutch roll mode, the chief lateral-control problems are to (1) obtain a tight control of roll in the presence of asymmetrical winds; (2)

stabilize the spiral mode of the airframe; (3) provide for turning the aircraft when commanded via a radio link. A simple and straightforward solution to these lateral-control problems is shown by the block diagram of Fig. 12.16.

In this system a vertical amount gyro and a roll rate gyro are used as basic sensors to provide the basis of a high-gain roll loop. With proper system adjustment, the roll axis of the craft is tightly stabilized against the effects of disturbances. A simple relay setup can be used to introduce a roll angle command resulting in a constant roll angle and hence a steady turn rate.

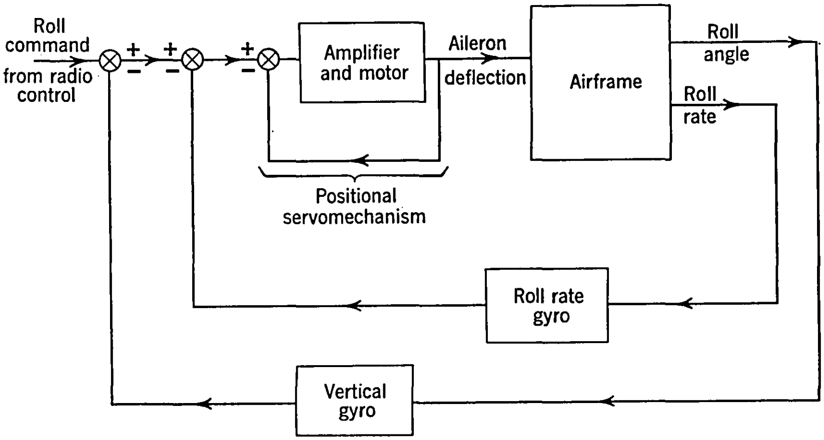


Fig. 12.16 Lateral-control block diagram

In the longitudinal axis the major system requirements are even simpler than those of the lateral axis. Here the objectives are to (1) stabilize the attitude of the craft; (2) provide for climbing and diving the aircraft when commanded via the radio link. If the short-period and phugoid motions are both stable and well separated in frequency, the pitch signal from the vertical amount gyro is often all that is required. If the airframe is more contrary, a pitch rate gyro is added for stabilization at the higher frequencies, e.g., increasing the damping of the short-period mode. This latter system is shown in Fig. 12.17. The configuration of the longitudinal system shown is similar to that of the lateral controller, and climb-dive commands can be introduced in an analogous fashion.

The simple missile control and stabilization system discussed here has a minimum of equipment and is capable, for some airframe configurations, of performing the desired tasks without additional accessories. As the aircraft flight regime is broadened, gain changers and command limiters variable with flight conditions become necessary. In some instances it may even be found that the entire system layout requires revision for satisfactory operation between the entire flight boundaries. For example, a very unstable Dutch roll mode would usually require the addition of a rudder axis controller, with additional necessary complications being required for the proper introduction of turn commands.

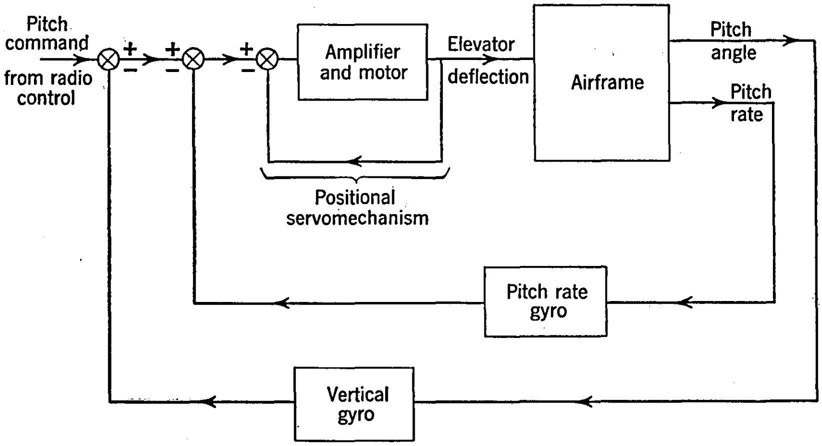


Fig. 12.17 Longitudinal control block diagram

12.8 REFERENCES

1. William Bollay, Aerodynamic Stability and Automatic Control, *J. Aeronaut. Sci.*, *18*, No. 9, Sept. 1951.
2. H. R. Hopkin, Automatic Control of Aircraft, *Trans. Soc. Instrument Technol.*, *5*, No. 2, June 1953.
3. C. S. Draper, Flight Control, *J. Roy. Aeronaut. Soc.*, July 1955.
4. Dunstan Graham and Richard Lathrop, Automatic Feedback Control and All-Weather Flying, *Aeronaut. Eng. Rev.*, *14*, No. 10, Oct. 1955.
5. Northrop Aircraft, Inc., Methods of Analysis and Synthesis of Piloted Aircraft Flight Control Systems, *BuAer Rept. AE-61-4I*, March 1952.
6. Northrop Aircraft, Inc., Dynamics of the Airframe, *BuAer Rept.*, *AE-61-4II*, Sept. 1952.
7. Northrop Aircraft, Inc., The Hydraulics System, *BuAer Rept. AE-61-4IV*, March 1953.

13. Automatic Production of Electronic Equipment

L. K. LEE

General Mills, Inc.
Minneapolis, Minnesota

13.1 INTRODUCTION

We have all seen significant progress in automation made by the oil, chemical and similar fluid-handling industries, but there has been relatively little progress made in the fabrication industries. In general, the processing industries have been able to develop continuous processing techniques and thereby automate the entire industry; but the fabrication industries have been using batch production methods and have lagged quite a bit behind in the development of automatic fabrication techniques. It was not until well after the second World War that much effort was expended by the electronics industry in developing automatic production techniques for the production of electronic equipment.

The electronics industry, which has aided in the mechanization or automation of so many industries, is in the role, we might say, of the cobbler's son who goes barefoot. The recent concentration of effort to develop automatic production techniques to reduce costs and increase the availability of electronic equipment is, of course, the result of the increased importance that electronics occupies in today's economy.

The Armed Forces, with its dependence on electronics for guided missiles, proximity fuses, communications, automatic fire control, and endless other applications, would, in an all-out national emergency, demand more production than present facilities can provide. A figure of at least sixty billion dollars a year has been estimated. This is ten times our present production rate, and we shall no longer have two and a half years to reach our maximum production rate as we did during World War II; hence, the military's interest in automation stems from the requirement of immediate availability of equipment without the disadvantage of storing large quantities of equipment which may become obsolete before they are ever used.

The stringent military requirements for reliability and performance under adverse conditions have created a major problem in respect to design and production. In addition, the reduction in size of electronic assemblies has required more precise placement of components and much more skill during many stages of the fabrication process. The limitations of the human assembler can become a very important factor in determining the availability of the product. Not only are we facing a situation in which the demand for electronics is increasing, but, by comparison, the complexity of electronic equipment is becoming even greater. In the military field, a B-36 depends upon about 2000 vacuum tubes for its navigation and control. The F-86 has about 600 vacuum tubes. Interceptors equipped for all-weather flying and landing require even more than 600 tubes. A modern battleship uses nearly 9000 vacuum tubes. Another factor contributing to the expansion of the electronics industry is its habit of infiltration into new fields such as biophysics, and many others; for example, the Atomic Energy Commission alone spent over $50,000,000 for essential electronic controls in 1950 and 1951.

The growth of the civilian market has been just as phenomenal. Ten years ago, television was practically nonexistent; last year, approximately seven million sets were sold for about two billion dollars. A survey of the electronic data-processing market shows that in less than ten years it will become a billion-dollar-a-year business. Electronics has just begun to make itself felt in the field of industry control.

Faced with this future, the industry also is coping with increased labor problems and the need for more efficient utilization of its present plant layout. The complexity of electronic equipment has increased the cost of training personnel, the incidence of rejects on the production line, and maintenance problems in the field. The requirements for the production of electronic equipment have far outgrown the early stage when it was simply a matter of designing a circuit and soldering component leads together. Today, equipment fabrication requires a closely integrated procedure of materials selection and processing, mechanical and thermal design, and correct component application.

One answer to providing quantity and quality at minimum costs appears to be the application of automatic production techniques. Aside from reducing costs from acceleration of production rates, automation of the production line would (1) improve and simplify quality control by maintaining uniformity of products; (2) reduce manpower requirements, particularly the skilled workers needed to put the assemblies together; (3) produce around the clock; (4) aid miniaturization by reducing tolerance and allowing components to be placed closer together, since all components and leads are in precise orientation and location; (5) reduce production time lag; (6) conserve materials and increase effective production by minimizing rejects and waste of materials.

Once management decides that automation may be a possible solution to gaining or keeping a share of the market, this decision influences the entire mode of operation of the organization from top to bottom. Much tighter coordination must be maintained, for now management decisions are much more far-reaching and much less flexible. With the advent of automation, the three areas of production—product design, materials processing, production machinery employed—become even more closely interrelated and must be considered as a single mechanized production system with the closest working harmony maintained between the three areas. The experience of one fairly large manufacturing company shows that in order to get the necessary coordination between the various departments, the president of the organization had to set up a facilities committee consisting of all department heads affected to study the market for its various products; then the committee related expected profits to the capital investment necessary to produce these products by an automatic production line. This committee studied general plant design and procurement of facilities and all phases of manufacturing from market surveys, styling trends, product design and development, production scheduling, to warehousing.

This steering committee found it was necessary to set up a separate working group consisting of styling specialists, research personnel, product development engineers, methods and materials people, tests and specifications engineers, and procurement and scheduling personnel. This "super" engineering department, we might say, had control of the tooling budget, which meant that it had cognizance of production engineering and could control the purchase of equipment and production machinery. We do not usually think of procurement and scheduling as an engineering function, but when an automatic production line is tooled up, it becomes a catastrophe if only a single part is missing; the entire line must be shut down because an automatic production line does not have the flexibility of partially completing an assembly while waiting for the missing part. Even if this were possible, the rates of production usually associated with the automatic production line are such that it would be completely impractical to warehouse the partially completed assemblies. One large television manufacturer mentions that during a good week their mechanized production line, which is still rather crude compared to the automatic lines used in other industries, could produce 20,000 sets. Thus it becomes necessary to ship the products directly off the end of the production line; otherwise there would be acres and acres of television sets.

Perhaps the most difficult part of thinking in terms of automatic production is recognition of the fact that the product or the fabrication process, or both, must be completely changed in order to make automatic production feasible. It appears that the most productive approach is to observe the function of the product and attempt to build a product that will perform this function by the use of automatic production machinery.

Before automatic production of electronic equipment becomes feasible, it is necessary to establish standards for the fabrication processes, the design and construction of the product, and the materials used. The machine designer must know what sizes and configurations of components the various machines must handle. When the product designer designs electronic equipment, he must know the range of capabilities of the various fabrication machines. The components manufacturer must know how to design his components so that they will be acceptable to the electronic-equipment manufacturer who uses automatic machinery. Organizations such as the Radio-Electronic Television Manufacturers Association are struggling with the problems of providing standards for the industry. The machine-tool manufacturer, electronic-components manufacturer, materials supplier, and the electronic-equipment assembler are trying to arrive at suitable compromises on standards for (1) manufacturing tolerances for component lead and body dimensions; (2) indexing mechanisms; (3) materials for printed circuit plates; (4) circuit layouts; (5) automatic inspection and test methods; and a multitude of other considerations.

It appears that the synthesis of an automatic production system can best be achieved by applying the same means used in building modern electronic equipment, such as digital computers; i.e., by using unitized functional packages as building blocks. The automatic production line or system can then be composed of functional unit machines interconnected by work-handling systems. The basic requirements for a practical approach to the development of an automatic production system to meet the economic and technical requirements may be summarized as follows.

(1) The functional machine units must be suitable for insertion into existing production lines gradually in the evolutionary process toward complete automation; i.e., building blocks for the automatic line must be versatile machine tools.

(2) The functional machine units must have sufficient flexibility so that new materials and components may be used, e.g., the substitution of a printed resistor for a prefabricated component.

(3) The production line must be flexible in handling a range of work and materials so that there can be rapid conversion of the line from the manufacture of one package to another; that is, the automatic line should be able to produce subassemblies of radios, television sets, and public-address systems, as well as walkie-talkie sets and radar and fire control equipment, with a minimum of retooling and shutdown.

(4) The production system must have sufficient versatility to accommodate a wide range of production lots economically, e.g., 100 radar intermediate-frequency strips to 100,000 table-model radios.

(5) A production system must have the maximum of reliability.

(6) The controls for the production line must be sufficient to insure high-quality performance of the product.

(7) The economic considerations relating the first six requirements to the cost of production must be sufficiently attractive to develop managerial interest.

13.2 APPROACHES TO AUTOMATION BY THE ELECTRONICS INDUSTRY

In the past the electronics industry has been very reluctant to automate its operations. Since the electronic-equipment manufacturing industry is one of low capitalization, comparatively speaking, it has been very reluctant to exchange girls, soldering irons, and wooden benches for expensive and complicated machinery. With the coming of age of the electronics industry, it is just beginning to harken to the possibilities of automatic production techniques. So far, in the automatic production of electronic assemblies, no one method of applying circuits or components has shown overwhelming advantages over others; but on one point there is agreement—practically all approaches to automatic production use as a basis the graphic-arts processes such as etching, embossing, stenciling, plating, and others for the fabrication of the conductive patterns and resistive elements. In general, there are two distinct approaches to the automatic production of electronic assemblies used in radio and television, automatic control equipment, and radar, which make up 60 per cent of all electronic equipment. These electronic assemblies handle signal level or the low-power range, direct current to 100 megacycles, and require low voltage. One approach is taken by the equipment assembler and the other by the component manufacturer.

The main emphasis of equipment assemblers has been on eliminating the wiring operations. Since manual wiring represents a large portion of the assembly costs and becomes an intricate and delicate operation for miniaturized equipment, it is not surprising to find that the first attempts at mechanization are toward removing this operation. A patent was issued in 1929 for a dip-soldering process. The component manufacturer has, in general, used his knowledge of processing raw materials such as carbon, resin, ceramics, and many others to form circuit elements and package them into an electronic subassembly as part of the assembly operation. Thus, silver paste is stenciled on a ceramic material, then fired to form the conductive patterns, and carbon and resin mixtures are stenciled and baked to form the resistive elements.

The equipment assembler's approach to automatic production, that is, mechanized or manual insertion of prefabricated components into a printed circuit board, has the advantage of versatility of product design and application in that a wide variety of tried and true components are available. It has the additional advantage of requiring the minimum amount of equipment to start automatic production, for individual machines may be inserted into the manual production line whenever it becomes economical to do so. The line control and quality control requirements are relatively simple. In many cases, these line controls are no more than interlock switches and relays to insure that the component is correctly inserted before the circuit board is moved on to the next station. Although some systems claiming more versatility do require complex control equipment, it is doubtful that very many production lines to build electronic assemblies would require complex computer and programming systems. Either the simplicity of the operations or the complexity of the production equipment required to use such computers makes it uneconomical for most commercial applications. The manufacturer must be careful not to overdo in using electronics; simplicity and economy must be the watchwords.

Preliminary economic evaluations of this first approach have shown that the assembly cost has been reduced by 60 to 80 per cent compared to the manually assembled unit. This represents a maximum savings of 24 per cent of the manufacturing costs since the materials cost of the assembly is approximately 70 per cent of the total cost. Thus, the savings effected by the mechanized insertion of prefabricated components onto printed circuit boards is only a good start toward the savings that will be ultimately realizable. In some cases, the increased cost of materials practically makes up for the savings effected in assembly.

To make further savings, we must automate the fabrication of these electronic components which make up most of the costs for the manufacturer. The problem must be solved by a basically new approach to component fabrication rather than by automating present production machinery. The components engineer, the electronics engineer, and the machine-tool manufacturer, working together, have pointed the way by coming up with products that are equivalent or superior to the conventional components and are designed for automatic production. Some examples of this new approach are the stacked tubes designed so that the tube elements are essentially discs stacked one on top of the other between spacers. The disc-shaped tube elements can be fed from magazines which orient and index them for machine assembly. The sliced-coil technique for the fabrication of low-power transformers may revolutionize the transformer industry. No amount of automating present coil-winding machinery would have so simplified the winding of coils to the same degree as slicing cylindrical rolls of copper foil interleaved with an insulating sheet into thin spiral coils which can be machine assembled for the transformer coils. As each component is developed and proved satisfactory, it may be substituted for the component originally designed for manual assembly. Thus the advantage of using prefabricated components is maintained; that is, the component may be preselected and pretested before insertion. By such means as these just described, the costs for components can be greatly reduced.

The second approach of fabricating the electronic component as part of the assembly process appears to be a possible long-range approach for most electronic assemblies. Such fabrication methods as screening silver paste on a ceramic base for the conductor and carbon mixtures for the resistive elements have the advantage of low materials costs and simplification of the production processes; also, individual adjustments can be made as the component is fabricated. Materials in critical shortage are not required. For the present, adequate production techniques must still be worked out, and the selection of components is limited. These production techniques are now being used primarily in consumer products where production tolerances and operating environments are not too stringent. The development of more stable components will no doubt extend the use of these types of printed circuits.

As the electronics industry has matured it has come to recognize its responsibility for providing reliable service. In fact, the industry's future growth depends on this. Thus, all new components and techniques should be closely scrutinized. This brings up the most critical of problems in automatic production, that of quality control and production testing or inspection. This is the heart and brain of the automatic line. We must develop methods for quickly and automatically checking the quality of materials or components in process and testing the performance of the completed electronic package. The difficulty of the problem lies not so much in the building of automatic test or control equipment, but in determining what physical characteristics are to be measured to provide a key to the reliability of performance of the component or the assembly. The RETMA and many manufacturers of test equipment are working on the problem of what to test and how to test the materials or electronic package while in process.

As we devote more effort to studying new materials and applying fruitful thinking unencumbered by conventional practices, it is not beyond the realm of reality that we may be able to form the circuit elements on an insulating material guided by machines which check each element and make the necessary adjustments to produce an electronic assembly that will perform reliably and have an operating life that is no longer measured in hours but in years. All this will be done in a production line untouched by human hands.

For the time being, let us give up our dreaming and return to reality to see what industry has been doing. At the present there are over twenty major programs on the development of automatic production techniques for electronic equipment. It is rather difficult to select just a few of these for description, but I have chosen the following on the basis given. The Sargrove automatic production machine was chosen because it is the first modern attempt toward the automatic production of electronic equipment, and we can learn quite a bit from the trials and tribulations that Mr. Sargrove had to endure. The work done at the Stanford Research Institute was selected because it is the most comprehensive program on the study of automatic production techniques and will serve to bring out the many problems associated with automatic production. The program at the General Electric Company will serve to illustrate the problems of versatility for small-lot production versus complexity of the machine. The assembly lines at General Mills, Inc., Mechanical Division and United Shoe Machinery Corporation illustrate the most simple and popular approach being taken now by many manufacturers, such as Admiral Radio, Sylvania Electric Products, and others now developing their own production machines. The Tinkertoy project at the Bureau of Standards illustrates the approach taken primarily by the components manufacturer. Among the other major pioneers in the field is Melpar, Inc., which is developing a component-attaching machine for small-lot production. This machine is not as complex as the General Electric equipment but more complex than the United Shoe equipment.

13.3 THE STANFORD RESEARCH INSTITUTE STUDY OF AUTOMATIC PRODUCTION TECHNIQUES

In spite of the intense interest of large segments of the electronic industry, each potential user of automation will approach this new venture with caution and will have to be convinced that the savings in time, money, and manpower and the improvement in quality of the product are really worth while. In the final analysis, the desirability of any automatic process is determined by its techno-economic advantages. This consideration was borne in mind during the planning of the program on investigation of automatic production techniques for electronic equipment sponsored by the Wright Air Development Center at the Stanford Research Institute. The general objectives of this program are:

(1) To develop interest and cooperation among the industries involved to solve the problems inherent in automating the production of electronic equipment.

(2) To help the electronics industry to recognize the applicability of the technological developments made in other industries such as materials processing and graphic arts.

(3) To determine requirements for an automatic production system that would minimize production time lag for military procurement and would also manufacture consumer products economically.

During the initial phase of this long-range program initiated in 1951, a questionnaire was sent to more than a hundred industrial concerns engaged in the manufacture of electronic components, materials processing, electronic equipment assembly and design, and the manufacture of machine tools, and to many major electronic-equipment users to determine the degree of automation in use and to review some of the problems. By the end of 1954, most of the military agencies concerned and over a hundred manufacturers were visited to discuss the problems of automating the production of electronic assemblies.

In the past, mechanization was achieved by determining the processes and materials used to manufacture a single product. Thus the possibility of rapid obsolescence deters general acceptance of automatic production techniques. In the approach taken by the Institute, special emphasis was given to those construction techniques which were adaptable to mechanized construction of a variety of consumer products and premium-quality equipment. In the attempt to establish technological feasibility supported by economic justification, all the elements of production must be studied as a system. The best approach to evaluating the related factors of fabrication method, materials and components used, and machinery employed is to choose a series of package designs or trial assemblies representative of widely used military and commercial equipment and consider the problems encountered in fabricating these by the hypothetical line. The configurations chosen consisting of thermionic tubes with associated resistors, transformers, capacitors, and relays not only represent radar intermediate-frequency strips, fire control video amplifiers, and walkie-talkie sets, but must also represent table-model radios, television sweep circuits, and industrial control equipment as well.

One of the very basic decisions affecting the electronic package design was the determination to use components now available or those having only minor modifications. This approach to automation is justified not only from the considerations of using simple machinery and making gradual changes, but also from considering the production tolerance and reliability of prefabricated electronic components now available. At the present state of the art of component manufacturing, it is necessary to prefabricate and inspect components before they are delivered to the automatic production line.

Although the emphasis of this program was primarily concerned with automating the fabrication or production operations, to achieve the truly automatic factory, equal emphasis must be put on automating the supporting functions which are concerned with data handling. This opens up the practically untouched field of mechanized data processing and paper handling for inventory control, materials dispatching, production scheduling, accounting, and a multitude of other monitoring activities.

Figure 13.1 shows the first trial assembly or product design to be evaluated. The circuits in this assembly may be considered as typical for low-power, video-frequency range, and for signal-level applications using only resistors and capacitors as circuit elements. Although there was no special effort made to miniaturize this assembly, it weighs less than three-quarters the weight of the conventional unit and its volume is about one-half that of the conventional unit. This package configuration was determined primarily by the method of air cooling and the use of miniature tubes. By arranging the tubes on top, the high-heat dissipation elements are effectively isolated from the rest of the assembly.

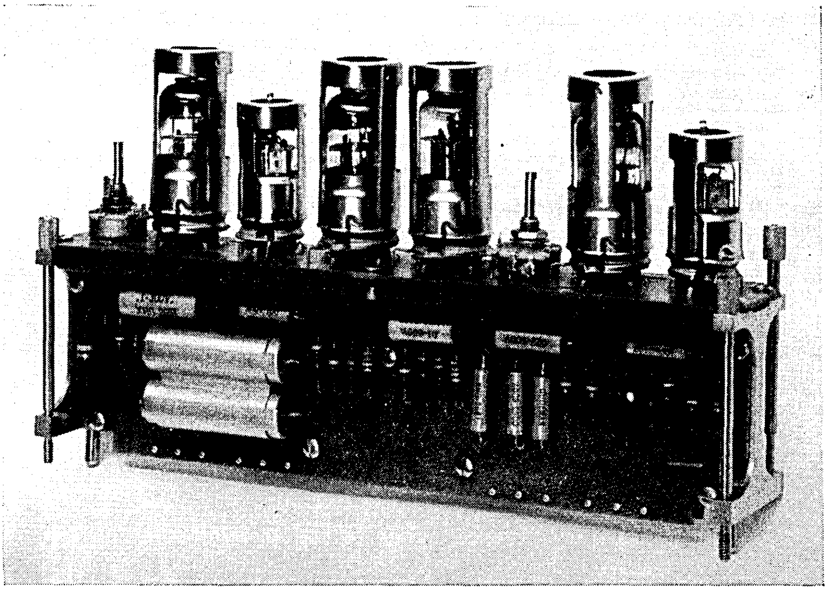


Fig. 13.1 First trial assembly or product design used to set up problems for the evaluation of an automatic production line—the design of the product, the handling of electronic components

The major aspects of a complete production line have been worked out on paper. Such a line would include preparation of etched or printed circuit boards, component attachment, soldering, assembly, testing, and packaging. The initial operation consists of cutting the copper foil and plastic laminate to form the baseplate; then holes for mounting components and indexing are punched. At this point the baseplate appears as shown in Fig. 13.2. An etchant resist of a wax and solder flux mixture dispersed in a volatile liquid is heated and sprayed onto the copper foil through a steel stencil held in place by a magnetic chuck. The steel stencil is cleaned and recirculated for

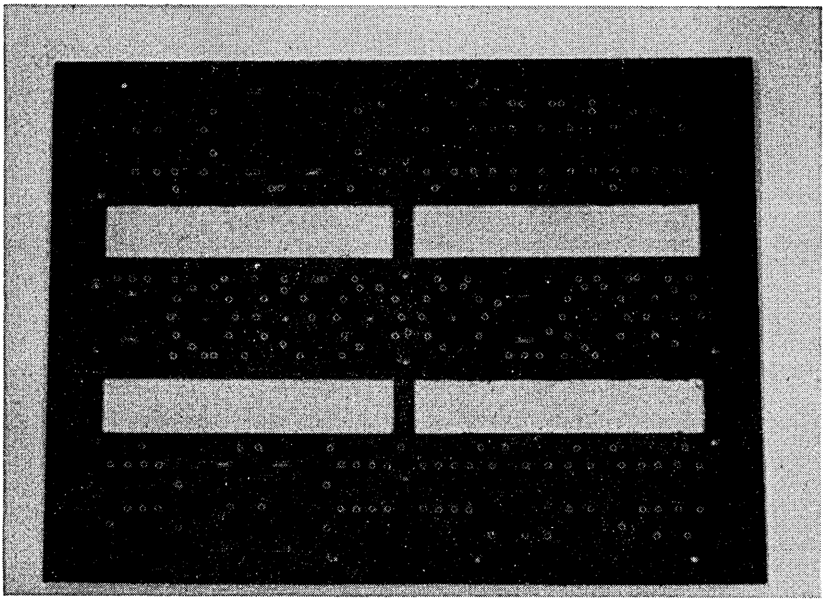


Fig. 13.2 Circuit baseplate with all holes for mounting components punched

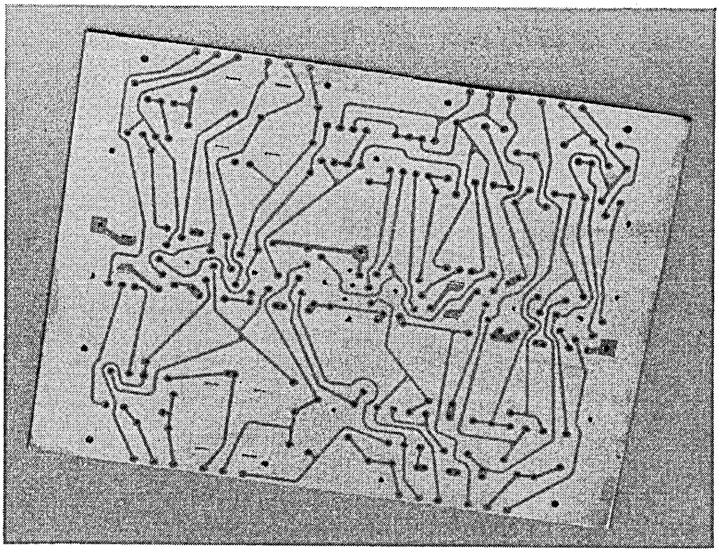


Fig. 13.3 Etched circuit board formed by the wax flux resist method

use again. The baseplate is drawn through the acid bath and the unprotected copper is dissolved. Then the baseplate is rinsed and passed through a neutralizing bath. At this point the circuit pattern is complete as shown by Fig. 13.3. Next we come to the heart of the automatic line, that of attaching the resistors and capacitors in the circuit. The components inserted by the automatic component-attaching machine (CAM) were tested 100 per cent and loaded in the magazines automatically; then the magazines were delivered to the component-attaching machine by a conveyor. Figure 13.4 shows a design of an attaching head. A trigger mechanism places the components in correct sequence on the component conveyor so that they are fed to the inserting forks in the right order. The component-attaching head trims and forms the leads, inserts them into position on the baseplate, and clinches the leads to hold the components in place at the rate of two components per second. After the base plate passes through the

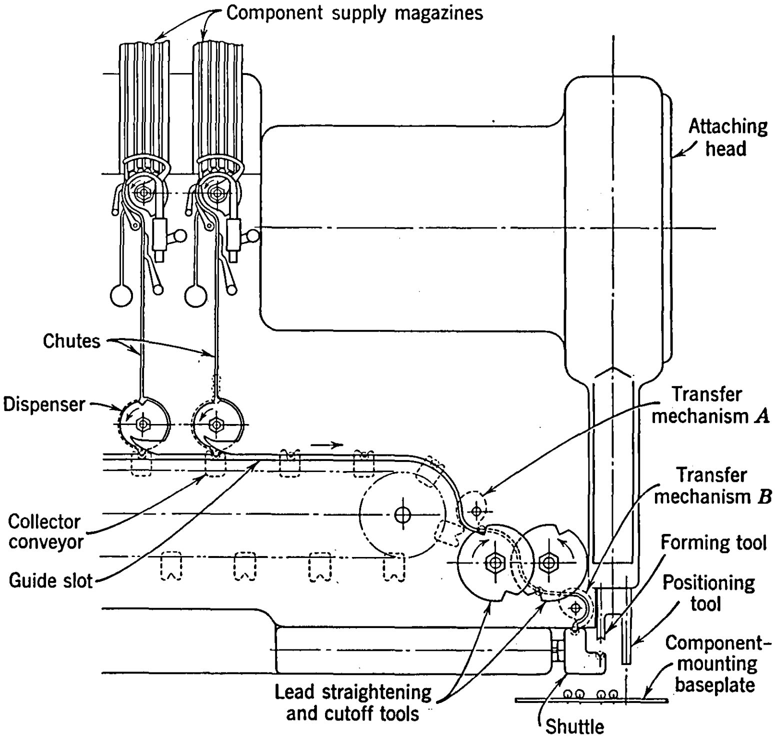


Fig. 13.4 One design of a component-attaching head which will insert a sequence of components on the circuit board

first several CAMs, all the resistors and capacitors are installed. The tube sockets and potentiometers are then riveted in place. When all the components have been installed, the baseplate is transferred to a dip-soldering tray which runs on a trolley. The trolley lowers the assembly into the molten solder bath and effects soldering of all the leads at one time. This operation takes from 3 to 10 seconds, depending on the size of the circuit assembly. After soldering, the excess flux and wax are cleaned off. Circuit continuity is checked and assembly is transferred to a workholder which clamps the assembly in a single plane. A punch press cuts off the tabs in the baseplate, thus allowing for folding later on. After application of protective masks, the protective coating and fungicide are sprayed. The mask is removed

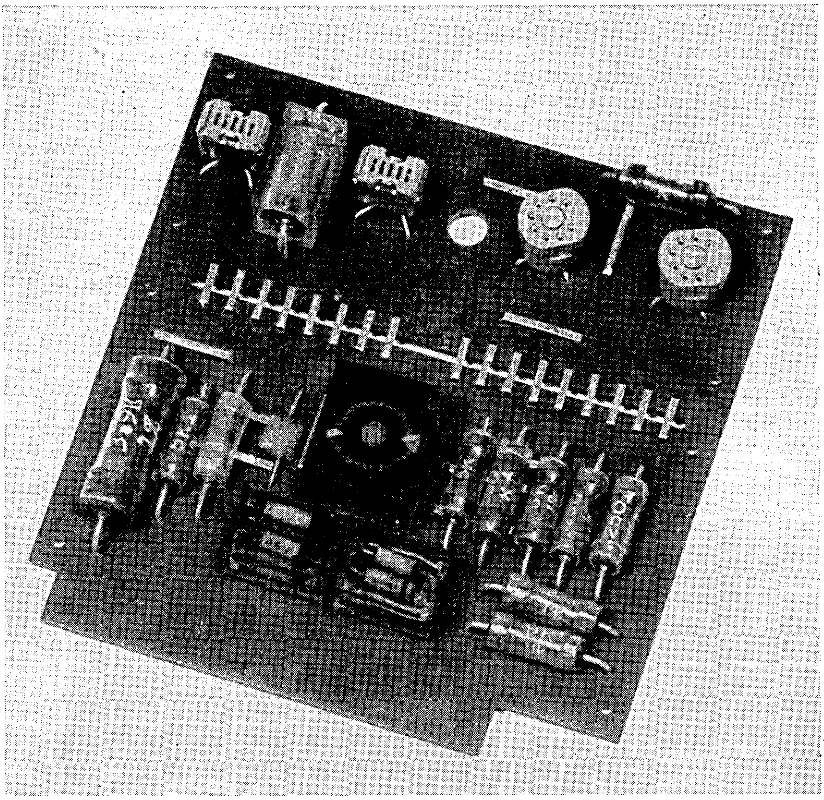


Fig. 13.5 Miniaturized airborne assembly showing types of components handled; the circuit board is laid out flat to facilitate installation of components by the attaching machine

and the frame is riveted onto the baseplate. The folding machine rotates the rigid laminate 90 degrees and rivets it to the frame. At this point all of the fabrication has been completed and the subassembly is ready for the oven where the protective coating will be cured. External mounting screws are installed and the tubes are inserted into the sockets. Tube shields are installed and the entire assembly is given a performance test and sent to the packaging machine. Although this description is for the fabrication of the trial assembly shown in Fig. 13.1, to show the versatility of the line to accommodate changes in product design and to handle widely different component sizes and shapes, the miniaturized assembly shown in Figs. 13.5 and 13.6 and the assembly shown in Figs. 13.7 and 13.8 can also be fabricated with only minor relocations of the equipment and retooling of the automatic line. Such multilead components as relays, right-angle tube sockets, delay networks, and pulse transformers can be indexed and inserted without difficulty.

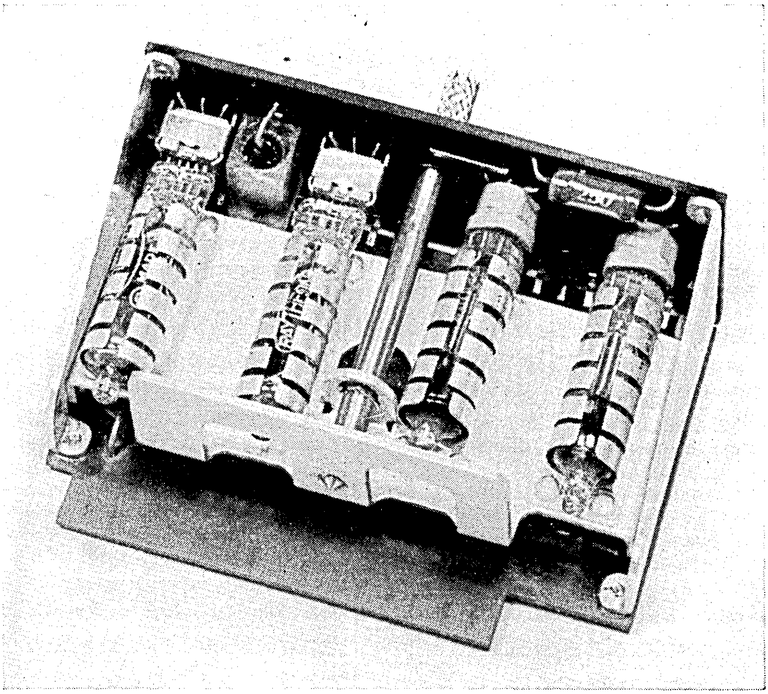


Fig. 13.6 The miniaturized assembly is completed by installing the frame and folding the circuit board into an L-shape for mounting; to facilitate dip soldering, the circuit pattern is on the side away from the components

As we can see by the number and types of operations described, there are a multitude of problems to be solved before we have an automatic production line. Effort has been concentrated in four areas of this production line. First, a wax flux etchant resist (a protective film spray) was developed for spraying through a steel stencil or for use with a silk-screen printing machine because available

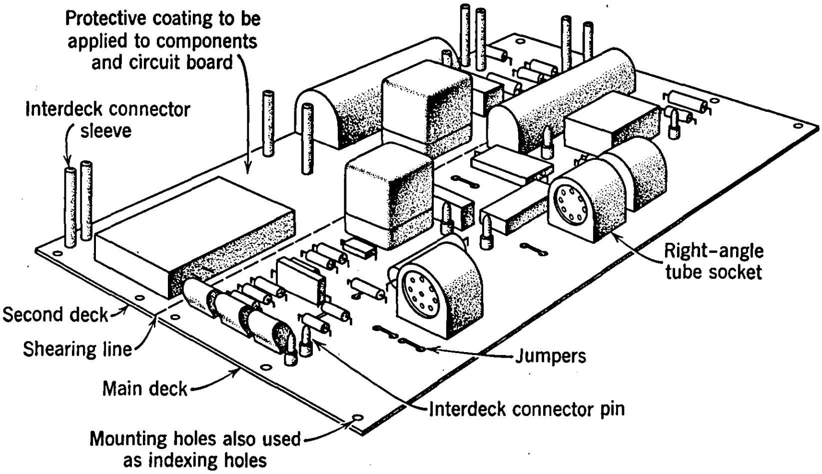


Fig. 13.7 Components installed on flat circuit board, components include delay line package, relays, and right-angle tube sockets

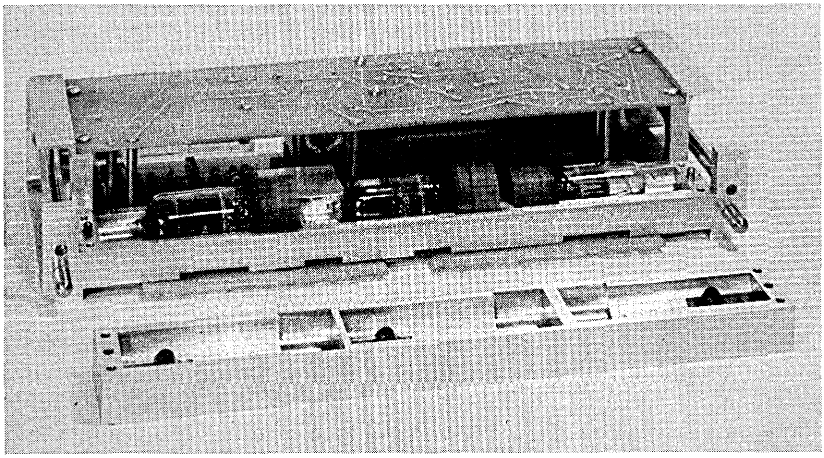


Fig. 13.8 Circuit boards mounted on frame which is also used as heat conductor to a sink for conductive-type cooling

resists were not suitable for use with an acid etchant and are difficult to remove. Wax hardens quickly and will melt in the dip-soldering operation, and the flux facilitates soldering. The second area of effort was in establishing the chemical-equilibrium requirements for a continuous etching process which has the features of etchant regeneration and copper recovery. A laboratory demonstration model

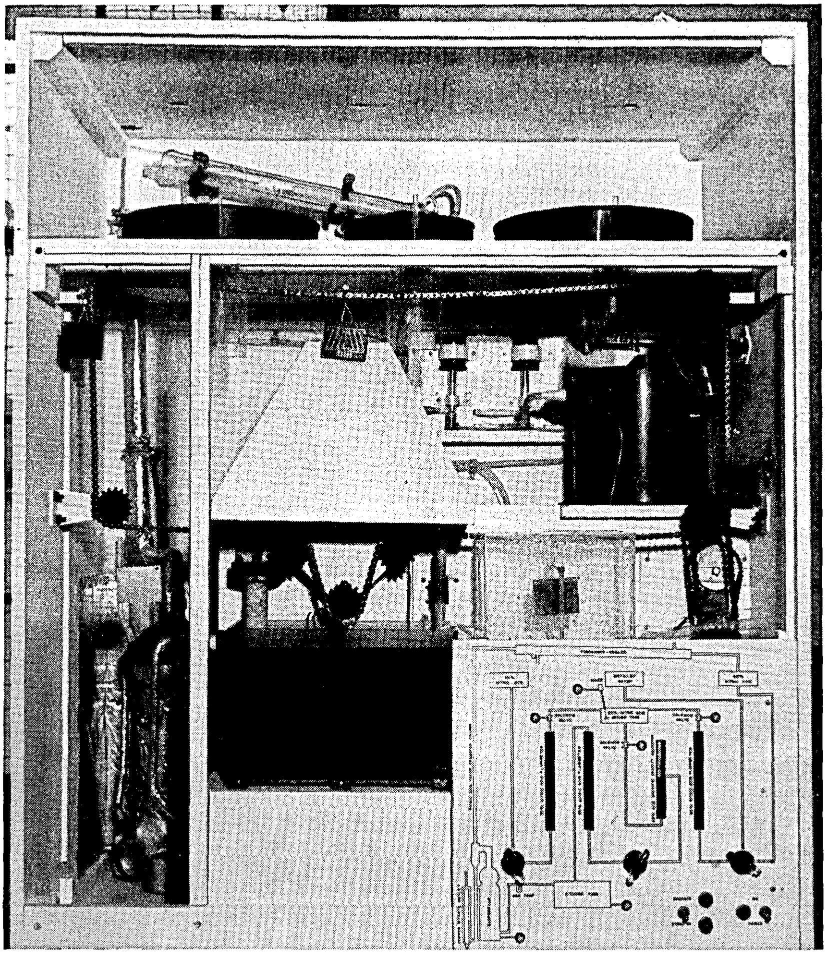


Fig. 13.9 Laboratory model of continuous etching equipment containing evaporation to recovery copper in the form of copper nitrate and to regenerate the nitric acid for reuse; chemical equilibrium is maintained by controlling the concentration and flow of acid into the etching tank and by controlling the rate of circuits into the etching tank

of the continuous etching process is shown by Fig. 13.9. Preliminary economic evaluation has shown that it is quite desirable to regenerate the acid etchant and to recover the copper. As the circuit is being etched in the tank, a given amount of spent etchant is drawn off to the still, which concentrates the nitric acid by a distillation process. The copper is recovered as copper nitrate, which is drawn off as a slag at the bottom of the still. The regenerated nitric acid is mixed with distilled water in the right proportions to be metered into the etching tank. A supply of concentrated nitric acid is added to the mixing tank with additional distilled water to make up for any losses. The acid regeneration process is about 80 per cent efficient. Level controls are used in all the storage, etching, and mixing tanks with a concentration control also installed in the mixing tank. As the liquid level in the still runs low, a control automatically shuts off the heat to the still. At present, several industrial organizations are looking into the feasibility of building such a machine for the market.

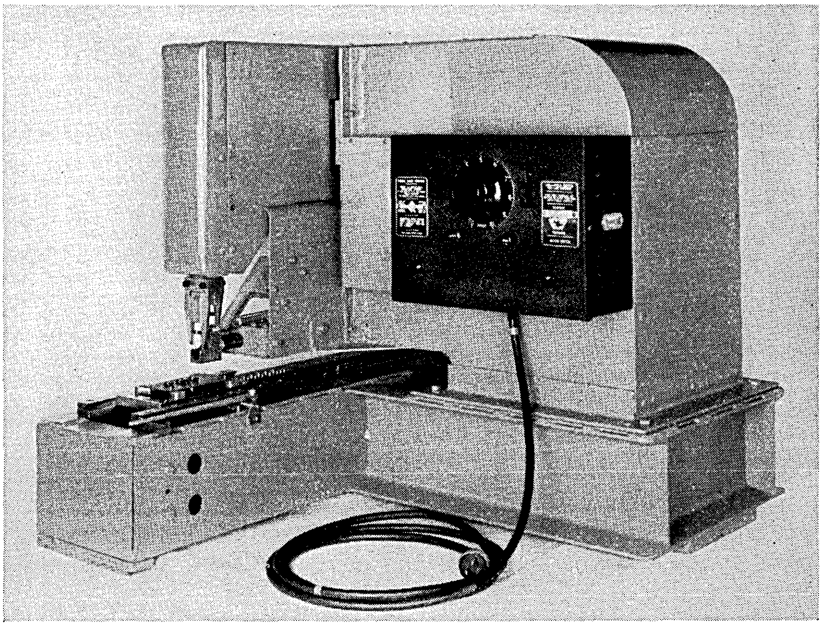


Fig. 13.10 Simplified component-attaching machine

A third area of effort is a study of the requirements for a machine to attach now-available components on a flat surface. Figure 13.10 shows a simplified version of the CAM built to study the problems of materials and components handling and the techniques of component attaching. The attaching head shown by Fig. 13.11 consists only of the basic mechanisms—the insertion forks, a workpiece-holding carriage, a simplified indexing system, and a component feed chute. A complete machine may have, in addition to the basic mechanisms shown, storage magazines for different values of components, a selective component feed system, and a versatile indexing mechanism. The close-up view of the attaching head shows more clearly how resistors are fed from the chute to the anvil of the shuttle where the leads are formed into a U shape and then inserted into holes on a baseplate. A clinching tool underneath the plate clinches the leads to the etched wiring pattern. Any desired setting of component locations is easily selected by half a turn of the screws at the left. This machine, completed about two years ago, has influenced several manufacturers to build similar equipment. The fourth area of concentration is that of the overall systems evaluation from a techno-economic approach.

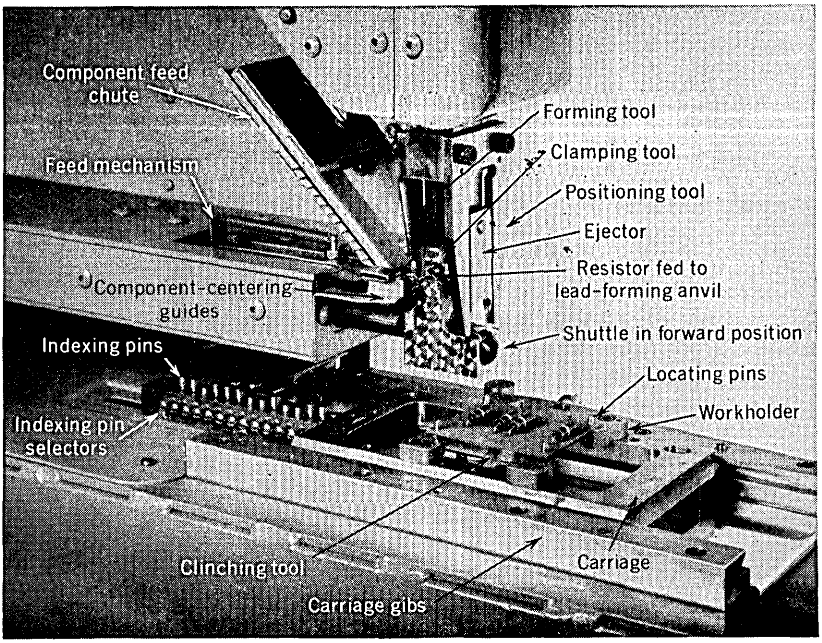


Fig. 13.11 Close-up of attaching head showing insertion fork installing 1-watt resistor; the simplified indexing mechanism can be adjusted by turning the pin selectors to raise or lower the indexing pins to vary the insertion pattern

From the very beginning of the program when various processes were being evaluated for use in the automatic production line, the economic aspects of each process such as cost of equipment, the time required for the process, and the materials used were considered. Owing to the lack of detailed cost information, common sense had to be relied on for the initial economic evaluation. With the establishment of the sequence of operations, as well as the detailed requirements for these operations, an analytical investigation of the economics of automatic production was made. As an initial step in this study, a comparison was made between the costs of producing the trial assembly previously illustrated by conventional methods and those of producing it by the automatic line being studied.

After a preliminary determination of which costs should be compared, it was apparent that differences in cost should be sought in material requirements, product engineering, tool design and manufacture, and all the steps of factory production. Therefore it was decided that the comparison should encompass all costs incurred beginning with the engineering decision whether the product should be production engineered for conventional or automatic production and ending with the shipment of the product. Consideration of the comparison between the cost of manufacturing the six-tube trial assembly in an automatic assembly and that of manufacturing an equivalent assembly in a conventional factory leads to the following conclusions. Before presenting them, however, it is important to recognize that the findings are limited in that they apply directly to but one assembly produced under various selected conditions. Obviously, there is no justification for applying the findings of this study to other situations without extensive revision to correct for new values. At the same time, the quality and quantity of data used and the basis for computation recommend valid indicators of comparative costs and of the parameters which favor automatic production.

In comparison with the conventional factory, the automatic factory shows cost advantages down to daily operating rates of under 500 units and lot sizes below 300. In selecting a suitable operating rate for the automatic production line under study, the reduction in cost diminishes rapidly when the operating rate is increased beyond the rate of 4000 per day.

If the demand for products from the automatic factory were lower than its capacity, it would be idle part of the time. Figure 13.12 shows that substantial savings over conventional production would result, even if the automatic production line were operated at half rate, or even if it were held in a stand-by condition for nine months or more in the year. This means that it may be economical to maintain stand-by facilities for production of military equipment, even if such facilities are ordinarily operated only part time.

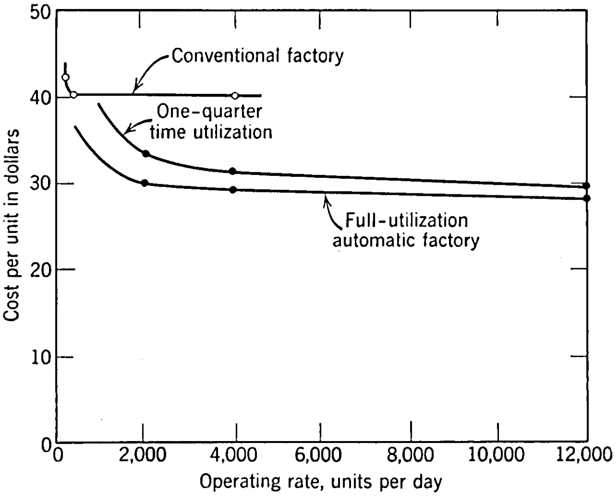


Fig. 13.12 Unit cost for conventional factory compared with automatic factory at full utilization and one-quarter time utilization

The manufacturing cost by use of the automatic factory is substantially lower than that by use of the conventional factory at all but the lowest rates and amounts of production. For 5000-unit lots at optimum operating rates the unit costs are shown in Table 13.1. A saving of $11.38 per unit—28 per cent—is indicated by the use of automatic techniques instead of conventional ones. If only the in-

Table 13.1 Comparison of Unit Manufacturing Costs for a Six-Tube High-Reliability Electronic Assembly

			Savings by Use of	
Cost Element	Conventional	Automatic	Conventional	Automatic
Material	$29.94	$26.55	...	$ 3.39
Labor	9.10	.55	...	8.55
Production engineering	.43	.43	...	....
Special tooling	.27	1.08	.81	....
Facilities	.35	.10	...	.25
Total	$40.09	$28.71	$0.81	$12.19

plant costs are considered by excluding material, a saving of $7.99 per unit—79 per cent—is indicated. The largest cost element is material, which represents some 73 per cent of the conventional-design manufacturing cost and 92 per cent of the machine-fabricated unit. A large saving is shown for automatic over conventional by reduced material costs, principally by substitution of a printed circuit for conductors, terminals, and connectors. The dominant position of materials in total cost should lead to further emphasis on application of new, cost-saving components such as packaged combinations, which are adapted to automatic production techniques.

Labor costs show the greatest savings for automatic over conventional methods; these costs represent about 23 per cent of total conventional manufacturing cost, but less than 2 per cent of automatic cost.

Production engineering costs are identical for automatic and conventional products, but special-tooling costs are somewhat higher for automatic production.

In the case of both conventional and automatic production, the unit cost of production facilities is between 1 and 0.3 per cent of the total. Even though large differences may be indicated in the initial cost of facilities under various conditions, the annual costs are spread over enough units of production that the initial differences do not have a large effect on unit cost.

The initial capital investment indicated for the automatic factory studied is somewhat less than $1,500,000 and provides operating rates up to 12,000 units per day. The investment for a conventional plant with a capacity of 400 units per day is about $300,000, but if a capacity of 4000 per day is needed, this investment increases to over $2,000,000. Most of the investment in the automatic factory is in equipment (about 85 per cent). In the conventional factory, it is mostly in buildings and land (about 70 per cent). Little additional investment is required to attain higher operating rates in the automatic factory, within the capacity of the basic production line. In the case of conventional production, however, higher operating rates mean more personnel with correspondingly greater space requirements.

The factor of reliability in service is one of great economic importance and is of special value in military applications. Although no evidence was presented to support the contention that equipment produced by automatic techniques would prove more reliable than that produced by conventional techniques, this possibility exists and its effect should be considered in any comparative evaluation of production techniques.

The results of this study show directly that the automatic factory is economically feasible for medium- to high-volume (lot size above 300 units) production of electronic equipment, as presented by trial assembly No. 1.

The high rates of production attainable by automatic production, as compared with conventional, raises the question whether there is sufficient demand to absorb the output of an automatic factory and make it economically feasible. How much equipment is susceptible to automatic production was not definitely ascertained, but it may be noted that the annual manufacturing costs indicated for the automatic factory operation at the maximum rate studied would be about \$104,000,000 or 3 per cent of the total requirements for military electronic equipment in 1953. In regard to commercial production, the economical operation rates for the automatic factory fall below those of the median-level manufacturers.

13.4 THE SARGROVE AUTOMATIC MACHINE

The machine developed by John A. Sargrove in England is of historical interest since it was the first automatic machine built to manufacture electronic assemblies. The machine shown in Fig. 13.13 was built in 1947 to produce one radio receiver every 40 seconds. The plastic panel with depressions molded to form the conductors, capaci-

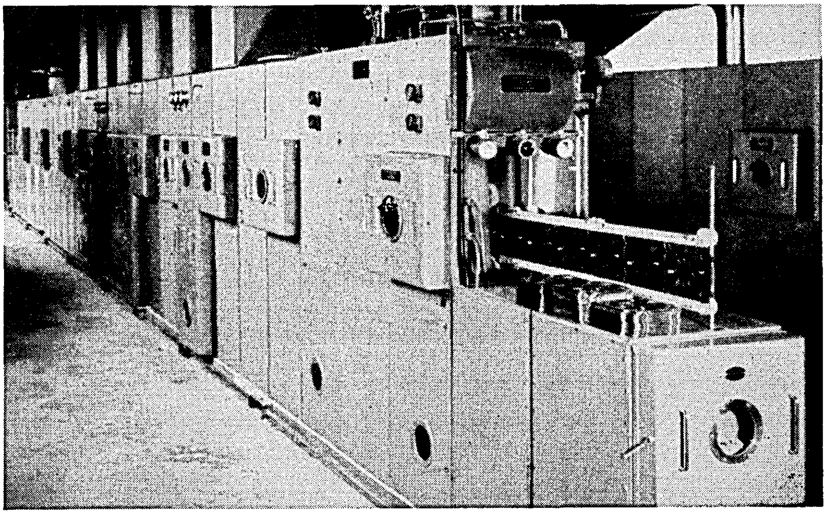


Fig. 13.13 The Sargrove electronic circuit-making equipment (Photograph by Fenno Jacob-Fortune)

tors, and inductors is successively subjected to a hot zinc spray and a high-speed milling process which is electronically controlled.

Any single resistor can be sprayed within ±5 per cent tolerance, but the others will fall somewhere within ±20 per cent. The capacitors are made by corrugating the area to be metalized to increase the surface area so that capacitances up to 500 micromicrofarads may be achieved. Automatic tests are constantly applied to the plate while in process, and if two plates are rejected in succession at any test station, the whole line is automatically stopped. The two plates required for a two-tube a-c–d-c set produced for sale in Asia are shown in Fig. 13.14. These two plates can be made for about $1.00. The completed set of a later model is shown in Fig. 13.15.

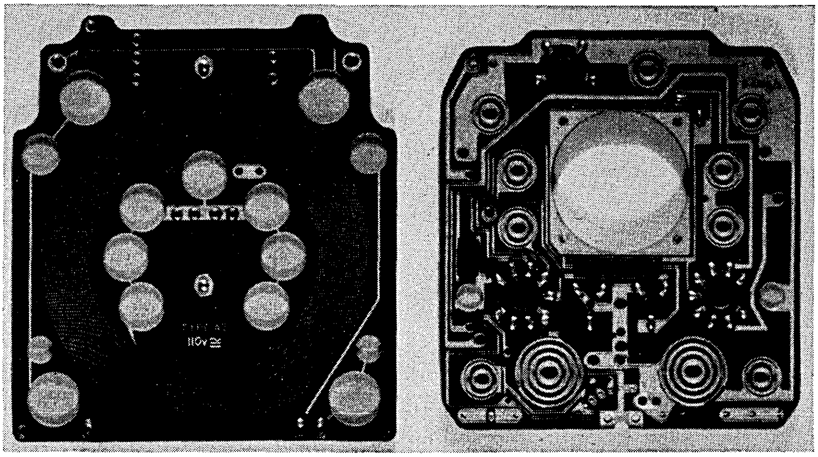


Fig. 13.14 Circuit plates for two-tube a-c–d-c radio (Photograph by Fenno Jacob-Fortune)

Tube sockets and other components are inserted automatically and held in place by combined riveting, welding, and soldering. Speakers and electrolytic capacitors are assembled to the chassis by hand.

A single machine can produce 500,000 sets per year and costs about $250,000. Compared with present-day costs of manufacturing radios, it would take a minimum production lot of 20,000 sets to be competitive. This includes tools and amortization of the plant. It would be possible to handle lots of 5000 economically by using temporary brass mounting tools.

The Sargrove machine is an example of going "too far too soon." This machine lacks the flexibility desirable in the present state of mechanization and standardization of electronic assemblies. When the Asian market collapsed after the war, the machine was never put in use because it was uneconomical to adapt it to manufacturing a range of products. However, it does forcibly illustrate the savings in time and manpower which may be achieved during the manufacture of large numbers of a particular item.

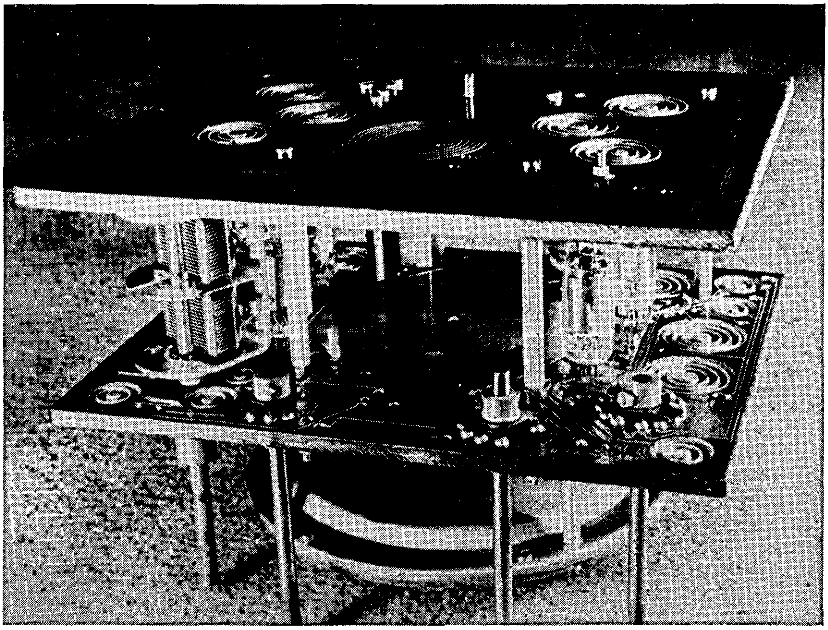


Fig. 13.15 A later model set manufactured by the Sargrove machine

13.5 GENERAL MILLS AUTOFAB

Typical of one industrial approach to the automatic insertion of components into printed circuit boards is the line built by the Mechanical Division of General Mills, Inc. Figure 13.16 shows the 24 component-attaching heads mounted in a line with the component magazines in a turret mounted above the attaching head. The present line is designed to operate at a speed of 20 to 30 completed assemblies per minute or nearly 10,000 in an 8-hour day. A standard machine can be adjusted to accommodate baseplates of various thicknesses in any rectangular size from 2 inches by 2 inches to 10 inches by 10 inches. The attaching heads can be rotated and adjusted transversely and laterally so that components can be attached at any angle and at any position on the baseplate. Adjusting or retooling the attaching head can be done in 3 to 8 minutes. A 24-station line, controlled electrically, can normally be operated by two operators and a supervisor. Figure 13.17 shows the circuit baseplate-feeding mechanism and magazine at one end of the line. In operation, the baseplates stacked into the automatic feeder are dispensed from the bottom when two pushers connected to the conveyor belt trigger a latch. The pushers then move the baseplate to each component-attaching head.

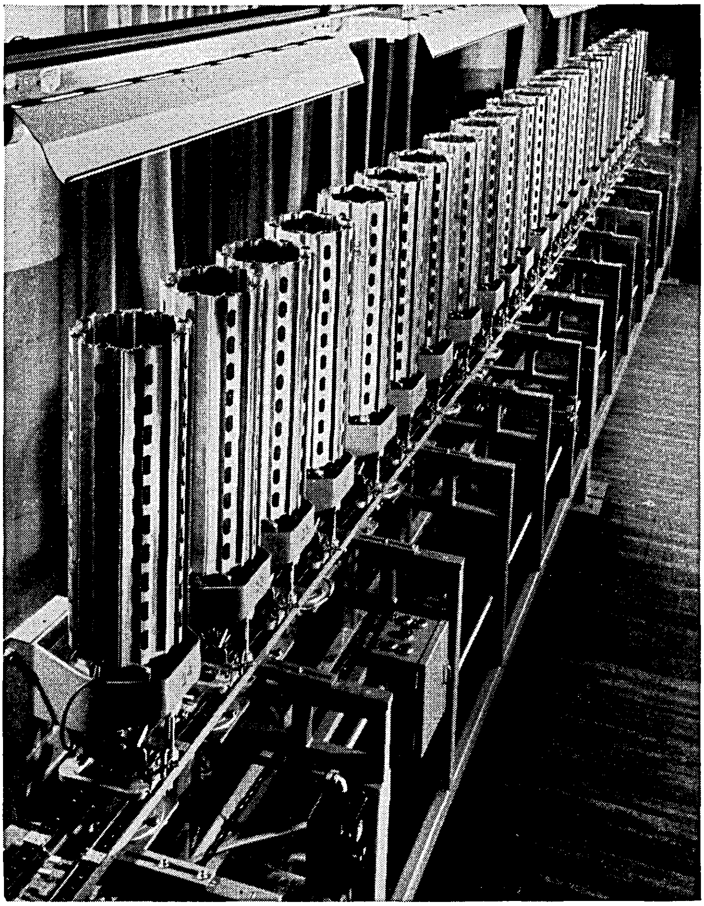


Fig. 13.16 Autofab by General Mills, Inc.—a 24-head automatic component-attaching machine

As the baseplates approach each attaching head, the conveyor slows down, and a locating stop on each head is raised as the conveyor stops. The baseplates are located and held under the attaching heads by the same indexing surfaces that were used in punching or drilling the holes for the leads. A conveyor system spring-loads the baseplate into position against these surfaces during the attaching cycle.

Figure 13.18 shows the pneumatically powered component-attaching head inserting a tubular pulse transformer onto the board. Each attaching head bends the leads to the proper spacing, inserts the leads into the baseplate holes, and makes a mechanical connection between leads and baseplate. To eliminate damage to the components, all attaching forces are applied directly to the component leads. Thus there is no stress of the junction between the lead and the component body. The attaching cycle completed, the stops are retracted and the conveyor moves the baseplate to the next station where the operation is repeated. After the last component has been attached, the plate is fed out of the machine ready to be moved to the next sequence in manufacture. Figure 13.19 shows components inserted

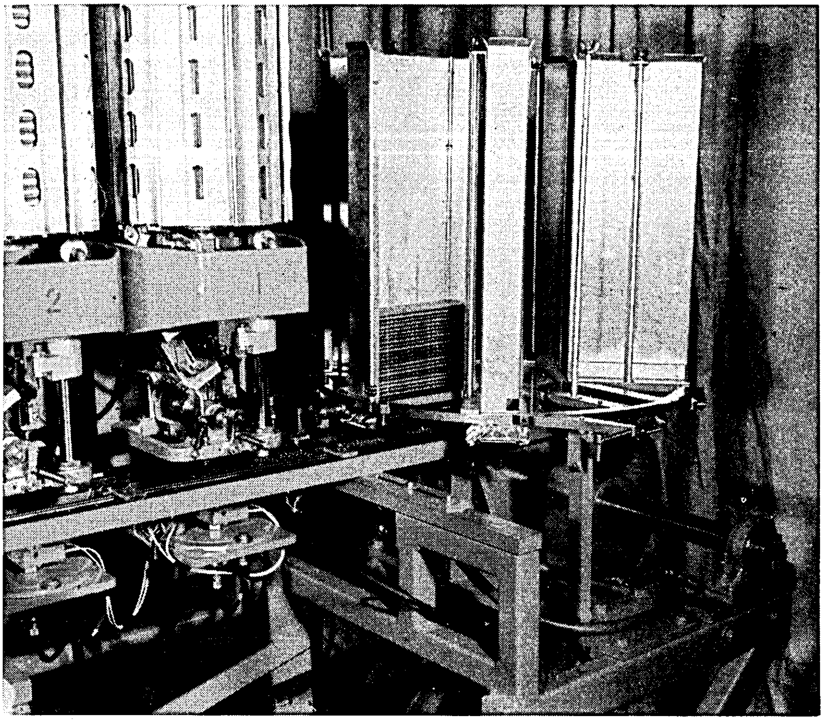


Fig. 13.17 Circuit baseplate storage and feeder—component magazine turret mounted on top of attaching head automatically indexes to next full magazine when empty

with the bodies against the circuit board. This is the usual method of installing the components. Figure 13.20 shows similar types of components, i.e., cylindrical bodies with axial leads, installed with the body off the board as may be required to facilitate cooling for high-heat dissipation.

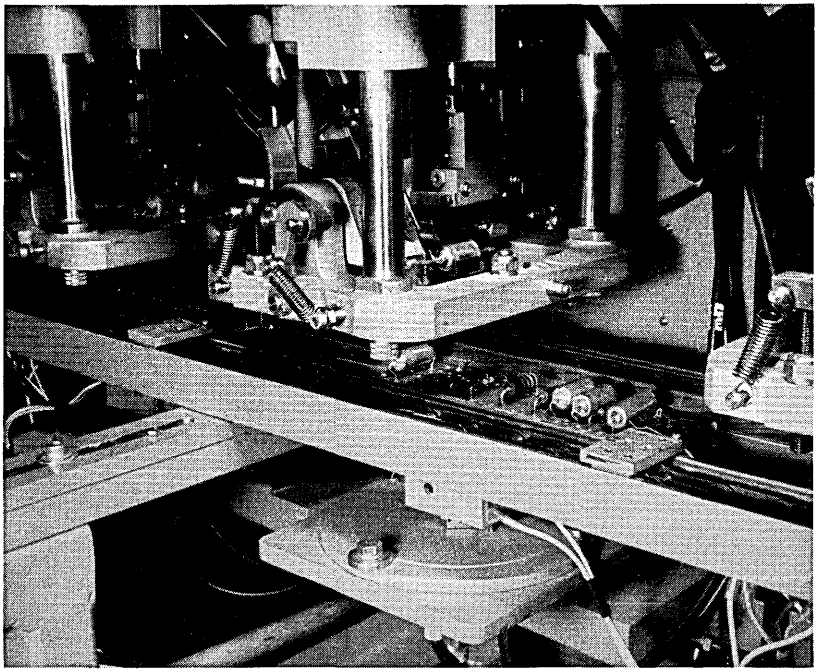


Fig. 13.18 Close-up of attaching head inserting transformer onto circuit board

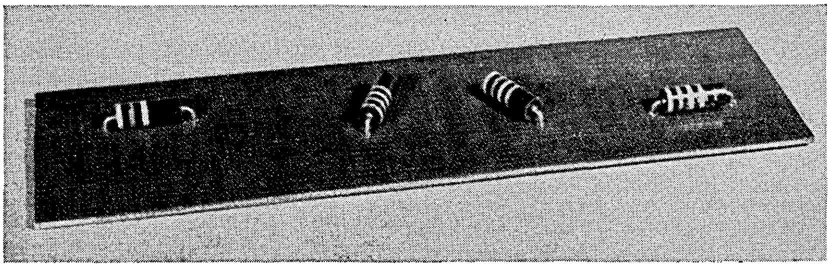


Fig. 13.19 Components installed on circuit board

Automatic indexing of the multiple-magazine turret as it is shown in Fig. 13.16 makes available a continuous supply of components.

An interlock system is provided on each attaching head so that (1) if no baseplate is in position under an attaching head, that head will not be actuated to insert a component, and (2) if a baseplate passes from an attaching head with the component missing, the machine will stop and the empty head will be indicated by a light. This allows the line to be started and stopped with no waste of electronic components, and at the same time each baseplate is certain to receive a full assembly. This same interlock operates if a component is not inserted properly because of a broken lead or for any other reason.

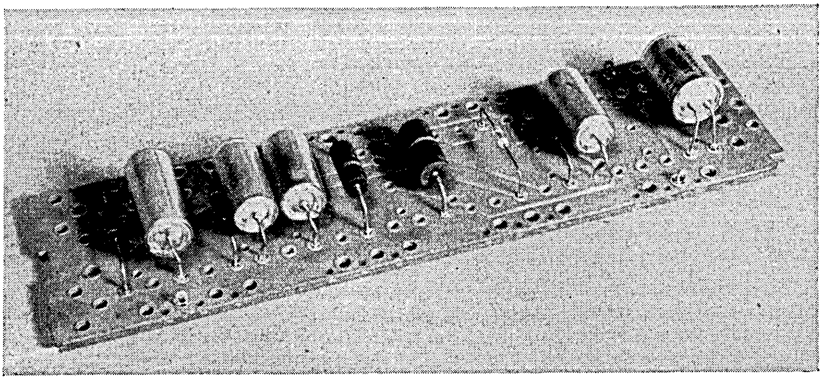


Fig. 13.20 Components installed on circuit board with body off the board

Both leads of the inserted component must be seated properly before the conveyor will move the assembly to the next station.

An automatic dip-soldering machine is being prepared for market. This dip-soldering machine will accept the printed circuit cards as they come off the attaching line and solder all the component leads in one dip.

13.6 THE UNITED SHOE MACHINERY CORPORATION DYNASERT

The original automatic line at United Shoe Machinery to insert components on a printed circuit board has been in operation since October 1954. Figure 13.21 shows the ten-station line equipped primarily to insert ½-watt resistors. This machine is in daily use and has processed well over a quarter of a million circuit board assemblies. Recent reports have indicated that faulty component insertions, i.e., those missing the hole or breaking the component lead, have been as low as 0.1 per cent. Frequently production runs of 1000 or more circuit boards are run without malfunctioning of any type.

Circuit boards which were manually loaded on the conveyor pallets are now automatically loaded and dispensed onto the conveyor. As the pallet approaches the attaching head, ejector arms move the pallet off the conveyor and butt it against a stop which serves as the indexing mechanism. Accurate indexing is achieved by inserting shot pins in the locating holes in the circuit plate just before the component leads are inserted and the leads are clinched by an anvil in much the same manner as a wire stapler.

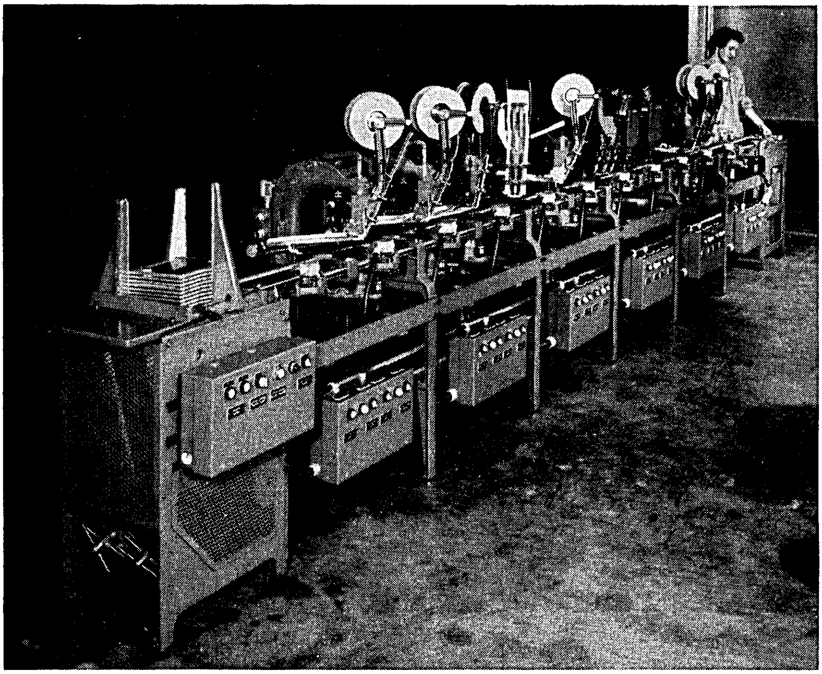


Fig. 13.21 The United Shoe Machinery Corporation assembly line

A preparation machine for axial lead components straightens the leads and applies paper tape to the ends of the leads to form a belt. This belt or chain feed method allows a much more compact attachment head design and is the primary reason for the difference in appearance between the United Shoe Machinery and the General Mills equipment. Figure 13.22 shows the belted resistors being fed into the attaching head. A sprocket provides positive component feed to the mechanisms that form and trim leads and is a feature that the manufacturer claims accounts for the reliability of insertions. The lead forming and inserting is done by a double-fork arrangement that forms the lead on an anvil in much the same manner as in most insertion machines. A fault detection device, which is no more than a continuity checker, is used to insure that both leads are inserted correctly and are not broken before the lead is clinched; otherwise, the machine stops and an indicator light flashes.

The pallet transporting of the circuit boards is a unique feature of this system. This is a conservative design feature to simplify the indexing problems. These pallets will accommodate a maximum board size of 5 inches by 8 inches. Instead of indexing by its edges or by an edge and a notch, as is used in most systems, the board is indexed by locating holes spaced along the board. Thus, hole misalignments from warpage and other dimensional instabilities of the phenolic boards are minimized. The pallets at the end of the line are returned by the same conveyor which transported the pallets to each machine. If the finished pallets are not removed, the conveyor automatically ceases feeding; thus the conveyor system cannot be jammed. The present assembly rate is fifteen circuits per minute.

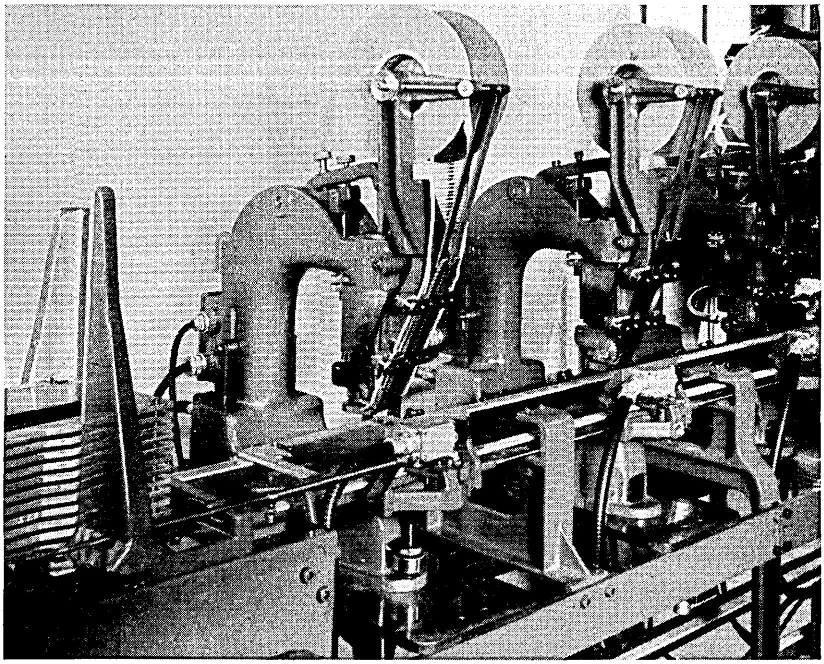


Fig. 13.22 Close-up view of pallet feed and attaching head inserting ½-watt resistors from a belt

Jumper wire and tubular capacitor-inserting machines have been constructed along the same lines as the resistor inserter. In addition, a semiautomatic model of the single attaching head is available. Figure 13.23 shows an early model which required manual indexing and circuit board feed. A foot pedal initiates the inserting operation.

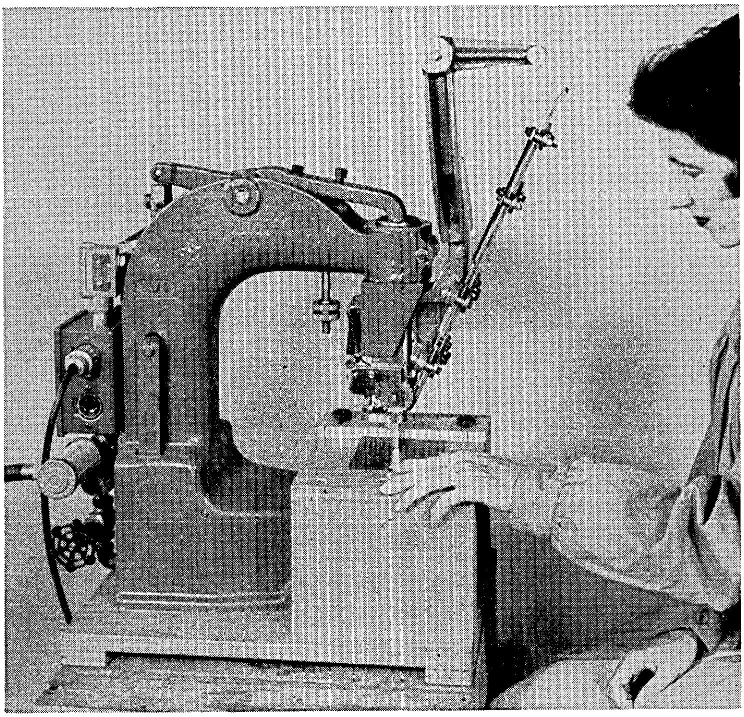


Fig. 13.23 Single attaching head requiring manual indexing and actuating of inserting fork

Currently in development are inserting heads for tube sockets, tubular capacitors with radial leads, disc capacitors, and intermediate-frequency transformers.

Both the General Mills' Autofab and the United Shoe Machinery Corporation assembly system use single-component CAMs; that is, each machine attaches one component on the circuit board, and as many machines are required as there are components on the board unless the boards are recirculated. The sequence of installation is set by the position of each machine in the line. These single-component CAMs have the advantage of simplicity and therefore low cost and possibly reliable operation. Because of their lack of flexibility, these machines are most suitable for large production runs where the setup and retooling time is distributed over a large number of units. Although the change-over time allotted for each machine runs from 3 to 8 man minutes, assuming an average of 5 minutes for each machine and 20 machines in the line to be operated by 2 people, it would take 50 minutes to make the change-over. Allotting a few more minutes to start a pilot run and make some final adjustments, this adds up to approximately an hour for each production lot. It would not take very many change-overs before the assembly cost would be increased to the point where it would be economically untenable.

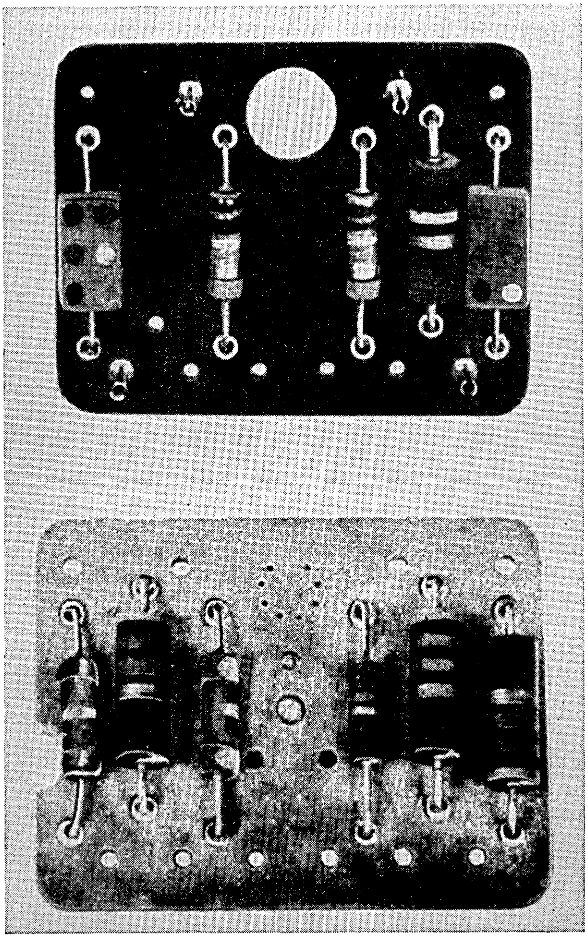


Fig. 13.24 Typical (2 inches by 1½ inches) circuit board handled by the Mini-Mech component assembly unit

13.7 PROJECT MINI-MECH

Project Mini-Mech, sponsored by the Bureau of Ships, United States Navy at Melpar, Inc., is essentially (1) the development of a mechanized construction technique for miniaturized subassemblies, and (2) the development of a machine to attach prefabricated components on to the circuit baseplate. Figure 13.24 shows two circuit boards typical of the assemblies handled by the component assembly unit. These boards are relatively small, in the order of 1½ inches by 2 inches. The circuit pattern on the baseplates may be fabricated by any of the commonly used methods such as etching or stamping. In the system of automatic assembly proposed, the machines under study are (1) the baseplate- or wafer-loading machine, (2) the component assembly unit, and (3) the assemblying and testing machine. The function of the baseplate-loading machine is to feed a continuous supply of circuit baseplates in correct sequence to the component-inserting machine.

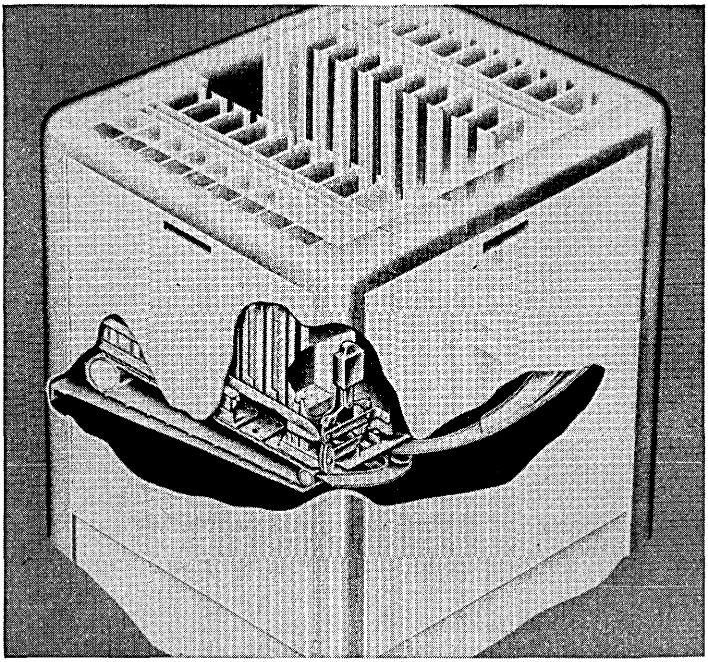


Fig. 13.25 Baseplate-loading machine

The baseplate loader shown in Fig. 13.25 makes available up to four types of wafers in desired order. On each cycle of operation, the plates are sequenced to drop from the magazine to the conveyor which carries them to the component assembly machine. It was visualized that each miniaturized package would have no more than four circuit plates. The assembling and testing machine shown in Fig. 13.26 is visualized to form the baseplates into a subassembly and give it a performance test. The assembly mechanism receives the baseplates and forms the deck construction in the succeeding stations, installs the necessary hardware, and then finally tests the performance. A partially assembled unit is shown in Fig. 13.27. The deck construction allows freedom of geometrical movement in three dimensions, thus giving greater compactness and minimizing size. One limitation of the product design is the difficulty of providing support for the tube in the deck structure and the difficulty of providing adequate cooling when the tube is buried in the midst of the assembly. It is estimated that this automatic assembly system will produce 1000 units per 8-hour day of a two-baseplate assembly. Preliminary analysis indicates that

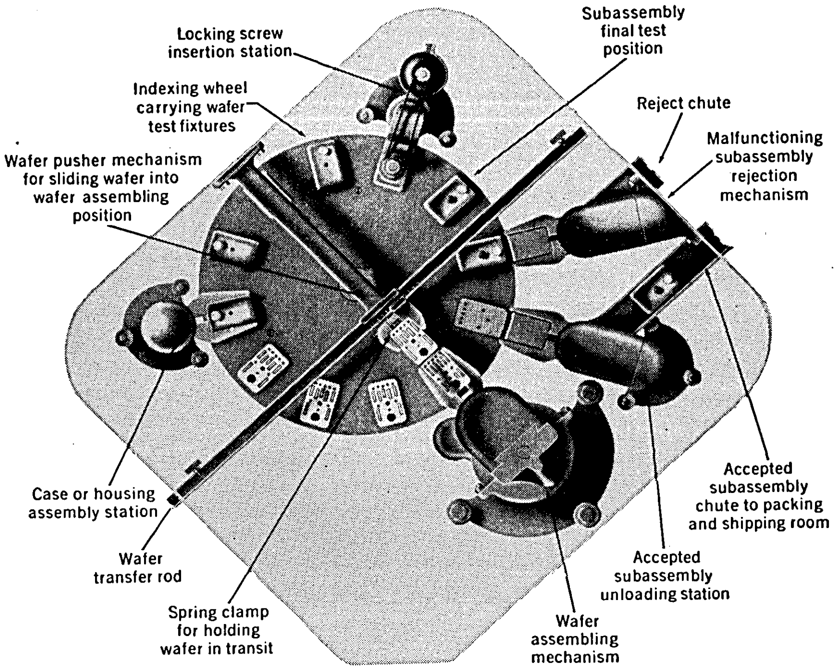


Fig. 13.26 Assembling and performance-testing machine

it may be economical to produce as few as 100 of these units between change-overs.

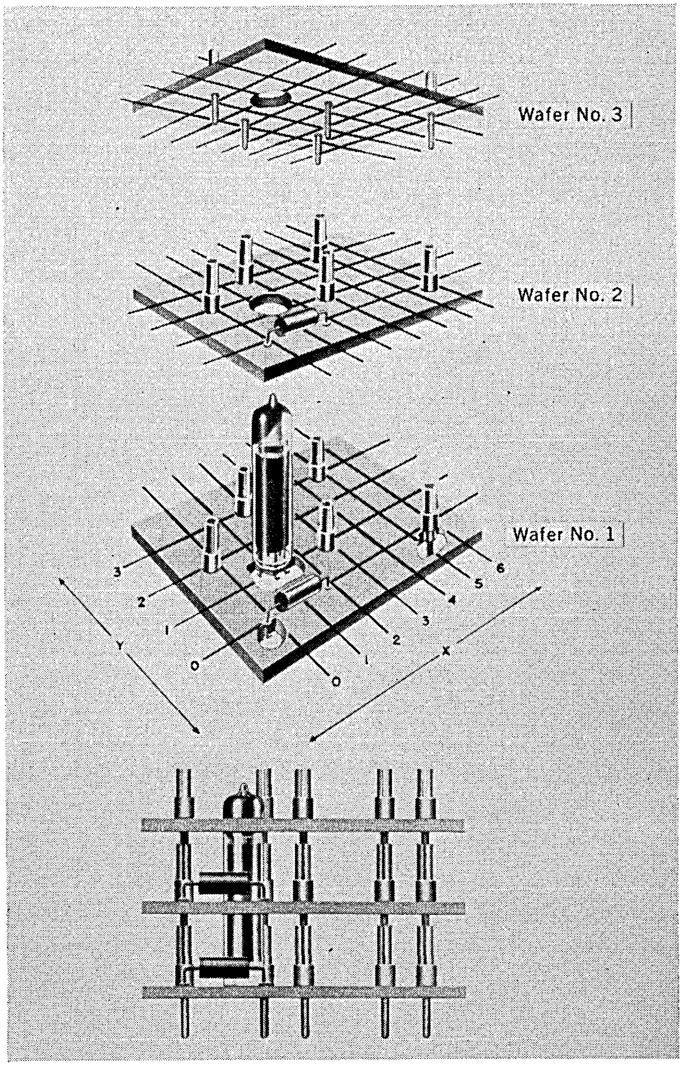


Fig. 13.27 Partially assembled unit to show deck construction and tube mounting

At present, the concentration of effort is on the development of a multicomponent insertion machine. The unit shown in Fig. 13.28 consists of a component storage magazine, the component-sequencing mechanisms, the insertion tool, and the component lead-clinching anvil.

The lower portion of the cabinet contains the pneumatic and electrical power supply and valves. Components with two axial leads already trimmed to length are stored in six separate hoppers which may contain as many as eighteen magazines so that each machine will insert and solder a maximum of six components. Figure 13.29 shows magazines and indexing mechanism. On reception of a control signal, the hopper indexing mechanism transports the components into a ready position and discharges one component as required for each sequence of insertions. The hopper indexing machanism consists of two endless bead chains and a suitable chain sprocket driven by a synchronous motor controlled by an electric clutch and an electric brake. As one magazine is emptied, the following one is opened by the hopper indexing mechanism and the components are advanced to the ready position. The components are discharged from the hopper indexing mechanism into one of two pneumatic slides shown in Fig. 13.30 which transport them to the forming blocks and inserting head. Each of the two pnuematic slides consists of a V-groove having approximately eighty 20 degree air ports at the bottom. These jets propel the components along the groove, thus eliminating the need for grasping the components with comparatively complex and slow-moving mechanisms. Essentially, the resistors are fed to the forming block by gently blowing the resistor along the groove. Mounted in the forming blocks are a light source and a photocell which detect the presence of a resistor before the inserting head is lowered. The reciprocating forming blocks remain in a closed position until the insertion head has formed the leads of the components; i.e., the leads are bent at an angle of 90 degrees

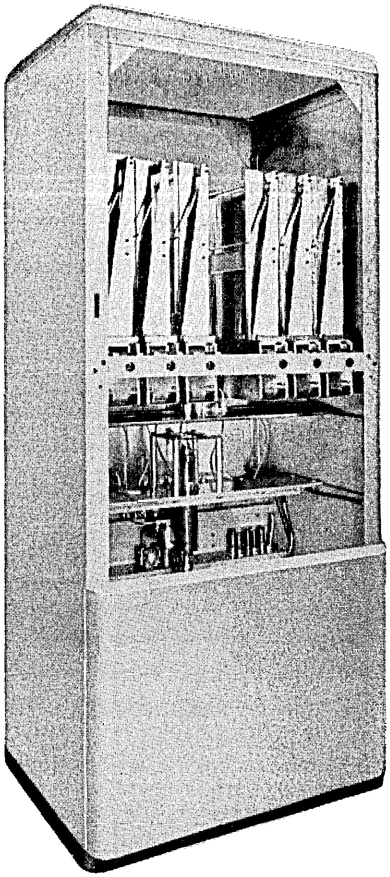


Fig. 13.28 Multicomponent assembly unit

with the body. The forming blocks will then open to permit passage of the insertion head with the component so that the component may be installed and the anvil raised. The anvil is a heated forming tool which will staple the component leads to the wafer and simultaneously solder the leads to the eyelet. As the anvil reaches the top of its stroke, the insertion head retracts. When the insertion head passes the forming blocks, the blocks close again. Closing of the forming blocks

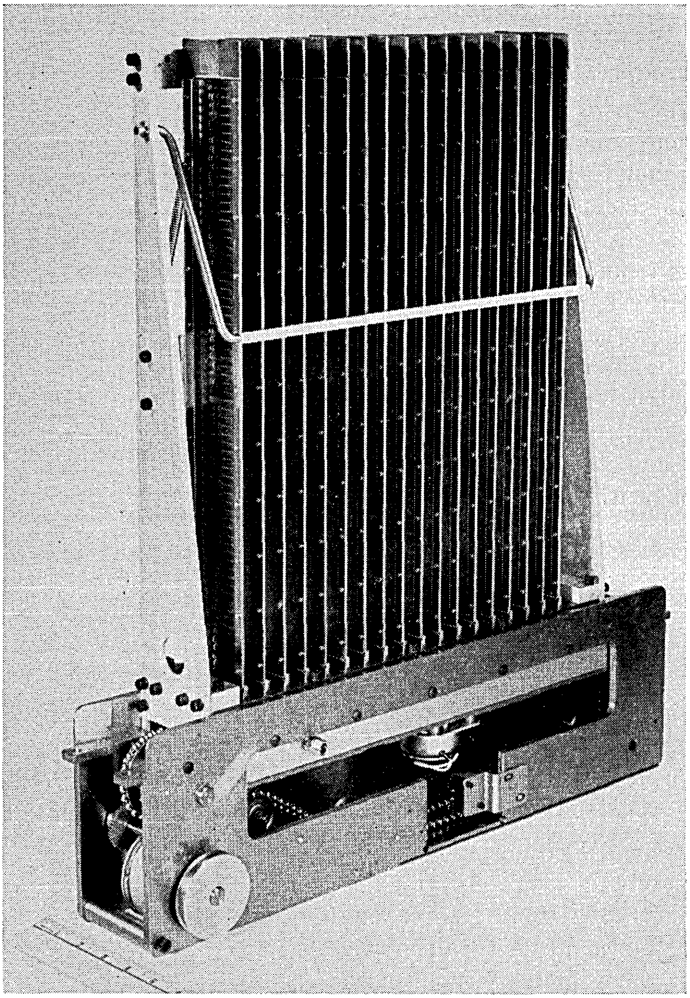


Fig. 13.29 Component magazines and magazine indexing or selection mechanism with bead chain component conveyor

resets the anvil, thus completing the cycle. Figure 13.31 shows a laboratory model of the component-inserting head.

The machines under development at General Electric Company and Melpar, Inc., are multicomponent CAMs. These machines require an indexing mechanism, hoppers or magazines, and a sequencing mechanism for various components; in addition, they must provide storage and indexing for the circuit boards. These additional mechanisms greatly complicate the component assembly machines and make them much more expensive. It is interesting that both General Electric and Melpar, who are under contract with the Armed Forces to develop a machine suitable for the production of a relatively small number of units of any particular subassembly, thereby requiring maximum versatility and flexibility of a machine, came up with multicomponent CAMs to minimize setup and change-over times.

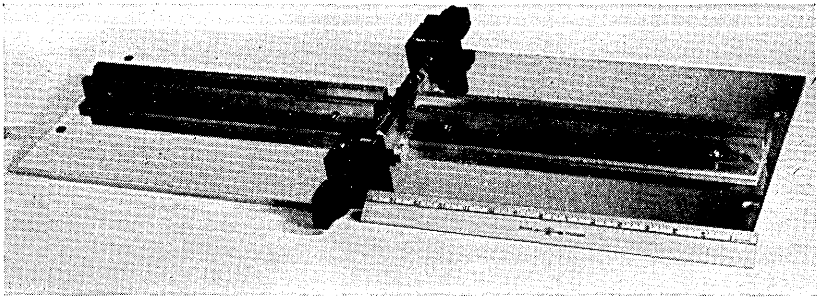


Fig. 13.30 Pneumatic transport, component feeder, and retractable component-forming block in the center

13.8 THE GENERAL ELECTRIC AUTOMATIC ASSEMBLY SYSTEM

The General Electric Company is taking a rather different approach to the development of an automatic assembly system. The object of the program, which is sponsored by the Signal Corps, is to develop a method for the automatic assembly of commercially available components on printed circuit boards to form subassemblies of military electronic equipment. This means that the production lots for each assembly or equipment may be only tens of units. Thus, special considerations are given to small-lots production in the design of this system. The work is being done for the Signal Corps as an industrial preparedness measure to provide know-how and production facility designs which can reduce lead time in the event of full mobilization.

Figure 13.32 shows an overall diagram of the automatic component assembly system. Considerations for the design of the automatic component assembly system have been divided into essentially five steps: (1) component preparation, (2) component and circuit board transport, (3) component assembly, (4) inspection or test unit, and (5) the overall control.

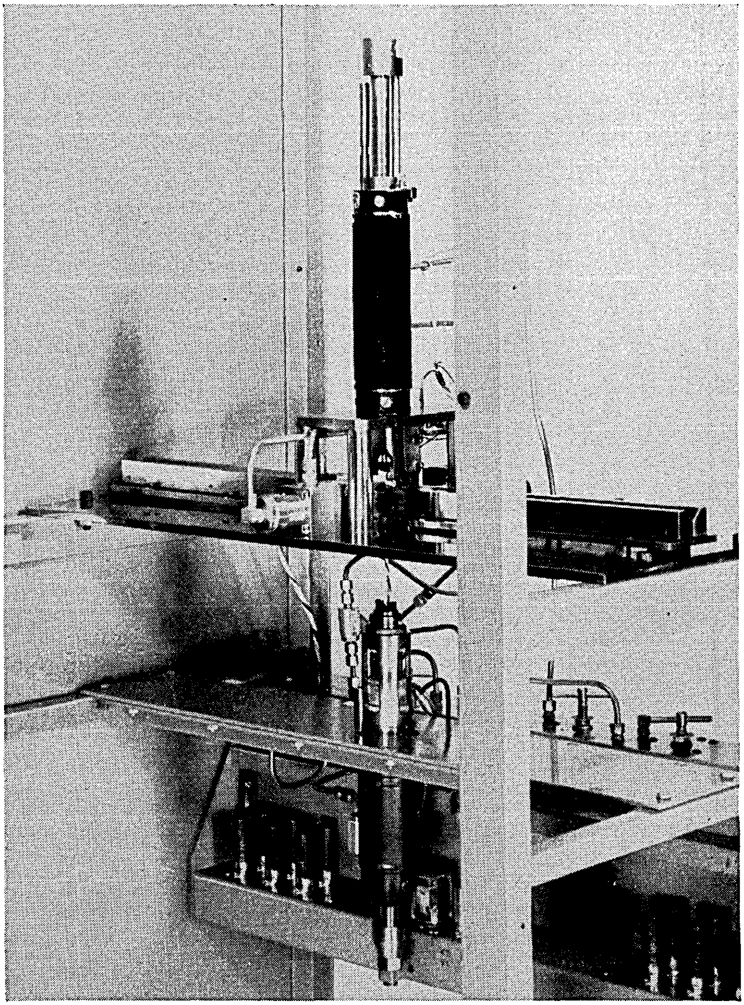


Fig. 13.31 Laboratory model of component-inserting head

Component preparation includes such functions as: (1) straightening

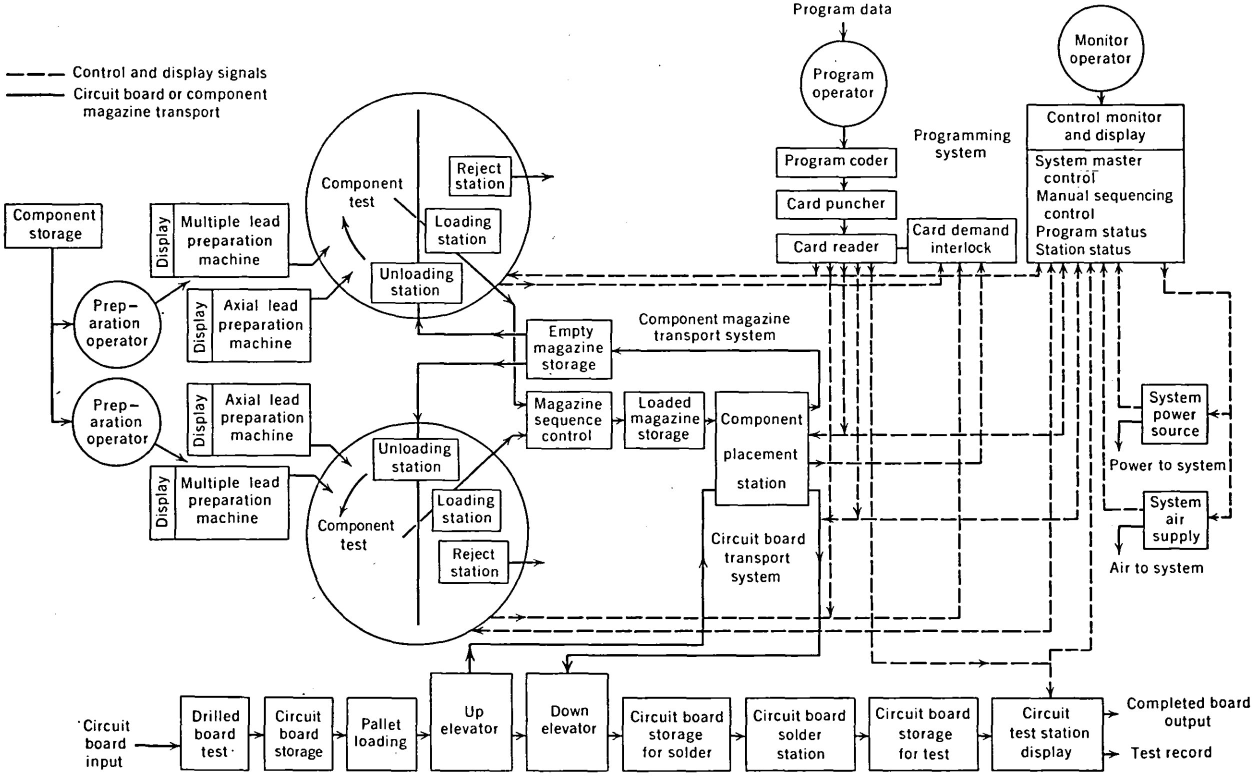


Fig. 13.32 The General Electric automatic assembly system block diagram

leads, (2) component inspection, (3) lead trimming and forming for insertion into the printed circuit board, (4) loading of components into magazines or component carriers. At present, two machines are being developed for the preparation of components, one to handle components with two axial leads, and the other for multilead components such as tube sockets and printed circuit subassemblies. The component inspection machine consists of a sixteen-station indexing table with eleven test stations which may be programmed to test resistor, capacitor, and tube characteristics.

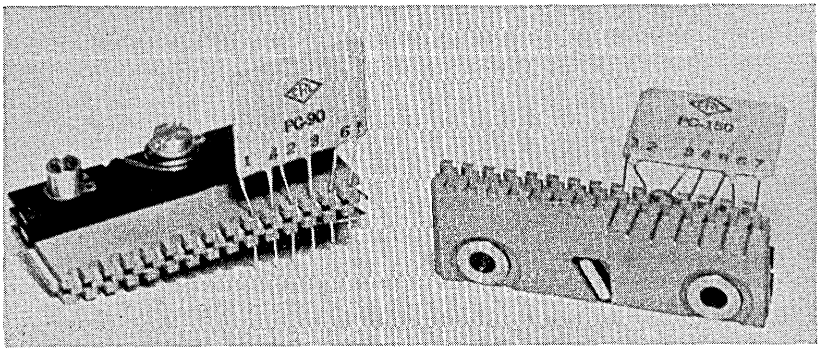


Fig. 13.33 Multilead component mounted on carriers; the component carrier has indexing holes and can be coded for automatic loading

The component and circuit board transport functions include: (1) receiving the component magazines from the testing machines in batch lots sequenced for each unit assembly; (2) transporting the component magazines to an elevator storage system; (3) transporting the circuit boards to the pallet feed; (4) loading the circuit boards on pallets; (5) transporting the pallets to the component assembly or placement machine; (6) recirculating the circuit boards through the component placement station until all the components are installed; and (7) returning the empty pallets to the pallet feed station.

Figure 13.33 shows a component mounted in its carrier, and Fig. 13.34 shows resistors mounted in their carriers and the combination inserted into a magazine.

Figure 13.35 shows a late model of a placement head which is electro-servo-controlled for the motions crosswise to the direction of work transport and also in rotation. The insertion jaws are also servo-controlled to adjust for different size components. Several servo systems have been studied for positioning the component placement head and the printed circuit board pallet which will accommodate a maximum board size of 8 inches by 12 inches. An accuracy of 1 part in 12,000 is required to position the placement head to an accuracy of 0.001 inches anywhere within its 12 inches of travel. An accuracy of 1 part in 8000 is required to position the board pallet to an accuracy of 0.001 inches anywhere within its 8 inches of travel. Each of these requirements is well beyond the accuracies that can be obtained by ordinary analog systems. In order to approach the 0.001-inch posi-

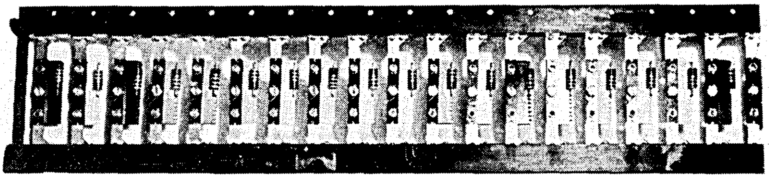


Fig. 13.34 Resistors mounted in their carriers and loaded in magazines ready for the component placement machine

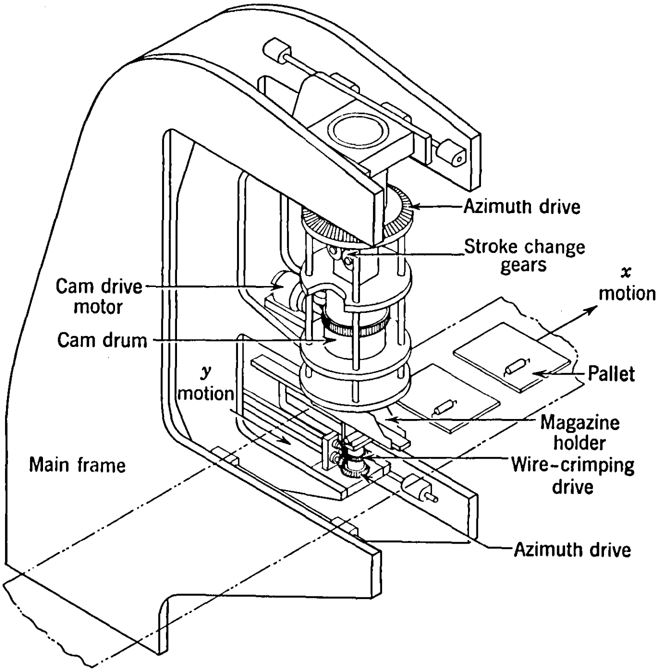


Fig. 13.35 Electro-servo-controlled component placement head

tional accuracy required, some modified analog systems and a digital system are being considered. The component insertion operation is pneumatically operated. To understand the operation of the component assembly machine, a cycle of operation of the placement machine is as follows.

Referring to Fig. 13.36, on command from the program control unit a batch of, let us say, twenty printed circuit boards are transported to the board storage unit of the component placement machine. At the same time, sufficient component magazines, each loaded with twenty components, are sequenced and transported to the placement machine. The number of component magazines is dependent on the number of components to be inserted on each circuit board. The first board is moved into position for placement of component number 1, and the placement head is also moved in translational and rotational motion for the first component. The component placement forks are automatically adjusted to component number 1. The circuit boards are circulated until all the number 1 components are installed. Magazine number 1 is ejected and magazine number 2 is inserted into the placement head and the placement forks adjusted for component number 2. The circuit boards are fed through the machine again. This process is repeated until all the components are installed; then the batch of twenty circuit boards is transported to another operation.

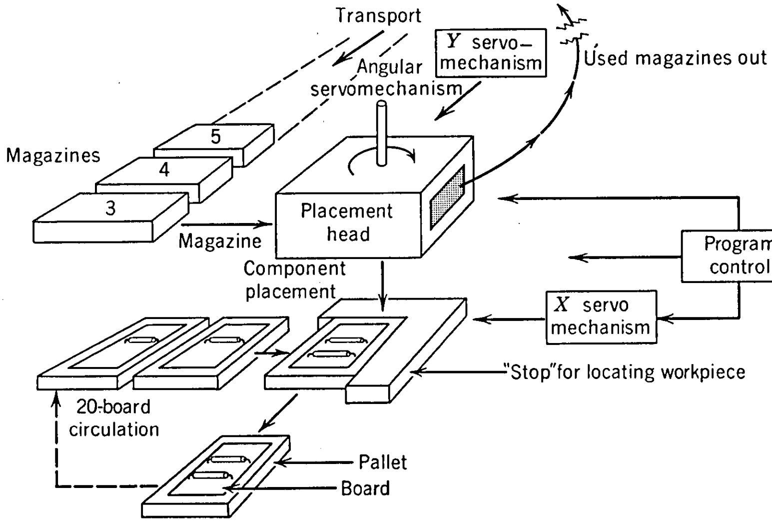
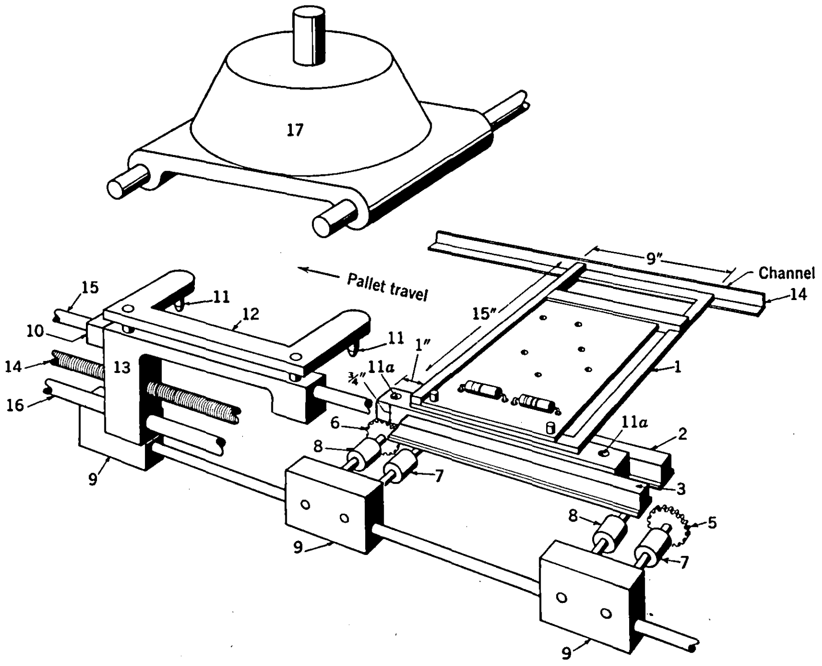


Fig. 13.36 Batch component placement cycle

The placement machine is ready for the next batch of circuit boards and component magazines which may be used to fabricate an entirely different assembly. For instance, on command from the programming unit, the magazines and circuit boards are circulated until all the components are inserted for the radar intermediate-frequency amplifier. Upon a second command, the component magazines are sequenced and the circuit boards are circulated until the radar trigger generator is completed. Each succeeding group of assemblies built will be a different unit. By building a few of each subassembly to complete a system, storage problems will be greatly simplified because it will not be necessary to wait until each lot of all the packages is



1. Pallet
2, 3, 4. Pallet guides and support
5. High-speed gear
6. Low-speed gear
7. High-speed clutch
8. Low-speed clutch
9. Gear box and power drive
10. Position sensor
11. Positioning shot pins
12. Shot pin holder
13. Shot pin guide
14. Position reference leadscrew
15. Main guide rod
16. Torque bar
17. Component placement head

Fig. 13.37 Work pallet positioning system for precise indexing completed before starting to assemble an equipment system. This is especially advantageous when a piece of equipment has a large number of different assemblies. Figure 13.37 shows the circuit board on the work pallet being recirculated to the component placement head for the third component. As the work pallet approaches the correct position at relatively high speed, a low-speed gear is engaged in the rack on the underside to slowly move the work pallet toward the final seating position, at which time shot pins engage reference holes on the work pallet for the precise indexing.

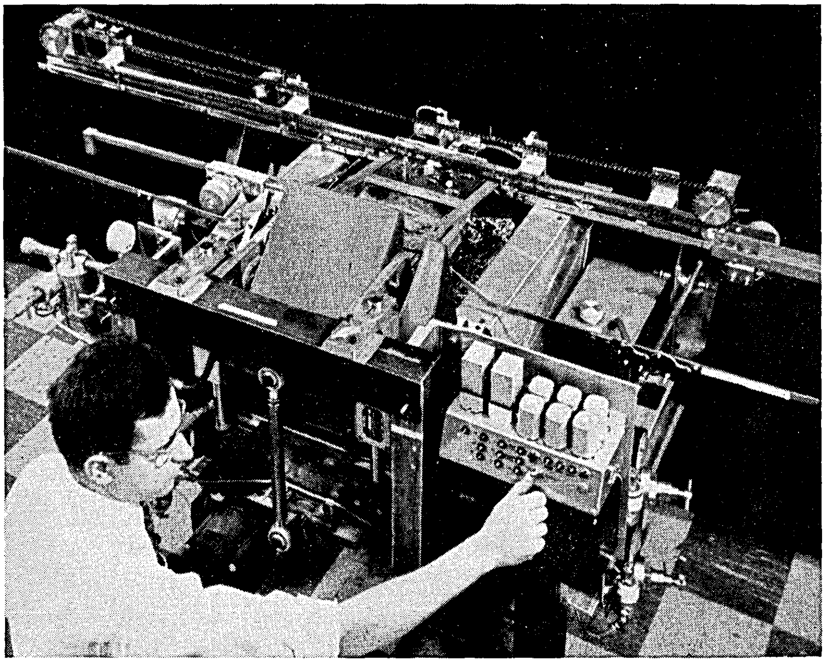


Fig. 13.38 Automatic dip-soldering machine

After all the components are inserted, the circuit boards are dip-soldered in the next operation. The dip-solder station is shown in Fig. 13.38. The position of the board is shown immediately before dipping. The completed printed circuit board in the pallet has been transported from the circuit board storage immediately beyond the component placement station by a conveyor chain. As the pallet and board assembly move toward the solder station, they activate the flux station just preceding the solder pot which sprays flux on the underside of the circuit. Immediately before the board enters the solder station, the dross wiper wipes the accumulated oxides above the solder ladle which is immersed in the molten solder. With the arrival of the board, the ladle is lifted vertically and the pallet is brought down onto the surface of the solder in the ladle to accomplish dip soldering. A rolling motion of the circuit board allows the flux to flow away from the line of contact between the solder miniscus and the board, thus preventing voids in the soldering. The board is vibrated while in contact with the solder to facilitate soldering. Then the board is rotated away from the solder bath and vibrated to remove excess solder while the solder is still molten. After the solder has solidified in about 25 seconds, the board is moved to the pretest storage. A complete cycle time for soldering is about 38 seconds minimum.

After assembly and dip soldering of the printed circuit board, the assembled circuit is ready for finished board testing. Since all components have been tested before assembly, and the component transport has been interlocked to prevent misalignment of the components during their placement cycle, there are three conditions that would cause failure of the printed-circuit assembly. These are (1) the component leads have not gone into their holes; (2) the component leads have not been soldered; and (3) solder has shorted out printed wires on the printed circuit board.

The primary function of the final, board test station then is to locate the boards in a batch which contain one or more of the defects above and to identify the type and location of the fault. Two types of testers are required to determine the acceptability of the printed circuit board. These are a shorts tester, which measures an extremely low impedance between any one printed wire and the others, and a continuity tester, which is capable of contacting the component leads and measuring a low impedance between one lead of a component through a solder joint along the printed wire and through the lead of another component soldered to the same printed wire. Thus, the first test detects shorted conductors; the second detects whether or not a lead has gone into a hole and whether it has been soldered, but the test does not determine the quality of the solder bond.

To refer to Fig. 13.32 again, the master programming unit or card reader controls and synchronizes the entire operation. The master programmer is divided into two parts, the card-punching section used to make up the system-programming punched cards, and the card-reading section used to sense the coded data on the cards and distribute these data to the machine stations in the system.

The function of the Remington Rand card-punching machine is to transform all necessary system control data into punched-card coded form. The system control data consist primarily of the following: (1) complete specifications for the components to be prepared and assembled, including data on the various preassembly tests to be formed on the components; (2) data to control the order of preparation and assembly of each circuit element; (3) data specifying the number of printed circuit boards to be assembled in a particular batch; (4) data specifying the accurate location of each circuit element on the printed circuit board; and (5) data specifying the tests to be performed on the completed circuit board assemblies. The card-reading section of the master programmer fulfills the function of sensing data on each program card. Transmission of these data sets up the control circuits of the machines to cause them to perform according to the command data on the cards. Figure 13.39 shows the monitor control and display console which provides means for an operator or supervisor to override the automatic controls in all the system variables, in addition to controlling the power system and the air supply system to the automatic component assembly system.

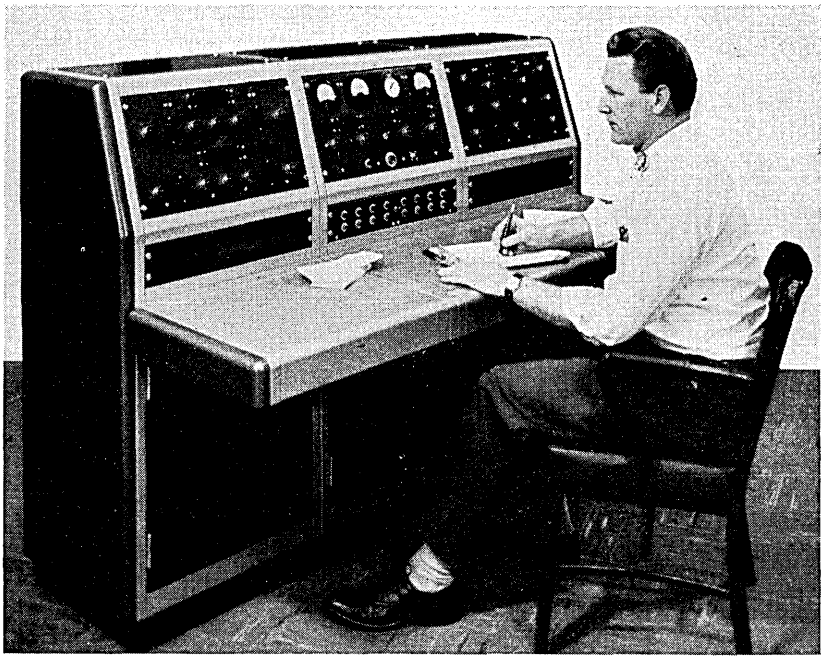


Fig. 13.39 Monitor control and display console for the supervision of the General Electric automatic assembly system

It is evident from Fig. 13.39 that the General Electric automatic assembly system is much more complex than the ones previously described. The fact that the component assembly machine is a multicomponent CAM, or a machine which can install sequentially several types or values of components, and one which has an index system, makes this machine a more complex one basically. To date, preliminary system studies to evaluate the advantages of the single-component CAM compared to the multicomponent CAM have led General Electric to select the latter. The claims of greater systems flexibility for production changes, simpler materials-handling problems, less storage space needed for assembled units, and versatility to adapt to small production lots have been made for this approach. It is difficult to perform a sufficiently thorough economic and technical analysis to determine the advantages of the one approach over that of the other at this early date without sufficient details such as performance data, reliability of equipment, maintenance costs, and initial costs. The question, "Is it cheaper and are the production problems fewer when a large number of simple machines are used in an assembly line, or is it better to use fewer, more complex, and more expensive machines?" remains unanswered until there can be a thorough evaluation of the many details which become apparent only when some of these production lines are built and operated. Another question which all electronic-equipment assemblers are asking is, "What production volumes and lot sizes must be maintained before it is economical to use one or the other component assembly systems?" Again, we can only await additional operating experience.

13.9 PROJECT TINKERTOY

Project Tinkertoy was sponsored by the Bureau of Aeronautics, United States Navy, at the Bureau of Standards. A series of some twenty machines have been constructed to provide a demonstration of feasibility of mechanizing an electronic-equipment production facility. The construction techniques used to fabricate the electronic package by automation production are quite different from most others. Instead of using the evolutionary approach and working with available electronic components as most of industry has done, Project Tinkertoy started from scratch and carried the manufacturing process from the raw materials such as blending steatite powder for the circuit wafers through the completed module or package, except for such items as tubes and components not suitable for the process to fabricate. Electronic components such as resistors, certain capacitors, and inductors were fabricated as part of the overall assembly process. Figure 13.40 shows the circuit wafers, capacitors, tape resistors, tube socket, and completed module.

We may consider the production line as having essentially four main branches operating in parallel which fabricate and feed components such as (1) circuit wafers, (2) tube sockets, (3) ceramic capacitors, and (4) tape resistor elements to the main production line, which then assembles these components to form the completed module.

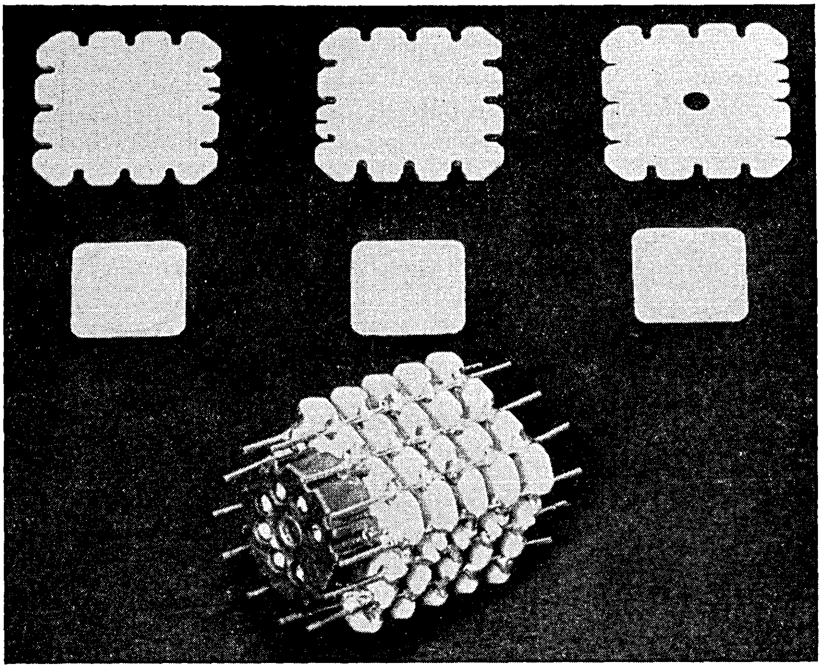


Fig. 13.40 Tinkertoy module, circuit wafers, and components

The circuit wafers are pressed from a mixture of ceramic raw materials which have been previously blended. The raw wafers are fired by being passed through a tunnel kiln at about 2300° F for 9 hours, after which each wafer is mechanically gaged for proper dimensions. The wafers are then passed through a silk-screening machine, shown in Fig. 13.41. The ceramic wafers fed from six vibratory feeders are conveyed to the stenciling screen which applies the patterns simultaneously on one side of the wafer to form a part of the conductive or inductive circuit. The silver paint is then fired by passing the wafers through a low-temperature furnace at about

800° F. The wafers fed from the vibratory feeder, which also orients them, are given a circuit pattern check by the machine shown in Fig. 13.42. This machine is set up to test each pattern by inserting a punched card into the programming unit on the left. The printed wafers are then transported to the main production line for the installation of components.

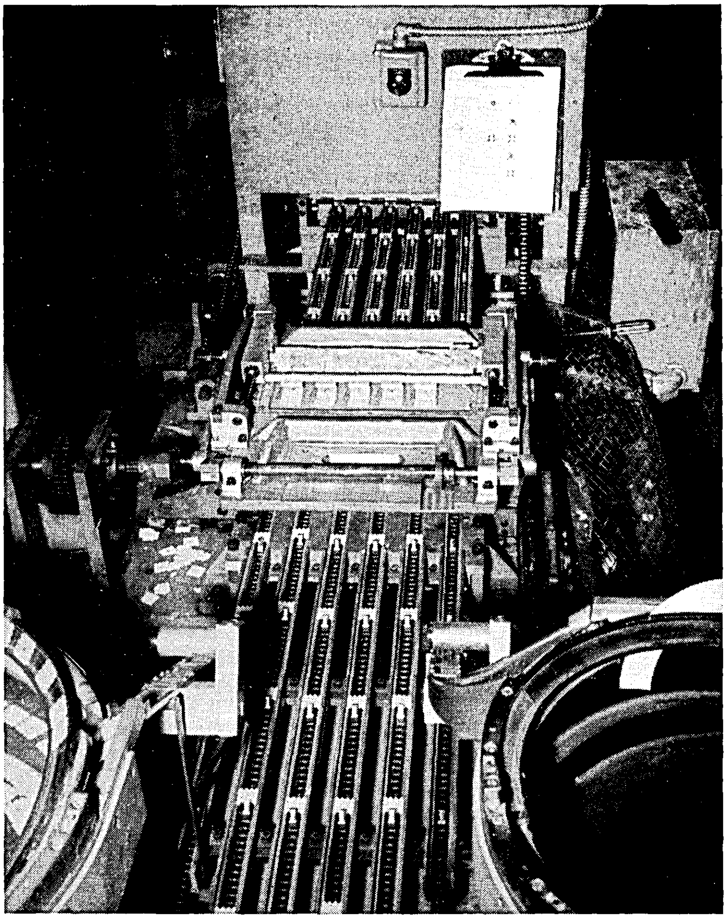


Fig. 13.41 Screening machine which simultaneously prints circuit pattern on six ceramic wafers indexed and fed from vibratory hoppers

With almost the same processes, the titanate capacitor bodies are pressed, fired, and silvered. The finished capacitor is 0.5 inch square by 0.020 inch thick with a capacity range from 7 micromicrons to

10,000 micromicrofarads. The capacitors are inspected by automatic machines for production tolerances in capacity, leakage resistance, and voltage breakdown. Present production techniques give a yield of about 80 per cent. The capacitors are then transported to the main production line.

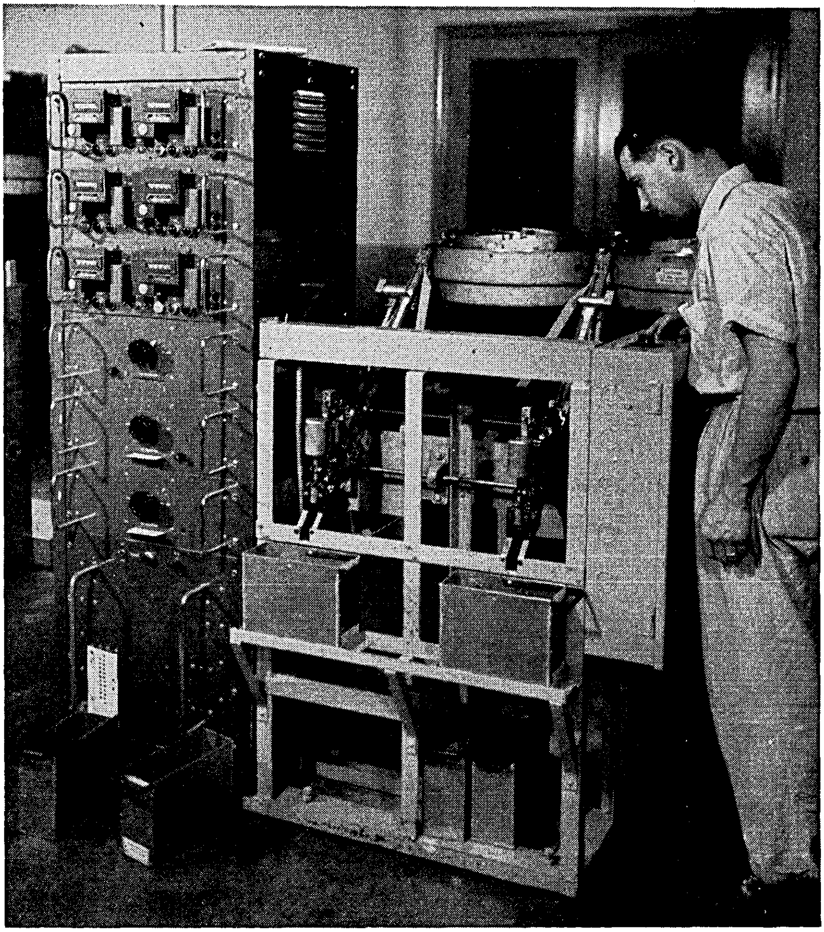


Fig. 13.42 Machine for testing circuit patterns—the test is set up by inserting a coded punched card into the programming unit on the left

The resistive element of the tape resistor is formed by spraying a thin film of resistive ink on asbestos paper tape. The ingredients of the ink include carbon, resin, and a liquid vehicle. Resistors produced from these tapes are within a production tolerance of ±10 per cent up to nearly 400° F and can dissipate 5 to 10 watts per square inch of surface in the usual applications. It is claimed that these resistors will pass most, and in some cases exceed, JAN-R-11 requirements. After the tape is dried, it is supplied to the main production line in reels ready for application on the circuit wafers.

Tube sockets molded with the same steatite ceramic as that used for the circuit wafers are also supplied to the main production line as a prefabricated component. The main production line then assembles these four types of components to form the module. The ceramic wafers are first taken to an automatic dip-soldering device which tins the side slots in preparation for subsequent circuit attachment. The resistor attachment machine applies the tape from the reels, cuts the tape to length, and bonds the resistive elements on to the ceramic wafer by curing the adhesive. Figure 13.43 shows a machine which feeds and orients the ceramic capacitors and the circuit wafer for induction soldering of the assembly.

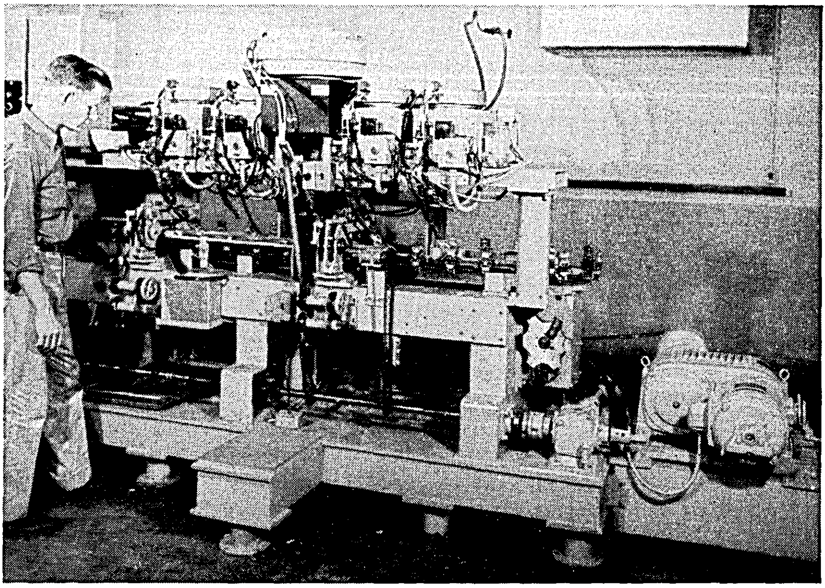


Fig. 13.43 Machine for assembling and soldering capacitors to the wafers—the two small vibratory feeders on either side of the large feeder deliver capacitors positioned for soldering; the center hopper feeds the circuit wafer into position and the assembly is induction soldered

As many as six different circuit wafers may be handled by the joining machine which performs the final assembly of the wafers into a module or subassembly. Six feeders carry the wafers to a slotted loading device which indexes, sequences, and holds the wafers in an upright position. Next, a chain drive carries the wafers in a jig to the soldering position where three wires on each side are guided into the pretinned notches in the sides of the wafers. Instant-heating soldering fingers are brought in contact with the wiring and pretinned notches, bonding the wiring to the wafer. In a subsequent operation,

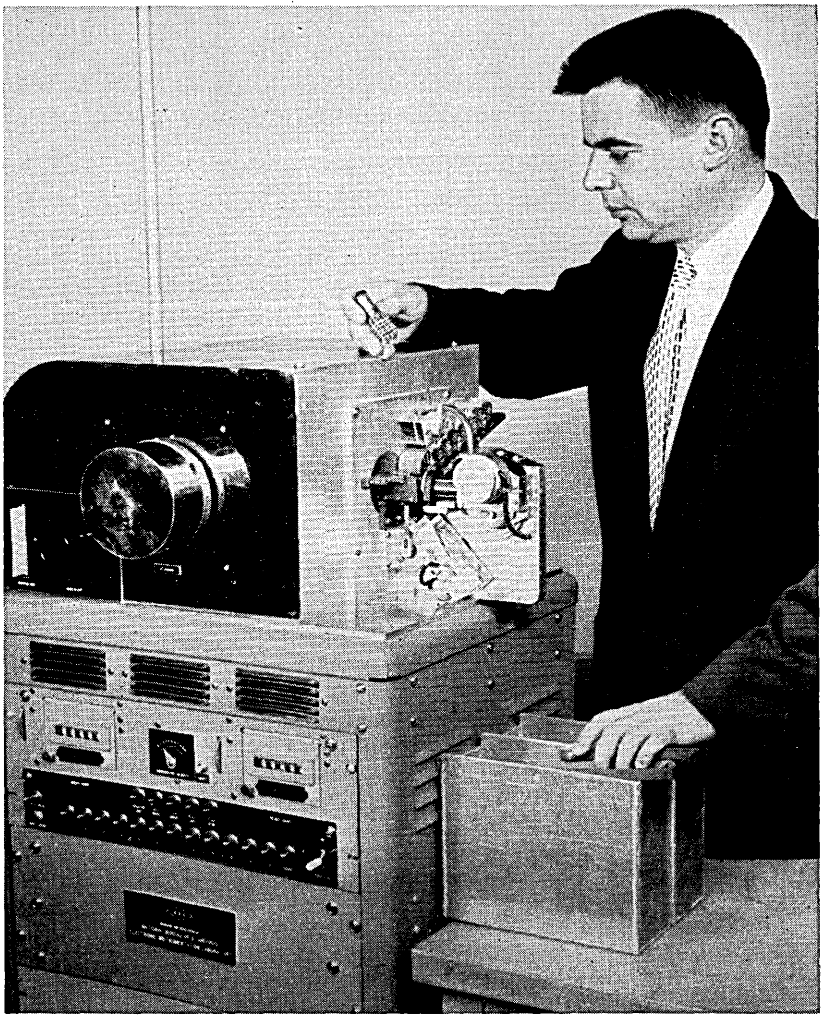


Fig. 13.44 Circuit checker for completed module; the tester is set up by a punched-card programming unit and tests the module through the tube socket

after rotating the wafer edges 90 degrees, the same machine transports the module to its next station where the complete procedure of soldering is repeated so that all four edges of the wafers are tied together as in a ladder. The completed module is then transported to the next machine which automatically checks the circuitry through the tube socket. This machine is shown in Fig. 13.44. A punched card programs the checker, and the electronic system inspects each circuit or component mounted on the wafer and compares it with a standard.

Obviously, the Tinkertoy approach is quite different from all the others. First of all, Tinkertoy fabricates the electronic components as part of the module assembly process instead of using available components. Secondly, Tinkertoy has taken an approach that is revolutionary in its effect on present production methods. Thirdly, the production design and the product configuration are such that components other than the flat tape resistor and thin ceramic capacitor are very difficult to install. Of course, one advantage of a revolutionary approach is the fact that it is not encumbered by past practices. The second advantage of the Project Tinkertoy approach is that it uses relatively low-cost materials to form both the components and the circuit wafer, and these materials in general are not in short supply. However, the difficulties of obtaining sufficient economic data to justify an extreme approach often prevents its industrial acceptance.

13.10 CONCLUSIONS

In evaluating the many approaches taken by industry to mechanize the fabrication of electronic assemblies, we recognize the far-reaching effects of automation in all phases of industry. Not only must the technologists work closely together and become expert in materials, processing, machine design, and electronics, but management, sales, and production must work hand in glove to insure a minimum of disruption to the production procedure once the machinery is set in motion to produce a given item. In spite of all efforts of the technologist, a mechanized line is less flexible than a manual line in the sense that it is not economical to frequently modify or change the machines in the production line to encompass changes in the product. Aside from the managerial problems, which are formidable indeed, the most difficult problem of automation is techno-economic in nature. This problem is concerned with the degree of flexibility, hence complexity and cost of the machinery in the automatic line. The industrial approaches to the solution of this problem are as varied as there are pioneers in this field. The final answer must await actual operating experience and the passage of time.

For the present, the following general conclusions may be drawn.

(1) The close integration of materials processing, product design of the electronic package, and its fabrication by machinery is gradually removing the distinction between many electronic-component manufacturers and equipment assemblers.

(2) The effective use of automatic production techniques requires much closer working relations within a given plant, between the market and economic analyst, the researcher, the design engineer, and the production specialist.

(3) The production line must be synthesized by the use of versatile and functional machine units to allow gradual assimilation of these units in the evolutionary process toward complete automation.

(4) Automatic production lines for military electronic equipment must be suitable for the production of commercial electronic equipment; then these production lines can serve as stand-by facilities for the rapid acceleration of production, in time of emergency, with a minimum of conversion and manpower needs.

(5) There is much need to bring to the attention of the machine-tool, materials-processing, metal fabrication, and other industries the problems of automating the electronic industry.

(6) Since the cost of materials and components accounts for 75 per cent of the manufacturing cost of an electronic assembly, much additional effort must be expended to reduce the cost of components. Developments such as printed-circuit television-tube deflection coils are pointing the way toward this goal.

(7) The most critical problem in automatic production is that of quality control and product testing or inspection. We must develop methods for quickly and automatically checking the quality of materials or components in process and testing the performance of the completed electronic package. The difficulty of the problem lies not so much in the building of automatic test or control equipment, but in determining what physical characteristics are to be measured to provide a key to the reliability of performance of the component or the assembly.

(8) Perhaps the most difficult part of thinking in terms of automatic production is recognition of the fact that the product or the fabrication process, or both, must be redesigned in order to make automatic production feasible. We must look at the function of a product and think about building it by automatic machinery.

(9) Although many technical problems must still be solved, and the detailed economics must await additional information, automation of the electronics industry is not just around the corner; it is here. The first attempts may be crude but, with sufficient thought unencumbered by convention, rapid strides are being made in the development of new fabrication techniques and the application of new materials. Printed circuits is but a faltering first step.

13.11 REFERENCES

1. Automatic Component Assembly System, Interim Reports 1 and 2, Contract DA-36-039-SC-30250, Electronics Division, General Electric Co., Syracuse, N. Y., Feb. and Nov. 1953.

2. Automatic Component Assembly System, Quarterly Reports 1–5, Contract DA-36-039-SC-30250, Electronics Division, General Electric Co., Syracuse, N. Y., 1954.

3. Automatic Component Assembly System, Special Report, Supplement to Quarterly Report No. 2, Contract DA-36-039-SC-30250, Electronics Division, General Electric Co., Syracuse, N.Y., Feb. 1954.

4. Mechanized Production of Miniature Electronic Packaged Subassemblies Using Modular Construction, Final Report, Contract NObs-5235, Melpar, Inc., Alexandria, Va., Feb. 1953.

5. Automatic or Semi-Automatic Production of Electronic Packaged Subassemblies, Interim Reports 9–12, Contract NObs-5235 Addendum-1, Melpar, Inc., Alexandria, Va., Aug. 1954.

6. R. L. Henry and H. H. Rosen, Summary of Modular Design of Electronics and Mechanized Production of Electronics, Vol. I, PB-111275, Contract NBS 12.05-20-5532, National Bureau of Standards, Washington, D. C.

7. H. S. Horiucki and R. Burroughs, Hand Fabrication Techniques and Photographic Processing for Modular Design of Electronics, Vol. III, PB-111277, Contract NBS 12.05-20-5532, National Bureau of Standards, Washington, D. C.

8. Electronic Packaged Subassemblies, Using Mechanical Methods for Electrical Interconnections by the Use of the Stewart-Warner Electric Mechanical Assembly Technique, Interim Reports 1–4, Contract NObs-5236 (1733), Stewart-Warner Corp., Chicago, 1951–1953.

9. W. H. Hannahs and J. A. Caffiaux, Automatic Assembly Techniques for Miniaturized Transformer Coupled Circuits (Mixer Amplifier AM 142/A1C), Technical Report No. 6628, Contract AF 33(038)-9350, Sylvania Electric Products Inc., Bayside, N. Y., Oct. 1951.

10. Electronic Equipment Construction, Final Report, Contract N7onr-32103, Stanford Research Institute, Stanford, Calif., June 1950.

11. Electronic Packaged Subassemblies, Automatic or Semi-Automatic Production of, Using Printed Circuits, Final Report, Contract NObs-5234, U.S. Electronics Corp., Oct. 1952.

12. F. Hom and L. K. Lee, Development and Application of Automatic Assembly Techniques for Miniaturized Electronic Equipment, Final Report, Contract AF 33(038)-18976, Stanford Research Institute, Stanford, Calif., Feb. 1952.

13. F. Hom, E. Gamson, and R. Newton, Development and Application of

Automatic Assembly Techniques for Miniaturized Electronic Equipment, Interim Reports 1–8, Contract AF 18(600)-141, Stanford Research Institute, Stanford, Calif., 1952–1954.

14. E. H. Wavering, Big Step Toward Automatic Assembly, *Factory Management and Maintenance,* Oct. 1952.

15. A Fortune Round Table, The Automatic Factory, *Fortune,* Oct. 1953.

16. L. Lee and C. Brunetti, Automatic Assembly of Electronic Equipment, *Elec. Manufacturing,* 97, July, 1953.

17. L. Lee and F. Hom, Automatic Production of Electronic Components, *Radio and TV News,* Dec. 1953.

18. W. Hannahs and W. Serniuk, Miniaturized Amplifier Designed for Automatic Assembly, *Elec. Manufacturing,* 88, Aug. 1951.

19. Harvard Business School, The Automatic Factory, *Elec. Manufacturing,* 246, March 1952.

20. *Proceedings of Symposium on Automatic Production of Electronic Equipment,* sponsored by Stanford Research Institute and United States Air Force, April 1954.

14. Process Control in the Petroleum and Chemical Industries

C. G. LASPE

Shell Oil Company
Wilmington, California

14.1 INTRODUCTION

Automation has been the key to the remarkable growth of the present-day process industries, among which are included the petroleum, the chemical, and the petro-chemical. Existing complex processes such as catalytic cracking in the petroleum industry or the atomic power reactor in the nuclear field would be economically unfeasible, if not physically impossible, without the use of modern automatic control equipment. This chapter will endeavor to present a brief background of the problems existing in these industries as well as to recount the developments that have resulted in the present high degree of automation in the modern processing plant. Recent developments foreshadowing the automatic processing plant of tomorrow will also be reviewed.

This chapter will consider typical process variables which must be measured and controlled. These variables may be placed in one of two categories. In the first group are found the operational variables. In the second are placed those that define a physical or chemical property of the product and may be termed definitive variables. To maintain these variables at a desired value, elaborate feedback control systems have been devised. Systems commonly encountered in the process industries range in complexity from the simple single-variable system to the highly complex, coordinated control system. After a description of the various control methods is given, this chapter will touch on recent applications of servo techniques to the analysis of these control systems, and examples will be given showing the accuracy to be expected. The remainder of the chapter will be devoted to the discussion of recent developments in automatic control equipment. This discussion will include recent innovations in pneumatic devices as well as a description of complete electronic control systems. The integration of data-handling equipment into the process control loop will be presented. The final section of this chapter will describe the application of continuous analyzers to determine both product composition and/or product quality.

14.1.1 Characteristics of the Process Industries

The process industries are those that handle bulk solids, liquids, or gases in some form and modify these materials either by physical or chemical means to produce a finished product with the desired properties. Many of these processes are strictly physical in nature. Some are chemical and some are combinations of the two. The commonly encountered physical processes include fractionation, extraction, blending, filtration, etc. The typical chemical reactions include those of cracking, isomerization, hydrogenation, neutralization, etc. As stated before, however, most processes include both physical and chemical changes.

14.1.2 Equipment

To get a better understanding of the control problems encountered, a brief consideration of the equipment peculiar to the process industries is in order. First, vessels are required for storage of the materials in the process, such as the raw feed, the finished products, and the intermediate products. Reaction vessels are required to provide the reaction or holding time necessary to process the material. Other vessels are needed for physical separations such as fractionating columns, settlers, flash chambers, etc. A second class of equipment may be termed heat transfer equipment. In this list we would find heat exchangers, coolers, furnaces, and kilns. Equipment is also needed for transferring materials from one section of the process to the next, through heat exchangers and through fractionating columns. This transfer equipment takes the form of pumps, compressors, and conveyors. Lastly, the interconnecting piping through which the material flows is needed.

Each item of this equipment is an integral part of the process. Each plays its own part in the control loop. Therefore, for adequate control system design, the characteristics of each piece of equipment must be accurately known. In view of the foregoing, it is apparent that the definition of a process must be carefully reassessed. Very simply the process is a cumulative effect of the equipment plus the material to be processed plus the nature of the chemical or physical reactions that are to take place.

Henceforth, when speaking of the process, it will be considered in the light of the definition above. Most processes can be described for purposes of control as having one or more time constants, dead time, or other nonlinearities, and an overall process gain. Representative values for typical processes will be presented in Section 14.2.

14.1.3 Process Types from the Operational Standpoint

Processes can be classified either as batch, continuous, or semicontinuous. As the name implies, continuous processes are those in which the material continuously moves into the plant, through the processing section, and continuously out of the plant. In general, all parts of the plant are in stable equilibrium and do not change with respect to time. If the stable equilibrium is altered owing to some unforeseen load change, the transient return to stability is known as a plant upset. Generally speaking, continuous-process equipment is small. Little intermediate storage is required, but the process control and regulation must be good.

Batch operation on the other hand refers to that type of operation in which a fixed amount of material is charged to the equipment, in which a desired reaction or physical separation is carried out. After completion of the desired processing, a fixed amount or quantity of the product is removed from the equipment. The requirements for batch processing, especially equipment-wise, are generally opposite to those for the continuous process. The equipment is larger, which of course results in a greater initial capital outlay for these items. The control requirements are generally somewhat less severe than in the continuous process. However, this may not necessarily be true. One feature of the batch process is the isolation provided between separate processing units. This is the main reason for less stringent control requirements of the batch process, since a disturbance in any one unit is effectively blocked before reaching a second unit.

As this chapter evolves we may note the emphasis placed on applications in the petroleum industry. This is the field in which the author is most familiar. However, since similar problems are encountered in the chemical industry, the same basic control theories are equally applicable. Actually the petroleum industry may be considered as a specialized branch of the chemical industry.

14.2 OPERATIONAL VARIABLES MEASURED AND CONTROLLED

The term operational variable refers to the actual physical or environmental conditions existing in the process equipment. These physical conditions include the various flow rates of the materials entering or discharging from the plant; the pressure existing in a vessel or fractionating column; the temperature existing therein; other environmental conditions such as humidity. All of these are referred to as operational variables. Last but certainly not least, is the control of liquid level. Good regulation of this latter variable is extremely important in keeping a stable operating condition in the continuous plant.

A wide gamut of process time constants and process orders is encountered among the operational variables. A comparison of the control characteristics for typical systems is given in Table 14.1. Note

Table 14.1 Control Characteristics of Typical Processes

Process	System Order *	Typical Range of Time Constants
Temperature	Generally second	30 sec to 10.0 min
Pressure	First or second	1 sec to 10.0 min
Flow	First	0.1 to 5.0 sec
Liquid level	First	1.0 to 10.0 min
Rotary speed	First	0.2 to 10.0 sec
Viscosity	First	0.5 sec to 5.0 min

* Refers to order of describing differential equation.

that liquid flow processes are generally characterized by an extremely short time constant and may be represented as a first-order system. The process would be infinitely fast were it not for the elasticity of the piping and the compressibility of the fluid flowing. A small amount of dead time or transportation lag results from the physical separation of the measuring point from the controlling point. This dead time is equal to the speed of sound in the fluid media. Liquid flow rates are generally controlled by maintaining a fixed pressure drop across an orifice or some other type of restriction in the process piping. The secondary or sensing element takes on the form of a differential pressure-measuring device such as a mercury manometer, or more recently, the mechanical-type transducers.

Although the manometer or differential-head meter is most commonly found in the process industries, other means are available for fluid flow control. Positive displacement meters find wide application where precise flows must be measured and maintained. The common meter at the gasoline pump is an example of this. Further, the recently developed mass flow meters are finding application where mass rates must be accurately measured and controlled. These latter type meters should be particularly adaptable where two-phase flow is present, such as in slurries or in vapors entrained with liquid.

A second major operating variable, pressure, is generally controlled by the rate of material addition or removal from process vessels. Hence, the time constants of pressure processes depend on the size of the vessels, as well as the flow rates into and out of them. These time constants will vary between fractions of a second, for extremely fast processes, to several minutes for the slower ones. If several vessels are connected in series, then several time constants may be involved. This will result in a control problem that is somewhat more difficult to solve.

Perhaps one of the most important variables from the operational point of view is that of temperature. Temperature generally dictates the amount or degree of reaction in the chemical process. Again in the physical process, temperature may effect to a very great extent the degree of separation. This is especially true in fractionation. This important variable is controlled by the modulation of heat or material input to the process. Heat input implies a heat transfer from a heating source to the process fluids. Generally, owing to the thermal resistance at the heat exchange surface, this heat transfer will be a slow process. Also, the thermal capacity of both the supply and the demand enter into the rate of heat exchange. If the supply capacity is large, then the time constant of a thermal process may be considered as equal to the product of the thermal-resistance and the thermal-demand capacity. It is evident then, that the time constant may vary between a few seconds to, perhaps, several minutes. When the heat transfer rates are high, as with clean liquids at high velocities, we expect to find short thermal time constants. Conversely, with slow-moving, viscous, or dirty liquids, a high thermal resistance limits the rate of heat transfer, which may result in a long time constant.

However, most thermal systems are not adequately represented by a single time constant. In order to insure protection to the primary measuring element it is often placed in a protecting sleeve called a thermowell. Thus, a second time constant may be added to the system in series with the first. Means for eliminating this second time constant will be discussed in Section 14.5.

The importance of liquid-level measurement and control is obvious. In a reaction vessel, a level is required to provide holding time. In an accumulator which provides surge capacity, liquid-level control is required to prevent the vessel from running over or from becoming completely emptied, resulting in a loss of pump suction. The method of control is also obvious—control is effected merely by changing the addition or withdrawal rate of fluid to the vessel. Holding times for various level systems may lie between a few seconds to many minutes.

14.2.1 Definitive Product Variables

Definitive product variables are those that describe properties of the product such as specific gravity, color, vapor pressure, end boiling point, refractive index, viscosity, *p*H, etc. Most of these variables are physical in character and are included directly as part of the product specification or are correlatable with the product specification. Many of these variables are now determined by laboratory-type instruments which operate on the batch-wise principle. This is particularly true of end boiling point measurements and vapor pressure measurements. To be employed in the process automatic control loop, however, these variables must be measured on a continuous basis by continuous flow-type analyzers. In many cases analyzers cannot be built which will measure the product property in exactly the same manner as the laboratory-type instrument. In these instances, an indirect method of determining the desired property is employed. Even though these latter continuous-flow analyzers do not measure product specification directly, very frequently the instrument output can be correlated directly with laboratory determinations. Sections 14.6 and 14.6.1 of this chapter will discuss recent developments in these analyzers.

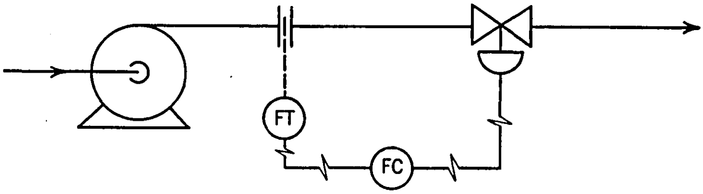
Product composition may be considered as the ultimate variable that must be measured and controlled. Many new and ingenious analyzers are being developed and applied to the continuous measurement and control of composition. Although product composition is generally considered a chemical property and is determined by chemical means, in many instances it may be inferred from a physical measurement. To name a few of the more important physical methods we find infrared spectrophotometry, ultraviolet spectrophotometry, mass spectrometry, nuclear magnetic resonance, emission spectrophotometry, and the radio-frequency spectrograph. All these instruments determine a physical characteristic, either of the product or of the predominant impurities. There are other instruments such as the titrimeters, the operating principle of which is based on the measurement of chemical properties of the product. These analyzers will be considered in more detail in Section 14.6.1.

All of the definitive variables are controlled by changing one or more of the operating variables. The ability to control depends upon the existence of a definite relationship between the various operating variables and the product quality. The relationship above, among other things, is a function of the feed stock or raw-material character as well as a function of the equipment and its limitations.

To review briefly, control of any one of the operating or definitive variables can be considered as the regulated modulation of either energy or material input into the system. This concept of energy modulation is a very important one and will be encountered again and again during the study of Section 14.3, which deals with basic forms of control systems as they are found in the process industries.

14.3 SINGLE-VARIABLE CONTROL SYSTEMS

The single-variable control system is basic. It can be considered as a building block upon which all other systems are built. It is the most common in industry today and represents perhaps 90 per cent of all control systems which are encountered. This basic system may be defined as one in which a single variable is measured, which in turn provides the control signal to modulate a single energy input to the system to maintain that variable at the predetermined set point. For example, in Fig. 14.1 a typical orifice meter flow control system is shown. The primary element, an orifice or other restriction, in the main line produces a pressure drop which is a function of the flow rate. The measuring element, which is generally some form of manometer, measures the differential head produced across the primary



FT–flow transmitter.
FC–flow controller.

Fig. 14.1 Typical single-variable control system element. In this particular example the measuring element has associated with it a transducer which converts the differential head into a proportional pneumatic signal. This pneumatic signal then is fed into a pneumatic controller where it is compared with the reference input or set point. The controller then operates on the error signal thus produced and provides an output signal in the form of a pneumatic pressure change. This modified error signal or control signal, as it is called, becomes the input signal to the final control element which in this instance is a control valve. Process piping then completes the control loop.

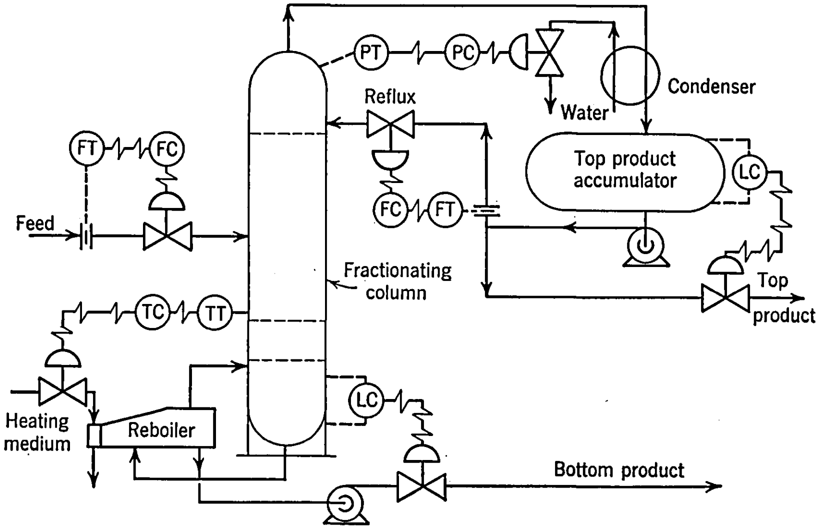
Note that in the example above, control was effected by modulation of energy input. In this particular instance, kinetic energy was involved. Factors affecting the control of the single-variable system include both static and dynamic quantities. The static factors include those of overall system gain; system nonlinearities, such as limit stops, or saturation; and the process characteristic known as self-regulation. Dynamic factors would include the time constants, dead times, and the system order—in other words, the frequency variant part of the system transfer function.

When two or more single-variable control loops affect the static or dynamic equilibrium of a process, the control problem becomes somewhat more complex. Perhaps the best example of a complex process in the petroleum industry is the fractionating column as shown in Fig. 14.2.

A note of warning must be injected at this point. The instrumentation shown in Fig. 14.2 does not represent a fixed or inflexible control scheme for all fractionating columns. Each problem must be assessed in the light of the particular job which the column must do. Hence, the instrumentation scheme of a neighboring column may not bear the slightest resemblance to that shown in Fig. 14.2.

Returning to Fig. 14.2, note that there are six single-variable control loops in this particular process. The feed to the column is maintained by a single-variable flow control loop. Likewise, the reflux to the top of the column is regulated by another flow control loop. Both the top- and bottom-product flows are determined by level controllers, first, on the top-product accumulator, and second, at the bottom of the column. Temperature is maintained in the bottom of the column by modulating the heat input again by a single-variable control loop. The pressure in this particular column is maintained by throttling the cooling water to the vapor condensers. This represents a case of throttling or modulating the energy removal.

Please note that in a complex process of this type, a change of any one of the six operating variables would momentarily upset the stable equilibrium conditions existing in the column. For example, if the operator were to increase the feed rate, the temperature in the bottom of the column would decrease, the pressure would decrease, the bottom level would increase, the top level would decrease. These variables would remain in the transitory state until such a time as the column reached new equilibrium conditions. Because of the interdependence of the variables, such a system is said to be cross-linked or interacting. If the exact relationship between the variables is known, then this interaction or cross linking may be effectively neutralized by the addition of several degenerative feedback loops. A similar method has been suggested by W. E. Phillips for the control of turbojet engines (6).



PT–pressure transmitter.
PC–pressure controller.
LC–level controller.
TT–temperature transmitter.
TC–temperature controller.

Fig. 14.2 Typical fractionation column control system

14.3.1 Cascade Control Systems

Although entirely adequate for many industrial applications, the single-variable control system suffers certain deficiencies when the operating variable is subjected to load changes and must be controlled precisely within limits. The deficiencies of the single-variable control system have, in many instances, been removed or minimized by the cascade control system. By definition a cascade control system is one in which the set point or reference input signal is the control or output signal from another controller. An example will best serve to illustrate this principle. Figure 14.3 shows a temperature control system on a product stream leaving a direct-fired furnace. The temperature is measured at the exit of the furnace and is compared to the reference or set point signal at the controller. The output of the temperature controller instead of directly modulating the energy input serves as the input or reference signal to a secondary controller which maintains a constant gas pressure at the burners. Such a system is relatively independent of any load changes which may occur in the gas header pressure upstream from the diaphragm motor valve. Without the secondary control system, any changes in gas header pressure would result in a change in energy input to the furnace. Some time later, determined by the dynamic response of the furnace, the temperature controller would then sense the heat input change. It is readily apparent then that the cascade control system has a vast superiority over the single-variable control system in this particular application. In general, it will be found that cascade control will be of value when load changes are possible in the energy or material supply to the processing unit. In essence, cascade control eliminates the effect of specific disturbances upon the major control loop.

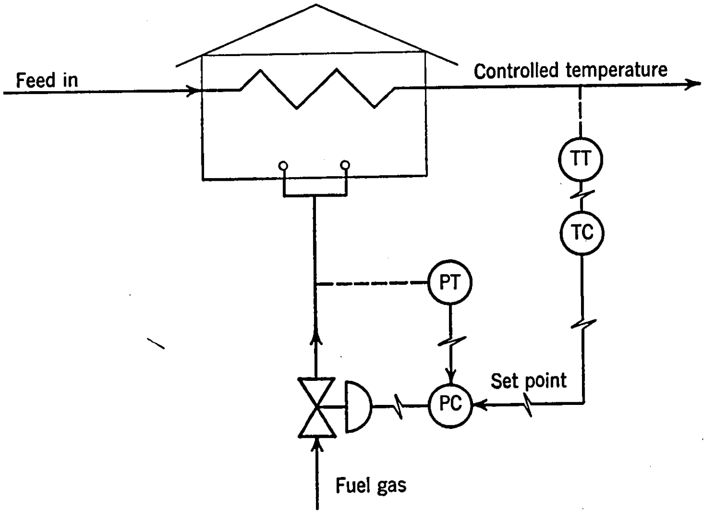


Fig. 14.3 Cascade control of a fired heater

Another example of a cascade control system is shown in Fig. 14.4. Temperatures at the top of the fractionating column are generally maintained by varying the reflux flow rate to the top of the column. In this particular example, the column-top temperature controller provides the set point signal to the reflux flow controller. If the flow controller were not in the temperature loop, then the temperature controller would operate the reflux valve directly. Any load changes in the reflux rate caused by pressure surges either upstream or downstream of the control valve would result in a change in tower-top temperature. Now, as we learned from Section 14.2, that temperature processes are relatively slow, this would result in a corrective action that would be long in taking effect. However, with the cascade control system, any load change due to a pressure surge would be immediately attenuated by the fast flow control system. The overall result is a temperature control system which is superior and capable of maintaining the tower temperature within close limits.

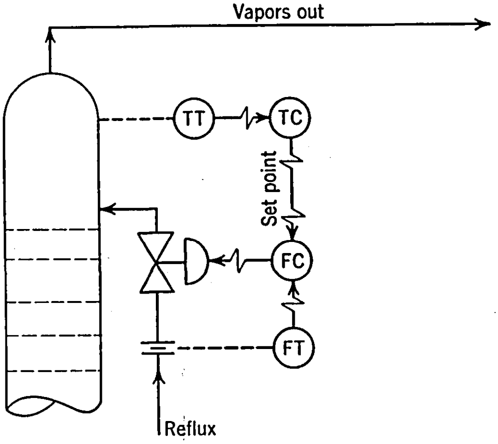


Fig. 14.4 Cascade control of temperature at top of fractionating column

Although the control systems discussed thus far are adequate solutions for a good majority of control problems encountered in industry, a more sophisticated answer is needed in some instances. This leads directly to the coordinated control system which will be studied in the next section.

14.3.2 Coordinated Control Systems

As the name implies, coordinated control systems are those in which the set point or reference input signals to several controllers are provided by a single master controller or computer. In this light it may be considered as a modified parallel-cascade system. Of the control systems studied thus far, the coordinated controllers are capable of the highest degree of precision and stability.

Modern combustion control of present-day steam boilers is a good example of a coordinated system. A typical scheme is shown in Fig. 14.5. Note that the master controller, in this instance the steam header-pressure controller, provides the reference input signal to both the gas flow control system and also to the air flow control system. Thus one reference input signal is supplied to each of the secondary controllers. In this case, the signal is the demand for a change in energy input to the boiler. Since there is approximately a one-to-one relationship between the energy input and the air required for combustion, only a single reference input signal is required. On the other hand, if the relationship between fuel flow and air flow were other than one to one, a separate input or command signal would be required by each controller.

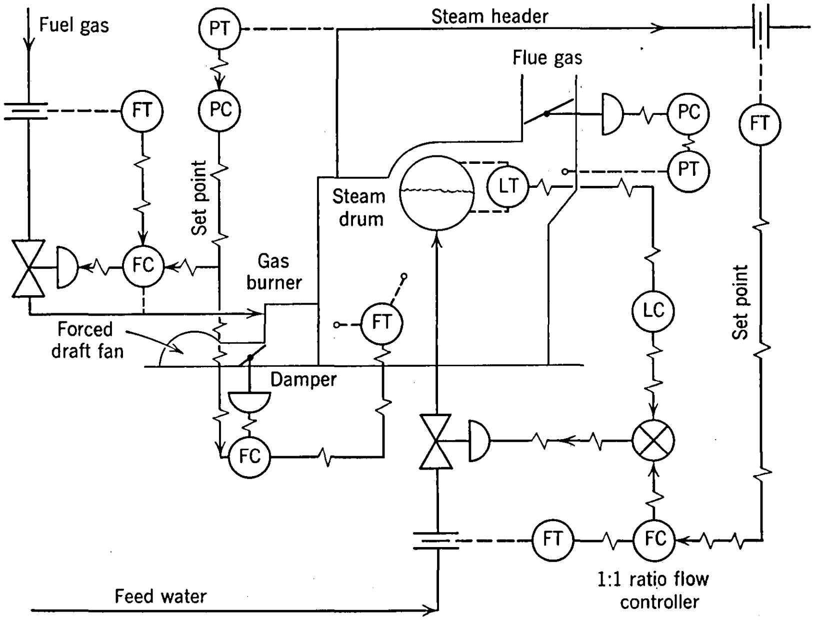


Fig. 14.5 Coordinated control on a modern steam boiler

Also associated with this particular boiler are several single-variable control systems. The draft controller maintains a given fire box pressure by operating a damper in the flue gas line. Note also in the fuel system that a pressure connection is made at the burners and is brought back to the fuel flow controller. The function of this particular connection is to operate a low-pressure limit control in the fuel flow circuit. Were it not for this low limit control, it would be possible for the fires to be completely extinguished, which of course would create a hazardous condition when the demand again called for the gas to be turned on.

Although not a true coordinated control system, the boiler feed water regulator shown in the same drawing is an interesting example of a multiple-element cascade control system. Note that a single-variable, feed water flow rate, is being manipulated. The main or primary control loop regulates the feed water. The set point to the flow controller is generated by the steam flow recorder-controller. This is an example of ratio flow control in which there is a one-to-one ratio between the two flows. The output signal of the ratio flow control system, instead of operating directly as the final control element, is modified by still a third controller—in this instance the steam drum water-level controller. In effect, the system operates as a one-to-one ratio flow controller with a trimming action superposed by the steam drum water-level controller.

14.3.3 Supervisory Control Systems

Supervisory control systems are those that limit the actions of the main process controllers. Included in the supervisory category are the alarms, overrides, limit stops, and also the safety shutdown devices. Most process variables have a definite limit beyond which it is undesirable for the variable to traverse. It then becomes necessary to provide some device for either initiating a control action or warning a human operator of the abnormal condition. If the abnormal condition does not result in a safety hazard to the operating personnel or in a completely ruined product, then a simple alarm is generally sufficient. However, where definite safety hazards are involved, process control systems may be equipped either with limit stops or safety shutdown devices.

Practically all of the supervisory control systems are complete on-off control loops in themselves. The output of these devices is in the form of a yes or no command to the main process controllers dictating whether they should function or not. These systems must be carefully engineered so that any automatic shutdown will proceed in an orderly and safe manner.

Later on in Section 14.5.3 of this chapter, the use of data-handling equipment to perform many of the supervisory functions will be discussed.

14.3.4 Computer Control Systems

The fifth, and most highly developed control system considered thus far, is termed computer control. In general the computer control system will be a combination of the first four types of systems and may contain one or all of them. In addition, a computer will be used to provide the several reference signals to the individual controllers. The position of the automatic computer in the overall control loop can best be illustrated by studying existing control practice. Figure 14.6 shows a fictitious processing plant in which several variables must be controlled in order to provide the optimum product yield and quality. In order to determine intelligently the set points or reference input signals for the individual variable controllers, the operator must have a complete knowledge of the feed stock properties as well as of the instantaneous properties of the product leaving the plant. In addition to these definitive variables, the operator must have at his command a knowledge of the relationship between his plant variables and product quality. With just this amount of information the plant operator can then produce the product that will meet the buyer's

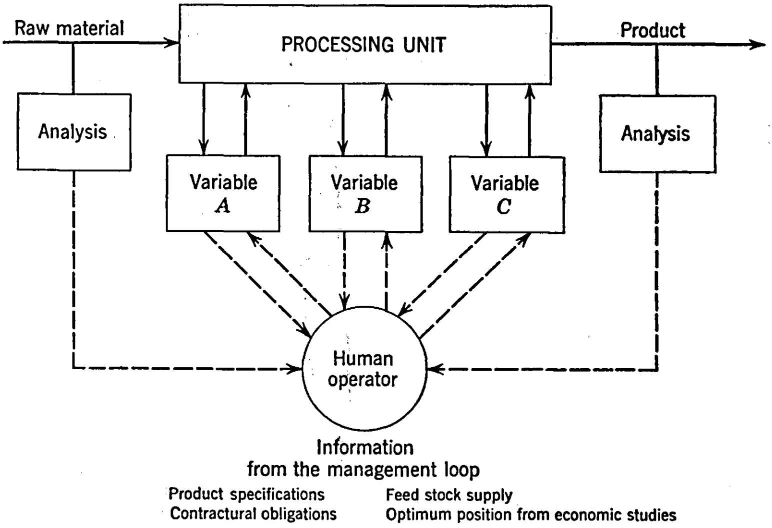


Fig. 14.6 Present-day control of a processing unit

specifications. Thus it is apparent that the first application for a computer would be that of replacing the human operator in so far as the plant relationships could be reduced to some definite formula. Note also in Fig. 14.6 that certain other pieces of operating information are utilized by the operator. These include instructions from higher management dictating the amount and kind of product to be produced. Also included would be information concerning equipment limitations, feed stock availability, and contractural obligations.

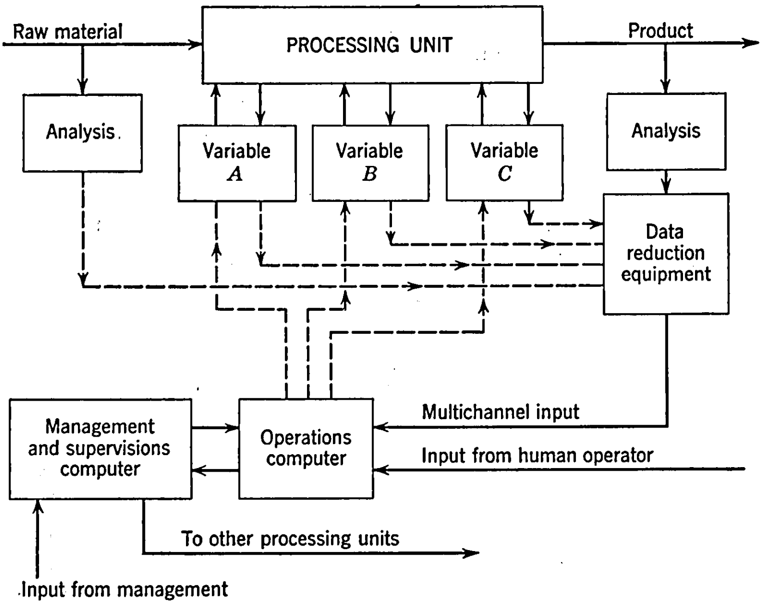


Fig. 14.7 The automatic processing plant

These latter data are derived from a management function which deals with a study of the existing as well as of the projected market demands. A second application for a computer, then, would be to replace certain management functions which deal with market analysis and optimum economic plant operation. It is interesting to note that several large oil refiners already have purchased digital computers for this very purpose. The outputs of these computers will still be in the printed form which requires a human operator to form a part of the control loop.

Figure 14.7 shows a block diagram of the completely automatic processing plant where routine decision making is taken from the human operator and placed in the computer. As has been stressed in previous lectures the computer-controlled plant will not be entirely devoid of the human operator. He will still be necessary for emergency operations as well as for startup and shutdown procedures.

An interesting point about the computer-controlled plant as described in Fig. 14.7 is that there is a definite possibility of feeding ticker tape directly from the stock market to the management computer which will allow continuous analysis of market conditions. Feedback instructions from sales surveys can also be fed into the management computer to predict market trends which will allow timely adjustments to be made in production rates to meet the market demands. It should be pointed out that there is in existence today no completely automatic processing plant. However, the technology is available.

14.4 SERVO TECHNIQUES TO EVALUATE THE DYNAMIC CHARACTERISTICS OF PROCESS EQUIPMENT

Within the past two years, there has been a rapid advancement in the application of servo techniques in the study of process control systems. Table 14.2 lists the various system applications which have been investigated to date. Included in this list are direct-fired furnaces, flow control systems, heat exchangers, fractionating columns, compressors, pneumatic transmission lines, and various and sundry control devices. It is interesting to note that controller manufacturers themselves are utilizing frequency response techniques in the design and development of new control instruments. In fact, it will not be in the too far-distant future that control equipment transfer functions will form a part of the control equipment specifications as set down by the instrument purchaser.

Table 14.2 Application of Servo Techniques to Process Control Systems

Process	Reference *
Liquid flow system	14
Heat exchanger	13
Pneumatic transmission line	8
Furnaces	Unpublished
Fractionating columns	12
Compressors	Unpublished
Instruments (controllers)	11, 15

* See bibliography at end of chapter.

To appreciate the magnitude of work represented in Table 14.2, let us briefly review the factors and equipment involved in obtaining the frequency response characteristic of a typical process. First, there is needed a sine wave generator compatible with the control system at hand. Most applications in the process industries require a generator whose output is in the form of a pneumatic sine wave. Before 1953, at least to the writer's knowledge, there was no commercially available pneumatic sine wave generator. All such equipment had to be hand-built and calibrated. This lack has been rectified by the recent appearance of several commercial generators. However, the cost of these instruments is still outside the range of many potential users. An important requirement of the sine wave generator is that it be continuously adjustable over an extremely wide range of frequencies from perhaps 1 cycle every half hour to perhaps 60 cycles per second. The amplitude of the sine wave must also be adjustable as well as its average pressure level. The latter adjustments are necessary to limit the magnitude of process changes occurring while the frequency response data are being obtained in the plant. Second, various transducers are required to convert the process variables into electrical or pneumatic signals capable of being handled on standard recorders. For fast processes it is imperative that electrical recorders be used to minimize the effect of transmission lines upon the observed data.

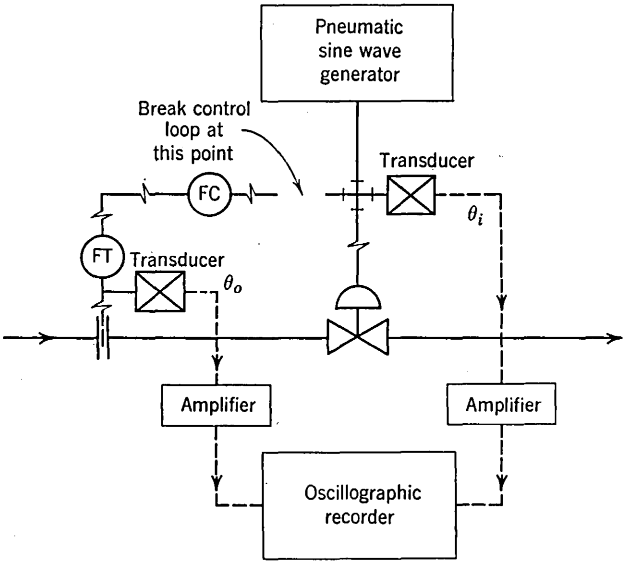


Fig. 14.8 Application of servo techniques to investigation of process dynamics

Data are normally obtained while the plant is in actual operation and producing a specification product. The manner in which the equipment above is interconnected with the control system is shown for a flow process in Fig. 14.8. After the equipment has been completely connected, the tests are ready to begin.

When the process is characterized by long time constants, relatively low frequencies must be investigated. If an approximation to the process time constant can be made, the corresponding break frequency may be calculated from the following formula:

$$f_b = \frac{1}{2\pi T} \qquad \text{cycles per second if } T = \text{seconds}$$

The "break frequency" is defined as the frequency at which the limiting asymptotes intersect in the Bode diagram. For the first-order system this intersection occurs at a point where the amplitude has been attenuated by 3 decibels and the phase shift is 45 degrees.

The frequency range of interest then will range from one-fourth of the break frequency to three or four times the break frequency. Thus, for a thermal system which has a time constant of 3 minutes, the break frequency will be on the order of one-twentieth of a cycle per minute or 3 cycles per hour. This means that a considerable amount of experimental time must be expended to obtain the frequency response characteristic of the thermal process. Often when cycling a process at the rate of 3 cycles per hour, many load changes will occur in the process during the period of 20 minutes. This will render interpretation of the data rather difficult. Therefore, the tests may have to be repeated several times and a statistical average taken of the data in order to arrive at a meaningful result.

Figures 14.9 and 14.10 show the frequency response characteristics of a typical Dubbs cracking furnace. These two figures together constitute what is known as a "Bode" diagram. In Fig. 14.9, note the intersection points of the straight-line asymptotes from which the process time constants may be evaluated.

Figure 14.11 shows a schematic diagram of the Dubbs furnace and also presents the transfer function as derived from the data appearing in Figs. 14.9 and 14.10. At the other end of the frequency spectrum lie the fast processes such as flow control systems and certain pressure control systems. At the high frequencies, limitations of the measuring equipment may be encountered. For example, above 10 or 20 cycles per second many of the pneumatic sine wave generators in use today become nonlinear with a resulting high degree of harmonic distortion.

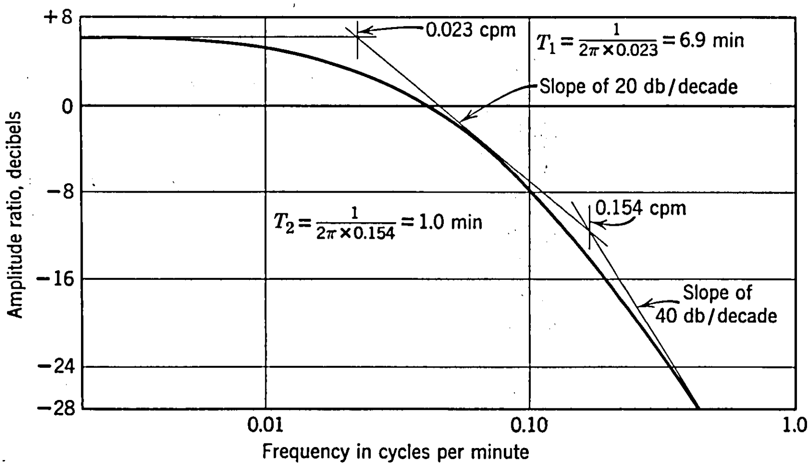


Fig. 14.9 Open-loop frequency response diagram of a cracking furnace—response measured from the fuel inlet valve to the furnace outlet product temperature

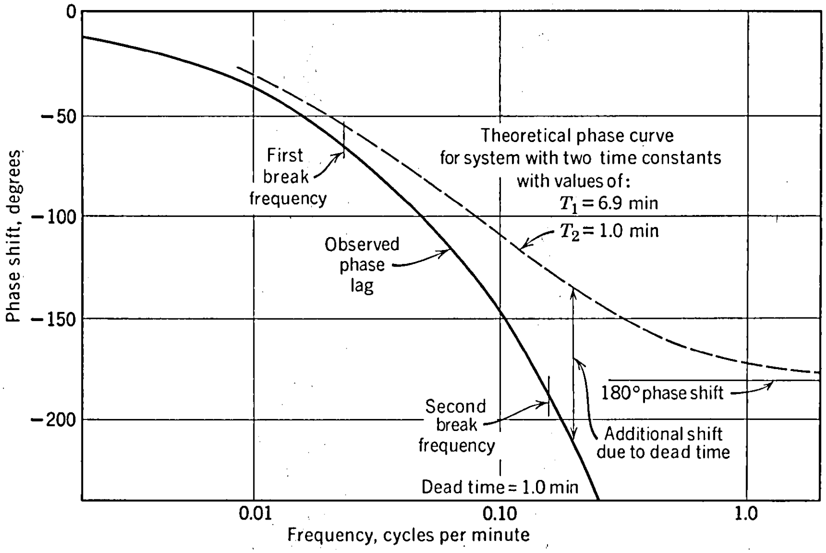


Fig. 14.10 Open-loop frequency response diagram of a cracking furnace (phase curve)

The output also may be attenuated below a useful level. The characteristics of the recording equipment and the transducers must also be considered when used at the higher frequencies. Therefore, the frequency response data obtained on fast processes are meaningful only when a limitation upon them has not been imposed by the measuring or recording equipment.

The operation of the plant is another factor to be considered when making frequency response measurements. Management generally will not allow cyclic operation of the plant to such an extent that product specification tolerances will be exceeded or that dangerous operating positions will be reached. In these cases it is necessary to limit the amplitude of the cyclic input signal. Often the permissible amplitude of this cyclic signal is approximately equal to the normal noise level in the process. In these cases, a series of data runs made at the same frequency can be analyzed statistically to improve the signal-to-noise ratio.

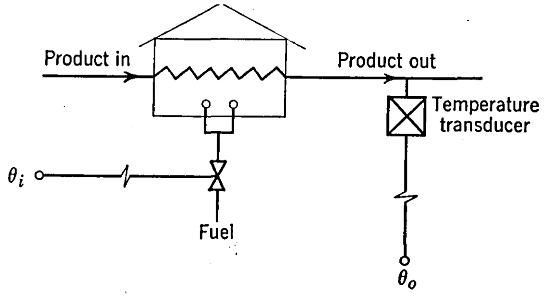


Fig. 14.11 Cracking furnace and transfer function derived on the basis of time constants and dead time obtained from figs. 14.9 and 14.10.

$$\frac{\theta_o}{\theta_i} = \frac{2.1e^{-1.0j\omega}}{(1 + 6.9j\omega)(1 + 1.0j\omega)}$$

The ultimate goal in the application of frequency response techniques to control equipment and industrial processes is to catalog all the transfer functions of each component part of the many industrial control systems. This catalog, then, will provide a ready reference for the systems engineer, allowing him to predict ahead of time, while the process is still on the drafting table, all of the control functions which will be needed, how the proposed control system will operate, and finally the cost of the control system. Today many data are available on the instruments themselves such as the measurement devices, the controllers, the final control elements, and the pneumatic transmission lines. As was shown in Table 14.2, a start has been made in accumulating response data on processes and their associated equipment. Much more work needs to be done along these lines before this catalog can be finally compiled.

This section concludes the review of the developments that have brought about the automatic processing plant of today. In summarizing, some of the characteristics and problems which are peculiar to the process industries have been considered. Briefly the first part of this chapter has dealt with the operational variables as well as with the product variables, which must be measured and controlled. We have studied the various forms of control systems from the simple single-variable to the complex computer-type control system. And lastly, a brief description has been given of the application of servo techniques in analyzing the preceding control systems. The remainder of this chapter will be devoted to recent developments in control equipment which will lead to the automatic processing plant of the future.

14.5 RECENT DEVELOPMENTS IN PNEUMATIC CONTROL SYSTEMS

The need for fast and precise control systems is being felt by the chemical and petroleum industries. Transfer line cracking, thermal flash cracking, and grid tray fractionating columns are examples of processes characterized by time constants on the order of 1 second or less. For control systems involving these processes, limitations are reached in the response speed of the pneumatic controllers themselves. This fact has spurred the development of pneumatic devices whose frequency response characteristic encompasses a greater frequency range. For example, some of the mechanical manometers operating on the force balance principle are capable of measuring differential pressure changes as rapid as 20 to 30 cycles per second. There is some attenuation of the signal at these frequencies to be sure; however, the useful range of these devices has been extended many fold over the older-type measuring instruments. Bear in mind that a response of 30 cycles per second by a pneumatic device is quite remarkable indeed. The frequency range of pneumatic controllers has been extended upward past 10 cycles per second in some of the more modern controllers. This too is remarkable when we consider that both the input and output signals are pneumatic in character. Several of these latter controllers have been developed by instrument manufacturers by applying the servo techniques as described in the previous section.

A second development of importance among the pneumatic devices is the relatively recent introduction of narrow-span primary sensing instruments. Narrow span may also be referred to as zero suppression. There are at least two commercial temperature-measuring instruments with a pneumatic output which provide a span of 50° F. The importance of narrow-span elements becomes apparent when it is realized that the inherent accuracy and sensitivity are generally a certain fixed percentage of the overall range or span of the instrument. Thus, an instrument with an inherent sensitivity of 0.5 per cent of the span will detect a 0.25° F change if the span is 50°. Now, for example, if the temperature being measured were at 1000° F and the measuring instrument had a span of 0 to 1000°, a sensitivity of 0.5 per cent would be equivalent to 5°. Thus, the greater superiority of the narrow-span instrument is readily apparent and in this instance is better by a factor of 20 to 1. Several of these improved temperature-measuring instruments have incorporated derivative or rate action in the measuring circuit. As will be recalled from Section 14.2, during the discussion of thermal control systems, it was often found necessary to provide protection to the primary thermal element. This added a second time constant to the control loop which in many cases was found to be the limiting factor. The addition of rate action in the measurement circuit now allows us to cancel one of these constants. Thus, by the addition of a proper amount of rate action, the thermal process can be simplified from a second-order to a first-order system.

Pneumatic pressure transmitters that permit the selection of any 20-psia span between the limits of atmospheric pressure and 1000 psia have also been developed. This, then, extends the field of narrow-span instrumentation to pressure systems also.

Perhaps the most interesting group of pneumatic devices currently available are the pneumatic computing relays. Relays are available which will add two numbers in the form of pneumatic signals. Others will subtract, and still others will add and subtract simultaneously. Still other relays are available that can multiply by a constant or determine the ratio of two pneumatic signals. To round out the list of pneumatic computing devices, there are the pseudo-derivative unit and also the pseudo integrator. By the proper interconnection of the units above, a pneumatic signal may be modified in any manner involving the mathematical processes of addition, subtraction, multiplication, integration, and differentiation.

We note that the recent developments have been primarily concerned with speed of response and extension of the frequency capabilities of the control equipment. One limitation imposed upon control system stability when using pneumatic controllers has been the transmission line. This is especially true when long lines are used and the time lags introduced by the lines themselves may be equal to or greater than the time constants of the process. For this reason, on fast processes there has been a trend toward mounting of the controller at the process in order to eliminate the long transmission lines which may be involved. This has necessitated the development of new types of controllers which will allow placement of the controller at the process and extension of the control indications back to the control board where the operator may keep in close touch with the process. Also, because of the field-mounting requirements, most present-day controllers are of the plug-in or quick-disconnect type. Mounting of the controller at the panel requires only two transmission lines, one from the primary measuring element, and one to the final control element. Controllers mounted in the field, however, require additional transmission lines for remotely changing the reference input signal to the controller.

In summarizing the developments in pneumatic control devices, we find the wide application of narrow-span primary sensing elements. High-speed controllers and measuring elements are now capable of keeping up with most industrial processes. Also available are pneumatic computing relays that can be utilized to provide any degree of sophistication required by the control problem. And lastly, pneumatic transmission lines can be removed from the control loop by the remote mounting of the pneumatic controller. Thus, for most applications modern pneumatic control devices are capable of producing a degree of control with acceptable speed and accuracy.

14.5.1 Electronic Control Systems

As was pointed out in the previous section, pneumatic control systems are capable of coping with relatively high-speed systems and provide a satisfactory degree of control. There are a few instances, however, where even greater speed of response is required. It is in these instances where electronic control systems found their beginning. Once the possibility of applying electronic controls had been definitely established, many other advantages of electronic control were discovered. It is, of course, obvious that there are virtually no transmission lags to contend with in electronic control systems. Secondly, the frequency response characteristics, not only of the transducers but also of the control equipment, have been vastly improved over those of the corresponding pneumatic devices. Perhaps one of the most significant contributions of the electronic control system is the elimination of dead time. Dead time, which is characteristic of most mechanical and pneumatic devices, is a large factor in limiting the highest frequencies at which the pneumatic or mechanical device is usable. Since control engineers need no longer be concerned with transmission lags, an entirely new concept of building processing plants has opened up. All control equipment can now be centralized in a single control room. The space requirements will be considerably less than in present-day plant layouts where several control centers are scattered throughout the processing area. Second, combined with other modern electronic control aids, the electronic-controlled plant will require less supervision and less operating labor. This results in a greater saving in capital outlay as well as in savings in operating and supervision personnel requirements.

There are other advantages to electronic control systems. The transmission medium now in the form of electron flow is not adversely affected by ambient conditions, such as temperature, humidity, etc. This is in direct contrast to the other forms of controllers requiring a pneumatic or hydraulic supply. In the case of pneumatic control systems, a clean, filtered air supply is required. This in itself represents considerable capital outlay. If the air is not completely dry, then the possibility of corrosion in the pneumatic piping is possible. This causes malfunctioning of the pneumatic control equipment. Also, the presence of water in the instrument air supply is especially dangerous in those sections of the country where freezing conditions may be encountered in the wintertime. For this reason, elaborate drying equipment is an absolute necessity in the colder climates.

Electronic control systems are not entirely without disadvantages. The initial cost of the control equipment is somewhat greater than that of the corresponding pneumatic equipment. However, newer manufacturing techniques are slowly closing the gap. Secondly, electronic control systems utilizing vacuum tubes are susceptible to all the vagaries of the vacuum tube and may require frequent maintenance. However, the reliability of electronic control equipment is being increased tremendously through the use of transistors and the printed- or etched-circuit techniques. Another disadvantage of electronic control systems is that a new class of instrument maintenance personnel would be required. This means that a training program would necessarily be undertaken by the industry. However, since technological advances appear to be headed in this direction, this is not an unreasonable task to ask of industry.

One advantage of the electronic control system which has not been mentioned is the fact that the control signals are readily adaptable for interconnection with data-handling equipment and computers. This is true since both the measured variable and control signals are either a-c or d-c electrical quantities. In a pneumatic control system these signals are in the form of proportional air pressures. To utilize data-handling equipment with the pneumatic system would require the use of many expensive transducers. An alternative to this is the use of one transducer on a time-sharing basis.

Today in the United States and in Europe there are at least several manufacturers who are in a position to supply essentially all the components necessary for a complete electronic control loop. There are also several United States manufacturers currently supplying just the electronic controller. In general, these controllers contain the same functions as the pneumatic-type controllers. Proportional action, integral action, and rate action are available in the present-day electronic controller. In keeping with modern trends, most of these electronic controllers have been miniaturized so that they require a minimum of space on a control panel board. Some of the controllers are of the plug-in or componentized type for ease of maintenance.

In present-day practice the only part of the electronic control loop which is not electronic is the final control element and its operator. A diaphragm motor valve still requires air or hydraulic fluid for its operation. A transducer is provided to convert the electrical signal coming from the controller to a proportional fluid pressure which is supplied to the top of the diaphragm control valve. This pilot transducer, or pilot relay as it is sometimes called, is mounted directly on top of the control valve in order to minimize any transmission lag.

Two separate applications of electronic control systems to a processing unit have been reported in periodical literature. The first of these units, in England, has been on stream since 1951. The second plant utilizing electronic control has been on stream since early 1954. In both instances the experiences have been good. Much tighter control has been realized than would be possible with pneumatic control equipment. Secondly, it has been found that less maintenance time has been required on the electronic system as compared with conventional pneumatic control equipment. On the basis of these two cases it must be concluded that both the accuracy and reliability for industrial process control can be provided by electronic control systems. The future of electronic control is unlimited. There is the definite possibility that the many single-variable controllers in the process may be replaced by a single electronic computer which will perform all the requisite control functions in an orderly and consistent manner. Thus, there is the definite possibility of simplification of our present-day control systems. This in itself would result in a considerable reduction in initial capital investment in control equipment. Secondly, the improved precision of control is unquestionable.

14.5.2 Graphic Panels

One of the interesting developments along the road toward automation has been the graphic panel. As the name implies, the graphic panel is a picture or flow sheet of the process laid out on the control panel board which shows all pertinent pieces of processing equipment and interconnecting flow routes. The miniature recorders and controllers are then located on the panel board in such a position that they are schematically superposed on the variable which they actually control or record. The graphic panel can be considered as an aid that will allow a greater and more rapid comprehension of the process by the operator. The graphic panel is valuable for several reasons: first, it allows the rapid training of new operators; second, it provides an easy and rapid picture for diagnosis by the operator when the plant is in an upset condition.

Graphic panels have been directly responsible for instrument miniaturization. By its very nature the flow sheet requires clustering of many instruments in a small space. The very purpose of the graphic panel would thus be defeated if the large-cased instruments were used. Certain operational difficulties have arisen through the use of graphic panels and miniature instruments. For example, should any one particular piece of equipment require maintenance, then a mechanic working on that instrument from the front of the panel board would block from the view of the operator perhaps four or five other control instruments. This very reason was a large contributing factor in the development of the plug-in or componentized instrument. Now the instruments are blocked from view only as long as it takes the instrument technician to unplug one control unit or recorder and to plug in a spare in good working order. This has actually been a boon to the maintenance program since it removes the urgency of maintenance and allows repair of the defective component to be made in the shop where more desirable working conditions are available.

The use of the graphic panel and miniature instruments has resulted in a saving in control room space requirements. This is a direct savings in capital investment. The use of the graphic panel has also brought about a trend toward the centralization of control equipment. As was pointed out previously, the centralization of control will reduce the supervisory and operational requirements. This trend toward centralization will undoubtedly be accelerated with the availability of reliable electronic control equipment.

The graphic panel also provides a very convenient place to mount the supervisory control systems such as alarms and automatic shutdown equipment. By the placement of colored lights directly in the process flow diagram, the operator can be instantly alerted to any abnormal condition and can direct his attention to that section of the plant requiring correction. Thus the graphic panel not only provides an overall picture of the operation when it is in stable equilibrium but it also provides a picture of any abnormalities existing therein.

It is the opinion of the author that graphic panels and panel boards in general, as we know them today, are merely a transitory phase in the road to complete automation. As the industrial processes become more and more complex and require decisions to be made more rapidly than the capabilities of a human operator, these functions will then have to be relegated to automatic computing equipment. This leads naturally then to a consideration of the next subject—that of data-handling equipment.

14.5.3 Data-Handling Equipment

Digital data reduction and data-logging systems are two of the stepping stones to the computer-controlled processing plant. Data reduction is a process of bringing together or correlating all of the information derived from the measuring instruments and automatically translating such information into a form which may be suitable for automatic computing and logging equipment. The importance of data-handling equipment in the automatic plant can hardly be overemphasized. First, its use would result in the elimination of the time-consuming and costly manual logging of plant operating data. As is generally the practice, an operator is required, at least once an hour, to read and record the important operating variables in his plant. Many times when a plant is in an upset condition, the operator may postpone his readings or may completely eliminate a set of readings while he is attempting to bring the plant back into stable equilibrium. It is at these times when data become very important since they will be used in the analysis of and in the subsequent elimination of plant operating difficulties. A second advantage of the data reduction equipment is the complete elimination of human errors and gross mistakes. When these data are to be used in plant efficiency studies or for cost accounting purposes, errors become intolerable and may render useless the remainder of the data. This is especially important when new processes are put into operation. It is then that huge quantities of data must be logged and stored away until such a time that a thorough analysis of the operating plant may be made. This applies particularly to pilot plants which are generally small models of the large processing unit. Since the main function of the pilot plant is that of data gathering, it would appear that data reduction and logging equipment is an absolute essential.

After the data have been logged or recorded they must then be analyzed. Certain analytical functions can be and are being built into commercially available data loggers. Some of these functions include (1) the continuous integration of all process flow rates to give a total daily production figure; (2) an hourly as well as daily average-yield balance around the plant; (3) efficiency balances around certain pieces of operating equipment may be computed and logged hourly. These latter functions provide the operating and supervisory personnel with up-to-date efficiency information so that the process may be run at its optimum economic position.

Data-handling systems, of which there are several available today, may be programmed to scan continuously the several hundred operating variables to be found in a modern processing unit. Once an hour, on the hour, or more frequently if desired, the instantaneous values of the individual variables are stored in the memory of the data-handling equipment. The data logger then scans its memory and converts each variable into a digital form which can then be logged by any standard read-out device. This is the first function of the data-logging equipment. A second part of the program is a continuous scan of the process variables with abnormal read-out. In other words, the data-logging equipment is continuously monitoring all the process variables. When any one of the variables exceeds certain predetermined limits, either on the high side or on the low side, this fact will initiate the read-out cycle. The variable that is abnormal or has exceeded these limits will then be printed out in red type, and the signal may also be utilized to actuate an alarm or other safety device. Thus the data logger can be used to fulfill certain of the supervisory or limiting control functions.

The use of data reduction equipment has a further implication in future automatic control systems. It is envisioned that very few if any process variable recorders will be required in future plants when data-logging equipment is installed. The log itself provides all the information necessary to operate the plant properly. Readings of all variables are given at definite intervals of time. If the process wanders outside of these limits, then such facts are noted on the data logger in red ink. If it is still considered that one or perhaps several variables are of sufficient interest that they must be continuously recorded, the control center of the future will probably contain a multichannel strip chart recorder for just such a purpose. If this is true, then the many single-function recorders which we find in today's processing plants will no longer be required. This in turn will result in a complete elimination of the panel boards as we know them today. The greatest outcome of this development, then, should be a considerable decrease in the initial capital outlay for instruments and control equipment.

There are at least three different systems being offered to industry today. Applications of two of these have already been reported upon in the periodical literature. The systems available are similar in these respects. The inputs to the data-handling equipment may be either pneumatic or electrical signals. These signals are generally analog in character, being derived from the many measurement transducers. The reduction equipment then operates on the input signals to convert them to digital form. Several types of outputs are available: the automatic typewriter, the perforated tape, or the punched card. A combination of outputs such as automatic typewriter and perforated tape is also available when it is desirable to further process the data by automatic computing equipment.

In light of the discussion above the role played by the data reduction equipment is self-evident. It is actually the link between the process operations and the supervisory computer which will be found in the automatic processing plant of tomorrow. In Section 14.3.4 on computer controls, a block diagram of an automatic processing plant was given in Fig. 14.7. There it was pointed out that certain information concerning the operation of the plant must flow to supervision and management. On the basis of this information, management is able to make decisions and provide proper guidance in the course of plant operation. It is one of the functions or duties of the data loggers to provide this flow of information to the management and supervisory control loop. Therefore, returning to the introductory statement in this section, data reduction and handling equipment form a very important stepping stone to complete automation.

14.6 CONTINUOUS COMPOSITION ANALYZERS

The emphasis thus far has been on the development of measurement and control equipment, data reduction and processing equipment, and computers. These great strides are of course very important to the overall control problem. However, a field which heretofore has been somewhat neglected is the development of continuous analyzers to measure the definitive variables of the process. Instrument manufacturers are slowly but surely filling the need for certain of these basic devices.

In the realm of continuous composition analyzers, several ingenious devices have made their appearance. Most of these devices are inferential in nature and depend for their operation on some predominate physical property of the product under investigation. Liquids or gases are the most easily handled by the present-day analytical instruments. The analysis of solids requires more elaborate sample preparation equipment. The most important analyzers used in the petroleum industry are the so-called spectrophotometers. These include the infrared spectrophotometer, the ultraviolet spectrophotometer, the mass spectrometer, and the radio-frequency spectrometer. Table 14.3 shows a list of applications to which each of the spectrometers above is especially suited.

Table 14.3 Typical Applications of Spectrometers to Process Control

Spectrometer Type	Measurement	Reference
Infrared	(1) Ethylene in ethyl chloride	26
	(2) Isobutane in hydrocarbon mixture	26
	(3) CO_2/CO ratio in flue gas	Unpublished
	(4) Ortho-ethyl toluene in isomeric mixture	28
Ultraviolet	Benzene in benzene raffinate	Unpublished
Mass	(1) Oxygen in ethylene	Unpublished
	(2) Oxygen in argon	
	(3) Light hydrocarbons in drilling mud	

The continuous infrared analyzer is particularly adaptable to the analysis of light hydrocarbon streams in which up to twelve separate components may be present. The analyzer is capable of determining the amount or percentage of one component only. In certain cases, the ratio of one component to a second component may be determined. Infrared spectrophotometers are based on the natural phenomenon of infrared energy absorption by resonant atomic configurations. Thus the infrared spectrophotometer provides a tool whereby chemical isomers may be distinguished. An isomer, as it will be recalled from chemistry, is a compound which has the same lumped chemical formula as another compound but which has a different arrangement or configuration of the atoms. A block diagram of a typical infrared analyzer is shown in Fig. 14.12.

Ultraviolet analyzers form a second class of spectrophotometers. The ultraviolet absorption phenomenon results when an orbital electron is excited from one energy level to the next. Since this transition is electronic in nature, we find associated with this transition process, frequencies which have wavelengths in the order of 200 to 400 millimicrons. All aromatic compounds such as benzene, toluene, naphthalene, etc., comprise a ring structure containing electrons which are easily excitable. For this reason these compounds absorb energy in the ultraviolet region. Certain sulfur and nitrogen compounds display a similar activity in the ultraviolet spectrum. Therefore, the ultraviolet analyzers will probably find greatest application in the analysis of streams where an aromatic is the main component under consideration.

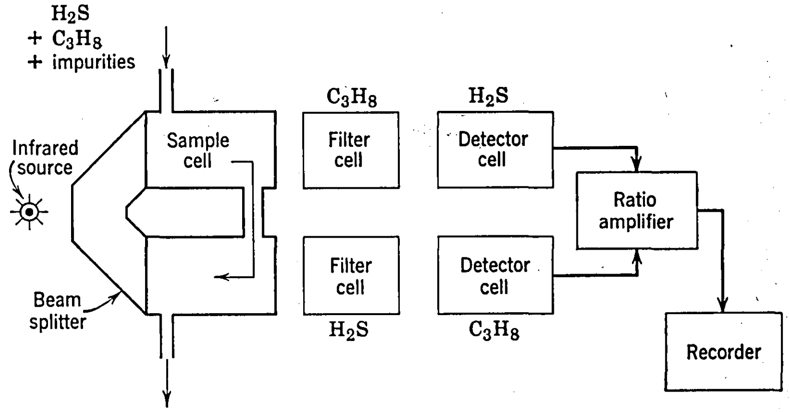


Fig. 14.12 Schematic arrangement of a continuous infrared analyzer (Luft type)

The mass spectrometer, as shown in Fig. 14.13, is a third type of continuous composition analyzer which functions in the following manner. A sample is vaporized and passed over a heated filament which ionizes the molecules in the sample. The ions and other molecular fragments are then accelerated in a circular path by an accelerating potential. For a given potential and field strength, only ions of a certain mass will strike the target or collector electrode. The number of ions reaching the target results in a measurable current flow which is directly proportional to the concentration of these ions. If the ionization patterns or spectra of the various components are known, the composition of the sample can be computed. Mass spectrometers are capable of handling all types of materials, providing these materials can be vaporized at the conditions prevailing in the ionization section of the ion tube. Since the mass spectrometer measures only mass, it cannot distinguish, by a single measurement, the difference between molecular fragments of the same mass.

The fourth composition analyzer to be considered is the nuclear-magnetic-resonance spectroscope. The principle of operation is based upon two physical characteristics possessed by most atomic nuclei. Specifically, these two nuclear characteristics are the gyromagnetic properties called "spin" and "magnetic dipole moment." To date this instrument has served primarily as a laboratory tool, although there is no fundamental reason why it cannot be adapted for plant use.

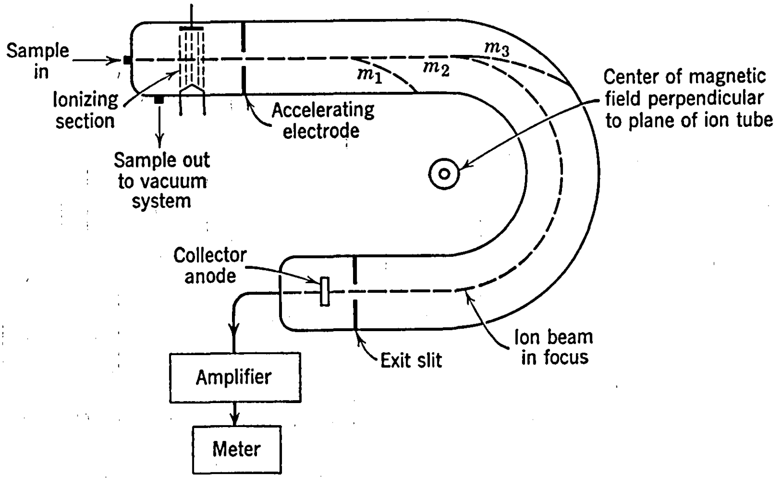


Fig. 14.13 Schematic arrangement of ion tube in a mass spectrometer

The main application of the nuclear-magnetic-resonance spectroscope has been for the determination of specific atoms. However, spectral perturbations caused by groups of atoms allow certain molecular structures to be identified. Liquid samples are most easily handled, although both gases and solids can be analyzed by nuclear-magnetic-resonance techniques.

The last analyzer in this group is the radio-frequency spectrometer which may be considered a special type of infrared spectrophotometer in that the absorption frequencies involved are a manifestation of molecular vibrations. To date, the radio-frequency spectrometer has not found as wide an application as have the other three types.

A second group of instruments, whose principle of operation is based upon a chemical character of the product, are the continuous titrimeters. The continuous mercaptan analyzer as described by Pompeo is an example of an instrument in this class. Generally these instruments are based upon an electrometric titration with continuous control of sample and reagent streams. The titration mixture is maintained at the equivalence point by a closed-loop control system which regulates the rate of addition of the titrating agent. Another example is the polarigraphic determination of dissolved oxygen in boiler feed water. Still a third is the Thomas autometer which continuously measures and records the amount of sulfur dioxide in the atmosphere. This latter instrument has found wide usage in atmospheric pollution studies.

14.6.1 Continuous Quality Analyzers

There have been developed many instruments capable of continuously measuring the physical properties of materials. As has been pointed out previously, many of these properties form a part of the product specification. Others of these properties sometimes can be directly correlated with product specification. Potentially, then, these instruments play an important part in the quality control loop.

Table 14.4 is a list of the various types of instruments commercially available for the measurement of physical properties. Space is too limited, in this chapter, to consider each one of these devices separately. However, a few general observations are in order. Most chemical and petroleum products must meet not only one but several specifications. This, of course, means that more than one of these instruments will be found in the quality control loop during the manufacture of a single product.

Table 14.4 Applications of Analyzers to the Continuous Measurement of Product Quality

Measured Quantity	Reference
Viscosity	27
Specific gravity	. .
*p*H	25, 30
Redox potential	25
Ionic conductivity	25, 29
Moisture	31
Refractive index	32
Liquid color	35
Vapor pressure	33
Chemical end-point (titrimeter)	34

All these instruments have one thing in common, i.e., the sampling system. Perhaps this has been one of the greatest sources of difficulty with today's continuous analyzers. Sampling systems include all the interconnecting piping and handling equipment necessary to condition or prepare the sample for analysis. When it is realized that most product specifications are set down at fixed and definite environmental conditions, then the magnitude of the sampling problem becomes apparent. For example, in the petroleum industry, specific-gravity measurements are generally referred to 60° F. If a particular product stream were at a temperature of 450° F, then adequate and regulated cooling equipment would have to be provided to prepare the sample for measurement. Viscosities of lubricating oils are measured at 122° F and 210° F to determine the so-called viscosity index. In this particular analysis problem, two well-regulated constant-temperature baths are required to condition the sample.

Perhaps one of the greatest disadvantages introduced by the sampling system is the transportation lag or dead time. The effect of dead time upon control precision and stability has been considered previously and will not be discussed here other than to say that its effects are extremely deleterious.

Often corrosion of the sample piping is a problem. This is particularly true when dealing with aqueous solutions or even in the handling of a simple material such as flue gas where moisture may condense when the sample is cooled. In these instances glass sample piping may be required. Due to its fragility, all glass sample connections and tubing must be adequately protected against accidental breakage.

Many of the product streams at the point of desired measurement contain foreign materials such as solids or possibly entrained water droplets. These foreign objects must, of course, be removed from the product before it can be accurately analyzed. Filters are generally provided for this purpose and these too have been a source of considerable maintenance since they require frequent cleaning.

Then, last, there is the perennial problem of obtaining a representative sample. This problem exists wherever two or more components are blended together to form the finished product. Sufficient time must be allowed for the complete mixing of all the components before the sample is taken. On the other hand the time allowed cannot be too great or the dead time introduced will seriously affect the control systems. Thus, we sometimes find ourselves on the horns of a dilemma.

In order to overcome some of the difficulties found with sampling systems, instruments are required that will measure the product property directly in the process at the conditions existing therein. The analyzer then will be required to compute the necessary correction to refer the measurement to the standard conditions at which the specifications were laid down. The Bendix Ultra Viscoson is an example of a continuous viscometer which partially performs this function. There are several conditions which must be fulfilled before this type of analyzer-computer may be utilized. First, there must be a known and definite relation between the product property measured under process conditions and this same property referred to the standard or reference conditions. Second, this relation should be fixed and should be relatively independent of product composition. For example, in the blending of fuel oils, the amount of any one particular component in the final blend will be determined primarily by its current availability. Since viscosity of the final product is the predominant specification that must be met and, further, since several possible formulations may have the required viscosity, the actual composition of the final product may differ from time to time. This would in all probability result in a different viscosity-temperature relationship for each formulation. In this instance, since there is no definite temperature-viscosity relationship that can be predicted, a simple computer-type instrument would be useless.

There is a distinct possibility, however, of developing a secondary instrument which would continuously measure the viscosity-temperature relationship. The information thus secured would provide an input signal for the main computer which refers the measurement back to the standard set of conditions. The secondary measuring apparatus would of course be burdened with the attendant sampling system and sample-conditioning apparatus. However, this would free the primary instrument from unnecessary time lags and allow much better precision in the main quality control loop. Since composition changes are generally gradual in nature, the system described above could provide a satisfactory degree of control.

Before leaving the subject of continuous quality analyzers, a few general remarks are in order. The field of continuous analytical measurement has scarcely been scratched. Many reliable instruments have already been developed. However, in other instances accuracy and reliability have yet to be provided. Maintenance of these analytical instruments will probably require the establishment of a new group or classification of service technicians.

14.7 REVIEW AND CONCLUSIONS

Automation has been largely responsible for the tremendous growth of the petroleum and chemical industries. These industries which deal with the processing of solids, liquids, and gases are confronted with control problems which are peculiar to the equipment used in the processing. Variables which must be controlled are temperature, pressure, flow, liquid level, and product quality. Today the predominant type of control system is the single variable. However, cascade and coordinated control systems are being used more frequently. To analyze control problems, the servo techniques have found wide application. Improved pneumatic devices are available. Complete electronic control systems have been tried. And now, the advent of data-handling equipment promises the last step toward complete computer control. Continuous analyzers which heretofore have been somewhat neglected are now coming into more widespread use. Integrated control systems are slowly but surely developing to a high degree. This foreshadows the automatic chemical processing plant which is now a definite technological possibility.

14.8 REFERENCES

Control Systems

1. C. H. Barnard, A.B.C.'s of Multielement Control, *Instruments,* Feb. 1949.
2. K. E. Coulter, Use of Pneumatically Set Instrument Combinations, *Instruments,* April 1949.
3. J. G. Zeigler, Cascade Control Systems, Texas A. & M. Instrument Symposium, Jan. 28, 1954.
4. R. J. Bibbero, Process Regulation with Analogue Control, *Automatic Control,* Feb. 1955.
5. R. J. Bibbero, The Analogue Computer and Automatic Control Applications, *Automatic Control,* Nov. 1954.
6. W. E. Phillips, Analogue Study of Controlled Interacting Systems for Turbojet Engine, ISA Paper 54–1–3.

Servo Techniques

7. R. E. Clarridge, New Tools for the Instrument Engineer, *Chem. Eng.,* May 1952.
8. J. C. Moise, Pneumatic Transmission Lines, *ISA J.,* April 1954.
9. Yaohan Chu, Study of Process Control by the Phase Angle Loci Method, ISA Paper 53–20–2.
10. D. W. St. Clair, Step Response as a Short Cut to Frequency Response, ISA Paper 52–9–3.
11. B. D. Stanton, Instrument Evaluation by Frequency Response, *Instruments,* May 1953.
12. B. D. Stanton and P. R. Hoyt, Process Control System Analysis Based on Field Test Data, Texas A. & M. Symposium on Instrumentation, Jan. 1953.
13. G. W. McKnight and C. W. Worley, Dynamic Analysis of Plate Heat Exchanger System, ISA Paper 53–6–1.
14. S. P. Higgins and G. W. McKnight, Practical Applications of Servo Mechanism Techniques to a Process Control Problem, *Instruments,* June 1952.
15. W. I. Caldwell, Single Knob Adjustment of Reset and Gain, *Instruments,* Feb. 1952.

Recent Developments

16. R. E. Clarridge, New Concept of Instrument Control, *Instruments,* 1950.

17. J. G. Zeigler and N. B. Nichols, Dynamic Accuracy in Temperature Measurement, ISA Paper 49–9–1.

18. C. O. Anderson, A Refinery Pioneers the Use of Electronic Controls, *Automatic Control,* Dec. 1954.

19. W. F. Marchment, Electronics Simplifies Process Control in Overseas Refineries, *Automatic Control,* Jan. 1955.

20. E. T. Davis, An Electronic Control System for Industrial Control Applications, *Instruments,* June 1951.

21. C. E. Mathewson, Advantage of Electronic Control, *Instruments and Automation,* Feb. 1955.

22. Automatic Logging System Speeds Data Reduction, *Automatic Control,* Aug. 1954.

23. A. F. Sperry, An Information System Consolidates Process Control, *Automatic Control,* Nov. 1954.

24. R. K. Stern, Digi-Coder, *Automatic Control,* Sept. 1954.

Continuous Analyzers

25. Fundamentals of Industrial Chemical Measurement and Automatic Control, *Minneapolis-Honeywell Tech. Bull., B-51-2.*

26. V. N. Smith, A Recording Infra-red Analyzer, *Instruments,* Mar. 1953.

27. Ultra-Viscoson, *Automatic Control,* July 1954.

28. L. W. Herscher, Infra-red Analysis of Liquids, *Petroleum Processing,* May 1954.

29. G. F. Rucker, Measurement of Electrolytic Conductivity, *Instruments,* Oct. 1952.

30. R. Buenger, pH Measurement, *Instruments,* May 1952.

31. R. E. Hadady, Moisture Measurement, *Instruments,* May 1952.

32. O. Barstow, New Recording Refractometer, ISA Paper 49–4–2.

33. R. B. Jacobs, Continuous Measurement of Vapor Pressure of Petroleum Products, ISA Paper 50–11–3.

34. Hallikainen and D. J. Pompeo, Continuous Recording Titrimeter, ISA Paper 51–2–1.

35. T. C. Fletcher, Application of Flow Colorimeters and Turbidimeters in Process Industries, ISA Paper 54–3–4.

15. Analog Computers in Industrial Control Systems

E. L. HARDER

Westinghouse Electric Corporation
Pittsburgh, Pennsylvania

15.1 INTRODUCTION

During the last ten years, computers have developed at least ten thousand fold in speed and power. They have largely supplanted human beings in the aiming and direction of large missiles and have made considerable strides in the processing of huge amounts of industrial, governmental, and scientific information. It is literally true that more large-scale computers have been built in any recent two-year period than in all the history of mankind up to that time.

It is to be expected, therefore, that in the next several years the computer will become a very important factor in our civilization, taking over much of the information-processing activity currently handled by human beings and eventually revolutionizing the activities of the workers of the industrialized part of the world. This conclusion is based on the observation that a large part of the activity of all industrial workers is now concerned with processing information. Computers and the mechanization of logical processes and data processing are nearly synonomous. They are overlapping phases of the same development, and we can recognize two paralleling developments which are becoming more and more closely interrelated. These have to do with the handling of man's physical work and his mental or clerical work and, in each instance, the emergence of the full-fledged computer or data-processing element is clearly evident.

(1) Physical work

(*a*) *The tool,* with man providing the power and intelligence.

(*b*) *The power tool,* with man providing the intelligence.

(*c*) *The automatically controlled power tool,* with primitive computing elements. Man provides pilot control, coordinates machines' activities, and provides all data processing.

(*d*) *The automatic factory or system,* or the complete process line control, in which groups of machines are automatically controlled and coordinated. The computer portion of the control is more mature and appears as a separate organ making intricate decisions at high speed.

(2) Mental and clerical work

(*a*) *The desk calculator,* adding machine, cash register, and simple aids to calculation.

(*b*) *The business machine*—masses of similar documents are processed by wiring into the machine instructions for the calculations.

(*c*) *The large-scale automatic computer*—complete sequences of mathematical calculations or elaborate business operations can be programmed. Man has tremendously greater resources at his command. The programmer still has a tremendous job of directing these facilities to do the work.

(*d*) *The self-programmed automatic computer*—the computer itself will be able to interpret simple commands into the elaborate programming required to make it flexibly approach all manner of logical operations of mathematics, business, industry, and various sciences.

Both analog and digital computers have a distinct place in industrial and process control, with the data processing of business operations being limited to the digital machine. The elemental appearances of analog computing functions in controls show up in the recognition of their transfer functions and modification of these functions to process the information so as to secure superior performance. If a regulating system is set up on an analog computer, we might say that it is only the computing or data-processing elements that are set up. Certainly the power-producing features are missing. We can then clearly see what computation the system makes and can analyze what further computations would be necessary for the control to perform automatically to produce output motions or results that are the desired functions of the information sensed.

The elemental forms of digital computers can be recognized in the automatic relay systems for complete automatic starting and running of machinery where the logical elements of "and," "or," "not," etc., are combined by suitable relay contacts which respond in a digital or "on-off" manner to set up discretely the operation performed, depending upon a number of "yes-no" answers. The director systems now coming into the picture utilize passive digital elements to perform these same functions with superior approaches to the design through Boolean algebra. They represent a further extension of digital techniques and again recognize the information-processing character of the majority of this control, except for the final power-handling contacts. As the information-processing requirement becomes more sophisticated, e.g., when complex analysis or optimizing operations are required as a stage of the control, then the full-fledged digital computer comes into the picture, as it already has in many military controls.

In this lecture, we shall confine ourselves chiefly to the analog computer techniques as applied to industrial and process control. We shall discuss the use of analog computers in designing the systems, their use directly in the systems, and the use of simulators in which the computer and the device are connected together. The theme will be developed by considering a number of examples.

15.2 EXAMPLES OF USE OF ANALOG COMPUTERS IN DESIGNING INDUSTRIAL CONTROL SYSTEMS

The differential equations describing the performance of the main motors, generators, and loads involved in an industrial process can frequently be written. These can then be represented on an analog computer. The equations describing various alternative forms of controls to form a closed-loop regulating system for the major equipment can likewise be written and represented on the computer. It is then possible to optimize the performance of each alternative arrangement and to determine its limitations and characteristics. The performance required or considered to be most desirable for the particular process is frequently known or obtainable by a careful study of the process and tests made thereon and from past experience. Thus, although all factors are not known, the use of the analog computer makes possible the determination of the best possible utilization of the data that is known and usually results in the development of a regulating or control system far superior to what could be obtained otherwise.

15.2.1 Steel Mill Tandem Cold-Rolling Mill Controls

In this instance, steel strip is to be passed through the five stands of a rolling mill and reduced in gage at each stand. The requirements include running at a low threading speed, accelerating to running speed without loosing or greatly changing the tension between stands during the process, in spite of differing inertias of the rolls and motors, and precise speed regulation at high strip speeds, sometimes exceeding a mile a minute. Such a regulating system can be represented on an analog computer by setting up analogs of the gains around the main control loop and the several feedback loops, and of the various time constants and compensation ratios involved.

The problem is characterized by the requirement of high performance over a wide range of system parameters. For example, the mechanical time constants of the rolls vary over a wide range as the motor field currents are changed to secure higher running speeds. The generator time constants and amplifications change, owing to saturation at higher voltages. This dictates a control system that is inherently very stable and will closely follow any changes in the reference speed setting. The control must also correct for disturbances due to mechanical irregularities in the sheet without overshooting, which would cause a loop to form in the sheet and subsequent breakage when it straightened out. Whereas a few of these conditions might be analyzed longhand for a few control arrangements devised by an ingenious designer, the assurance of an optimum control under the wide variety of operating conditions for a mill requires the analog computer.

In setting this problem up for study, it is sufficient to analyze the control for one typical stand. Synchronization is obtained by use of a master speed reference which each individual stand control follows in going through the cycle of threading speed up to running speed, back to threading speed, etc. The precision and stability with which the control causes the speed of a particular stand to follow the reference pattern, including inertia compensation and speed droop, is sufficient to judge the merit of the control in the complete system. Generally speaking, quick response with a minimum overshoot will best accomplish these objectives, except that the acceleration and deceleration characteristics must be matched.

The four steps in the solution to this problem by electrical analog computer are illustrated by Fig. 15.1 (a), (b), (c), and (d). The problem here is the determination of speed versus time for impact loads and for sudden changes of the speed setting. Solutions are required for a

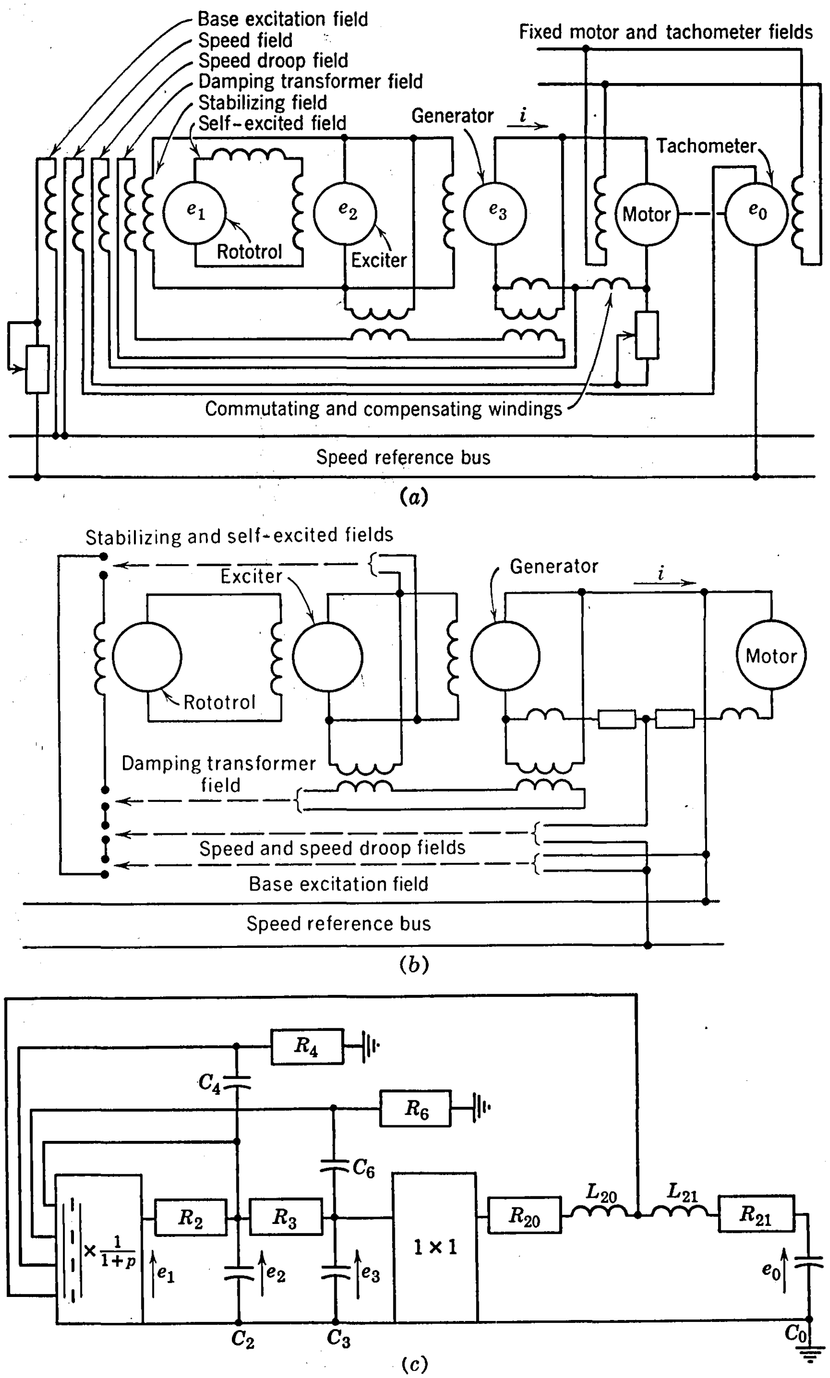


Fig. 15.1 Speed regulator analog: (*a*) regulating system, (*b*) simplified equivalent, and (*c*) schematic analog

large number of values of the parameters, and so the analog computer method is to be used.

The system shown in Fig. 15.1*a* consists of a d-c motor with fixed field supplied by a generator whose field is controlled by a rototrol through an intermediate exciter. The rototrol responds to deviations of the speed (tachometer voltage) from the reference voltage setting or pattern. Speed droop with motor load is provided by feeding back the commutating and compensating field drop into a separate rototrol field. Stabilization is afforded by degenerative feedback from the exciter and by damping transformers which feed back derivative functions of exciter and generator voltage. Base excitation is supplied through still another rototrol field to reduce the "unregulated speed error" and hence increase the accuracy of regulation for a given regulator gain.

A machine and regulator theory requiring a specialist in this field are necessary to reduce this system to its equivalent, Fig. 15.1*b*, and its corresponding equations. It is an important matter, sometimes overlooked, that between the actual physical problem and its mathematical expression, with such approximations as are permissible in the particular art, lies the whole field of knowledge with which the operator of the general-purpose computer will not, in general, be familiar. The mathematical expression of the problem is strictly a function of a specialist in the field of the problem, working, of course, in conjunction with the computer staff to express the problem in proper form and with sufficient information. Practically, there is considerable overlap in these two functions since analog computers have usually been developed as special-purpose computers for particular problems, and the computer engineer is, or soon becomes, skilled in the art in question.

Figure 15.1*c* gives the schematic analog which obeys the same differential equations as the system in question. It is made up by reference to tables of basic analogs that correspond to specific functional relationships. For example, the first block shows an operational amplifier which adds four inputs and produces a time delay. It thus has the same transfer function as the rototrol in Fig. 15.1*b*. The direct relation between the other analog elements in this diagram and the corresponding constants of the physical system are shown in Table 15.1. Note, for example, how the *RLC* circuit at the right is adjusted to agree with the motor-generator armature circuit time constants and the mechanical time constant of the motor. The motor and load mechanical time constant may be thought of as the time constant of the exponential approach of the motor speed to a new value if the generator voltage were suddenly raised and if this circuit had no induct-

Table 15.1 Analogous Relations in Speed-Regulating System of Fig. 15.1

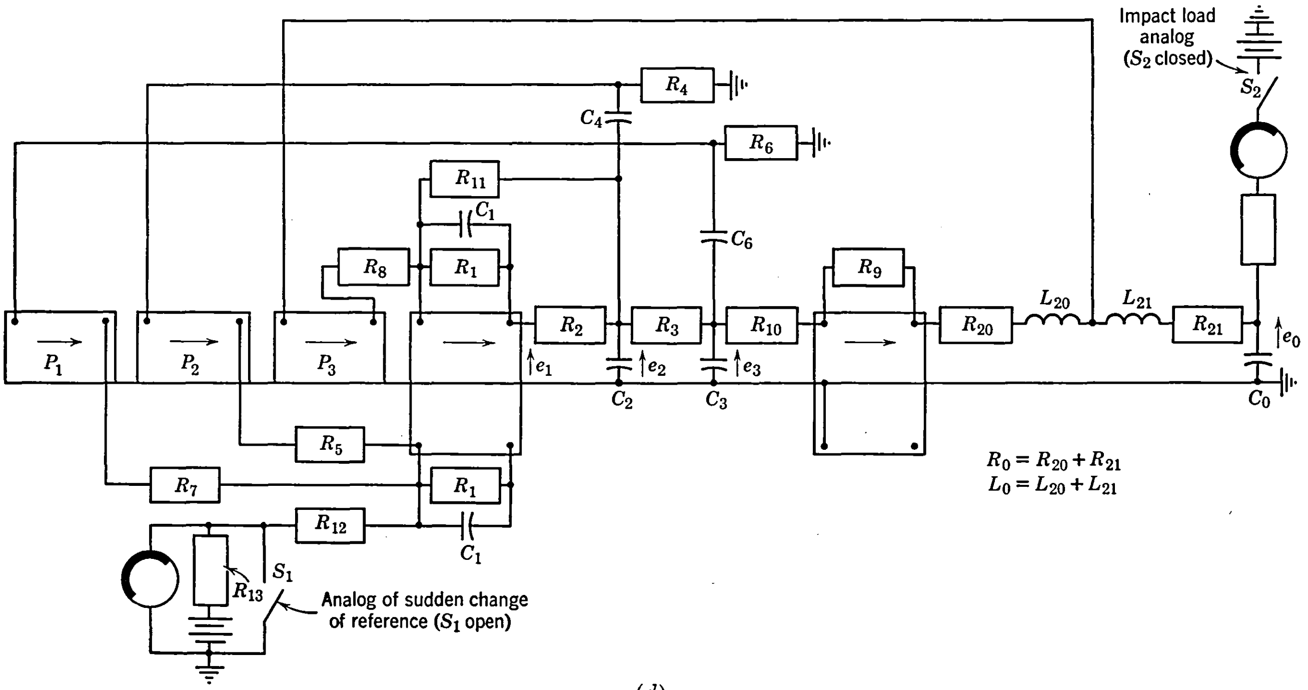
Quantity in Actual System	Description	Analog
T_0	Mechanical time constant, RI/bg * $R_0 = R_{20} + R_{21}$	R_0C_0
T_1	Rototrol time constant	R_1C_1
T_2	Exciter time constant	R_2C_2
T_3	Generator time constant	R_3C_3
T_4	Armature circuit time constant $L_0 = L_{20} + L_{21}$	L_0/R_0
T_5	Commutating and compensating field time constant	L_{21}/R_{21}
T_{p2}	Exciter damping transformer primary time constant	R_4C_4
t_2a_2	Exciter damping transformer mutual time constant times loop gain	$R_4C_4P_2(R_1/2R_5)$
T_{p3}	Generator damping transformer primary time constant	R_6C_6
t_3a_3	Generator damping transformer mutual time constant times loop gain	$R_6C_6P_1(R_1/2R_7)$
D/D_{mg}	Ratio of the speed droop, D, to the unregulated speed droop D_{mg}	R_{21}/R_0
A	Speed loop gain	$P_3(R_1/2R_8)(R_9/R_{10})$
N	Negative feedback loop gain (stabilizing and self-exciting fields)	$R_1/2R_{11}$

* R = armature circuit resistance; I = moment of inertia; b = torque per ampere; g = volts per radian per second.

ance. Note that this time constant varies inversely as the torque per ampere and the motor back voltage per unit of speed and thus varies inversely as the square of the field current as well as depending on the motor and load inertia. For this reason, this time constant varies widely as the field current is changed to raise the motor speed. This accounts for the wide variation in mechanical time constant that must be accommodated by the control system with high performance.

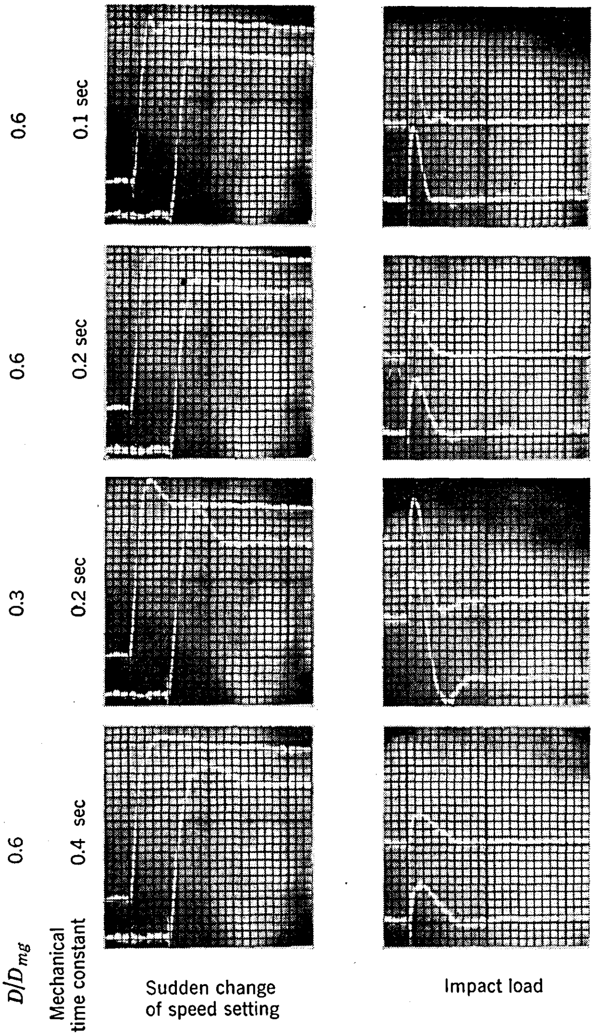
To return to Fig. 15.1*c*, this time constant is set up as R_0C_0 where R_0 is $R_{20} + R_{21}$. This still provides flexibility for dividing R_{20} and R_{21} proportional to the generator and motor armature resistances respectively, and for setting L_{21} to provide the proper commutating and compensating field time constant, T_5, equal to L_{21}/R_{21} and for setting L_{20} to provide the armature circuit time constant, T_4, equal to L_0/R_0, where L_0 equals $L_{20} + L_{21}$.

In practice, the motor speed might be measured by a tachometer



(*d*)

Fig. 15.1 Speed regulator analog: (*d*) complete analog connection diagram



(e)

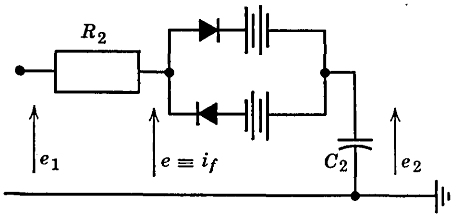
Fig. 15.1 Speed regulator analog: (e) typical generator voltage traces for regulating system which would correspond to making the feedback connection across the capacitor, C_0, or the generator voltage might be held constant, which would correspond to making this connection to the left of R_{20}. The latter would result in a speed droop due to load corresponding to the full IR drop of motor and generator. Thus, by moving this point of connection, any ratio of speed droop, D, to the unregulated speed droop, D_{mg}, can be represented.

The next step in going to Fig. 15.1*d*, is distinctly a computer job of developing the schematic analog into a connection diagram, taking proper account of the scale factors, computer time base, impedance levels, and the like. Figure 15.1*d* is the actual analog setup for this problem. It is a direct analog having a one-to-one correspondence between the 14 parameters which completely describe the system and the same 14 parameters in the analog.

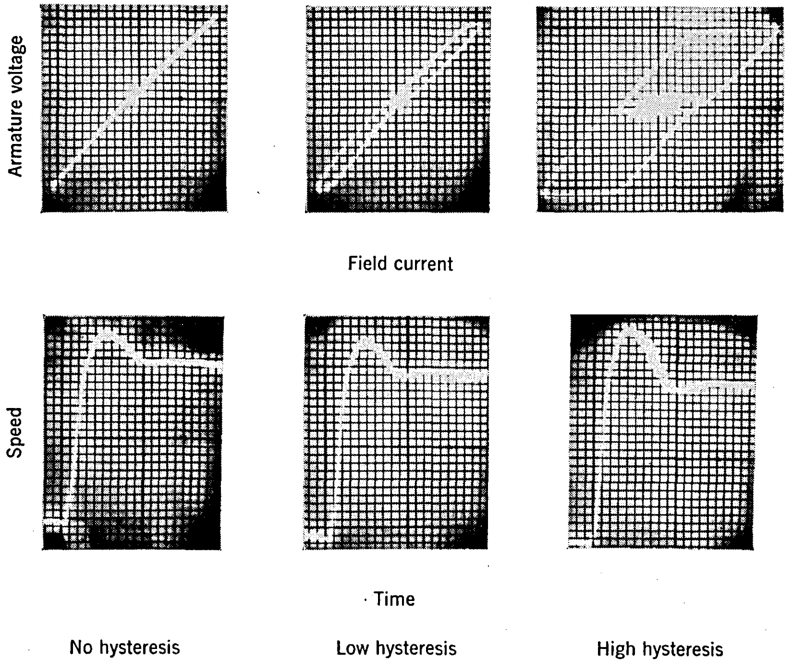
Impact load is simulated by a negative step function of current into the capacitor, giving an initial rate of decrease of its voltage which is analogous to speed. Sudden change of speed setting could be represented by inserting a step function voltage in the speed loop. However, it is just as accurately and more conveniently added to the speed signal in the first amplifier. Typical solutions for these two conditions are shown in Fig. 15.1*e*.

Hysteresis analog. As electrical machines usually have hysteresis to some degree, the flux and hence the generated voltage is not a single-valued function of field current but depends on past history of the iron also. Since this functional relationship can be expressed only very approximately, and since the effect is relatively small, a rather approximate analog is adequate to study the effects. The analog may be that shown in Fig. 15.2*a*, or alternatively a bridge of diodes with a single battery may be used. For rising field voltage, e_1, the voltage e_2, which corresponds to machine flux or armature voltage, is below the voltage e_1 by a fixed amount equal to the battery voltage. On falling voltages, e_2 is above the voltage e_1 by the battery voltage. Thus, whenever the time derivative of field current changes sign, a finite change in field current is required before the proportional current-flux relationship is resumed. This action is closely equivalent to the actual hysteresis in the machines. If the ratio of width to height of the resulting parallelogram-shaped hysteresis loop is made the same as for actual loops, a reasonably accurate solution for the effect in a regulating system can be obtained on the computer.

The solutions for three values of hysteresis are shown in Fig. 15.2*b*. The upper three oscilligrams show the flux-current locus traced during the oscillation following a sudden change of speed setting. The lower loops of these oscillograms correspond to the oscillation during return of the speed setting to normal, the complete operation being repetitive on the Anacom computer.



(*a*) Modification of the exciter analog in speed regulator Fig. 15. 1, to represent hysteresis



(*b*) Hysteresis and speed records taken on the Anacom

Fig. 15.2 Effect of exciter hysteresis on speed regulator transient performance

In the lower three oscillograms, which are the speed-time plots, the effects of hysteresis can be noted. A moderate amount of hysteresis in the exciter has very little effect, whereas a greater amount considerably increases the overshoot and recovery time. Note the presence of two distinct limiting solutions depending on the initial condition of the flux which can differ from zero by plus or minus the battery voltage, even though e_1 is initially zero.

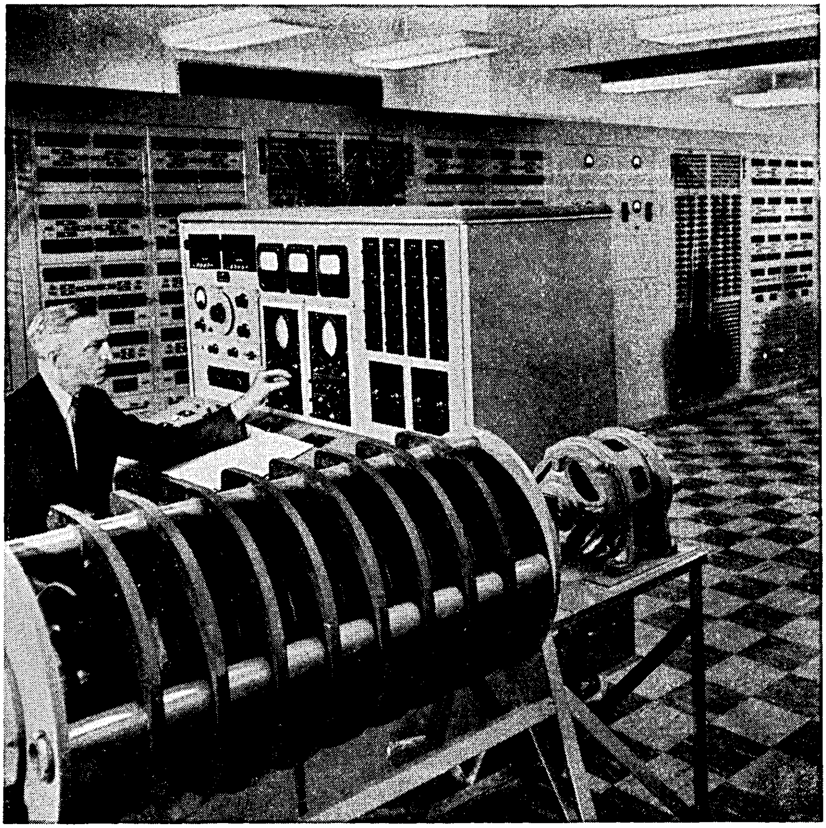


Fig. 15.3 The Anacom, a large-scale general-purpose analog computer located at the East Pittsburgh works of the Westinghouse Electric Corporation; it is used principally for design and application problems of industrial products (Courtesy Westinghouse Electric Corp.)

The Anacom, a large-scale general-purpose analog computer, on which this work was done, is shown in Fig. 15.3.

In the course of this study, very extensive field tests were conducted in which an opportunity was presented to compare the computer and field results, to eliminate any discrepancies, and to establish the computing method on a sound basis. In other words, the analytical representation of the various machines and control circuits was verified.

This paid handsome rewards in the next stage of development of this system which was incorporation of magnetic amplifiers in place of the rotating amplifiers shown in Fig. 15.1. The magamp tandem cold-rolling mill control is shown in Fig. 15.4. The design of this system was based completely on computer studies. The six-to-one improvement in response time and overall superior performance predicted by the computer studies were realized, and the system was placed in service with but a small fraction of the extensive field tests which had been required in connection with the original designs not using computers. This illustrates the paralleling developments that must take place in computation, namely, the development of methods of representation of systems as well as the computational methods of study. Successful use of analog computers hinges on equal attention to these two phases.

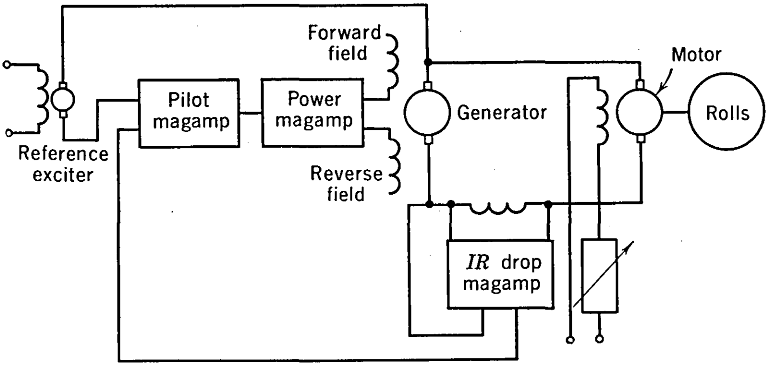


Fig. 15.4 Magamp tandem cold-rolling mill control

Two analog techniques are used on the Anacom. In one, the forcing function is repeated by a synchronous switch at a repetition rate of about ten times a second and the whole solution repeats as illustrated in Fig. 15.1*e*. In this technique, inductors can be readily used because of the reasonably high frequencies involved (100 to 1000 cycles for the principal frequencies in the solution). The repetitive technique also has an advantage, that the solution for any adjustment is displayed continuously, and it is a simple matter to optimize, using two or more parameters, by simply adjusting until the optimum solution is realized. This solution can then be recorded photographically.

In the second technique referred to as the "single-shot technique," the problem is set up on a time base such that the solution requires 1 or 2 seconds or more. This has the advantage that pen and ink recorders can be used, giving multichannel records. At these longer time bases, inductors can be used only with difficulty. A few high-inductance (several henries) inductors are incorporated into the computer for such use. Problems requiring a large number of inductors, such as shock problems of mechanical systems, ordinarily use the repetitive technique at the faster time base. Regulating system studies and controls more commonly use the single-shot technique, although the repetitive technique is used for certain general studies as in Fig. 15.1*e*.

15.2.2 Magamp Generator Voltage Regulator System for Turbine Generators, Waterwheel Generators, and Synchronous Condensers

Here a voltage-regulating system was to be devised suitable for application, with as few modifications as possible, to a complete line of electrical power machinery. Chief considerations were to be reliability and performance but a minimum number of sizes, simplicity of manu-

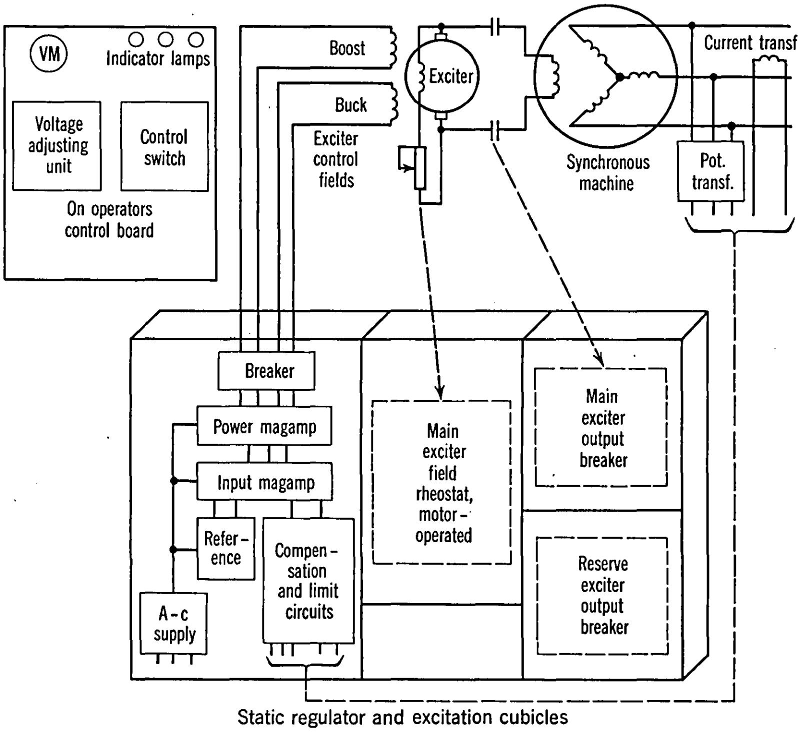


Fig. 15.5*a* A simplified schematic of the new regulating system; VM–volt meter

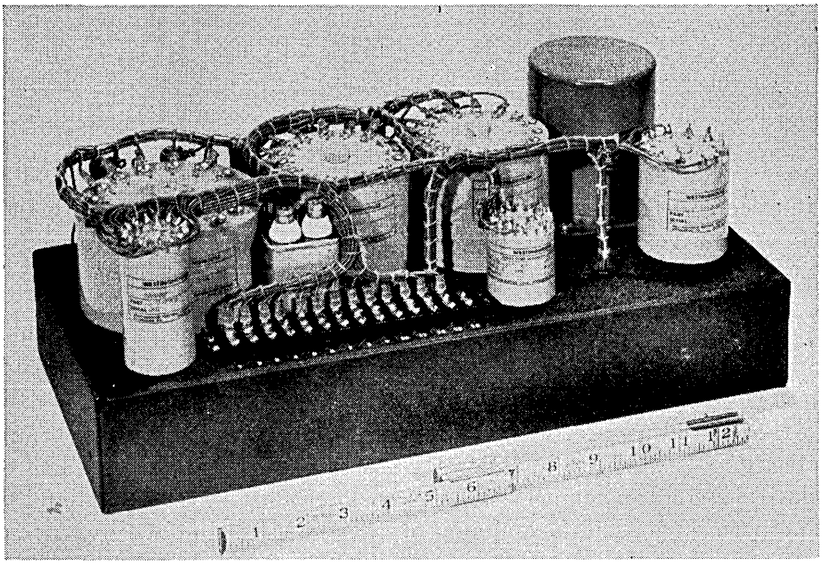


Fig. 15.5*b* Input magnetic amplifier assembly consisting of two stages of amplification, reference, and transformer (Courtesy Westinghouse Electric Corp.)

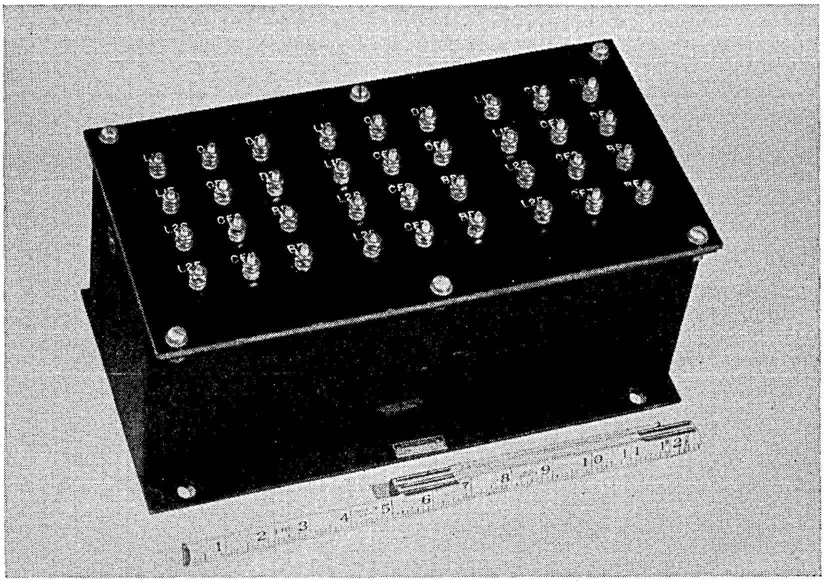


Fig. 15.5*c* Power magnetic amplifier, three-phase reactor assembly. Two such units are required to obtain polarity reversal (Courtesy Westinghouse Electric Corp.)

facture, ease of installation, minimum adjustments, and flexibility for incorporation of limits and other accessories were also to be considered. The analog computer function in this project could be divided into several phases as follows.

(1) Determination of the optimum performance obtainable if the regulating system introduced no limitation and comparison of this performance with the optimum performances obtainable with electronics, with magnetic amplifiers, and with rotary amplifiers. In this stage 400-cycle magnetic amplifiers achieved such a close approach to the optimum with a reliable, permanent, and static form of regulator that magnetic amplifiers were selected for further study and development.

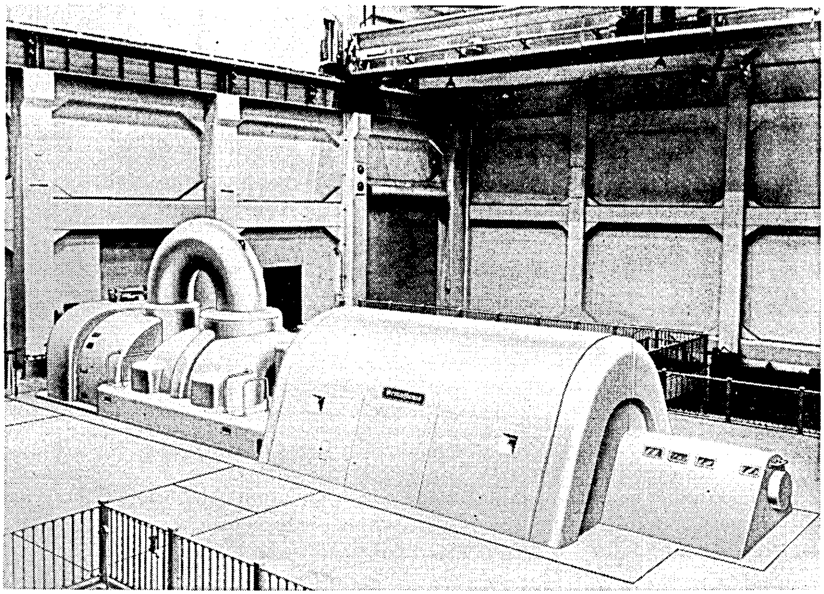


Fig. 15.6 A typical turbine-generator installation; the exciter is in the small housing at the right (Courtesy Westinghouse Electric Corp.)

(2) Studies after the basic design work had been carried out on various regulator components, to determine ratings and requirements, and to establish a line of component parts out of which the complete system could be built for all sizes and types of machines. This phase also included design of accessory limits such as upper and lower excitation limits and field stability limits and their incorporation into the system.

The resulting regulating system is illustrated in Fig. 15.5*a*. The input and power magnetic amplifiers are illustrated in Fig. 15.5*b* and 15.5*c*. Figure 15.5*a* shows that a composite signal taken from the potential and current transformers on the generator passes through computing circuits indicated as compensation and limit circuits. The signal is then compared with the reference and amplified through two stages of magnetic amplifier, through a rotating amplifier, the exciter, to the field of the generator. A typical turbine generator installation is illustrated in Fig. 15.6, the exciter housing being at the right end.

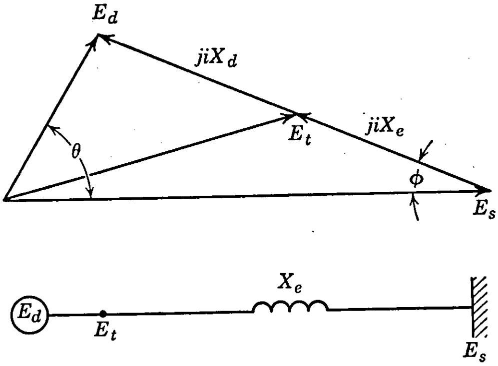


Fig. 15.7 Phasor relationships of voltages which determine stable operation of a generator connected to an infinite system through a reactance, X_e

The analog computation performed by this system may be indicated as follows. By referring to Fig. 15.7, it may be shown that for the generator, having an internal voltage E_d and a terminal voltage E_t to operate stably when connected to an infinite system, voltage E_s, through a reactance X_e, it is necessary that the angle θ of the vector diagram not exceed 90 degrees. In fact, to maintain a margin of stability, the function $E_s - I(X_e + X_d) \cos \phi$ must be maintained above a predetermined minimum value.

As indicated in Fig. 15.8, the analog computing circuits of the system compute this stability function. They also make certain compensating calculations on the voltage, for example, averaging the three phases, providing reactive droop compensation, line drop compensation, and subtracting the result from the voltage reference. An auctioneering-type circuit then performs the logical "or" operation in the selection device, thus enabling the voltage regulator to regulate for voltage so long as stability is not endangered but giving preference to the stability requirement if the field current becomes so low as to require this.

hird phase is the design of a generator-simulato
es the transfer function of the generator and excite
ent power level to provide inputs to the regulator and to a
the outputs therefrom for testing the regulator system in develop
and during installation. It will be described in Section 15.4.2.

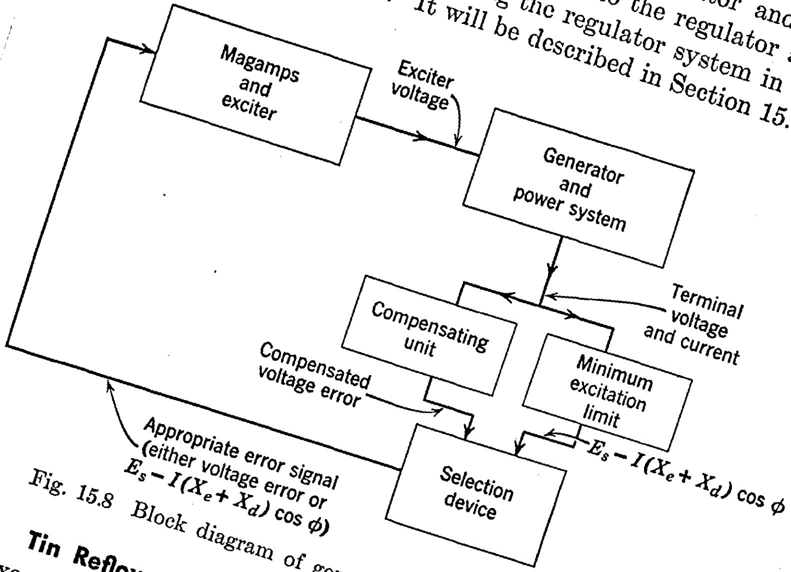
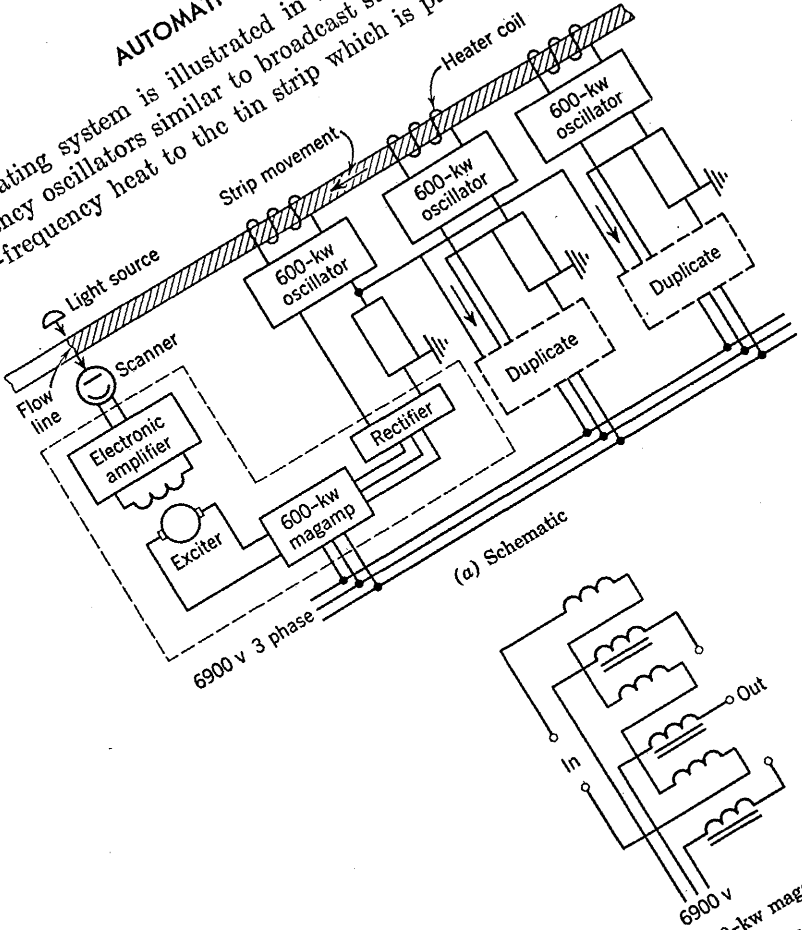


Fig. 15.8 Block diagram of generator voltage-regulating system

15.2.3 Tin Reflow Line

Re-examination of an industrial process of reflowing electrolytically
plated tin aided in establishing the correct transfer functions of parts
of the control for further designs and for improvements. As shown in
Fig. 15.9, the strip passes through three induction heating coils which
must be regulated to bring the electrolytically plated tin just up to the
flow point at the end of the third heater as recognized by a photosensing
device where the strip changes from rough to shiny. Built as a war
measure, when our tin supply was cut off and methods for making
vailable tin go farther were necessary, this control had never received
e full treatment of modern servo techniques and computer methods
s initial installation. Comparison of computer results with actual
rmance revealed that the transfer functions or information-proc-
characteristics of some of the elements were different from
ey had been believed to be. This led to tests on these com-
resulting in corrected transfer functions. The correspondence
ed and actual results then pointed the way to needed im-
in the system.

ating system is illustrated in Fig. 15.9, in which
ency oscillators similar to broadcast station equipment pro-
-frequency heat to the tin strip which is passing through at



(*a*) Schematic

(*b*) 600-kw magamp

Fig. 15.9 Tin reflow regulator

a very high speed. It would tin-plate a football field in about 10 m
utes. The left-hand coil is regulated to control the position of
flow line at the scanner, and the other two coils have similar regu
equipment but are slaves to the first one. The regulating eq
is interesting in several respects; it includes electronic, rota
magnetic amplifiers, whereas the present trend would be to
netic amplifier throughout or possibly electronic or tra

(3) The third phase is the design of a generator-simulator which reproduces the transfer function of the generator and exciter on a sufficient power level to provide inputs to the regulator and to accept the outputs therefrom for testing the regulator system in development and during installation. It will be described in Section 15.4.2.

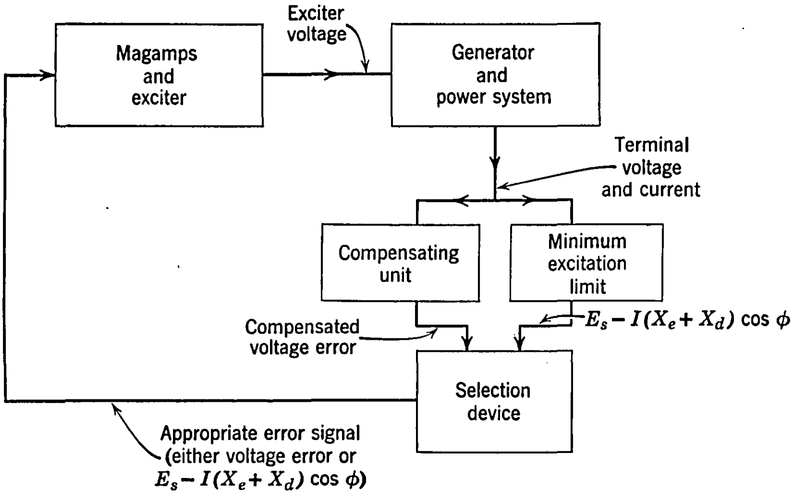
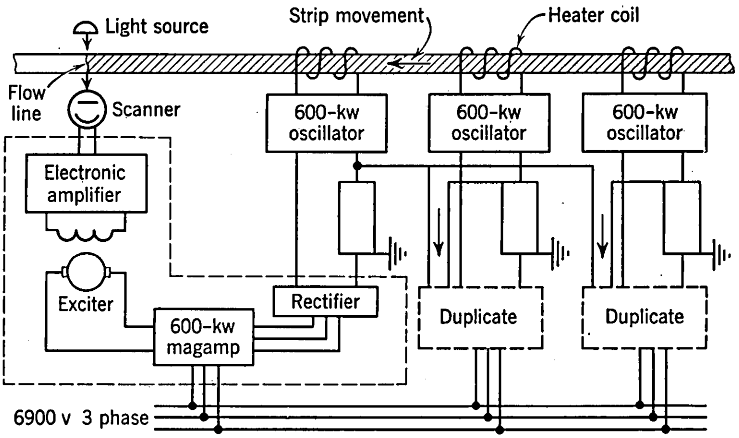


Fig. 15.8 Block diagram of generator voltage-regulating system

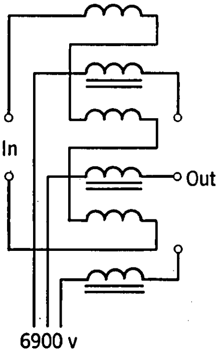
15.2.3 Tin Reflow Line

Re-examination of an industrial process of reflowing electrolytically plated tin aided in establishing the correct transfer functions of parts of the control for further designs and for improvements. As shown in Fig. 15.9, the strip passes through three induction heating coils which must be regulated to bring the electrolytically plated tin just up to the flow point at the end of the third heater as recognized by a photosensing device where the strip changes from rough to shiny. Built as a war measure, when our tin supply was cut off and methods for making available tin go farther were necessary, this control had never received the full treatment of modern servo techniques and computer methods in its initial installation. Comparison of computer results with actual performance revealed that the transfer functions or information-processing characteristics of some of the elements were different from what they had been believed to be. This led to tests on these components, resulting in corrected transfer functions. The correspondence of computed and actual results then pointed the way to needed improvements in the system.

This regulating system is illustrated in Fig. 15.9, in which three high-frequency oscillators similar to broadcast station equipment provide high-frequency heat to the tin strip which is passing through at a very high speed. It would tin-plate a football field in about 10 minutes. The left-hand coil is regulated to control the position of the flow line at the scanner, and the other two coils have similar regulating equipment but are slaves to the first one. The regulating equipment is interesting in several respects; it includes electronic, rotating, and magnetic amplifiers, whereas the present trend would be toward magnetic amplifier throughout or possibly electronic or transistor input



(a) Schematic



(b) 600-kw magamp

Fig. 15.9 Tin reflow regulator

and magnetic amplifiers. A similar drive using a single 1800-kilowatt magamp is shown in Fig. 15.10.

Both the magnetic-amplifier gain and time constant and the electronic-amplifier transfer function were changed after comparison of field results with computer results indicated that modifications were desirable. This resulted in improved methods of representation of these devices and, with these advances in the technique, further improvements were possible in this system as well as correct representation of these devices in later systems. This re-emphasizes the necessity of developing techniques of representation as well as computing methods, in order to obtain the expected advantages of computing of any type.

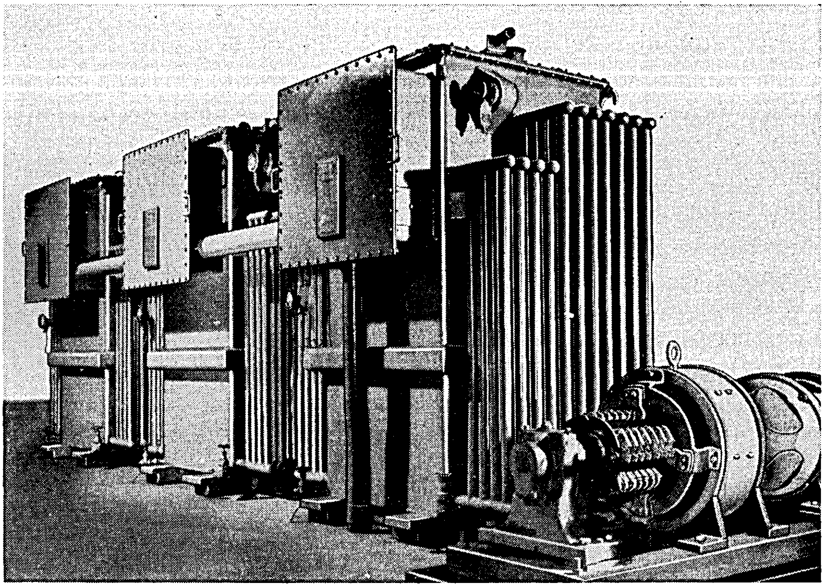


Fig. 15.10 An 1800-kilowatt magnetic amplifier for a tin reflow line (Courtesy Westinghouse Electric Corp.)

15.3 COMPUTER FUNCTIONS IN INDUSTRIAL CONTROLS

When a control or regulating system is set up on an analog computer, the information-processing aspects of it become very clear. Frequently, when a design has been completed, elements substantially identical to those employed in the analog computer may be utilized as part of the control. In other cases, the same transfer functions are secured in equipment of time scale many times removed from that on which the computer is set up and perhaps in a different physical system such as mechanical or hydraulic rather than electric as in the computer. In the past, many a device having extensive computing functions has been thought of as a device that performs a certain function, rather than as a device with both information-processing properties and useful power control or motions. However, it is helpful now to re-examine some of these devices with a view to separation of these functions.

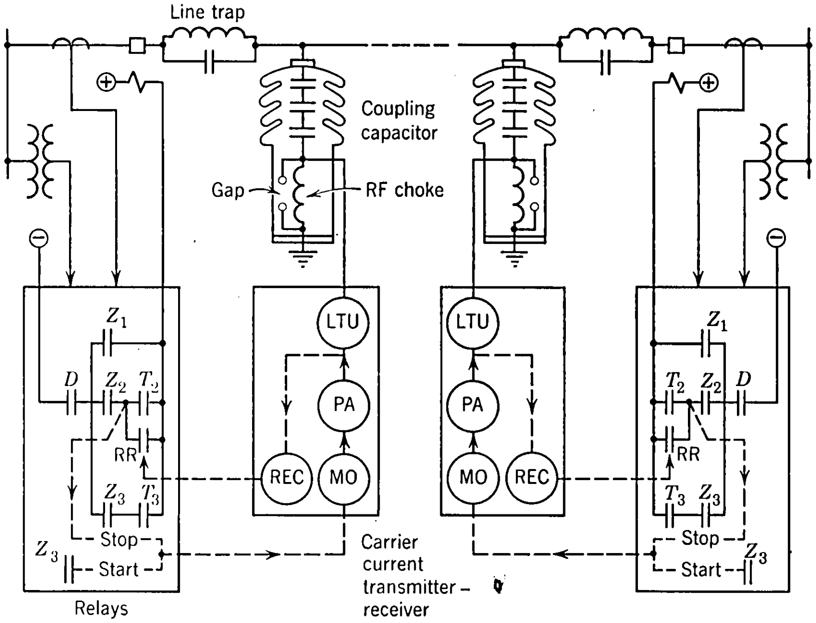
15.3.1 Protective Relaying for Electrical Power Systems

The initial protective relays recognized only overcurrent or perhaps undervoltage in a circuit as a criterion to be examined to determine the tripping action. The information-processing characteristic that can be incorporated in a protective relay system has advanced greatly. Today transmission line protective relaying of the directional-distance variety, utilizing carrier current or microwave circuits, processes the following information in evaluating whether or not the condition observed is a fault within the protected section or not, and hence determining the tripping operation.

The three-phase voltages and the three-phase currents at the relay location are utilized to determine whether the phenomena is of fault magnitude or of load magnitude, whether it is a phase-to-phase or phase-to-ground phenomena, whether it is within 90 per cent of the line length from the relay in question, or beyond this point. In the latter case it might be either in the far end of the section or in the next section of line. Finally, they determine whether it is in a portion of the system which should have been cleared by another relay but has lasted too long, indicating failure of another device.

The relay system also receives information from the other end of the line over the carrier, pilot wire, or microwave channel which settles the question of whether it is in the far 10 per cent of the line or not. Time is also taken into account, with proper recognition of the fact that if the disturbance remains after the far breaker has had a chance to open, it must be within the line section. All of this information is digested and processed within approximately one-sixtieth of a second. The relaying system then makes a correct determination of whether, and when, to open the circuit breaker which it controls. The distance measurements are, in effect, impedance measurements and require voltage-to-current ratios, utilizing balanced forces on a beam or a rotating disc. Some of the logic involved requires "and," "or," and similar logical decision circuits so that actually protective relaying is a mixture of analog and digital computing techniques.

Such a carrier–pilot relay system is illustrated in Fig. 15.11. Note that energization of the trip coil requires that contact D be closed, indicating that a fault is in the line direction; and that either Z_1 be closed, indicating that it is within the first 90 per cent of the line;



D–the directional element contact.

Z_1, Z_2, and Z_3–first, second, and third zone impedance or distance-measuring element contacts.

T_2 and T_3–second and third zone timing relay contacts.

RR–carrier receiver relay contact.

LTU–line tuning unit.

PA–power amplifier.

MO–master oscillator.

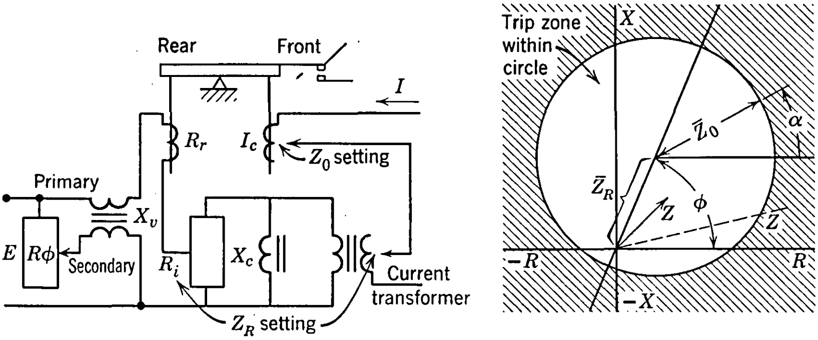
REC–receiver.

Fig. 15.11 Carrier-pilot relay system for a high-voltage transmission line, including relays, carrier transmitter-receivers, coupling capacitors, and line traps; dashed lines indicate symbolically the carrier controls

or that Z_2 be closed combined with a time, T_2, sufficient for the far breaker to have opened; or that Z_2 be closed together with the receiver relay, RR, of the carrier transmitter, indicating that the fault is somewhere on the line or possibly beyond the end of the line, but that the information from the far end of the line resolves that it is within the line section. Under this condition, immediate tripping can take place for faults anywhere on the line.

Typical relaying circuits and devices performing complex calculating functions are illustrated in Fig. 15.12a and 15.12b. Figure 15.12a

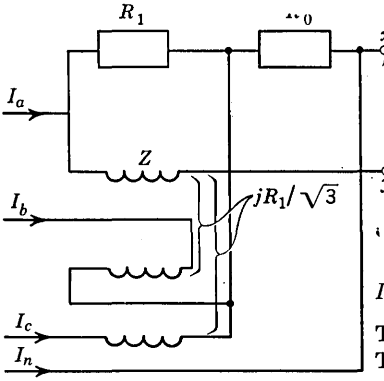
shows a modified distance relay which physically consists of balanced beam with coils on each end, associated with circuitry which includes potentiometers, and current and voltage transformers. With this combination of circuit and mechanical computing devices, fast closure of the relay contacts can be secured whenever the vector voltage and



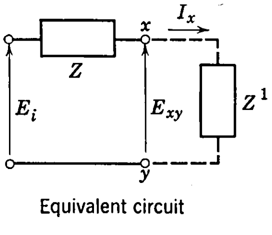
High-speed balance beam modified impedance element. Adjust impedance radius of circle, Z_0, by current coil taps, E, and core screw (air gap). Adjust angle of line along which center is shifted by taps on resistor, $R\phi$. Adjust impedance $\bar{Z}_R$ by which center is shifted by taps on current transformer and on resistor, R_i.

Displaced circle impedance characteristic.
Relay trips for all faults for which impedance, Z, seen by relay falls within circle.
$\bar{Z}_0$, $\bar{Z}_R$, and ϕ are adjustable.
Balance point locus: $Z = \bar{Z}_R e^{j\phi} + \bar{Z}_0 e^{j\alpha}$ for all values of α.

(*a*)



Combined positive sequence current and weighted zero sequence current filter



Equivalent circuit

$$I_x \frac{2R_1}{Z + Z^1}(I_1 + KI_0) \quad \text{where } K = \frac{3R_0 + R_1}{2R_1}$$

The internal voltage is $E_i = 2R_1(I_1 + KI_0)$.
The internal impedance is $Z = R_1 + R_0 + Z$ where Z is the impedance of indicated winding of 3-winding reactor.

(*b*)

Fig. 15.12 Analog calculations are performed by protective relays: (*a*) modified distance relay, and (*b*) filter for pilot-wire relay current applied to the device are such that their complex ratio $R + jX$ falls within the circular boundary shown.

Figure 15.12*b* shows a relaying arrangement in which practically all the computing is done in the circuitry, the output being used to operate a simple overcurrent element which then responds to the very complex function computed. This particular circuit is utilized in a pilot wire relay. This circuit produces an output proportional to the positive sequence current I_1, plus the weighted zero sequence current, KI_0, where I_1 is defined as

$$\tfrac{1}{3}(I_a + aI_b + a^2I_c)$$

and

$$I_0 = \tfrac{1}{3}(I_a + I_b + I_c)$$

where

$$a = -0.5 + j{\cdot}866$$

It can be demonstrated mathematically that this discriminating quantity, $I_1 + KI_0$, has but small variation for all ten types of faults that occur on transmission lines, viz. *a* to *b*, *b* to *c*, *ab* to ground, etc. As such, it affords an ideal discriminating quantity for comparing a single quantity at the two ends of the transmission line over a pair of pilot wires where this quantity is a good measure of the fault or load condition, regardless of its nature.

Thus, analog computing circuits have been in widespread use for relaying, instrumentation, and metering circuits for many years. The examples given above are typical of many hundreds of applications.

15.3.2 Typical Speed Control System

The tandem mill previously described is illustrative of a complex speed control in which analog techniques are utilized to provide the desired following of a speed pattern with matching of rolls, *IR* compensation, inertia compensation, and derivative feedback for stabilization. This is all accomplished by suitable measurements of d-c voltages, the use of damping transformers (mutual transformers having suitable predetermined time constants) and through a correlation of the time constants and amplifications of the control equipment with those inherent in the main power devices. A log carriage drive will also be illustrated to point up analog techniques for processing the information.

As illustrated in Fig. 15.13, a log carriage is a vehicle peculiar to the lumber industry and is used to transport logs back and forth past the saw as the log is reduced to lumber. The electric motor driving the carriage is stationary as towing cables are attached to each end of the carriage and pass over idler sheaves and are then wound on a cable drum mounted on the motor drive shaft. To obtain maximum production, quick reversal is necessary. A d-c adjustable voltage drive is ideally suited to such a task. It is equipped with a brain to obtain the shortest reversal time by forcing the drive equipment to its maximum capacity without exceeding its electrical, thermal, or mechanical limit. This brain is the magnetic amplifier that causes the drive to respond to the sawyer's master-switch signals so long as the resulting motor current is within safe limits. Signals calling for greater current are automatically modified by the magnetic amplifier to limit the current to a safe value.

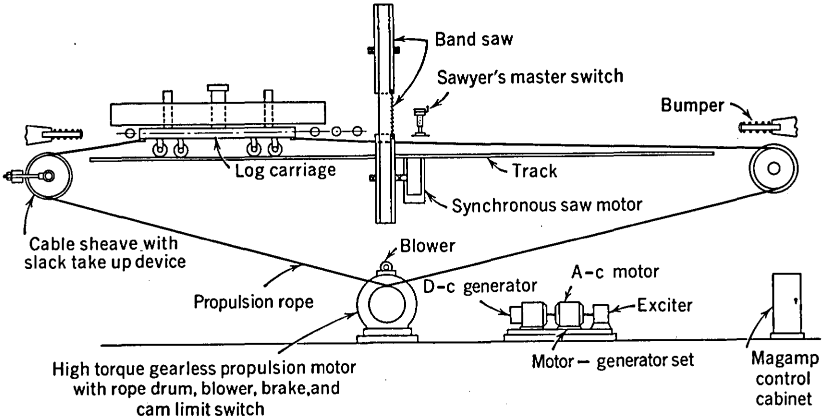


Fig. 15.13 Diagram showing principal parts of a gearless electric log carriage drive

The circuit for this drive incorporates both rototrols and magnetic amplifiers. The current limit circuit compares the drop across an armature circuit resistor with a fixed portion of the exciter voltage and provides overriding excitation on the magnetic amplifiers to overcome any other signals and prevent further increase whenever the armature current gets up to the present limit.

15.3.3 Economic Dispatch Computer for Power System Manual or Automatic Control

A full-fledged analog computer is used in dispatching an electric power system in such a way as to minimize the combined production and transmission costs. As a matter of fact, the computer function is so distinct in this instance that the resulting device can be used either for providing information to the operators for manual operation of the system in the most economic manner, or as a sensing device for an automatic control system which meets other requirements of controlling load or frequency but at the same time dispatches the long-term trend changes in generation in the most economic manner.

The general principle followed by this computer is that the total cost of fuel input to the power system is a minimum when the incremental delivered power cost is the same from every variable station. This requires adding the incremental production cost to the incremental cost of transmission to obtain the incremental delivered power cost for each station.

The computation is complicated by the fact that the cost of transmitting power from any one station to the load depends on how much power each other station is also transmitting to the load at the same time over the same interconnected system. It consequently requires the solution of a system of simultaneous equations, twenty equations if there are twenty stations and ties, in order to determine the economic dispatch at any time.

The computer contains well-known analog elements, servomechanisms, potentiometers, amplifiers, and is characterized by a rather complete display of the situation existing on the system at any time and into which upper and lower limits for the various stations can be set. Provisions are made for quickly changing the station cost curves as new units come on the line or are cut off. The unit thermal-efficiency curves remain constant over long periods but provision is made for rapidly changing fuel cost or other factors that may change the station costs. Thus the operator utilizing the computer either directly or through an automatic control works at a higher level in which he feeds the basic information on system conditions and fuel costs into the computer, but it takes over the detail job of determining the best dispatch based on these factors.

As used in an automatic control, raising or lowering pulses initiated by the load control equipment are routed by the computer so that raising pulses go only to stations which have delivered cost somewhat below the average and lowering pulses go only to stations which have costs above the average. In this way, all delivered costs are kept substantially the same so that the economic dispatch results. This will be illustrated by displaying the equations to be solved by the computer and describing its uses.

The m simultaneous equations to be solved by the dispatch computer for an m station system are:

$$C_j + 2\lambda \sum_{n=1}^{m} P_n B_{jn} = \lambda; \qquad j = 1, 2, \cdots, m$$

where $C_j = f(P_j)$ is the incremental production cost of station j; λ is the Lagrangian multiplier, here the delivered power cost to the average load; P_j is the power delivered by station j; and B_{jn} are constant coefficients of the transmission loss formula.

The computer must further determine the delivered power cost, λ, at which

$$\sum_{n=1}^{m} P_n = P_g$$

where P_g is the total generation set into the computer. Adding this relationship, there are $m + 1$ simultaneous equations to be solved. The resulting set of P's are the powers that should be generated at each station of the system.

The use of such a computer may be illustrated by four typical problems. The illustration, Fig. 15.14, shows diagrammatically the components of an economic dispatch computer.

First, the dispatch for a particular total generation is to be determined. Assume that station cost curves have been set by means of the plug board (see Fig. 15.14) and that the transmission loss coefficients pertaining to the transmission system have also been set into the computer. With the selector switch in the compute position, the total generation dial is set at the desired generation value: the several station power indicators then move under the control of the computer circuits to positions that indicate the proper amount of power to be dispatched from each station. The meter at the top indicates the average delivered power cost and the second meter shows the incremental cost at any station or tie line for this condition, by depressing the corresponding station cost button. As a variation of this problem, if the cost at station 1 is increased 10 per cent above that for which the station cost curves were calculated, the small dial under the station dial would be moved up to 1.1 and the computer would immediately show the correct dispatch for this condition.

Second, suppose the system consists of two generators and one tie line, and 50 megawatts of scheduled power are coming in over the tie line; the economic dispatch of the system generation under this condition is desired for a particular total generation plus tie power. The third station representing the fixed tie would be thrown to the fixed position (flip-flop switch), and set on 50 megawatts. The total generation dial is again set for the desired total and the two station dials positioned to show the correct power dispatch.

Third, assume that the incremental cost is desired for an additional block of 50 megawatts brought in over the tie line. The incremental value of this power to the system changes as the increment is increased from zero to 50 megawatts, and the cost desired is the average incremental cost. This can be obtained accurately by the computer, or approximated by adjusting the tie power setting to show half of the power increase and reading directly the incremental value at this point. This is the average incremental cost for the full block. It can be determined either with the additional power used to substitute for some of the running generation, or by taking on this additional block

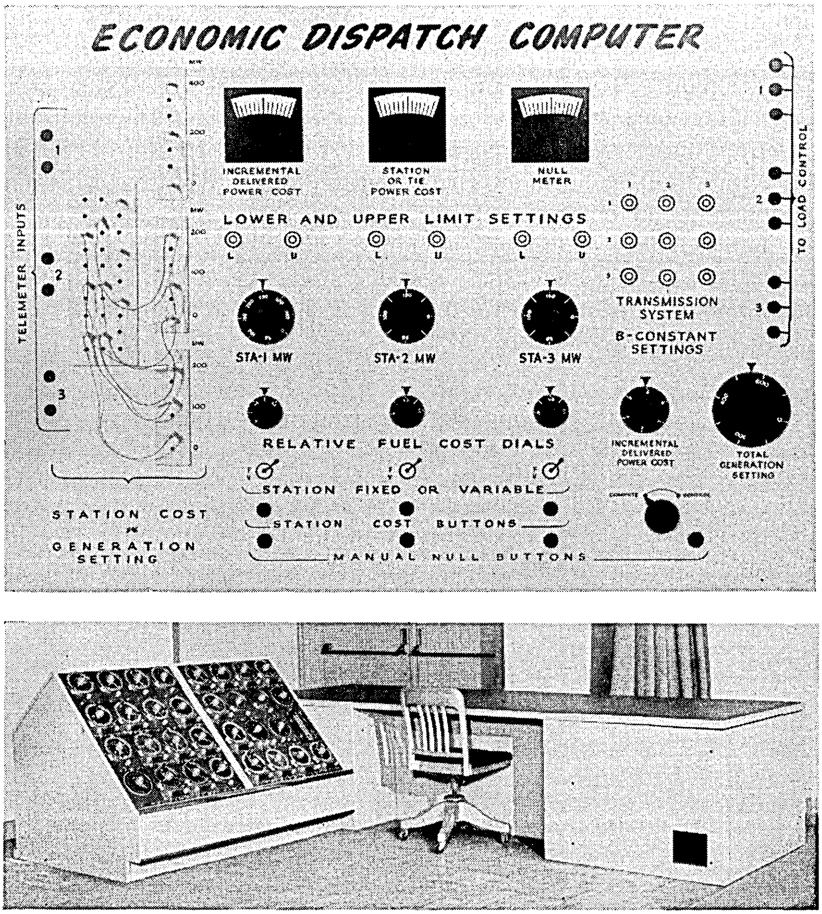


Fig. 15.14 Display model of an economic dispatch computer for an electric power system (Courtesy Westinghouse Electric Corp.)

to feed the next block of load on the system. The computer gives the answer to either question.

Fourth, suppose the computer is to be cut into automatic control of the system. The telemetered indications of generation at each station are brought into the computer, and the control switch is thrown into the "control" position, which makes certain changes in the computing circuits. The station indicators then show the actual generation at each station instead of the economic dispatch; contact circuits are set up to route the raising pulses of the load control system to stations having low incremental power cost and the lowering pulses to stations having high incremental delivered power cost. The pulses can be originated by any automatic load control system. The computer thus acts as the sensing element and does not interfere with the normal functioning of the load control system in assigning frequency and swing load contributions.

15.4 USE OF SIMULATION COMPUTERS

Automatic control equipment utilized with large expensive machinery or to control expensive processes poses a question of testing and development as well as security in placing the unit into service at any given time. Great reliance is placed on the control to avoid overloading or burning out of portions of the equipment or actual destruction of the major equipment from subjecting it to conditions beyond its strength or endurance. The simulator, utilizing either analog or digital computer techniques, has come into the picture to solve many of these problems and appears to be a very important tool in automation. Three examples will be cited to illustrate some of the uses and advantages. Most of the simulators to date are of analog variety. Simulators have been used very effectively in military development as well as industrial. Only industrial applications are mentioned here.

15.4.1 Transient Performance of Potential Devices and High-Speed Relays

High-speed protective relays connected to transmission lines and receiving their potential supply through coupling capacitors are expected to discriminate faults from loads correctly under a variety of conditions. The number of variables is tremendous. Faults can be at different locations on the system. There are four different types of phase-to-phase and phase-to-ground faults. There are various designs of coupling capacitors and potential devices used with them and various designs of relays to be considered. The loading on the devices differs from station to station. Although the device itself is substantially linear and might be calculated, the analysis of whether or not its transient output acting as one of several variables on the protective relays would result in their correct discrimination or not is almost hopeless. To test the complete system by placing primary faults on the power system can be done for a few check cases but is far too costly for any complete analysis. Furthermore, probability enters into the picture since the faults may occur at various points in the voltage wave in time; therefore a complete analysis requires varying this factor also.

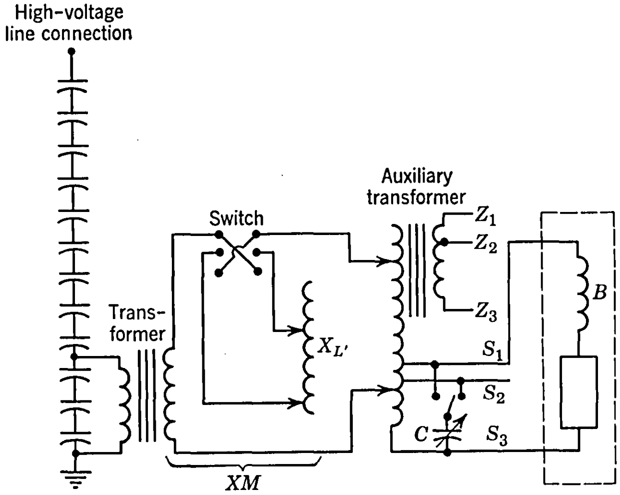


Fig. 15.15 Coupling capacitor potential device

The use of a large-scale computer as a simulator proved to be the best answer to this problem. The simulator correctly represented the power system, the coupling capacitor, and the tuned potential device and was set up on a power base that could supply the actual output potential to the relay together with the associated current to which it would respond. Tests could then be made under sufficiently controlled conditions to provide statistical answers to the probability phases of the problem and to determine the degree of performance obtainable. This also made possible a comparison of different relay characteristics under these transient conditions.

Figure 15.15 illustrates the circuit of a coupling capacitor potential device. The coupling capacitor is simply a small very high-voltage capacitor connected from a transmission line to ground. The lower section, about 4000 volts, is tapped, stepped down through a potential transformer, and connected to a relay burden, B, through suitable auxiliary circuits. A series reactor tunes the source capacitance to give the device good regulation. Such devices are used to supply potential to protective relays and to meters having burdens up to approximately 150 watts. They are considerably cheaper than a high-voltage potential transformer, which accounts for their use in spite of the apparent greater degree of complication. Three such devices would be required for a three-phase transmission line with the outputs interconnected to measure line-to-line or line-to-ground voltages as needed. Figure 15.16 illustrates such a device.

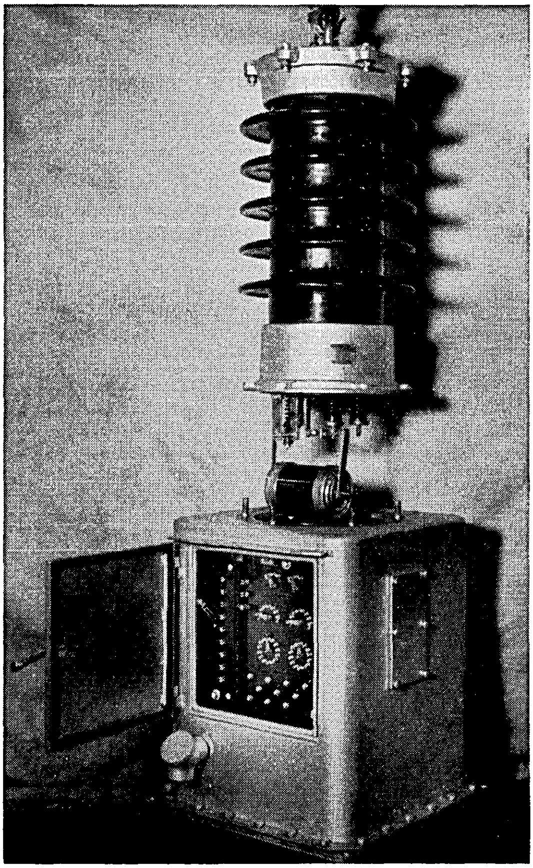


Fig. 15.16 Type PCA potential device with coupling capacitor partially removed and door open to show adjusting panel (Courtesy Westinghouse Electric Corp.)

The Thévenin equivalent circuit, Fig. 15.17, shows how the device can be considered as having a voltage proportional to the high-voltage line voltage and having an internal capacity, C, whose reactance is tuned out by inductance L at the normal line frequency. Loss is represented by resistors. Figure 15.18 illustrates the simulation circuit

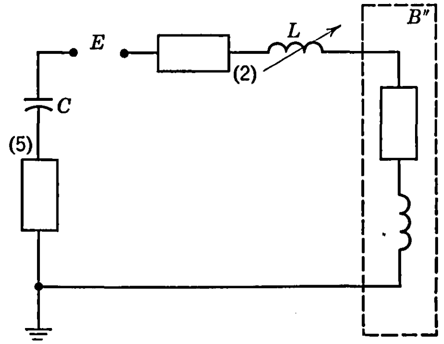
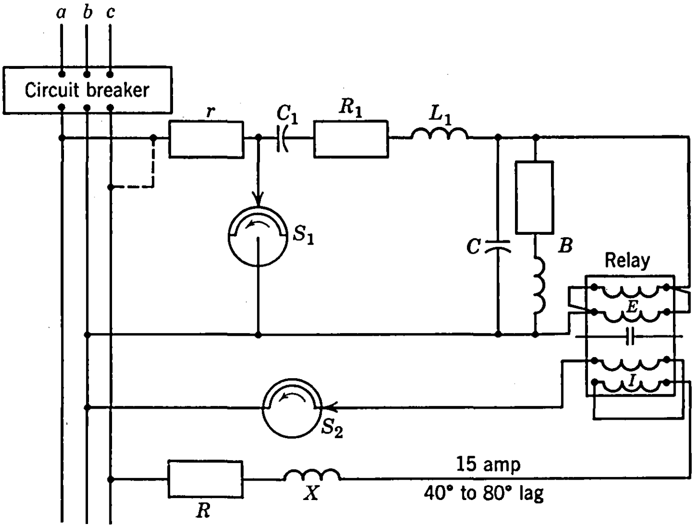


Fig. 15.17 Simplified circuit according to Thévenin's theorem; "B" is the equivalent burden



Phase rotation a-b-c.
B–auxiliary burden.
C–power factor correction capacitor.
As shown for three-phase fault, as dotted for line-to-line fault.

Fig. 15.18 Simulation of potential device and transmission line for combined fault tests with protective relay used to test protective relays on a low-voltage circuit and still include the effect of the capacitor-potential device and the transmission line. The two synchronous switches, switch 1 and switch 2, close simultaneously and can be adjusted to close at any point of the voltage wave. These correspond to the initiation of a short circuit on the power system. With the relay normally energized with three-phase, 60-cycle potential through the circuit breaker and the potential device and with the relay current initially zero, the closure of switch 1 reduces the input voltage to the simulated potential device to zero, *r* being simply a protective resistor. Thus the relay voltage contains the oscillations which would be encountered in such an operation under actual conditions owing to the tuned nature of the potential device. At the same time, switch 2 passes current through the relay current coil having proper phase position and transient characteristics since the line impedance angle has been represented by the *R* and *X* as shown.

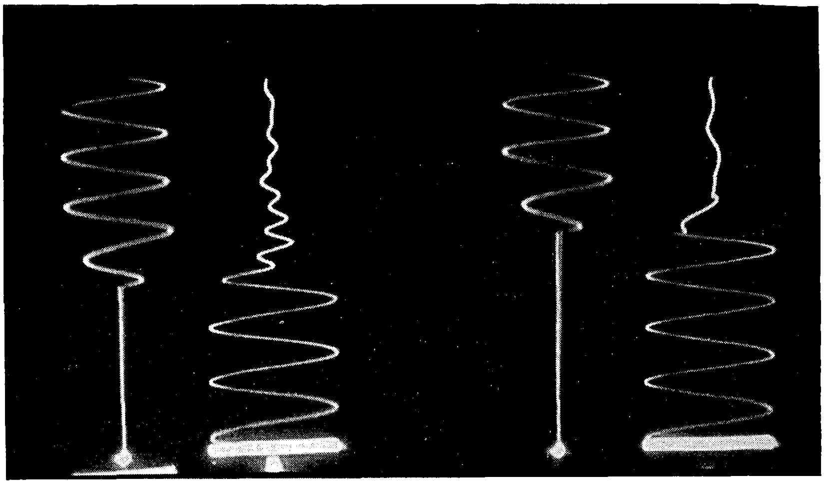


Fig. 15.19 Current (upper) and voltage (lower) traces from transient test using simulator

In the general study which it was possible to make on this simulated setup, the switching angle, the type of fault, the line impedance angle, the relay burden, the power factor correction, and the type of relay and rating of potential device could all be varied at will to obtain general data required for application purposes. This would have been entirely impractical on the actual system. Current and voltage oscillograms for different conditions are illustrated in Fig. 15.19, the upper trace being current and the lower trace being voltage in each case.

15.4.2 Generator-Simulator for Voltage Regulator Testing

To tie up a 100,000- or 200,000-kilowatt turbine generator for extensive testing of its voltage regulator system is economically most undesirable. Waterwheel generators, on the other hand, are first assembled in the field, which means that the application must be worked out quite completely before the regulating system is installed so that it can be placed in service with practically no testing whatever. The principal elements that must be represented by the simulator are the generator together with current and potential transformers. The generator is characterized by its time constant and amplification and a saturation curve and by relations concerned with its reactances and armature field coupling which determine the output voltage characteristic when the machine is suddenly loaded.

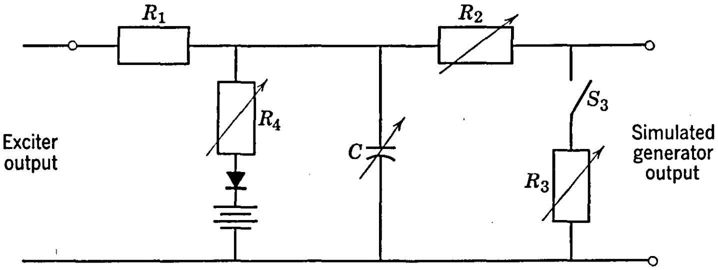


Fig. 15.20 Generator simulator for voltage regulator and exciter tests

The differential equations can be written for all parts of this system, and a simulator can be constructed to produce the same transfer functions and, at the same time, to provide sufficient output power to supply the regulator. This device is used for development of the regulating system and for checking specific regulators having particular time constants and gains characteristic of an application before connecting the regulator system to the main machines.

The general principle of the simulator is illustrated in Fig. 15.20. Although all of the relationships are established by making the mathematical transfer function of the simulator exactly equal to that of the generator, the general functioning may be explained as follows.

In testing the voltage regulator, it is assumed that the exciter is available. For example, the generator may be running on the spare exciter while the regulator exciter is being used for test or, in the case of waterwheel generators which are first assembled in the field, the exciter may be available in the shop for test with the voltage regulator. Thus, the simulator only need represent the generator from its field input terminals to its output. The simulator output corresponds to the output of the rectifier which averages the three-phase voltages.

The generator field time constant, unsaturated, is represented by R_1C. At voltages above saturation, the circuit R_4 with diode and battery cuts in, reducing this time constant and also the gain of the machine in proportion to that encountered in the actual machine. This obviously corresponds to representing the saturation curve by two straight lines—which is sufficiently accurate for this application. Application of load to the generator is simulated by closing R_3. It can be demonstrated mathematically that this results in the proper reduction of transient time constant and proper abrupt voltage drop (transient drop). Thus the simulation covers all basic elements of the generator transfer function, the gain, the time constant, the effect of saturation, and the reactance relationships that enter into sudden loading.

With this simulator, it is possible to test the voltage regulator in the factory without having the actual generator available or to test a regulator during installation, making sure that it is in proper operation and correctly adjusted before connecting it to the main machine.

15.4.3 Wind Tunnel Machine Simulator for Control Supervision

The main compressors for the Tullahoma wind tunnel involve 216,000 horsepower of motor capacity on a single shaft. There are two 25,000-horsepower wound rotor induction motors and two 83,000-horsepower synchronous machines. The machine is brought up to speed by means of the wound rotor motors. An installation of this size requires careful attention to many design features and limitations in the control equipment and rheostats which might be taken care of by overbuilding in a much smaller installation. Great dependence is therefore placed upon the control equipment in bringing the motor up to speed without exceeding thermal or mechanical limitations of any element of the system. The intelligence and all of the limitations are, in effect, programmed into the control system. However, two problems arise. First, in making adjustments it is undesirable to start and stop the huge machine many times; it is also a hazard since unless the adjustments are right in the first place, damage may be done to the equipment. Second, when the machine has been shut down and maintenance done on the equipment, some method of check is necessary to assure the operator that everything is in working condition before he entrusts the main equipment to the control. The simulator provides both of these functions and is a permanent part

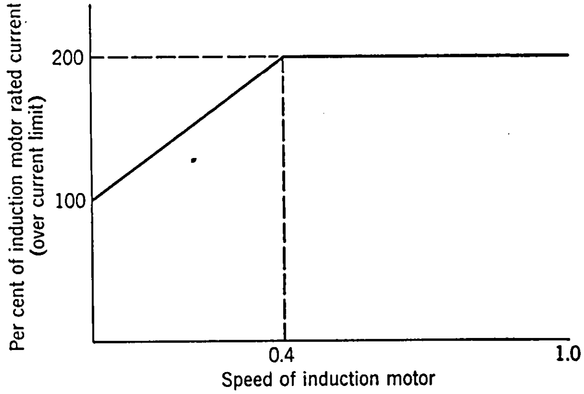
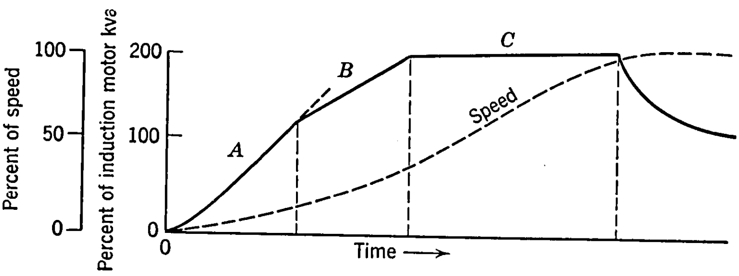


Fig. 15.21 Speed of induction motor over current regulator characteristic



A-kva rate limit—30 mva/min.
B-over current limit.
C-over current limit.

Fig. 15.22 Starting limit characteristic

of the installation, being utilized as needed to test the functioning of the control before entrusting to it the main machinery.

The operation of this simulator may be understood by reference to Figs. 15.21, 15.22, and 15.23. The permissible current as a function of speed is shown in Fig. 15.21. However, a kilovolt-ampere rate limit is also imposed by the power company so that as the motor comes up to speed as shown in Fig. 15.22, the current is initially limited by the kilovolt-ampere limit and then by the current limit; finally it drops off as the motor comes up to full speed.

The simulator in this case represents the rotating machine characteristic. All of the regulating equipment and the liquid rheostat for the wound rotor motor are to be used, although the rheostat will not be carrying full current. It will, however, assume the correct positions as the simulated motor comes up to speed, and the speed regulator and kilovolt-ampere and current rate limits will be effective. This simulator is illustrated in Fig. 15.23 in which a shaft shown at the top is driven by a series field motor to a position that corresponds to the motor speed at any time; thus speed readings are taken off by the two potentiometer circuits to the speed regulator and the kilovolt-ampere rate limiter. A current is produced in the liquid rheostat proportional to what it should be at that speed by a three-phase variac connected to the speed shaft whose output is applied to the terminals of the liquid rheostat. The resulting proportional currents through suitable current transformers supply, to the regulator current circuits, currents appropriate to that speed.

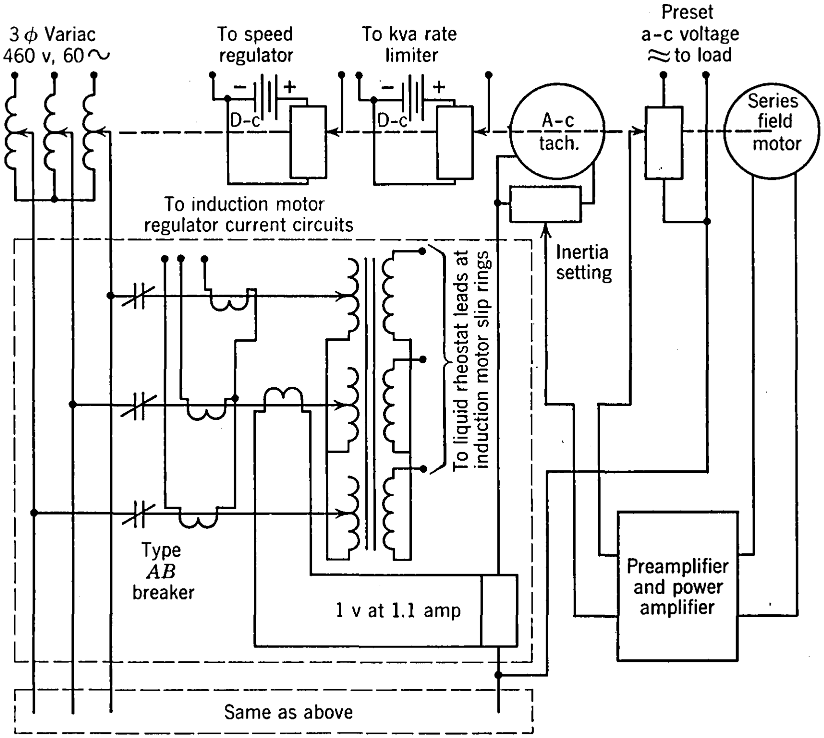


Fig. 15.23 Tullahoma drive simulator

The speed shaft is brought to the proper position at all times by a circuit that satisfies the torque equation. The torque balance requires that the load torque plus the accelerating torque equal the motor torque. The sum of these three quantities is supplied to the input of the amplifier which runs the series field motor to a speed and position at which the equality is satisfied. The tachometer produces a voltage proportional to the inertia torque since the shaft position corresponds to motor speed and hence the shaft speed corresponds to motor acceleration. Note that this inertia can be adjusted to correspond to connection of the supersonic or transonic compressors or various motor combinations. The motor torque is proportional to rotor current and appears as the drop across a resistor. The load torque is a function of speed and connected load, and these adjustments are provided.

Thus in use, the simulator makes it possible to completely test the regulator and all of its limiting functions, together with the liquid rheostat, for proper operation before it is connected to the main drive. It can readily be visualized that as systems become more automatic and more complex, such simulators of the associated equipment, needed to check out vital control parts, will become more and more important.

15.5 REFERENCES

1. E. L. Harder and W. E. Marter, Principles and Practices of Relaying in the United States, *Trans. AIEE,* 1948.

2. E. L. Harder and M. A. Bostwick, A Single Element Differential Pilot Wire Relay System, *Elec. J.*, 443, Nov. 1938.

3. E. L. Harder, W. E. Jacobs, R. W. Ferguson, and D. C. Harker, Loss Evaluation—II, Current- and Power-Form Loss Formulas, *Trans. AIEE,* 1954.

4. R. H. Travers, D. C. Harker, R. W. Long, and E. L. Harder, Loss Evaluation—III, Economic Dispatch Studies of Steam Electric Generating Systems, *Trans. AIEE,* 1954.

5. E. L. Harder, Economic Load Dispatching, *Westinghouse Eng.,* Nov. 1954.

6. R. E. Wendt, Jr., and J. T. Carleton, A-C Generator Simulator and Response-Ratio Meter for Testing Voltage Regulators, ISA Paper 55-17-2, Sept. 1955.

16. Digital Control of Machine Tools

JOHN L. BOWER

North American Aviation, Inc.
Downey, California

16.1 BASIC CONSIDERATIONS

16.1.1 The Economic Aims of Numerical Control

The objective of most engineering is the design of a device that will conserve human effort or will make possible a new useful function. These are also the objectives in the case of numerical control of machine tools. A properly designed numerical control system can often save enough in machining lots as small as one unit to provide ample justification for installation. Product designs are now generally restricted to those shapes that are easy to generate by motion along or about one of the axes of a machine tool. If a coordinated motion with respect to several axes at the same time could be accomplished, the possible shapes that could be generated would be much less limited. Further, by proper use of data on the known errors in the machine, at least the repeatable portion of those errors can be eliminated so that more precise machining becomes possible.

To accomplish the above, the machine controls should be designed to eliminate mental as well as physical operations on the part of the machinist. A surprisingly large part of the machinist's time involves no metal cutting. Layout is very time consuming because it requires a combination of calculation, measurement, and precise scribing.

Again, often overlooked is the fact that the routine sequencing of the controls governing coolant, drive motor, and change gears calls for a significant amount of mental activity, and a careful operator may spend more time on those functions than on cutting metal. The value of eliminating operator decision is especially apparent in the case of controls of a large machine such as a skin mill, which normally requires more than one operator. The high cost of a possible error and the complexity of the operating function for such a machine make it imperative for the operator to consider every step carefully if scrap is to be minimized.

In considering the design of a machine we cannot overlook the cost of preparing the data necessary for control. The number of controlled operations and the manner in which they are controlled are principal determinants of the operating cost of a system.

16.1.2 Interrelation of Control Functions

For most systems the control of each axis can be considered as a separate numerical control, with requirements and problems separate from those applying to other axes, until the preparation of data is undertaken. The latter operation, of course, calls for coordinating the functions of all controls. In our present consideration of means of assembling numerical controls, therefore, we shall regard the several axes as independent until the study of programming is undertaken.

In some machines it is most practical to allow one axis of movement to be driven independently by conventional means and to control the other axes as a function of the independent axis. In the case of a lathe, for example, the cross-head movement is made a function of travel of the tool along the axis of the workpiece. This principle of operation involves no important change in technique since we are substituting axial movement for time as the variable against which cross-feed is numerically programmed.

16.1.3 Position and Contour Control

If a machine is to be used only for drilling or boring, accurate control of position while in motion is not required. The accurate setting of the axes of a machine to carry out an operation at fixed location will be called *position control.* If routing or milling is to be done on complex curves, it must be possible to move the machine in two or more axes simultaneously, with the relative positions on the axes as defined by the inputs to the system. Such control of movement will be referred to as *contour control* to distinguish it from position control defined above. In some machines, it will be desirable to provide a combination of position control and contour control. As an example, a horizontal milling may require position control parallel to the rotation of the mill but contour control in two axes normal to the axis of rotation. For maximum speed of operation and for simplicity of command, in some applications it is desirable to provide both position and contour control on the same axes. The positioning instruction "Move to position defined by $X = \cdots$, $Y = \cdots$, $Z = \cdots$" does not require the definition of a path of movement and can be executed at a speed limited only by the capabilities of the drive of the machine; contour control, on the other hand, calls for both path definition and some limitation of speed.

16.2 A SIMPLE POSITIONING CONTROL

16.2.1 Requirements

To acquire some familiarity with simple techniques of numerical control in positioning, we shall consider the problem of positioning a drill press in two coordinates, X and Y, for drilling holes in plate stock. See Fig. 16.1 for location of two holes x_1, y_1 and x_2, y_2. We wish the machine to move to a hole position to be specified on a tape, and when both axes are positioned within the least count, to signal the drill head to feed the drill into the work. The drill then retracts, a new hole position is called for from the tape, and the cycle is repeated. We shall assume that the movement in each coordinate is limited to about 9 inches, and that positions are to be specified to the nearest 0.001 inch.

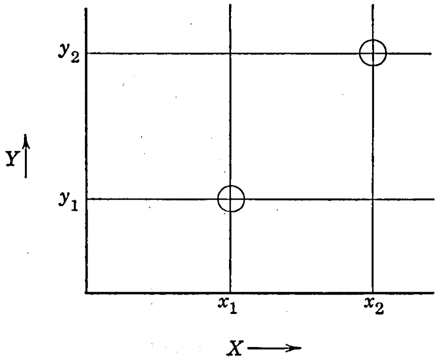


Fig. 16.1 Coordinates for position control

16.2.2 Solution

The rough layout of a design to meet the requirements is given in Fig. 16.2. A punched paper tape will be supplied with one channel specifying X position, another Y position, and the third one (at top)

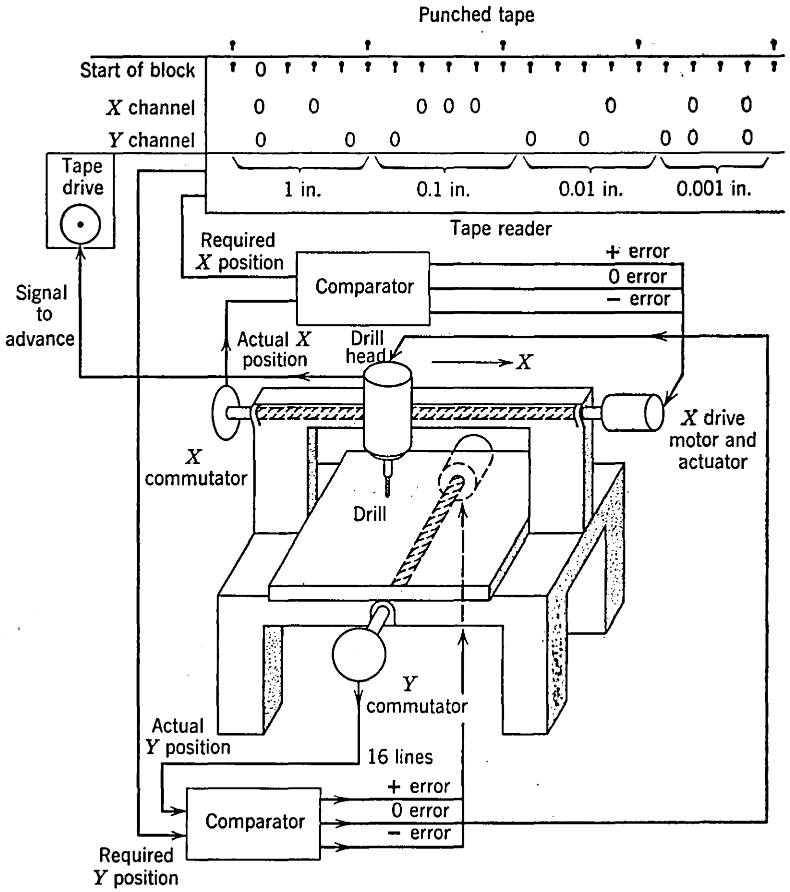
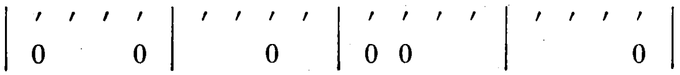


Fig. 16.2 Position-controlled drill press

used to indicate the start of a block of data. For an X position there will be sixteen holes, four for each of the four digits, the four on the left indicating the number of inches, the next four the tenths of inches, and so on. If in each group of four we code the hole positions as

$$\begin{matrix} 0 & 0 & 0 & 0 \\ 1 & 2 & 4 & 7 \end{matrix}$$

we could represent the location $x = 8.437$ inches as



This code calls for punching no more than two holes per digit.

The foregoing indicates how the machine will receive data. The holes can be read in several ways—by shining a light at the hole position and detecting with a photocell whether or not the hole is there, and by sampling each hole position with a pin that passes through to actuate a switch when a hole exists.

We now need to consider the means of moving the drill to the indicated coordinate. In Fig. 16.2 the table must be moved to meet the specified Y position, and the drill head is moved across the beam to arrive at the X position. We can drive each of these axes through a lead screw, gears, and drive motor as shown. To learn the actual position at every instant, we shall attach to each lead screw a commutator similar to one described in Chapter 9. The X output from the commutator is a coded number of 0.001-inch increments, as in the coding for the tape. The X output is fed to the "comparator" whose function it is to compare the coordinate specified on the tape with that obtained from the commutator and to indicate which is the greater. This unit is a variety of circuit like those discussed by Mr. Ware in his chapter on digital computers. Here, however, the only output required is the polarity of the difference or the fact of coincidence in value. When the number on the tape is greater than that from the commutator, the X drive motor is caused to move the drill head to a higher number; when the error is negative, the opposite movement is called out. When the difference is zero the drive motor stops.

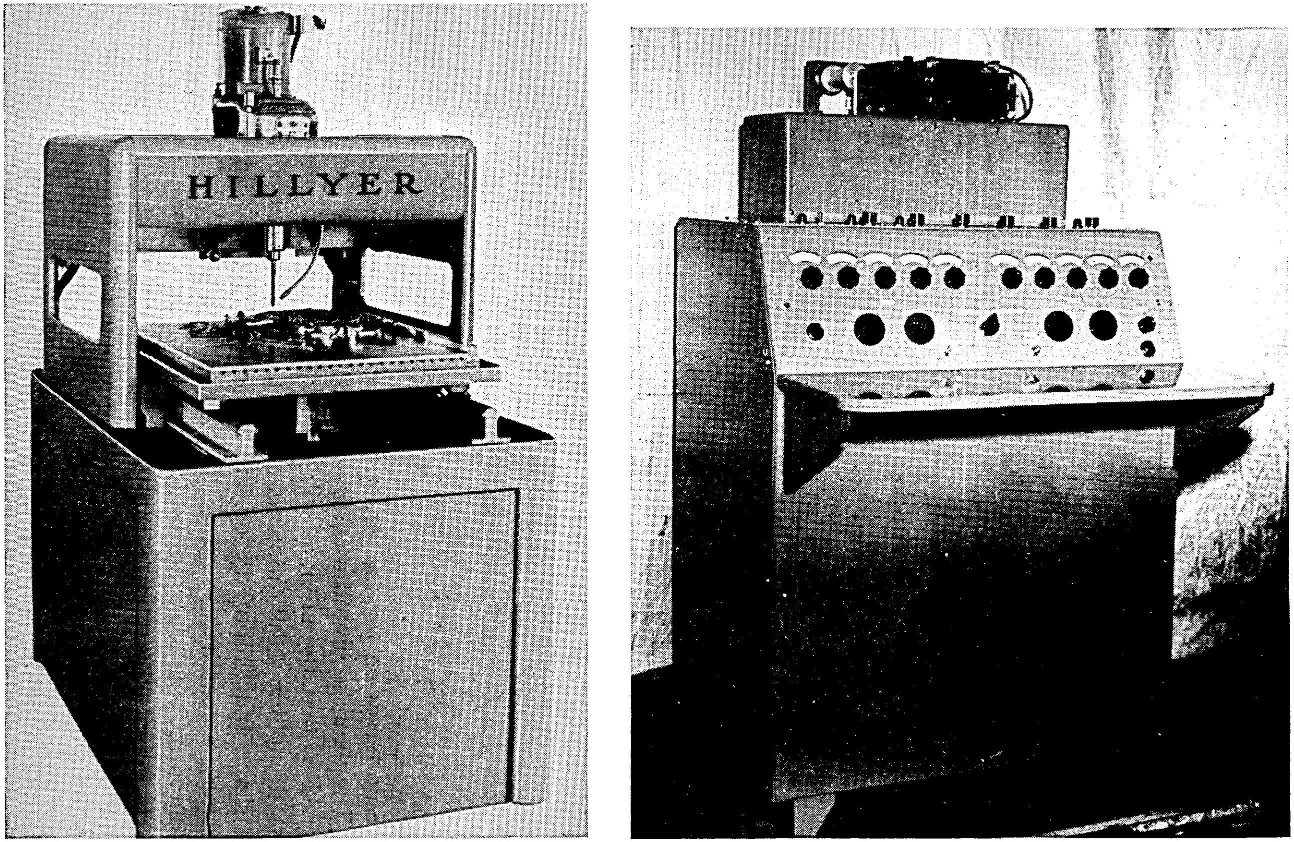
Thus the X position and similarly the Y position are brought into correspondence with the coordinates called out on the tape. When coincidence (zero error) is reached in both axes, the drill head is signaled, it drills through the part, retracts, and signals the tape reader to advance to the next drill position on the tape.

16.2.3 Example of Automatic Drilling Machine

The scheme of control described here resembles in certain ways the positioning system of the Hillyer drilling machine made by Hillyer Instrument Company, Inc., New York City. A photograph of the machine is shown as Fig. 16.3.[1]

---

[1] See R. H. Spiotta, Electromechanical Spot-Drilling Reduces Costs, *Machinery,* 229–231, April 1954.



(*a*) (*b*)

Fig. 16.3 The Hillyer numerically controlled drilling machine: (*a*) drill, (*b*) control (Courtesy Hillyer Instrument Co.)

16.3 CONTOUR CONTROL

16.3.1 General Methods of Control

The requirements for a system executing contour control are more severe than those for position control systems in several ways. (1) The contouring machine must be capable of moving continuously along a specified path with the required precision, while the positioning machine need be accurate only when stopped at an operation point.

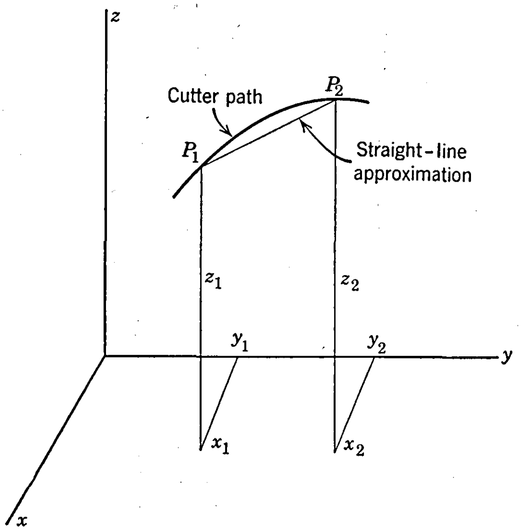


Fig. 16.4 Point-by-point specification of a path

(2) Side forces on the tool are generally more severe in contouring. (3) The data fed into the machine for contouring must be carefully planned so that the machine will not be called upon to make sharp turns or sudden stops at points where accuracy is important. (4) The shape and size of the cutting tool must be considered in programming a contour.

Contouring can, of course, be done in two or three dimensions. The control for all axes of either system is basically the same. However, the programming for three-dimensional cutting is much more complex than for two dimensions, since a specified surface, rather than a specified line in a plane, must be approximated by the cutter. We shall not concern ourselves at this point with the details of programming or with the differences between two- and three-dimensional contouring. Instead, we shall look at the problem in general to see how contouring might be accomplished. Data for the machine can be presented in many ways.

(1) See Fig. 16.4. The coordinate of each point on the surface along the cutter path might be specified. This requires a great many numbers per unit length of path (x_1, y_1, z_1; x_2, y_2, z_2; etc.).

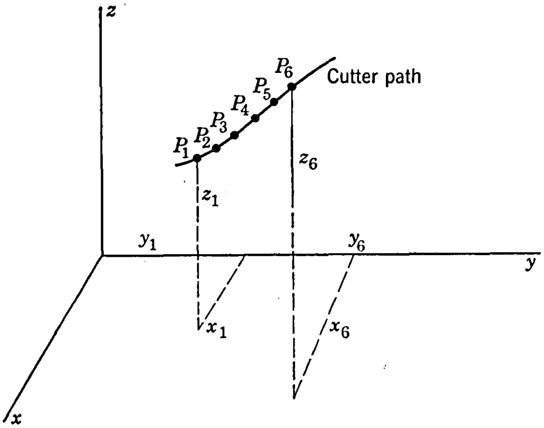


Fig. 16.5 Straight-line approximation to a path

(2) See Fig. 16.5. The cutter path might be approximated by a series of straight-line segments. For these it would only be necessary to specify the increments $x_2 - x_1$, $y_2 - y_1$, $z_2 - z_1$, and a time to go from P_1 to P_2. The information required here is much reduced.

(3) A higher-order approximation might be used, calling for fewer points than in the single positioning system but for more complexity in the control.

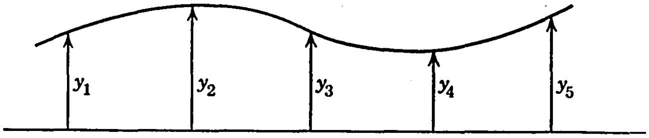


Fig. 16.6 Specifying a path by coordinates on a family of curves

(4) Some family of curves might be adopted and their parameters specified. For example, in a plane, a conic section might be used as an approximate to the desired curve and its five parameters given. As a further example, a curve of a certain order, say fourth, might be specified by five evenly spaced ordinates. See Fig. 16.6.

16.3.2 Mechanizations

If the first method is used, the control of the system can be very much like the position control of Section 16.2. Numbers representing coordinates can be read from a tape in rapid order, compared with existing coordinate data, and used to speed or slow the actuators to reduce the error to zero.

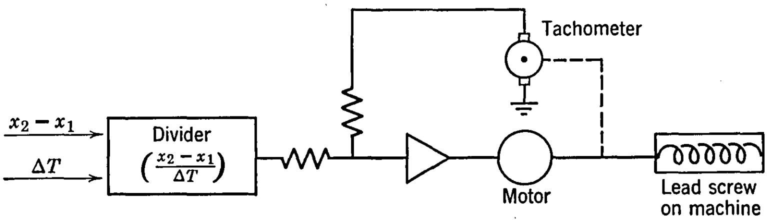


Fig. 16.7 Analog generation of straight-line approximation to a path

If the second method is used, we must ask more of the data-interpreting device since coordinates are not directly read from the tape or card input. Here, the length $x_2 - x_1$ and the time from P_1 to P_2 define a rate of movement required in the X axis. The values $y_2 - y_1$ and $z_2 - z_1$ similarly specify the velocities in the other directions. We

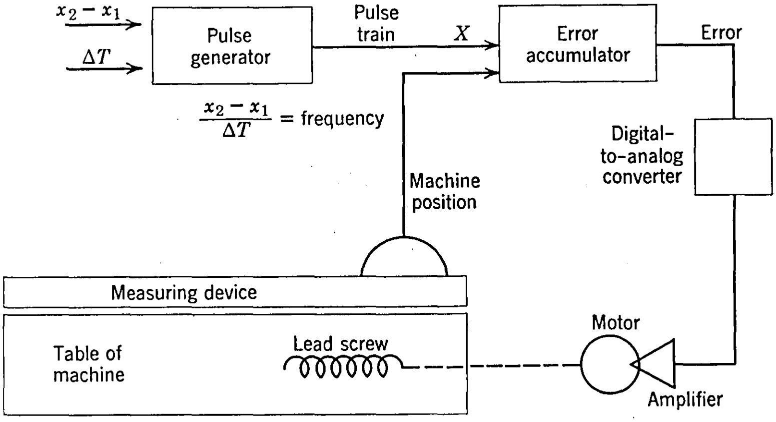


Fig. 16.8 Digital generation of straight-line approximation to a path

might consider controlling the speed as shown in Fig. 16.7. Here we use a divider to obtain $(x_2 - x_1)/\Delta T$ and cause it to control a motor on the machine lead screw. A tachometer on the lead screw shaft measures the rate in x and, by feeding back for comparison with

$(x_2 - x_1)/\Delta T$, causes the speed to assume the required value. The fault with this scheme is that it requires fairly accurate analog control or closely spaced points P_1, P_2, etc., to keep the errors between P_1 and P_2 from becoming excessive.

Our instinct tells us to consider some means of avoiding the analog speed control and to use instead a precise digital method of control along the segment. To do this we cause a pulse train to be emitted at the frequency $(x_2 - x_1)/\Delta T$. See Fig. 16.8 in which each pulse corresponds to an increment of required movement, say 0.001 inch. At the same time we receive from a measuring device on the machine a pulse for each increment of movement (0.001 inch). Plus and minus movements are properly identified and handled. The pulse train from the pulse generator is fed into a counter in the positive sense, and that from the measuring device is fed into the counter in the negative sense, with the result that the counter accumulates the net algebraic difference between movements specified by the pulse train and executed on the machine. The error is fed into a digital-to-analog converter and caused to speed or slow a motor driving a lead screw on the machine to bring the error to zero.

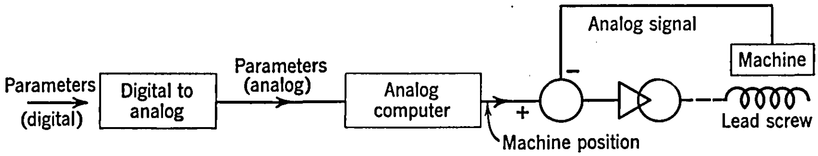


Fig. 16.9 Analog generation of path from parameters

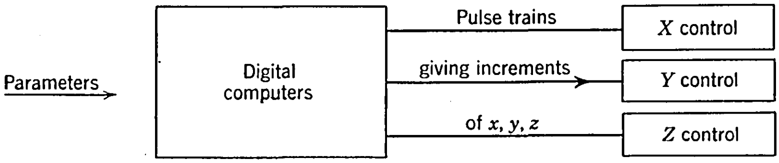


Fig. 16.10 Digital generation of path from parameters

It is easy to appreciate that, barring failure of the equipment, the accuracy of reproducing the specified segment will depend upon (1) accuracy of ratioing $x_2 - x_1$, $y_2 - y_1$, and $z_2 - z_1$; (2) accuracy of measuring devices; and (3) quality of the servomechanism in maintaining small error.

For the third and fourth methods we obviously require some form of computer to generate the path from the data given. Some possible arrangements are shown in Fig. 16.9 that cause the machine position to be generated in analog form from the given parameters. In Fig. 16.10, however, the accuracy of digital language is retained throughout and a digital computer to generate a proper pulse train and a suitable servomechanism controlled by a pulse train, as in Fig. 16.8, are required.

16.3.3 The Problem of Measurement

General. Regardless of the means of control, numbers must be converted to length or angle in a numerically controlled machine. There are several forms of these measurement devices, as represented in Figs. 16.7 and 16.8, and we shall here discuss a few of the more useful ones.

Linear measuring devices. Because of the low cost and precision of commercially available lead screws, they form the basic elements of most precise machines for the generation of measurement of lineal movement. By proper connection of some form of digitizer to the shaft of the screw, very satisfactory devices can be constructed either in a form suitable for moving a machine tool directly through a digital actuator, or in a form to give a digital indication of a movement that is actuated by other means. Although there are screws available approaching an accuracy of 0.0001 inch over a length of about a yard, the use of a conventional mechanical nut on the screw will degrade its performance materially. One jig borer of foreign manufacture makes use of optical reading of a very fine helical line on a cylindrical surface to avoid such loss of accuracy. Even with the elimination of nut errors, however, the screw bearings will contribute some error as wear occurs or as the thickness of lubricating film varies.

Even though a screw is entirely satisfactory for many control and measurement problems, other devices have generally been used when repeatable accuracy of location to 0.0001 inch is required. A bar of metal or glass, having graduations in some form suited to the reading means, is used in most cases. No movement is required except along the length to be measured. Such bars serve as government standards the world over and are commonly used in jig borers.

The problem in the use of the graduated rod as a basis of length measurement is that of obtaining a convenient indication of a movement of the least increment of length. There are two methods of reading the rod currently in use or proposed for use. One makes use of a grating having alternating transparent and opaque lines of the same period as the rulings on the graduated rod, both rod and grating being of glass. Movement of the rod relative to the grating with the grating lines parallel to those of the rod gives a variation in light transmitted through the two elements on to a photocell. The variations can then be counted in an electronic counter. Provision can be made to obtain direction of motion if two or more grating-photocell combinations are used simultaneously. See Fig. 16.11. This scheme has been proposed for use on machine tool controls with a least count of 0.0001 inch.

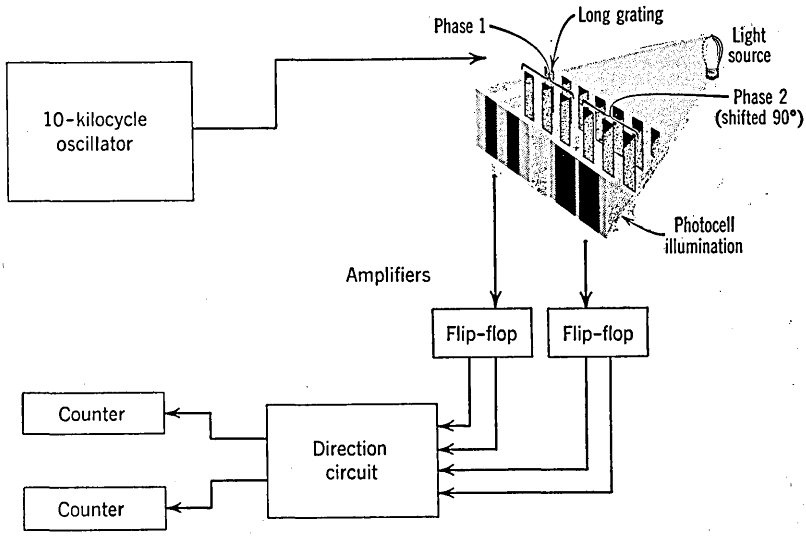


Fig. 16.11 Basic principles of gage based on optical grating

Another form of graduated rod consists of a flat silver coil etched on the face of the glass rod. The wavelength of the coil, of the order of 0.1 inch, is the basic graduation of the scale. Readings on the rod are taken by two shorter coils of similar construction phased from each other by one-quarter wavelength. The rod is excited in the kilocycle range, and the movement of the rod can be obtained from the two small stationary coils. By use of a resolver or voltage dividers, interpolation can be made within one wavelength to obtain an accuracy on the order of 0.0001 inch. (A device of this type is made by the Farrand Optical Company, New York.)

The importance of choice of material for the body of the rod is obvious. The principal considerations are long-term stability and temperature coefficient, although for some applications resistance to rough handling must not be overlooked. The long-term stability of many alloys of aluminum would exclude them from consideration, for example. Temperature coefficients of several materials are listed below for the temperature range around 20° C:

Invar	$0.9 \times 10^{-6}/^\circ C$	Plate glass	$8.9 \times 10^{-6}/^\circ C$
Steel	10	Fused quartz	0.42
Aluminum	25		

Fused quartz would make an excellent material for a measuring rod if it were available at reasonable cost.

Since the international standard unit of length is in terms of the wavelength of the red light emitted by cadmium under standard conditions, some mention of the merits of an interferometer as a basic measuring means is in order. Superficially, it would seem that such a type of gage would be devoid of the troubles of temperature compensation and machining errors and would provide a laboratory standard for the shop where it would do the most good. However, there are fundamental problems in the use of the interferometer. For one thing, the lack of long-term coherence in the light emitted by available sources limits the measurable length, although not enough to eliminate the method for distances required in most jig borer work. A second problem arises from the fact that the wavelength of light in air near standard pressure and temperature varies with a temperature coefficient several times that of fused quartz, for example. Although the method is usable, it is not without difficulty in practical application.

Angular measuring devices. The techniques of measurement of linear displacement have been used with success in angular measurement. First we might get some idea of the accuracy of angular meas-

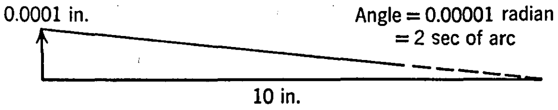


Fig. 16.12 Significance of small angles

urement required for precision work. See Fig. 16.12. At a radius of 10 inches, an arc of 0.0001 inch subtends an angle of about 2 seconds of arc or 10 microradians. It is difficult to measure to this accuracy, although some of the devices based on principles described in the paragraph above are claimed to have repeatability to 1 second of arc.

16.3.4 Actuators

In Chapter 4 the ideas of feedback control were introduced. It is clear from Fig. 16.8 that feedback is a centrally important tool in the control of machines. We are interested in the quality of such control because the cutting time on a great many parts is determined ultimately by the speed of response of the servomechanisms. This can be appreciated by examination of the requirements in following a signal of the type described in the second paragraph in Section 16.3.2, generated for a parabolic curve. It can be seen in Fig. 16.13*a*, that such a curve calls for steadily increasing velocity in y if we assume the speed along the path is constant. The required Y velocity actually increases in steps as shown in Fig. 16.13*b*. If now the position error in the servomechanism resulting from a step in Y velocity is as shown in Fig. 16.13*c*, it can be seen that the long "tail" on the transient error persists far beyond the interval of each step in Fig. 16.13*b*. Therefore, the several transients resulting from the succession of steps in Fig.

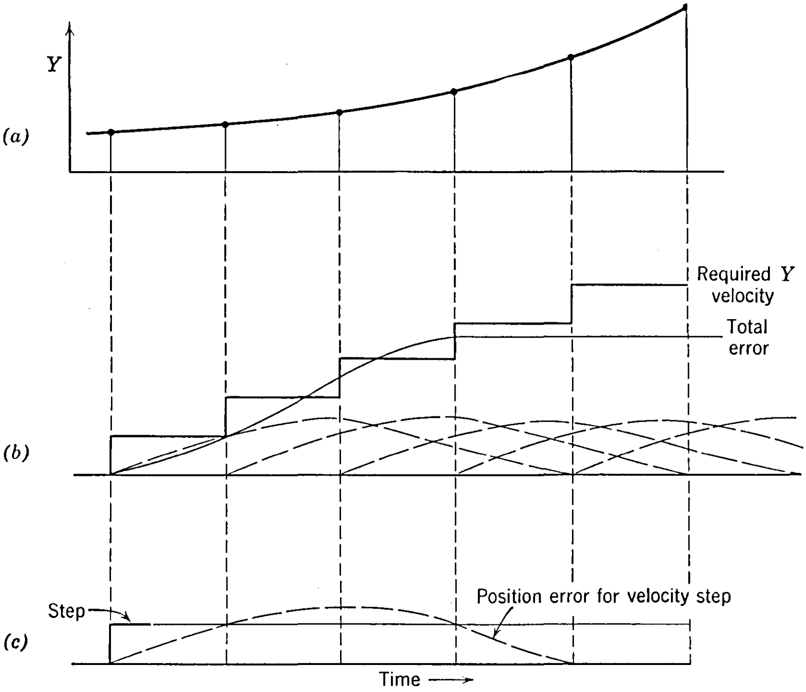


Fig. 16.13 Analysis of servomechanism errors in following a parabola

16.13*b* will accumulate for several intervals, giving the total error as shown.

Both the magnitude and duration of the transient error resulting from a step of velocity are important in determining the accuracy of following. As is logical to expect, the extent to which we can reduce the transient error depends upon the speed of response of the actuator that drives the machine itself: if it is fast, the transient error can be made small.

Some idea of the possibilities here can be obtained from the fact that an hydraulic control of speed of movement of a machine table weighing about 2 tons can be made to respond (open loop) to sinusoidal fluctuation at approximately uniform amplitude with variation in frequency up to 10 cycles per second. Electric drives, operating through lead screws, have been used successfully, although they tend to be slower.

It might be noted here that much of the servomechanism error could be compensated out by proper programming, provided the cutting were always done at the speed chosen for programming.

16.3.5 The MIT Numerically Controlled Milling Machine

The first outstanding example of a numerically controlled machine tool is the mill at the Servomechanism Laboratory of MIT. This machine makes use of the principle of control described in the second paragraph of Section 16.3.2, to provide contour control in three axes. The basic machine is a 28x60-inch vertical Cincinnati Hydrotel.

Data describing the components of a segment of path, $\Delta x = x_2 - x_1$, $\Delta y = y_2 - y_1$, $\Delta z = z_2 - z_1$, and the segment time, ΔT (refer to the second paragraph of Section 16.3.2), are punched on Flexowriter tape, using a redundant error-correcting code to reduce possibility of error. The data in binary form are fed into relay storage a block at a time, a "block" being the data describing a segment. Since the machine must proceed without pause from one segment to the next, it is necessary to store two blocks of data, that being used by the machine and that for the next segment.

The data stored in the relays is in binary form. For example, a relay locked in equals one; a deenergized relay equals zero and this data is easily used to control the pulse trains needed to define the motions of X, Y, and Z servomechanisms. The method is shown in Fig. 16.14. A clock oscillator drives a string of flip-flops in series. The noncarry pulses (those not used to drive the next flip-flop) are fed to a row of gates. At each gate is a pulse train with one-half the frequency of that on the gate to its left. It turns out that no two pulses at these gates are coincident in time. By starting with all flip-flops synchronized, if the clock feeds in 32 pulses, the last flip-flop will deliver out $2^0 = 1$, the next $2^1 = 2$, the next $2^2 = 4$, and so on to the first, $2^4 = 16$. If we wish to deliver a pulse train of 10 pulses, therefore, or in binary notation 01010 pulses, we open the second and fourth gates. For the binary number of steps required in x, each gate is opened when its corresponding digit is one, and closed when it is zero.

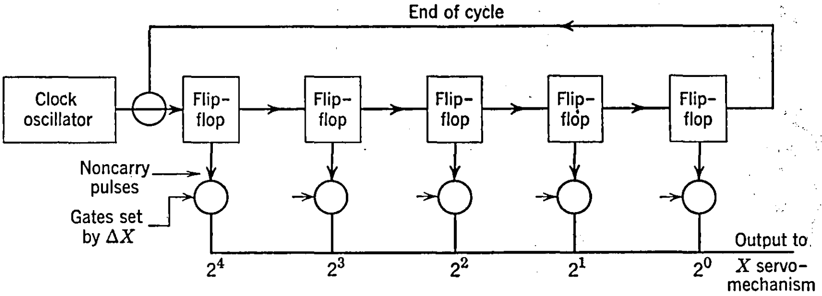


Fig. 16.14 Generation of pulse trains in numerically controlled milling machine at the Servomechanisms Laboratory, MIT

To terminate the segment, the carry pulse from the last flip-flop stops the feed from the clock and throws control to the next block of data. The same technique is used simultaneously to control the other two axes, employing a common set of flip-flops and clock. Only the gates are duplicated for the several axes.

To permit use of conventional power drive servomechanisms, the pulse trains derived as above are used to drive decoding servomechanisms by the general technique of the second paragraph in Section 16.3.2, and the power servomechanisms on the machine tool are then linked to the decoding servomechanisms by means of a selsyn circuit. See Fig. 16.15. Each decoding servomechanism uses a small motor driven by vacuum tubes and coupled to a three-brush commutator to give the direction and amount of motion. One degree of selsyn rotation is geared to give 0.0005 inch of machine travel; 0.0005 inch is one unit of movement in each axis.

The power drives are conventional hydraulic pump and motor servomechanisms, coupled to the machine through gears and lead screws. The selsyns are driven through rack and gears. Their drives have been a principal source of error in the system.

Although this machine is not designed for production use, it has provided a clear demonstration of the capabilities of numerical control and has served to remove many doubts about the practicability of such control.

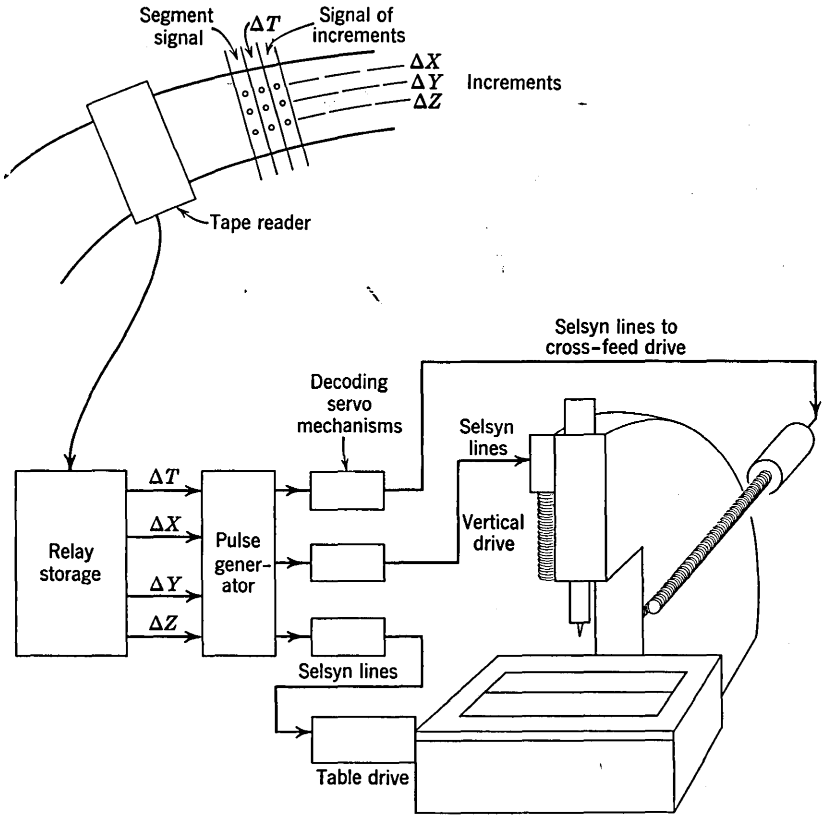


Fig. 16.15 Basic arrangement of numerically controlled milling machine, Servomechanisms Laboratory, MIT

16.4 PROGRAMMING

The problem of programming the tape for a machine tool includes (1) study to determine tools and fixtures required for each cut to be made, and selection of the proper order of operations; (2) calculation of the exact cutter path for each cut; and (3) coding of cutter path into the system for which the control is designed.

It is clear at once that the difficulties of programming and machining are much affected by the design of the part. There are many operations, without doubt, that are better carried out on a conventional machine at present. As numerically controlled machines with several degrees of freedom (as many as five per cutter) are introduced, however, it should become possible to do more and more machining with fewer setups. Any great advance toward this end will, of course, depend upon the extent to which designers work to exploit the full capabilities of numerical control.

In this discussion we are concerned with the problems peculiar to numerical control; we shall have to pass over tooling problems and deal only with calculation and coding of the cutter path.

16.4.1 Calculation of Cutter Path

The computations to determine cutter path will vary greatly, depending on the form of data provided. In some cases, the surface to be cut may be defined by an equation, generated from given parameters. In others, it may be necessary to fair a curve through a number of points to arrive at the surface to be cut. The possibilities are infinite, but we can assume that by one means or another the desired surface is reduced to mathematical form. Then the cutter path must be derived for the defined surface. Where the surface to be cut is a right cylinder whose elements are parallel to the axis of the cutter, the determination of cutter path is simple, though not trivial. See Fig. 16.16. Assuming the use of a straight mill, the center line of the cutter path is then the line whose points lie at a distance *r* from the surface, along a normal to the surface. This simple relation applies to many parts cut from flat stock.

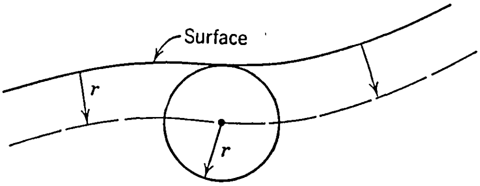


Fig. 16.16 Derivation of the cutter center path

Most surfaces are not so simple and call for a large number of passes of the cutter to provide a reasonable approximation to the surface. As an example, if we are cutting a surface that is approximately plane near some point, using a ball-end mill (with spherical end), the successive passes will leave ridges whose heights depend on the radius of the mill, *r*, and the spacing of cuts, *d*. See Fig. 16.17.

The height, h, of each ridge is approximately $d^2/8r$, or if a height h is permitted, the spacing must be

$$d = 2\sqrt{2rh}$$

For the case where $h = 0.0005$ inch and $r = 0.5$ inch, $d = 0.0447$ inch. Thus, to cut a swath 1 inch wide would take 23 passes, and to finish a 10x10-inch surface at 15 inches per minute would require about 2.5 hours. Obviously, the cutter should be changed to the largest radius that can be used to reduce this time.

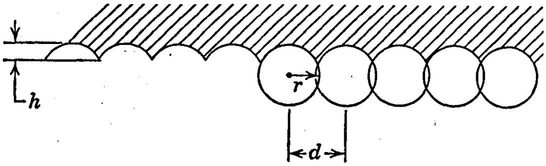


Fig. 16.17 Granularity of surface produced by ball-end mill

The problem of determining the cutter path from the surface required for the general compound curve involves more complex geometry, but it is basically similar to that of Fig. 16.16 for a ball-end mill.

16.4.2 Coding for the Machine

Up to this point, we have not been concerned about the language of the machine, since the cutter path depends only on the construction of the machine and the tools used. Here, however, we must adapt to the requirements of the machine. Let us consider for comparison, three machine languages, calling for, respectively, (1) position; (2) slope (rates in three coordinates); and (3) rates of change of slope.

Obviously the first requires the greatest number of points to be computed, since we are to give position each time one of the coordinates covers a distance equal to the least increment. For example, if the increment of measurement is 0.0005 inch, and the range of movement in each axis were 20 inches, three coordinates would have to be specified by 16 bits each, every 0.0005 inch of path. One inch of path would require about 96,000 bits of data. To place this on six-channel magnetic tape would require about 80 to 160 inches of tape. The number of computations would be large, and therefore tape preparation would require a great deal of labor and expense.

For the "slope" method, as used in the MIT machine, and as described in the second paragraph of Section 16.3.2, the computations are reduced since the number of points to be calculated on the cutter path are small. For example, to approximate a path having a radius of curvature, R, to an error not over Δ requires a length of segment $4\sqrt{R\Delta}$. For a radius of curvature of as little as 0.5 inch and $\Delta = 0.0005$ inch, the length of segment is 0.02 inch. Even for $R = 0.5$, calculations of the cutter path coordinate must be made only about one-fortieth as often as for the method of the foregoing paragraph. Of course the "slope" method calls for a more complex computation at each point, reducing the advantage somewhat, but the method is well justified where fine control is needed.

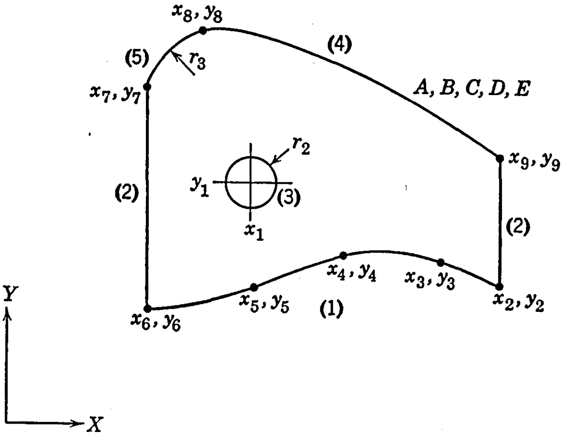
If the "slope" method is good, it is logical to ask why not go to a language involving even higher derivatives, say the second? The deterrents to use of the schemes involving rates of change of slope are in the greater complexity of the controls in the machine to use the data, the more involved computation to arrive at the proper arc length, and the frequency of occurrence of discontinuities requiring termination of an arc. Although some applications may justify the use of a more complicated approximation, the first order is a satisfactory and economical system for a great many applications.

16.4.3 Economic and Other Benefits of Computer Programming

We have here contemplated the execution of thousands of computations for the programming of a simple part. If each computation is to be done manually, there is no question that the preparation of tape will become such a factor that numerical control will be usable only for moderately large runs, or for work that can be handled in no other way. However, a modern electronic computer can reduce the cost of computation of around 100 segments to a few dollars. Even with the time necessary for programming the computer, the cost per part can be low enough to justify economically numerical control on lots of one part each.

In addition to the direct savings in the fabrication cost by computer-programmed numerical control, there are large savings in the cost and elapsed time in going from the designer's sketch of a part to its final form in steel or aluminum. The simplicity of the method is illustrated in the manufacture of a flat metal part shown in Fig. 16.18. The engineer merely prepares an order list as shown. This list is converted to cards or tape, as required for input to the computer. Programs already at hand, corresponding to the several order numbers, cause the computer to prepare a machine instruction tape based on the parameters supplied. The tape is taken to the numerically controlled machine and the part is made, with cuts normally requiring extensive layout work.

By these techniques the time from engineers' conception to finished part could be reduced to a few hours, rather than several days, and the preparation of extensive drawings might not even be necessary.



Engineering Order List		Meaning
Order No.	*Coordinates*	
3	r_2, x_1, y_1	Drill hole, radius r_2, at coordinates $x_1 y_1$.
1	$x_2, y_2; x_3, y_3; x_4, y_4; x_5, y_5; x_6, y_6$	Starting at x_2, y_2, mill a faired path through points given, ending at x_6, y_6.
2	x_7, y_7	Cut a sharp corner at x_6, y_6 and mill on a straight line to x_7, y_7.
5	x_8, y_8, r_3	Mill a circle of radius r_3 to x_8, y_8.
4	x_9, y_9, A, B, C, D, E	Mill a conic section with parameters A, B, C, D, E *to* x_9, y_9.
2	x_2, y_2	Cut a sharp corner at x_9, y_9 and mill on a straight line to x_2, y_2.

Fig. 16.18 Typical order list for a two-dimensional numerically controlled milling machine

16.5 REFERENCES

1. J. H. Brown, Measure Motion to 0.0001 Inch Without Friction or Wear, *Control Eng.*, 50–52, April 1955.

2. Electronics Research, *Aircraft Production,* 476–477, Dec. 1954.

3. A Numerically Controlled Milling Machine—Contract AF33(038)–24007, Servomechanism Laboratory, MIT, Cambridge, Mass., May 31, 1953.

4. E. C. Beandet, Non-Circular Gears Generated Automatically, *The Iron Age,* 122–123, Feb. 19, 1953.

5. Co-ordinate Control of Machine Tools, *Engineering,* 766, June 11, 1954.

6. Harry W. Mergler, A Digital-Analog Machine Tool Control System, *Proceedings of the Western Computer Conference, February 11–12, 1954,* AIEE, April 1954.

17. Manufacturing Automation

R. W. BOLZ

Editor, *Automation*
Cleveland, Ohio

17.1 INTRODUCTION

There is a significant saying that security—job, company, and national—is built on better *methods;* that nothing can so completely or surely destroy an established business or its profits as *new and better methods or equipment* in the hands of an enlightened competitor. Searching to the heart of it all, this is the primary reason why all industry today is vitally interested in automation. What is it? What does it offer in the way of security? Is it really something new and radical?

There has been a great deal of written and published material concerning automation in recent years. Little of this highly colored information actually paints a true picture. Most of the authoritative conclusions presented resemble to a striking degree the age-old fable of the blind men and the elephant. Each blind man describes vividly what he imagines it to be. The result has been interesting for the layman but of dubious value to industry.

Automation is the modern term which denotes manufacture, processing, or performing services *as automatically as economics permits or demands.* Although new in name, manufacturing automation basically comprises the application of principles which have developed steadily over a great many years. Wherever conventional manual methods of manufacture, processing, and distribution could not be carried out within acceptable quantity, quality, effort, and cost levels, mechanization developed; today we have the next step, automation.

The continuing effort to supply the needs of the American people with products, commodities, and services at acceptable and competitive prices has made automation a necessity. In a dynamically expanding market, no other development holds so much opportunity for a gradually rising standard of living. It is not just a matter of using or not using automation but rather one of economic necessity to make possible the broadest availability of utilitarian and luxury items. This fact is well illustrated by a simple example: One company alone assembles 500,000 automotive devices each day with an average of 72 individual parts in each unit. To do this without benefit of automation would present an insurmountable task from both a cost and production standpoint.

In another plant where some 230 employees are able to produce 1,200,000,000 lamp bulbs per year on continuous automatic equipment, 1927 methods and equipment would require 75,000 men to equal this output. At today's wage levels, few people would be in a position to buy under these conditions, even if men were available!

At this stage of development, it is relatively impossible to present a concise basic approach to automation engineering. However, to show what is being done and, perhaps, to sow a few seeds for development, we can cover some fundamental points while pictorially making an extended plant tour of many manufacturing industries.

Whenever demand is high and the product relatively or partly standardized, there is an opportunity for automation, regardless of the type of manufacture. Let us take a look at a modern production line that operates on a highly automatic basis.

Figure 17.1 shows a complete one-man automatic line for producing, from steel strip, six standard components used in carloading. A coiled steel strip 6 inches wide by 1/8 inch thick passes automatically through a 100-ton blanking and forming press; the stamped pieces go on to a de-oiler, drying conveyor, accumulator, and Wheelabrator blaster to a shipping box. The blaster operates with batches of 500 and 1000 pieces, depending on the part; the parts produced are counted by the press controls and accumulator pan timer. The blaster is push-button-controlled by the operator; on completion, the charge drops into a box. One boxload of 500 or 1000 identical parts comes off the line every 6 minutes. Change-over between different parts is simple.

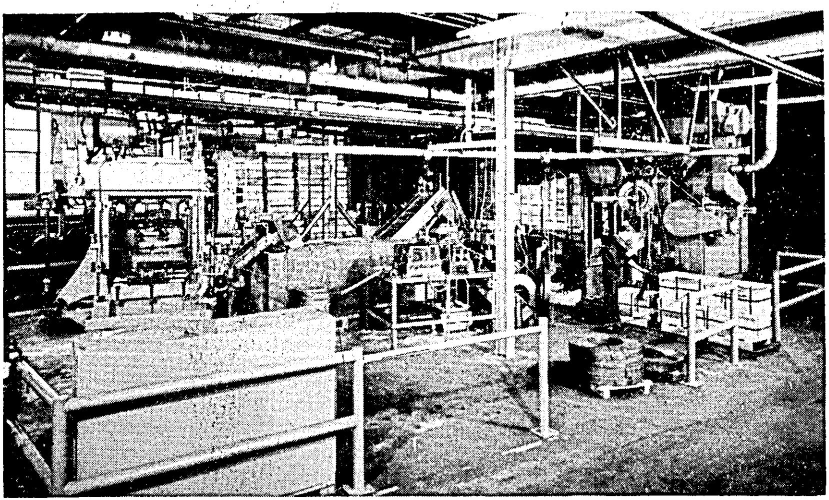


Fig. 17.1 Complete one-man line for producing six standard components at Signode Steel Strapping Company. (Courtesy American Wheelabrator Co.)

17.2 AUTOMATION AS A BASIC PHILOSOPHY

Basically, automation can be termed a philosophy or method of manufacturing. It may require transfer machines or automatic materials handling. Also, it may demand complex electrical control, use of instrumentation, and application of feedback techniques. Accomplishment of really economic plant automation requires a careful combination of equipment, machines, and controls, depending on the conditions of production.

Economic plant automation involves using, in proper degree and combination, such items as automatic machines, automatic mechanisms, automatic controls, automatic computers, automatic data-processing systems, and automatic devices for handling, conveying, processing, assembling, inspecting, and packaging. With this wide range of equipment, automation systems today are being developed to a high degree, in more complex operations and in larger plants than in the past.

An example which demonstrates a range of equipment and instrumentation is a typical transfer machine for automotive castings shown in Fig. 17.2. Transfer bar and chain carry the pallets of parts. The operator unloads the cleaned and finished part from the pallet fixture, loads a raw casting, and pushes a button to start the loaded fixture into the line. Although many transfer machines are single-purposed, the trend is toward more flexibility in their design. This type is composed of basic standard units which, when necessary, can be repositioned and retooled for product changes. Spaces are left for the addition of other units along the transfer stations.

The high degree of perfection which has been reached in all these areas of automatic operation, along with the tremendously expanding market of recent years, has made automation practical in many industries. Not only do the speed and accuracy of operation of much modern equipment demand automation, but the monotony of repetitive operations, heavy labor, and dangerous conditions in many instances creates problems that can be solved adequately in no other way. The sheer bulk and time-consuming handling of necessary paperwork for modern enterprises have led to automatic data-handling and computing systems to provide immediate information vital to economic operation. Flexible automatic control in the form of a tape into which complete coded information is recorded can be used to spearhead paperwork operations not only in one manufacturing office but throughout a series of widely separated manufacturing plants.

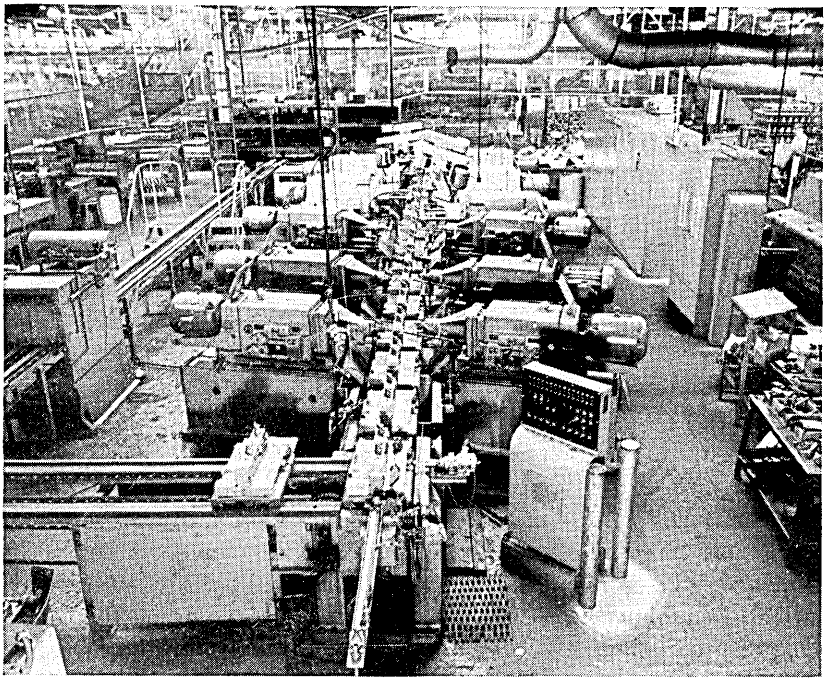


Fig. 17.2 Typical transfer machine for automotive castings (Courtesy Modern Tool Works Ltd.)

17.3 ECONOMICS OF AUTOMATION

As products grow more diversified and complex and market demands increase to fulfill the modern needs for better living, not only does the demand for machinery, electrical power, and control grow, but the basic production task of assembling the finished products becomes staggering. With a single item in one plant, this problem resolves itself into how to assemble 3,500,000 separate parts into some 25,000 similar units each day. The practical result today is automation. It is here and already at work for us.

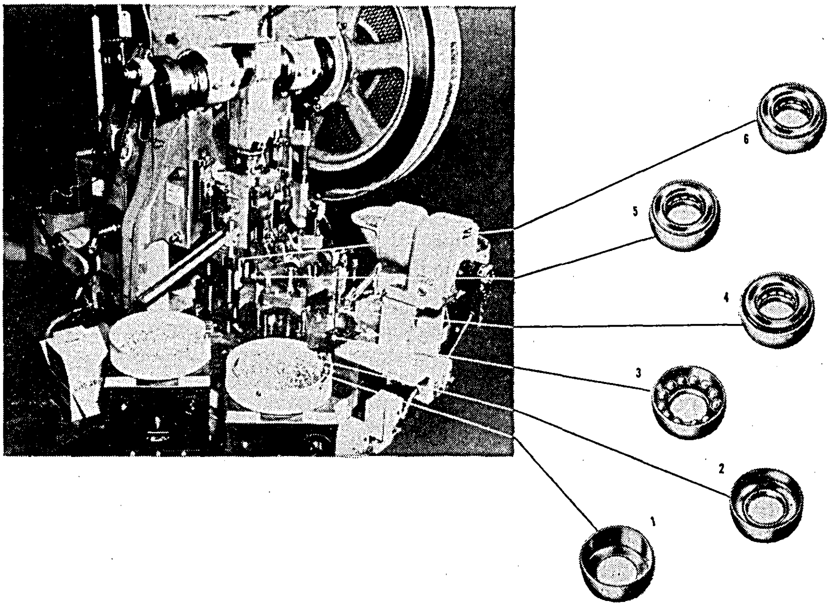


Fig. 17.3 Standard press setup to assemble low-cost bearings (Courtesy New Departure Division, General Motors Corp.)

From a practical standpoint, automation need not comprise gigantic multimillion-dollar projects as with some transfer machines. Applied in reasonable degree, even on an in-plant designed basis, partial automation can provide real dollar-and-cents savings. Standard automatic machines can be equipped so as to offer flexible automatic operations.

Figure 17.3 shows a special six-station press arranged to assemble low-cost ball bearings automatically. The shell, two rings, and balls are hopper-fed into fixtures for proper positioning. The operation sequence is (1) align finished shell and inspect; (2) insert first race ring; (3) insert balls and lubricate; (4) insert second race ring and inspect; (5) form shell around rings and balls; and (6) finish form for size and running clearance with upper ring. At the completion of the six operations shown, the bearings drop into the package.

Increased productivity, better quality, and lower labor costs greatly improve the wage and profit picture in plants with automation. Lower unit costs made possible with automation not only more than pay for the equipment but result in better satisfied customers and workers.

Many examples could be cited. One case indicates the trend. On one automatic press line, 1000 horsepower in electric motors and $250,000 worth of electrical drive and control equipment are used. The line saved $1200 per working day or the cost of the equipment yearly. Contrary to some ideas, however, roughly the same number of employees are required but productivity is much higher without the need for incentives.

There are many examples where automatic machines have produced phenomenal cost savings—reduced production costs on a fastener, for instance, from $12 to $2 per thousand; on another assembled component, reduced costs from $4 to 40 cents per thousand; on radio and television equipment, reduced costs on some operations of approximately 50 per cent; etc.

An important factor to recognize here is that small plants and job-lot operations can also reap the benefits of automation. Systems for automatic poultry packing have been installed for less than $15,000. Other small plant or small-lot manufacturing which have been automated are automobile radiators, special gears, auto headlight components, smelting and refining, tire mounting, appliance assembly, rubber molding, aircraft wing riveting, etc.

17.4 AREAS OF APPLICATION

Automation of plant operations may range from single lines of automatic equipment, such as presses which are fed and cycled continuously with automatic handling, to complete manufacturing facilities where components are fed, sorted, oriented, assembled, and tested automatically. Where job-lot operations are involved, flexible automatic control of standard machines is sought in order to achieve the economic results desired. Industries in which automatic operation is now used or is being developed include ordnance, primary metals, fabricated products, machinery, telephone and telegraphy, transportation equipment, food, tobacco, textiles, lumber and wood, furniture, paper, printing, chemicals, petroleum, rubber, leather, stone, glass and clay, mining, merchandising, and distribution.

An electronic servo-positioning system for the Wiedemann turret punch press is illustrated in Fig. 17.4. This view shows mechanical components with a schematic of the control system. Such semiautomatic control offers floor-to-floor savings of 10 to 20 per cent. To keep the servomechanism as simple as possible, full advantage was taken of pins used to lock the turrets in place. These made precise angular location as well as dead-beat positioning possible. The servo amplifier oscillates between the outer boundary of the dead-band region and maximum sensitivity, which corresponds to an error of less than 0.1 degree. A relay oscillator-limiter initiates and maintains this cyclic operation until registry is obtained. The oscillator section causes a relay to make and break on a 1-second cycle, charging and discharging a capacitor in the limiter section. During the capacitor-charging cycle, limiter contacts momentarily close. The circuit is so arranged that sensitivity is shifted to maximum when the limiter contacts are closed. Therefore, during the charging cycle, the drive motor is given a short burst of power. Essentially, then, the oscillator-limiter section is an automatic jogging control, somewhat analogous to a camera shutter. The error-sensing mechanism for the system consists of a pair of self-synchronous units, or synchros. The command unit is mechanically connected to a station selector and remotely coupled with the response unit, which in turn is mechanically connected to the lower turret.

In such areas as chemical and petroleum processing, flow of component materials is continuous from step to step under instrument control to final desired end-products and by-products. Automatic instruments maintain preset temperatures, pressures, flow, etc., and record and control the processing continuously. Materials may be weighed, fed, conveyed, and dispensed by automatic devices. Completed products may be fed directly to automatic bagging, packaging, canning, bottling, or other equipment and thence to storage.

In the food-processing field, automatic machines again are linked together to produce an automated line. Like many modern chemical and metallurgical processes which defy manual control owing to speeds, pressures, loads, and accuracy of measurement, so also does food processing because of sanitary and critical processing conditions. Automatic control devices monitor the conditions, flow, and movement of the product continuously as preset. Processing may be continuous and untouched by hand through to the final packaging and cartoning stages.

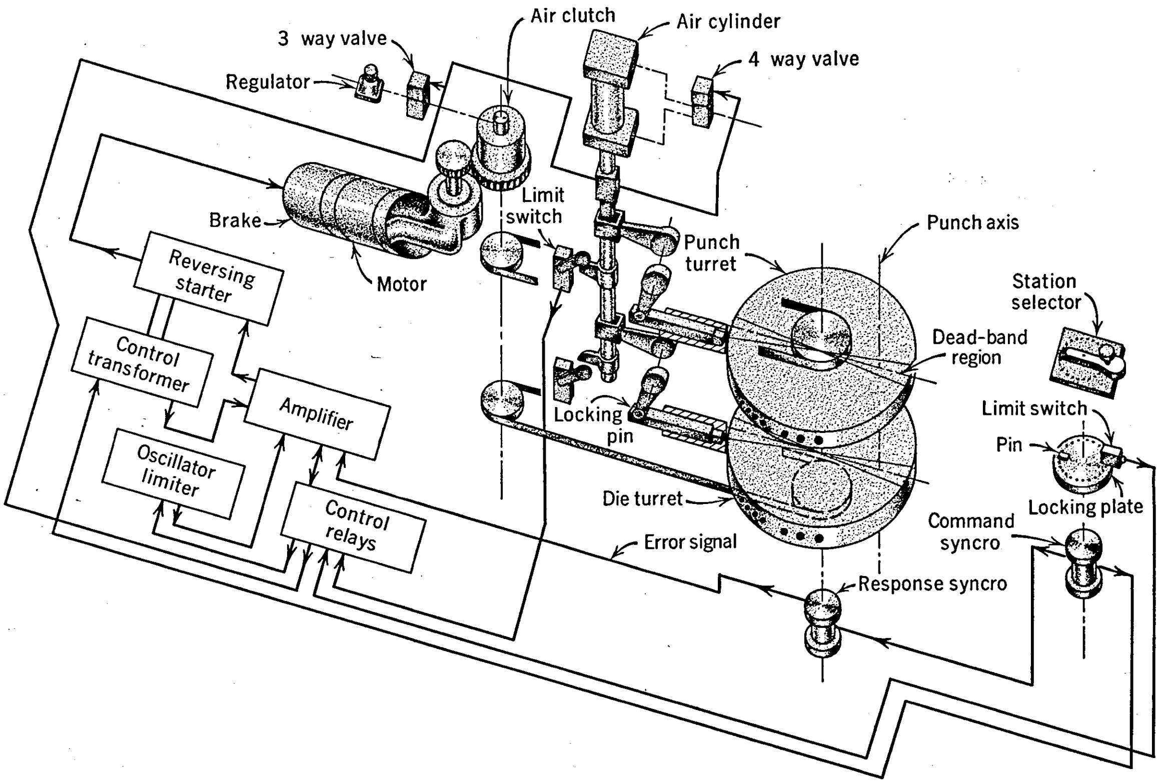


Fig. 17.4 Turret punch press servo-positioning system (Courtesy Wiedemann Machine Co.)

Here, owing to necessity, really economic processing has been achieved. Today, necessity is leading to similar achievements in other areas.

In metalworking today, automation ranges from simple units all the way to complete plants. Through automation the big problem of handling between operations is eliminated almost entirely with accompanying cost savings. Machining, stamping, heat treating, washing, degreasing, plating, and finishing operations as well as assembly of complete units are being done.

With the advent of digital control by means of tape, numerous machines are being developed. Coded tape used to reproduce information in complete form is being adapted for controlling various kinds of equipment. Today typesetting machines can be so controlled. Machine tools are in use that follow instructions punched into a tape or recorded by similar coding marks on a photographic film or magnetic tape. Among the types of tape being employed are common-language paper tape, 35-millimeter movie film, magnetic recording tapes, stainless-steel magnetic tape, and punched plastic tape in varying widths.

Figure 17.5 shows an experimental punched 35-millimeter-film-controlled milling machine developed by Industrial Controls Corporation. The photograph shows a photoelectric tape reading device, thyratrons, and three-step motors to produce relative table movements. The motors provide 108 steps per revolution but motors with more or fewer steps can be built.

Figure 17.6 pictures a Fellows gear shaper with tape and servo control of cutting actions for producing noncircular gears used in computing devices. Tape is printed 35-millimeter photographic film. Quantities of special gears from one piece and up can be produced with real savings in tooling costs.

Although relatively few machines today are being produced with automatic tape controls, most of the new models are being designed with such use closely in mind. One new aircraft miller now completed uses a combination of common-language paper tape and Flexowriter unit with magnetic tape for effecting control. However, most of the variety of machines in construction employ more conventional tracer techniques adaptable to digital control. It appears today that the most accurate tracing devices are a combination of electric tracing and hydraulic follow-up servo control.

What can be achieved with tape control, however, is well illustrated by a type of machine now used successfully in aircraft plants. This machine reads a tape to drill specific sequences and sizes of holes, select a suitable rivet from automatic feeders, place the rivets, buck them properly, and drive the rivet under precisely controlled conditions. As requirements change or the design is modified, a new tape and fixtures adapt the equipment as necessary.

Figure 17.7 shows a unit for automatic riveting of aircraft wing panels; this unit effects a 90 per cent time saving. Five-channel perforated 35-millimeter movie film is used for input data. Three films are used for control, the longitudinal one shown cuing the two others which control transverse functions. Tape holes are punched during first setup operation.

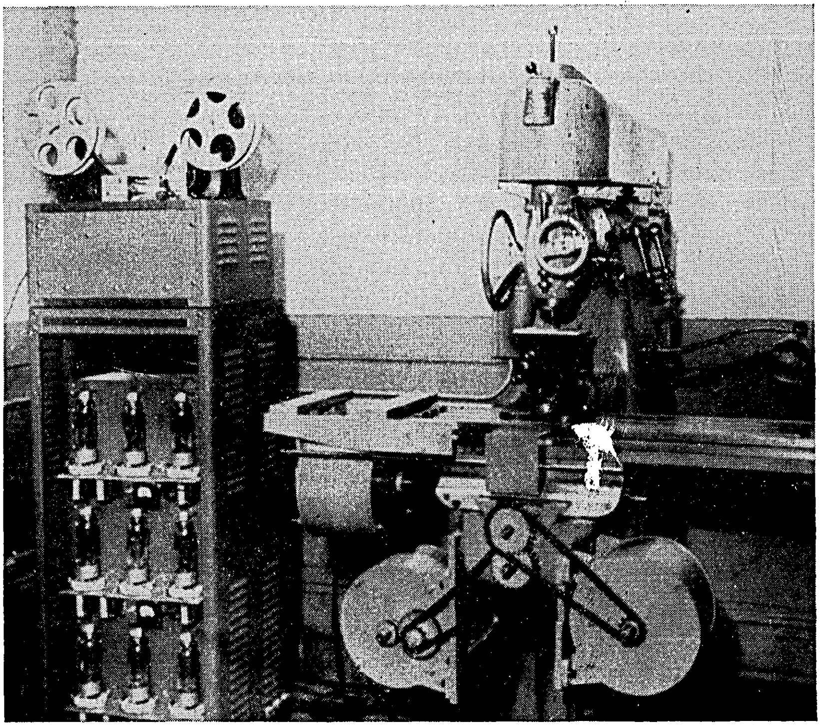


Fig. 17.5 **Punched-film-controlled milling machine powered by special step motors** (Courtesy Industrial Controls Corp.)

Tape control is still in the experimental stage as far as most machine manufacturers are concerned, and to date no definite trends have developed. Some data processing is swinging over to stainless steel tape at present, and developments are moving apace. For shop operation plastic tape is considered too fragile.

One area in manufacture where automation today is being developed to meet a critical need is in final handling and distribution. Here the delays and problems of handling products are being overcome through the use of automatic machines designed to handle, store, unload, and

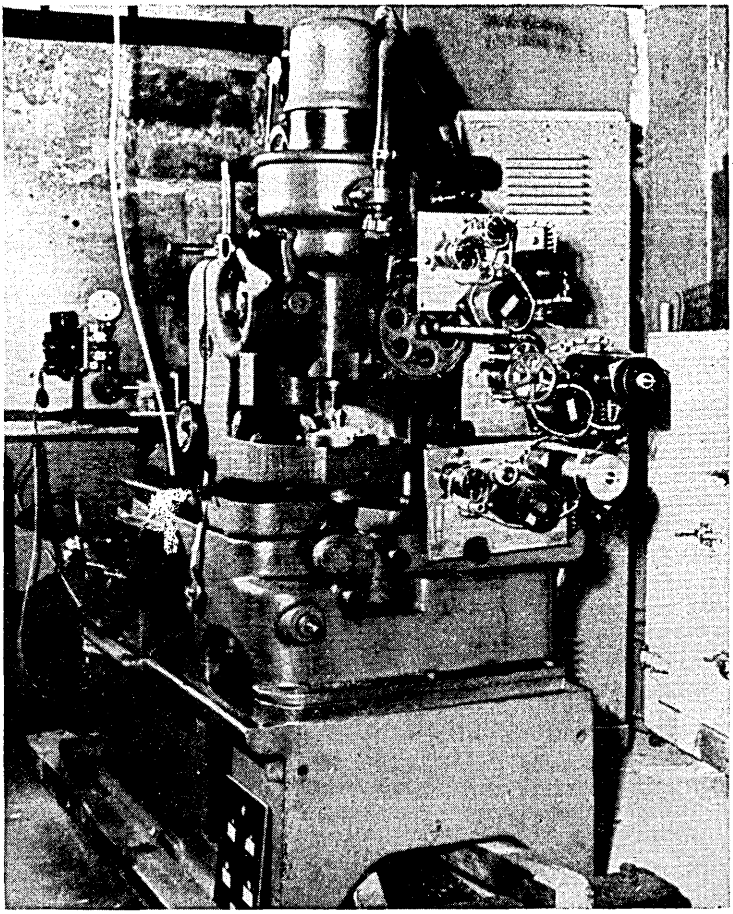


Fig. 17.6 Film-controlled Fellows gear shaper for noncircular gears (Courtesy Cunningham Industries)

load for shipment. Laborious handling of goods in-plant and in warehousing is carried out automatically with systems of machines and handling devices. Conveyors of belt, chain, plate, roller, screw, and overhead rail types automatically deposit, retrieve, and carry products under control of limit switches, photoelectric cells, and numerous other devices to dispatch them as needed. Central push-button control panels permit selection, coordination, and dispatching as desired. Among the most successful are those utilizing basic telephone control devices for serviceability.

Figure 17.8 is the relay control panel for a warehouse handling and sorting system. Automatic packaging is used at the end of the line.

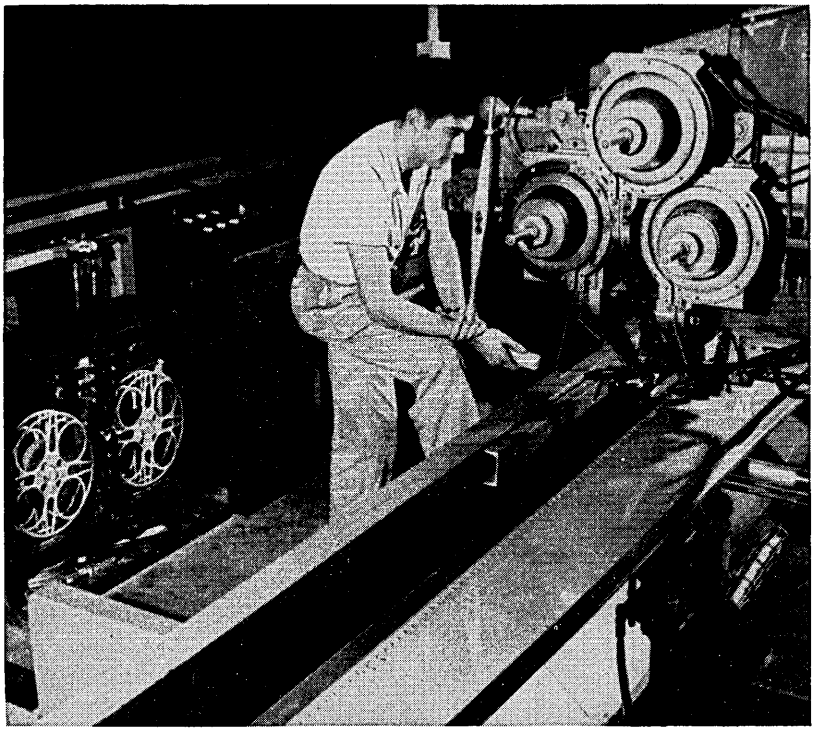


Fig. 17.7 General Riveters Corporation automatic aircraft wing panel riveter with punched-film control (Courtesy Convair)

Output was increased over 200 per cent and warehouse capacity over 100 per cent.

Figure 17.9 gives a view of the inner workings of the control center of this equipment. Standard communications control components are step switch banks, relay levels, indicators, and counters. The unit permits selection of carton destination, shows carton location and full racks by signals, and the electric counter records inventory with both addition and subtraction functions. In the operation the unit performs three basic functions. First, there is the selection of the container destination and addition to the inventory. One bank of step switches

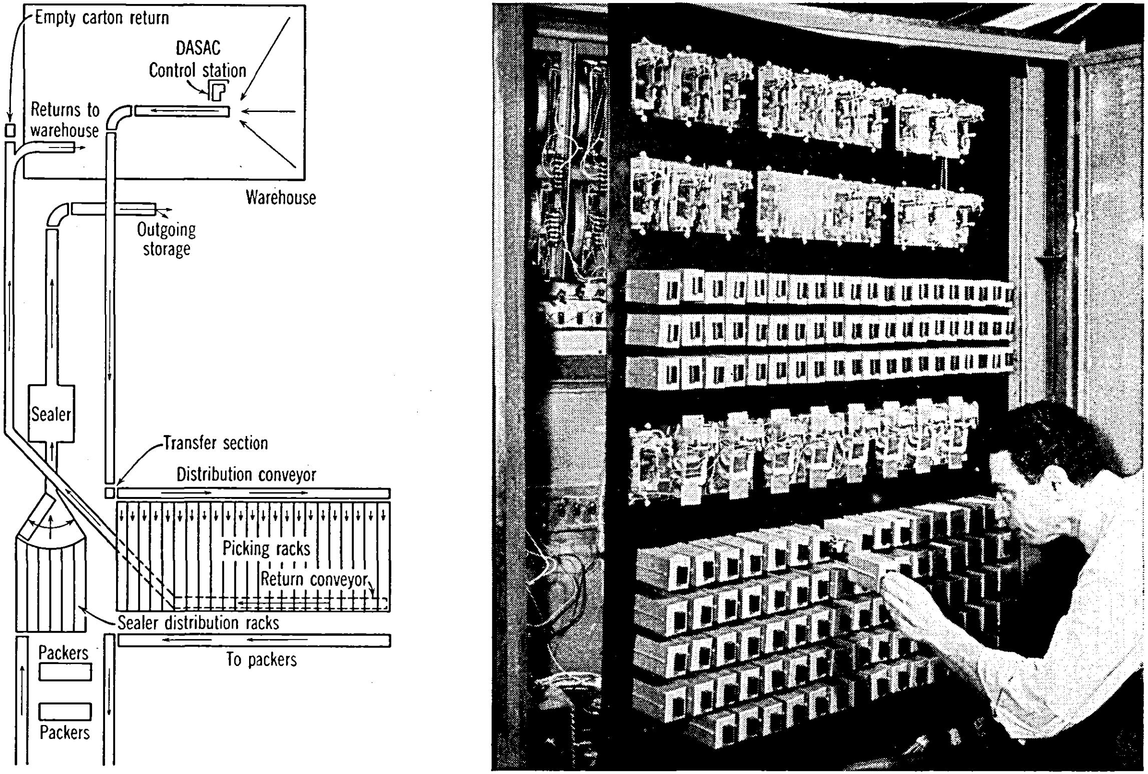


Fig. 17.8 Relay control panel for a warehouse handling and sorting system (Courtesy Dasol Corp.)

working with fast-acting relays performs this function. It is important to make the stepping of the relays and switches independent of the length of time it takes for the operator to push the select button. Special add-and-subtract counters are used to keep up-to-the-minute inventory. These are tied into both the previous select relay circuits and the tilt sections associated with the sorting racks.

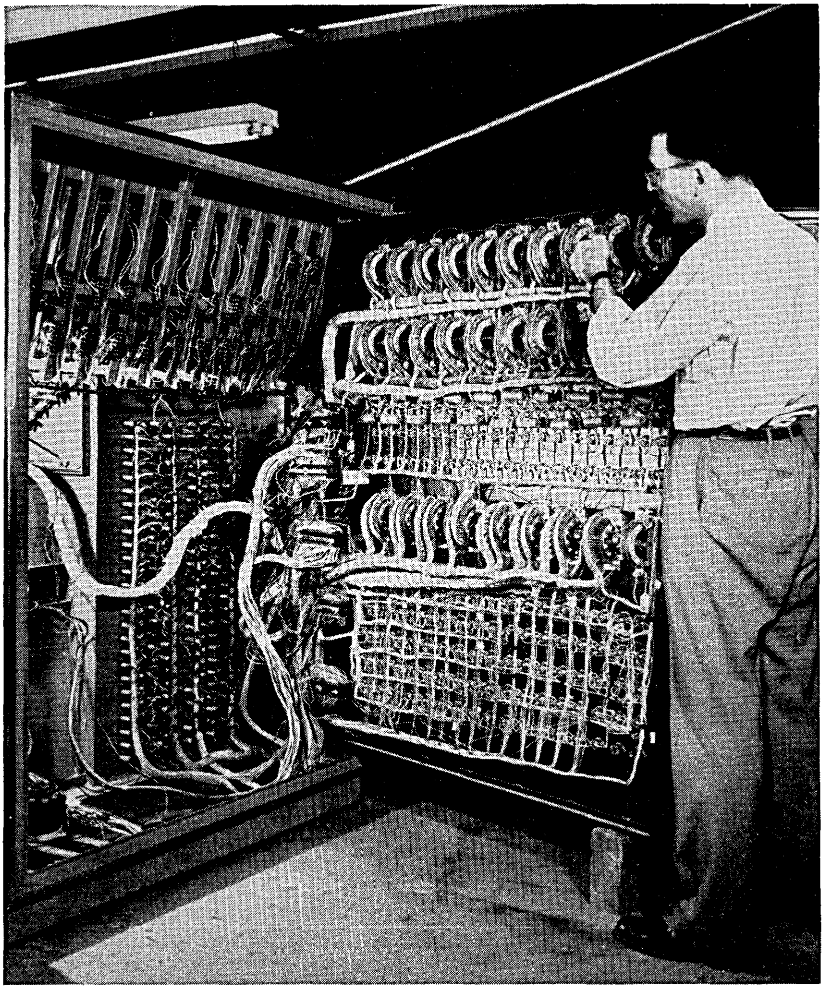


Fig. 17.9 Inside view of control panel of Fig. 17.8 showing components (Courtesy Dasol Corp.)

The second phase consists of memorizing the selections in consecutive order. Two banks of step switches under the control of relays pick up the calls and record them for future use. Should there be an error in selections, an independent relay circuit allows the error to be erased and deducted from the inventory of that particular rack.

The third and last phase is the read section. A group of step switches coordinates the information relayed by limit switches at the individual rack and the selections retained in the memory levels of the previous circuits. Motor control relay circuits are energized to stop the container at its proper discharge point and effectively send the container down its rack over a specially designed pull-off friction wheel drive. Erase levels then delete the section from the memory. The stepping levels of the read section wait for the signal that the container has passed the eye before carrying out the next selection. Separate levels in the step switch circuitry guard against asynchronous troubles by giving a visual indication of the causes as well as preventing any further error.

Today, controls of this type are being applied to machine tools to an increasing extent.

17.5 AUTOMATION IN MANUFACTURE

Automation in manufacture today is of real interest for companies seeking to improve the quality of their products as well as to reduce the cost for a competitive and broader market. And, it is in this area that some of the most difficult problems are being met. But today, even such complex items as parts of television and radio circuits, springs, radiators, appliances, engines, bodies, and a host of other products are being produced on automated lines.

Manufactured items, hardware, and other products are being produced partially or completely automatically or in a series of semiautomatic steps. Automatic machines linked together by feeding, sorting, conveying, orienting, stacking, and other automatic devices carry the parts through a series of manufacturing operations. Such operations as casting, machining, forging, stamping, bending, drawing, coiling, etc., are being done automatically in a series of steps or are included in a completely automated procedure. Much as instruments monitor the processing of chemicals and the melting, casting, and treating of metals, so in like manner inspection and gaging devices control the processing of pieces being manufactured or completed and assembled into a finished product.

It may well be asked what the difference is between automation and ordinary mass production as we know it. First, automation should not be confused with basic mass-production techniques. Mass production, as developed from the original ideas of Eli Whitney, comprises the basic science of manufacturing, without individual fitting, large quantities of products from groups of components produced in mass. The primary feature of mass production revolves about dimensional standardization of components to permit interchangeability in assembly. Manufacture may be entirely by hand methods from start to finish. Mass-production techniques, however, constituted a major step in creating the possibility of providing large quantities of complex products with superior quality and uniformity at lower cost in time and money. But today, hand methods often fail to fulfill the requirements because of economics, market demand, speed, labor, and other factors. The solution is the succeeding step in the manufacturing picture—automation.

17.6 TYPES OF AUTOMATION SYSTEMS

In the achievement of automation, groups or sequences of operations, automatic mechanisms or machines, and control and handling devices are brought into a single system to produce continuous or cyclic operation. Wherever two or more automatic machines are tied together with overriding automatic controls to create self-feeding, a self-initiating and self-checking process, an automated system is created. Material, data, or pieces can be introduced into the system manually or automatically, and the processing steps are carried out to a reasonable degree, without manual intervention, to completion. Control may be simply mechanical, electrical, electronic, or a combination. Generally, almost any process can be automated. It is possible to automate a single-stage operation, a sequence of operations, a whole department, or a plant.

One modern installation which indicates the handling possibilities is television-tube production shown in Figs. 17.10, 17.11, and 17.12. An early stage in the Sylvania television-tube automatic line is shown in Fig. 17.10. From the bulb-washing machines, tubes move through this settling room controlled by one man. Automatic suction arms lift bulbs to spill out the chemical solution and then deposit them on a screen drying conveyor.

After passing through a screen lacquering operation, the tubes are placed on this automatic aluminizing line where the funnel portion is coated (Fig. 17.11). Each bulb is evacuated and pure aluminum vaporized within the bulb to create a mirror surface.

The actual aluminizing operation is carried out by the equipment shown in Fig. 17.12. Following this the lacquer is baked out, units are exhausted, bases sealed, aged, and automatically tested. The automatic tube conveyor system employs some 400 limit switches, 700 relays, 75 timers, 100 motors, and 16 miles of control wiring.

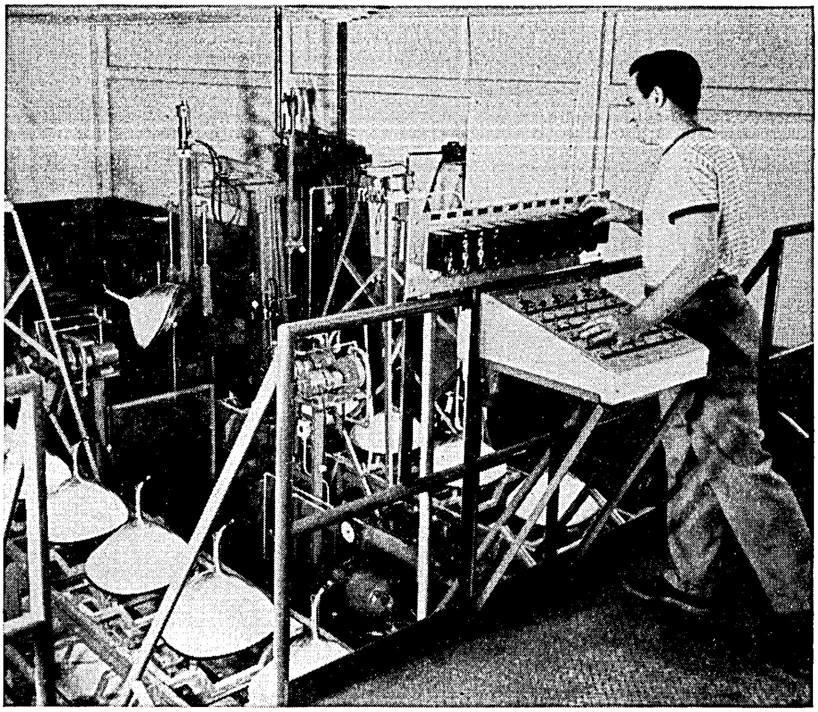


Fig. 17.10 Early stage in the Sylvania automatic television-tube line

Automation can be segregated into several types. It is possible to create an automatic batch or a continuous system. One process, for instance, may be more economically carried out in batches and another when operated continuously. Process characteristics as well as economics dictate the best method. On one hand, operations such as those including furnace treatment may require batch-type arrangement, although mass-production operations can often be carried out more economically on a continuous basis, keeping close control of the workpiece.

Figure 17.13 shows the schematic flow of the automatic batch production line for bearing races with integrated press forming, machining, annealing, washing, carburizing, and grinding operations. The raw material is first press-formed, machined, washed, annealed, washed, press-deep-drawn, washed, annealed the second time, washed, machined, gas-carburized, water-quenched, tempered, cooled down in soluble oil, and moved to final grind. All furnace charging is carried out through hoppers equipped with metering devices and charging mechanisms. By building up charges in these hoppers, any variations in flow from machine tools and other equipment will not interrupt the regular flow of parts required to keep the furnaces operating at peak capacity.

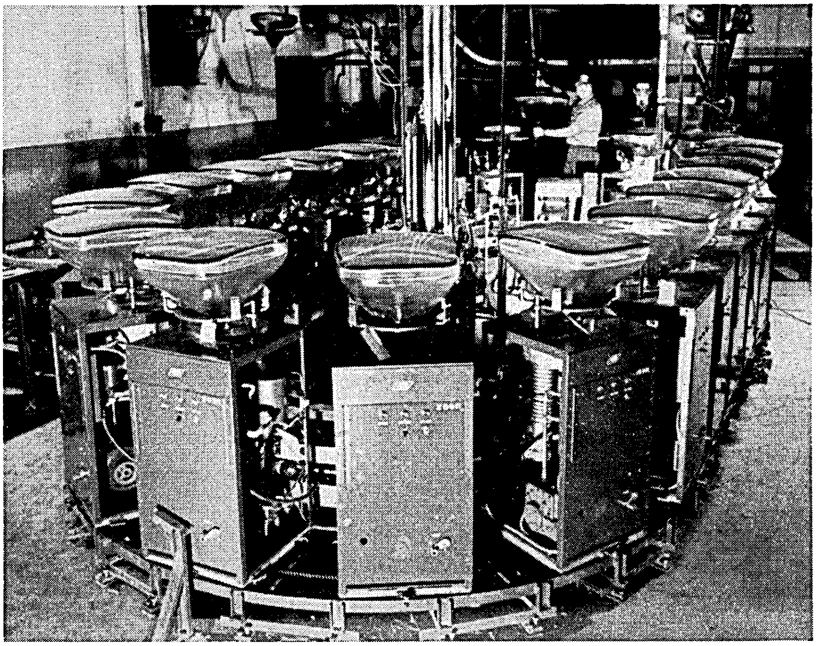


Fig. 17.11 Automatic aluminizing section of Sylvania television-tube line

Secondly, manufacturing automation systems may be set up with end-point control or in-process control. With end-point control, processing is completed before testing, checking, or gaging is done. This is suitable in many processes and feasible with others. However, where high precision or high-cost pieces are involved, in-process control is desirable for economic reasons. For in-process control, material being processed is under continuous control; metal parts, for instance, are gaged in the machine and correction or size control is accomplished during the operation. Gear cutting and shaving operations, for example, can now be carried out under continuous gaging and measuring instrument control, and necessary corrections are fed back to the equipment continuously.

Figure 17.14 is an example of end-point control. For existing internal grinders, a gage may be used which fits into the machine discharge chute. The completed workpiece is discharged by machine, rolls into the gage as shown, and is trapped for inspection. Good pieces are accepted and roll down the chute. Off-size pieces actuate the trap door shown to segregate good and bad production.

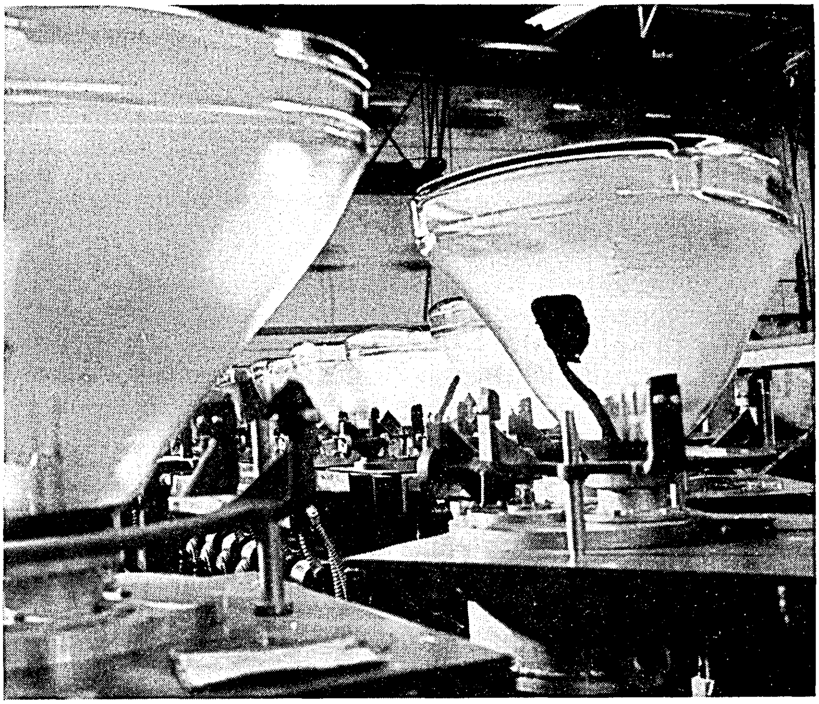


Fig. 17.12 Sylvania aluminizing operation close-up

Figure 17.15 is an example of in-process control. Automation installation shows parts on the conveyor passing over the top of the gage. When in position between grinder centers, the part is automatically picked up from the chain. The hydraulic cylinder built into the gages causes the gage to jump on to the work; the grinding wheel then advances to begin the cycle. When the part reaches an acceptable size, the wheel is automatically retracted and the cycle continues.

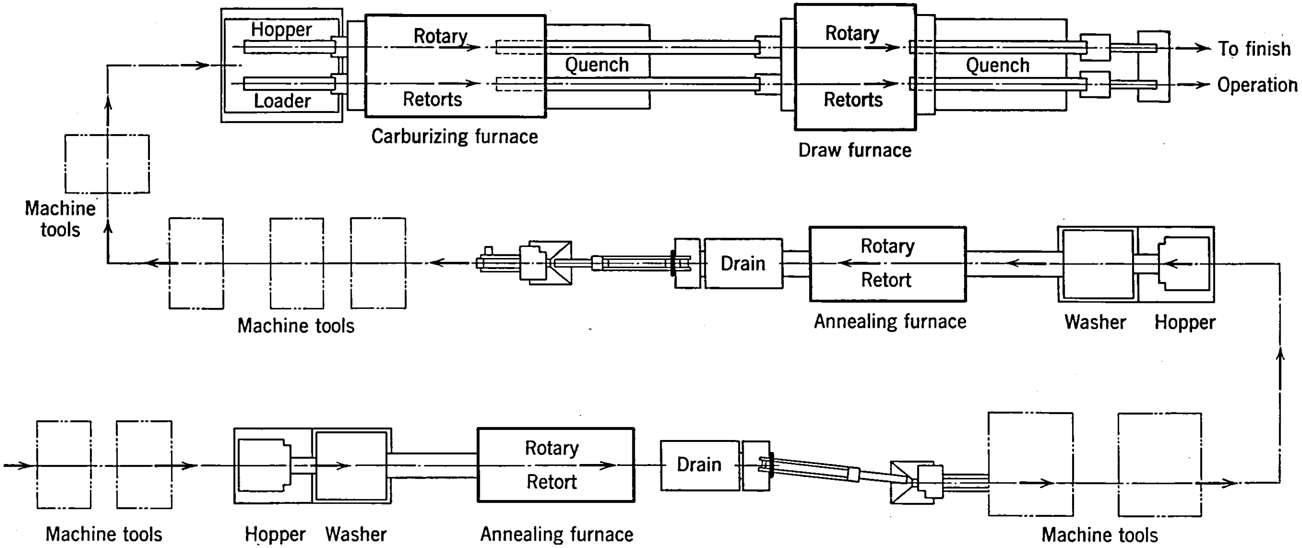


Fig. 17.13 Flow plan and machine arrangement for producing hardened finished bearing races (Courtesy Surface Combustion Corp.)

An example of a completely automated standard gear production machine is shown in Fig. 17.16. Work is fed into the machine from

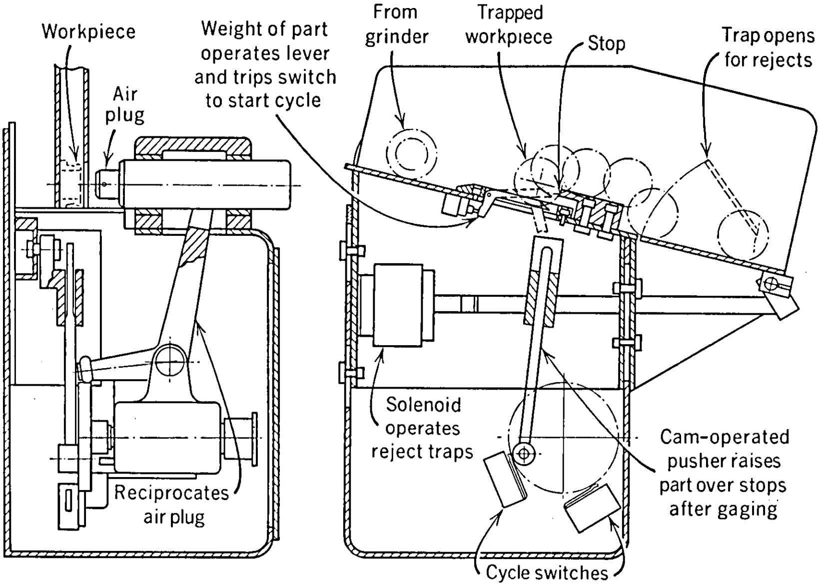


Fig. 17.14 End-point control device for grinders (Federal Products Corp.)

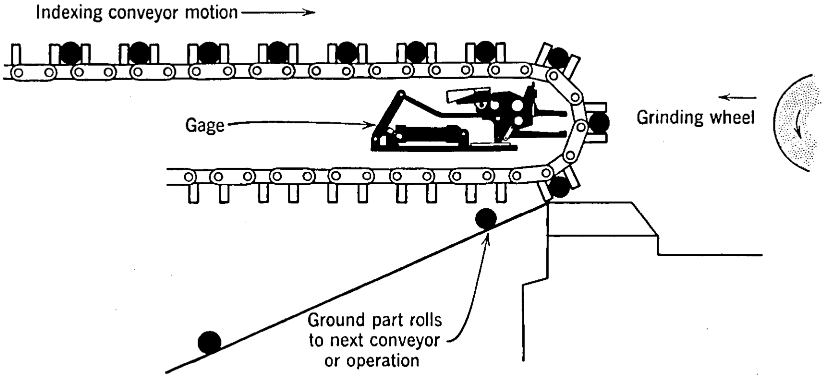


Fig. 17.15 In-process control for center grinding (Federal Products Corp.)

the hopper and chute at left; a three-way classifier in the center checks the gear for size; the controller at right automatically adjusts the center distance between cutter and work. The controller also keeps a statistical record of production and shuts down the machine after a predetermined quantity has been reached. Quality improvement and scrap elimination are considerable. With complex parts, end-point control offers the only practical solution.

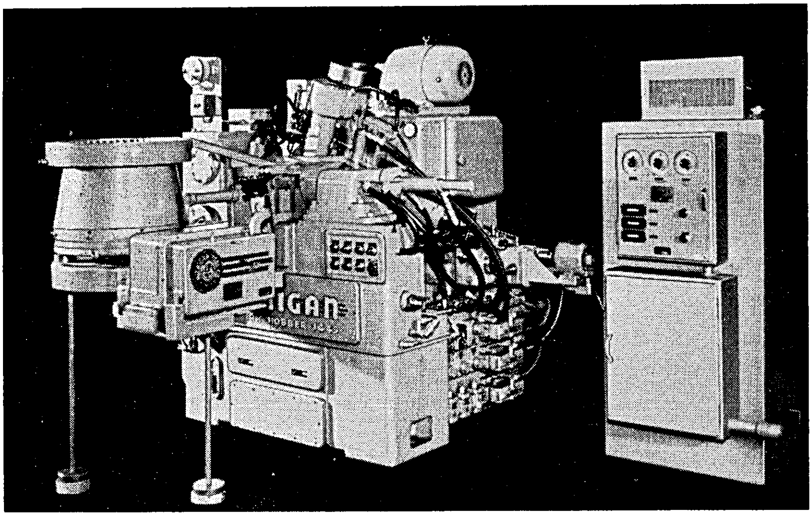


Fig. 17.16 Completely automated gear production setup (Courtesy Michigan Tool Co.)

17.7 QUALITY AND FEEDBACK CONSIDERATIONS

Another area that assumes key importance in automation is statistical quality control. Although desirable in ordinary mass production, it is neither always used nor an absolute necessity. Common methods of gaging after manufacture without machine control are still commonplace and still suitable to a large degree. To permit economic automatic assembly or automatic production of separate pieces, in-process continuous gaging to eliminate scrap and off-size is desirable and often necessary. Although not commonly recognized, feeding, chuting, orienting, and assembling operations in automation fail most often because of nonuniformity or nonconformity of parts.

Through automatic size and uniformity control of the parts being produced, production of bad pieces is eliminated. Economically of little importance in production for hand assembly (where scrap can be detected by the assembler), in-process gaging before continuous production assembly operations is important. Extremely small irregularities can jam intricate devices easily and, hence, crude operations may often be impractical to automate.

Here the design for economical automatic production assumes critical importance. Individual pieces can be designed for maximum facility in the reproduction process; in the gaging procedures necessary; in the handling, chuting, orienting, and feeding steps required; and in the actual assembly arrangement.

Where granular and fluid materials are processed, the same holds true—the quality of the product produced is dependent on the control of the ingredients. Manual control usually results in high rate of loss and wide variations. Quality can be improved and maintained by automatic control of ingredients. Cost is largely reduced by eliminating scrap and rerun and attaining faster output. In many cases, scrap reduction and quality control alone account for significant savings. In one instance, by replacing a hand inspection test with an accurate automatic test, not subject to operator variations, needless scrap was reduced almost 50 per cent. In another instance, automatic metering and weigh feeding of ingredients eliminated millions of pounds of rerun. Cost savings are worthy of real concern but in many of these cases, actual losses fail to appear as such or are accepted as a necessary evil associated with the particular product.

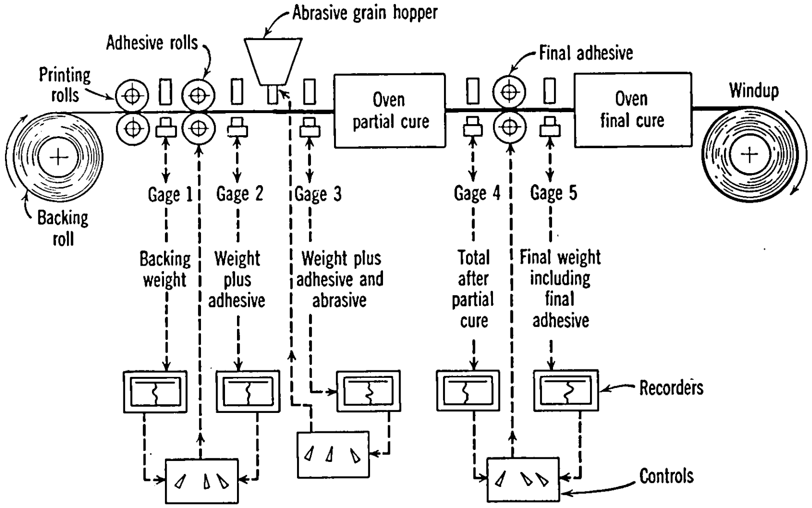


Fig. 17.17 Nuclear gaging system used for coated abrasive processing (Courtesy Industrial Nucleonics Corp.)

Figure 17.17 is the schematic arrangement of a nuclear gaging installation in coated abrasive processing. Until recently, gages used functioned merely as recording instruments. Control was exercised manually and the machine had to be stopped every 1000 yards to cut weight samples. Now the equipment includes five Industrial Nucleonics gages with feedback to permit automatic, precision control. Yearly savings were $90,000, exclusive of labor saving which was only $7500. Similar installations are being widely adapted in metalworking.

This then is the area of feedback. Regardless of the level of automation, information can be fed back through the control system to more or less control the accuracy of the job being done. Gages and measuring devices can control production operations; instruments and sensing devices can monitor processing operations to hold preset values or vary values as desired; mechanical controls can maintain preset or sequences of steps as needed. They can also shut down the process in event of trouble. In other cases the actual procedure itself forms the feedback loop—if it cannot be completed the operation stops—or an operator at the final controls forms the feedback loop. The main purpose is quality or accuracy.

17.8 DESIGN OF PRODUCTS AND AUTOMATIC ASSEMBLY

Real economy in mass manufacture of products has been closely associated with good design for production, but failure to insure the best design features has seldom denied satisfactory results. Actually, it has even been possible to defy good design criteria and depend on the skill of the men along the lines to supply the necessary correction factors. Not so in automation. Mere mass-production techniques must be superseded by modern production design for automatic manufacture. Although there are instances where production of components and assemblies has been automated without design change, in most cases significant savings await ingenious design modification.

This is especially true of automatic assembly. To attempt to assemble automatically without consideration of design modification for suitability would parallel the attempt to mass produce without regard to the technique of interchangeability. Today, with automation, simplicity in production design is the byword. Basic principles for assembly design are just now being evolved, and in the near future it will be possible to present specific data for design guidance. With some designs, even arduous labor and large expenditures may not provide the desired economic result. The much publicized Project Tinkertoy bears this out. After thorough economic studies were completed it was found that semiautomatic methods offered the most effective result, not complete automation.

In this analysis of the project it was found that the electronic amplifier of simplified modular design reduced manufacturing cost 38.5 per cent over conventional piecemeal design. This method consists in hand-assembling the simplified machine-produced components. Fully mechanized production of the same modular design reduced cost 44 per cent with roughly the same output per hour. However, seven of the process steps proved to be lower in cost with the hand method, and when substituted in the otherwise mechanized procedure the combination offered the lowest total cost per unit.

Significant factors in the project were that $665,000 was required to create the all automatic facility and $82,000 the semimanual. Little savings would accrue from increase in size of either facility. Limitations on the output, and hence economics, of the mechanized production installation were found to center around equipment cyclic rates, machine delay time, reject percentage, equipment costs, numbers and skills of operators, and equipment maintenance. This example merely indicates that automation is not a panacea. As always, a neat overall compromise in both design and production provides the best end result.

Today, in electronics assembly, the most successful approach appears to be that of using time-tested basic components with etched circuit boards. For example, the automatic electronic assembly machine built by General Mills, Mechanical Division (see Chapter 13, Figs. 13.16 to 13.20), inserts from one to twenty-four electronic components—assorted resistors, capacitors, pulse transformers, and diodes—into printed circuit boards. The normal speed is twenty assemblies a minute, nearly 10,000 in an eight-hour day. For assembly of an IBM computer circuit, the etched circuit has been designed as a standard. By inserting components into the proper sequence of holes, any one of a series of basic computer circuits can be produced. Standardization of the design has permitted automatic production of boards. All necessary circuits can be produced from one standard circuit board.

Overriding considerations in basic design detail which must be brought into play for automation can also, and often do, contribute direct additional benefits. Typical cases on record show that the constant drive toward suitability for automation and for simplicity result in products that are much superior from standpoints of performance, uniformity, stability, service life, and maintenance.

Actually, automatic assembly is not new but its development has been slow and laborious. The demand for assembled products in many areas was such that there was no possibility of reasonably producing them by conventional hand methods many years ago. As a result these items have long since been produced automatically. At present the following operations are included in automatic assembly lines: (1) machining—drilling, reaming, tapping, milling, boring, and burring; (2) press operations—stamping, riveting, upsetting, staking, piercing, trimming, and forming; (3) joining—welding, soldering, and heat sealing; (4) feeding—hoppering, feeding and driving screws, and applying sealers and lubricants; (5) inspection—weighing, sorting, measuring, and testing; (6) miscellaneous operations like burning off materials and cleaning operations; and (7) packaging—counting, package making and closing, and printing.

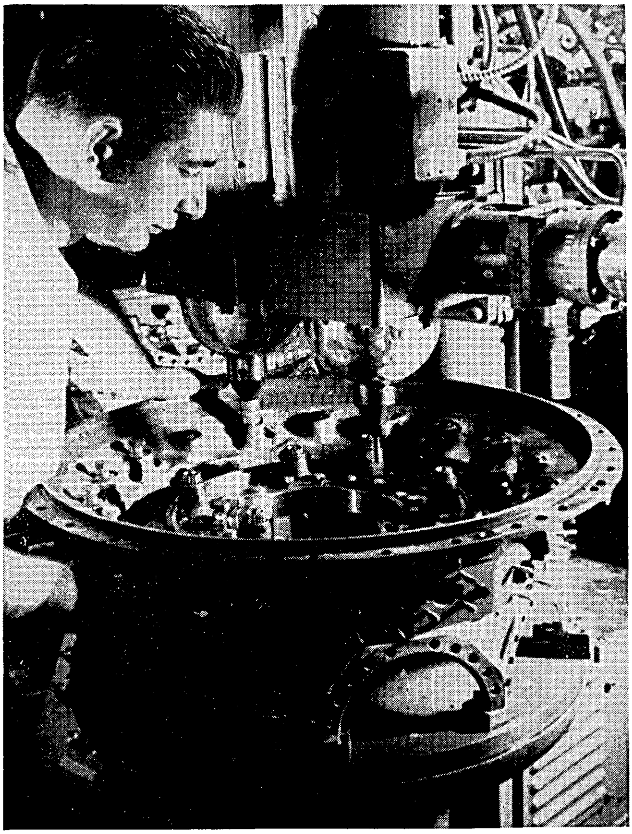


Fig. 17.18 Automatic bolt-positioning and torquing machine for aircraft engine housings (Courtesy Wright Aeronautical Division, Curtiss-Wright Corp.)

Today these principles are well established and any product can be assembled at least in part. But this equipment requires highly trained operators and setup people for successful operation. Present know-how indicates that a limited group of steps should be included in one assembly sequence—perhaps up to twelve—and for success each sequence should be arranged to permit dropping out of the line for short periods. On the basis of modest-sized components successfully handled so far, work is now in progress to automate the assembly of automobile engines. The semiautomation of aircraft engine assembly is already history.

The automatic machine shown in Fig. 17.18 positions and tightens bolts of the engine tappet housing to the rear aircraft engine crankcase section. The tappet housing is positioned between the power and rear section and houses the pushrod assemblies for operation of the cams and rocket arms through the cylinders. Each of the automatic machines for assembly is positioned in front of the assembler at working height and the assembly is done with as little human error as possible.

It is often necessary to develop new production techniques for economic automation. The Rockford Army Ordnance shell plant is a good example of an automation process design based on conventional production techniques. The line has two parts: (1) forging and (2) machining for 155-millimeter shells. Compared with a standard line the following factors have been pointed out. Product quality and uniformity are better. Operating cost is "average," but if equipment depreciation were figured in, cost would be higher than that for a standard line. First cost for the line is greater. Maintenance costs are higher and spare parts of a special nature are a problem. The same number of plant personnel are required but in all cases much higher skills are necessary. Today, this type of shell processing has been revamped for production by extrusion techniques which completely obsolete the earlier approach.

17.9 ENGINEERING AND MANAGEMENT

Automation creates a greatly increased demand for engineers. Not only does the product or service to be automatically processed require better and more competent preliminary design, but so also does the production equipment. Improved design for production is a must for economic reasons and for suitability to automatic processing. Greatly increased engineering time is absolutely necessary in the development of successful automation. Today, the engineering time and cost are paramount, and the actual equipment construction cost often becomes insignificant in comparison. The achievement of more widespread automation places a high priority on the supply of competently trained engineers and engineering technicians. The future rate of advance will be closely associated with their availability and the development of new techniques and know-how. Necessarily, advance will be gradual rather than overnight, owing to the sheer bulk of problems requiring solution in each instance.

Along with increased engineering staff and engineering activity, real achievement in automation requires a much closer coordination of top management, engineering, and all phases of manufacturing. Successful application of automation principles necessitates full appreciation of the numerous factors involved and a rational approach is needed.

Viewed as an engineering and organizational problem, automation can be basically termed a horizontal technology. Both functionally and industrially, automation requires a blending and coordination of many specialized vertical areas of engineering and industry.

Where continuous processing is concerned, two or more specialized industrial fields may be involved. Where previous practice permitted each to develop single facilities separately, under automation these generally must be tied into a single system. It is not unusual to find materials handling, chemical processing, heating, mixing, blending, pressing, machining, assembling, inspecting, and packaging combined into a single system.

Figure 17.19 is the layout of the Dodge continuous forging line for crankshafts. Total time for each part is 8 hours but shafts emerge at 30-second intervals.

From the previous forging line, crankshafts move to the heat treatment line (Fig. 17.20). Twin systems match this relatively slow process with the forging line production rate.

The major problem of automation consists in the integrated engineering of widely divergent manufacturing areas in an atmosphere where specialization has been the typical practice. Thus, the need has arisen for developing special automation engineering groups of broad background and experience. Before automation, few engineering departments were set up to cope with this problem. Today, however, such engineering facilities have been found to be a necessity, and many specialized consulting groups have been formed to supply a complete service for companies not in a position to maintain such a group.

With the development of engineering groups of broad industrial experience, know-how and techniques commonplace in one industry are readily adapted and transplanted into another. Instances have been all too numerous where one industry has been found to be working diligently to develop a method identical to one in use for many years in another industry. Automation, by virtue of its horizontal nature, is fast bringing to light many of these obscure developments and evolving an overall technology which makes tremendously increased demands on engineering acumen. The result will be seen in strong demand for increased engineering personnel in top management, in the engineering department, and in plant operations.

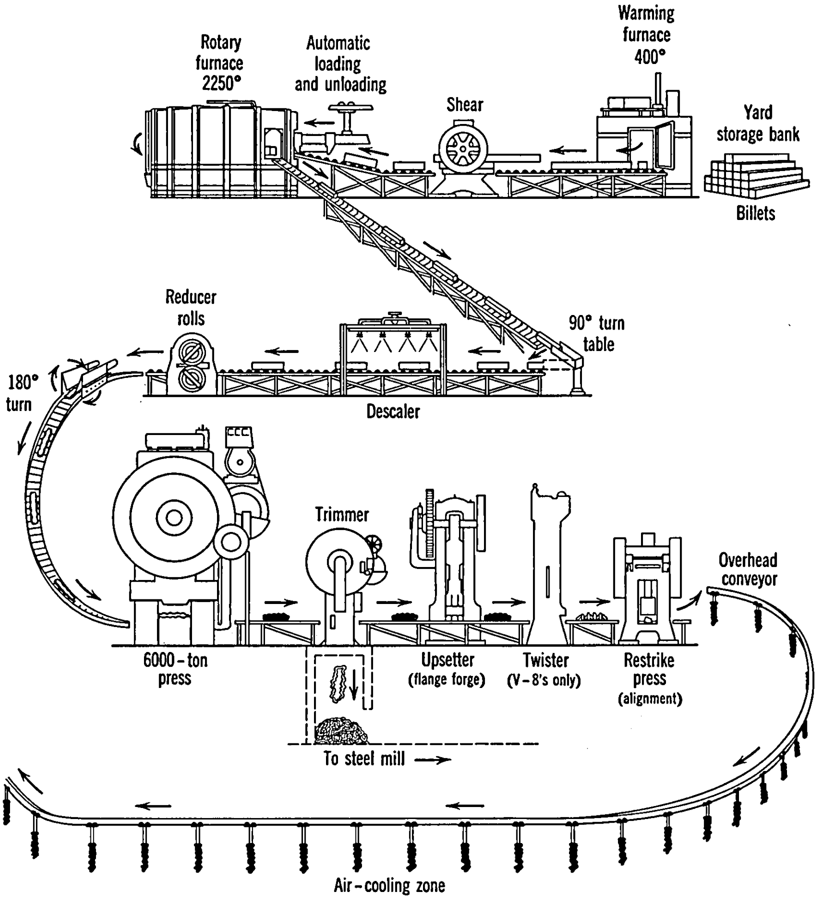


Fig. 17.19 Continuous forging line for automotive crankshafts (Courtesy Dodge Division, Chrysler Corp.)

In no case should the president or top management decree automation. Economically this is not feasible—results often will be either unsatisfactory or even unsuccessful. The many cases of real achievement on record which have stemmed from altogether unrelated causes emphasize the need for careful management study of the overall prob-

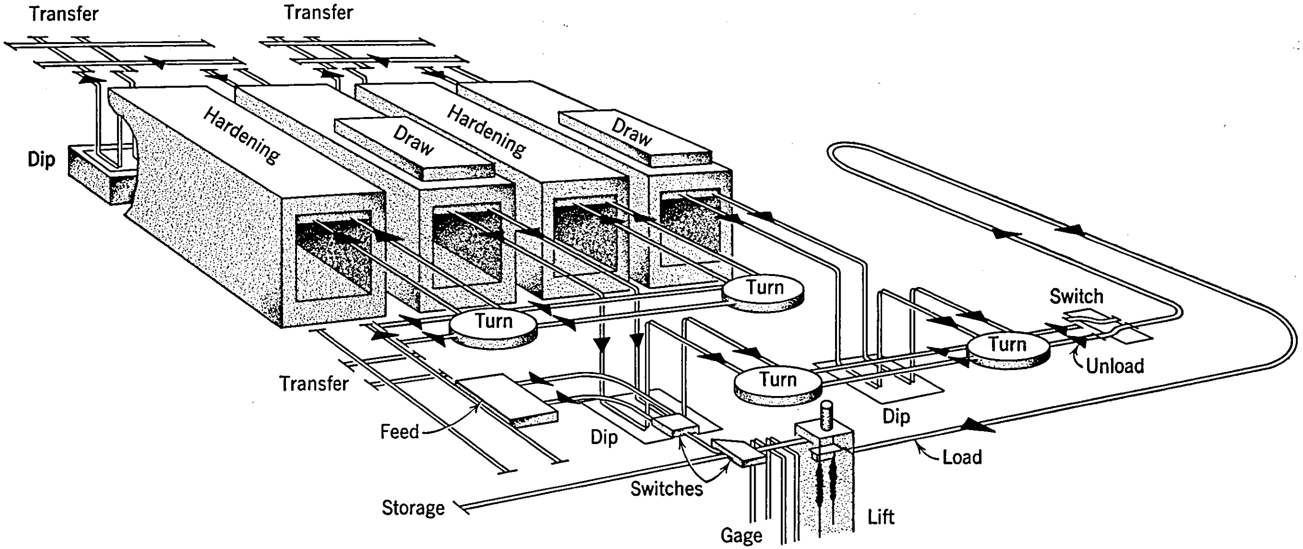


Fig. 17.20 Continuous heat treatment line for the forged crankshafts shown in Fig. 17.19 (Courtesy Dodge Division, Chrysler Corp.)

lem. To expect automation to develop "upward" from plant and engineering sources alone will fail to produce the full potential.

In many instances today the automation programs being instituted are set up so as to be sparked and fully controlled by management since the effective climate for automatic operations includes many areas of the enterprise outside the ordinary sphere of engineering. Thorough research by management can often reveal the presence, absence, or adaptability of critical factors necessary for realization of automation.

A survey of industries both old and new using automation reveals that it has proved uneconomic to attempt a direct jump to full automation. The very complexity of most installations forbids it—and seldom has it been done. On the contrary, it is most practical and economical for automation to evolve gradually. Automatic manufacture of various types of headlamps and other light bulbs provides an excellent example. Here it is found that evolution has been in process for perhaps twenty years, step by step, with each step requiring considerable engineering and development—and with no one person or group being the sole creator. Rather, many engineers and plant personnel are needed to produce the final successful product.

Thus, the rational approach proves to be one in which automation is entered into with considerable study, research, and economic evaluation from all vantage points. Every project must be carried out and completed under a master plan which includes full evaluation of automation possibilities, even if use is practical only in a minor degree.

Working with a program in which all phases of the existing organization are cooperating, management, engineering, and labor, leads to an understanding of what automation is trying to accomplish. Methods study men of yesterday sought to determine whether a given motion was really necessary, but the automation engineer of today is questioning every phase. The design of the product—is it suitable for manufacture? The manufacturing method—do we need this operation at all? And the operation itself—is it in the right sequence? Should it be done by automatic machines? Can the automatic operations be linked together to get automatic production?

17.10 THE FUTURE

Once the full perspective and potential of automation are apparent, its application in proper degree can result in real advance. Vastly improved working conditions and far greater development of potential interests and skills can be attained. The necessary elements for automation are largely available. The question today is not will there be automation but how much will there be and how soon will it spread across all manufacturing industry.

17.11 REFERENCES

1. *Automatic Control and the Automatic Factory,* Council for Technological Advancement, 120 S. LaSalle Street, Chicago 3, Ill., Nov. 24, 1952 (No. 6).
2. *Automation and Other Technological Advances,* American Management Association, 330 West 42 Street, New York 36, N. Y., 1953.
3. *Guides to Meeting Tomorrow's Production Needs,* American Management Association, 330 West 42 Street, New York 36, N. Y., 1953.
4. David G. Osborn, *Geographical Features of the Automation of Industry,* University of Chicago, Department of Geography, Chicago, 1953.
5. N. L. Bean, Automation, *Mech. Eng., 71,* 389–90, May 1949.
6. C. F. Hautau, Methods of Automation, *Cleveland Eng., 46,* 15–20, March 5, 1953.

18. Economics of Plant Automation

FRANK K. SHALLENBERGER

Stanford University
Palo Alto, California

18.1 INTRODUCTION

It is my intention in this discussion to consider the dollars and cents aspects of automation from the management point of view. It is important that we look carefully and critically at the economics of automation, for automation will stand or fall on the economic yardstick of cost versus payoff.

Automation is an industrial product, sold in industrial markets. It is new. It is expensive. It has many uncertainties. It will not be bought as a style or fad or gimmick. It will be bought only when it offers a better, more efficient way of doing the job. The economics of automation will make or break it—it must pay off.

In these terms our discussion should include a consideration of the economics of automatic equipment, the impact of automation on management, the application of data processing to plant management problems, and the impact of automation on the society of which industry is a part.

Although automation is a new word in our industrial vocabulary, it is not a new phenomenon, and a brief look at the past will provide clues to the significance and the acceptance of automation in the future. Therefore, I feel compelled to take time to try to build a frame of reference for the consideration of the economic and management aspects of automation.

Fundamentally automation is a natural and logical extension of the technological change that had its origin in the wheel, the crank, the screw, and other basic mechanical concepts. It is an outgrowth of the mechanization which characterized the first industrial revolution. It is a logical end to imaginative, uninhibited methods study and process improvement. Modern automation brings important new techniques to industry—programming, feedback, computers, servomechanisms, etc. It introduces the principle of self-regulating control, the concept of machines to run machines. It makes possible the replacement of human judgment in the control of industrial processes. But from the dollars and cents point of view of operating management, automation represents merely the addition of new techniques to the tools of mechanization which industry has been developing over the past 150 years. These new techniques will accelerate and greatly extend the substitution of machines for men. They will make economically feasible the mechanization of operations where mechanization previously was not practicable. But to plant management automation is not basically new, and it must compete with the many other techniques of cost reduction and product improvement.

Expressing automation in terms of our personal lives, we can say that we are today at the stage where our automobiles have power brakes, power steering, and automatic transmissions. All of these have been designed to help take the physical effort out of driving, but the operator is still required for control. Automation in the modern sense can "close the loop." It could make possible an automobile that could be driven on to the highway, and then, with the dial set for San Francisco, the occupants could read or watch television or look at the scenery or sleep until they arrive there. Modern control-type automation can make similar contributions in the plant, the office, the warehouse, and the farm.

Automation has been called the second industrial revolution, a revolution in which electronic controls will replace man's brain or sensory system in the same manner as power machinery replaced his muscles in the first industrial revolution. It has been said that the first industrial revolution provided machines; the second, machines to run machines; or that the first revolution was characterized by power and the second by control. We are interested here in the economic impact of automation rather than in semantics, but the point should be emphasized that automation is likely to be more of an evolution than a revolution. Like mechanization it will come slowly and gradually and not overnight. And as economic historians look back on this period of our industrial development and interpret it in terms of the grand sweep of economic history, they will note that in the mid-twentieth century, "something new was added," something which greatly accelerated the application of technology to industry, something which increased the rate of productivity growth from its traditional 2½ to 3 per cent per year to a rate perhaps double that. They will note that whereas extensive mechanization was feasible only in mass-production industries, the development of automatic electronic controls made possible the substitution of machines for men on a much broader scale than ever before.

18.2 THE GROWTH OF AUTOMATION

Before we term automation the second industrial revolution, we should recognize another contender for the title, the revolution in management philosophy and practice pioneered by Frederick W. Taylor at the turn of the century. This revolution was not evidenced by material changes such as automatic conveyors, transfer machines, and workerless factories; but Taylor's contribution of scientific planning and control as a substitute for off-the-cuff management was in many ways much more fundamental than automation. His proposal that management decisions be based on objective investigation and fact, rather than upon hunch and tradition, and his advocacy of determining in advance what management hopes to achieve and then establishing controls to make certain it is achieved, were extremely radical concepts at that time; in fact, unfortunately, they are still radical doctrines in some business quarters today.

It is interesting that from Taylor's basic concepts and study of the management job came not only what we now call scientific management but also the systematic approach to industrial technology that is today leading to the automation of production processes. Taylor's own development of high-speed steel is a classic example of the type of organized industrial research on which the technology of modern-day industrial automation has been built. And as we shall see later, Taylor's concept of management will reach its highest level in the age of the automatic plant, when long-range planning and control by management will be fundamental to successful operation, and when electronic data-processing equipment will make possible advance planning, logical decision making, and effective coordination and control to a degree that Taylor never dreamed possible. The popular example of early automation is Oliver Evans' flour mill built outside Philadelphia in 1784. Rather meager information regarding the construction and operation of this plant is available, but we are informed that by using water power to drive such modern devices as belt, bucket, and screw conveyors, grain was automatically processed into finished flour without human assistance.

There are numerous other examples of early-day automation. The "Victualing" Office of the British Navy is said to have operated an automatic biscuit factory in 1833. Jacquard developed and built his first punched-card-controlled loom for weaving figured fabrics in 1804. Elias Howe developed the modern high-speed sewing machine in 1845, and mechanical coal stokers were introduced in the same year. In 1854 the automatic telegraph message recorder was developed. Conveyors that automatically weighed and transported animal carcasses were introduced in the meat-packing industry in 1869. The Westinghouse automatic air brake was invented in 1872, cigarette machines were introduced in 1876, and automatic bottle-making machines in 1907. In 1920 the famous A. O. Smith plant was built, a factory designed and equipped to produce automobile chassis automatically. Strip steel was sheared, punched, formed, riveted, and painted by automatic equipment, ultimately at a rate of 10,000 finished chassis per day in a plant where the workers were primarily inspectors and maintenance men, not producers. Ford has had an Automation Department since 1947, the same year in which Sargrove built two machines in England for making radios automatically.

Our process and chemical industries have been largely automatic for many years; so also has been the production of bottles, tin cans, paper, light bulbs, radio tubes, cigarettes, phonograph records, and many other products.

Individual automatic machines are common throughout industry—automatic lathes and screw machines, automatic mills and drills, automatic punch presses, canning equipment, printing presses, and bottling plants. We have automatic molding machines in foundries which fill and ram molds at the rate of one every 20 seconds, draw, blow clean, place sprues and risers, roll over, close, shake out, and after pouring return the flasks to the molding station while the castings are automatically carried away, gated, and cleaned.

During World War II, the transfer-type machine was developed. The transfer machine consists of a number of more or less standard machine heads coupled together with automatic handling between stations, mounted on a common base and centrally controlled. One recently built performs 688 machining and inspection operations on 100 V-8 cylinder blocks per hour with one operator. Such machines typically have automatic conveyors to return work-holding fixtures to the loading end, automatic chip removal, and signals to indicate tool wear and malfunctioning. These units are becoming so complex and carry on so many functions that it is said that it can be statistically proved that they cannot function, that at any given moment they will be shut down for maintenance or tool replacement. Cyclic tool change, preset tooling, unitized plug-in circuits, graphic indicator panels to help localize trouble when breakdowns occur, and segmented automation are some of the techniques used to meet this problem.

Automatic inspection, testing, sorting, weighing, and packaging equipment is widely used. Simple components such as oil filters are now being automatically assembled. A $21.95 radio is reportedly being manufactured and assembled in 4 minutes, the whole process being completely automatic. A $60.00 vacuum cleaner is being made with only 60 cents worth of labor in it. A trackless, operatorless "sniffer"-controlled industrial truck can follow a wire buried in the plant floor. Color sorting is being carried on at speeds up to 200 scannings and sortings per second. Remington Rand is currently advertising that the UNIVAC high-speed printer can read out and print a full page in 9 seconds.

Through control automation, machines have been taught to observe, think, make decisions, and act accordingly. For example, most automobile engine plants now have crankshaft balancing machines which check the balance of a machine crankshaft, remove the right amount of material from the right place to correct any unbalance, and pass the finished shaft on to the next operation.

There is another type of development which has much the same impact as automation—shell molding, die casting, and powder metallurgy—processes that are not only ideally suited to automation themselves but also have the effect of eliminating subsequent operations or facilitating their automation by providing consistent dimensions which permit automatic handling, chucking, and assembly.

Well, if automation is not really new, if it is basically the same old stuff in fancy dress, then why all the excitement over it? Why do people like Ben Fairless, Douglas MacArthur, George Meany, and Walter Reuther go out of their way to comment on it, sometimes at great length? Why all the attention given to automation in the magazines, Sunday supplements, editorials, company publications, symposiums, labor-management meetings, and military groups? Why is it that we see the word in practically every issue of such magazines as *Fortune, Time, Business Week,* and the *Wall Street Journal?* Why do we have several new magazines devoted entirely to the subject?

Why are advertisements using it for a catchword? Why are others manipulating it as a trick word—"economation," "fleximation," "airmation"? Why are labor leaders making a scare word of it, using it as a fulcrum for wage increases and the guaranteed annual wage? Why have even the Russians climbed on the bandwagon with an article purporting to describe "the world's first automatic piston factory"?

There are primarily three phenomena responsible for this upsurge of interest.

(1) Mechanization or "Detroit automation," applied thus far in only a rather narrow segment of American industry, the production of automobiles, appliances, foodstuffs, and similar items of mass consumption. Mechanization is increasing at a rapid rate under pressure of increased labor costs, competition, expanding markets, high purchasing power, and easy capital.

(2) The technique of automatic, self-regulating control, the heart of the new plant automation, the direct outgrowth of the electronic technology of World War II. This is the application of the "little black boxes" which pilot aircraft, direct gunfire, and control the flight of guided missiles to plant problems of guiding the feed and cut of machine tools, controlling conveyors, inspecting, assembling, and packaging the finished product.

(3) The development of computers or electronic brains which can and will automate vast areas of clerical, statistical, and administrative work.

It should be noted that the devices of automation are not necessarily electronic, nor are they necessarily programmed, nor do they all incorporate the principle of feedback. Some represent only techniques of mechanization without feedback—cams, levers, air or hydraulic pistons, limit switches, relays, solenoids, etc. Such devices will be part of the automation package, too, the actuators which carry out the instructions of the control equipment.

The more recent additions to the automation kit, with which the plant engineer or machine designer works, are the electronic devices, offering programmed self-regulating control. The ability of these devices to follow a planned sequence of actions, to make a logical choice between alternatives, to monitor their own performance, to detect and correct their own errors, and to make certain that they have performed exactly as instructed—these are the contributions of modern automation. The versatility of electronic controls is almost without limit. There are few if any jobs in industry they could not perform if economically warranted.

18.3 ECONOMIC BENEFITS

The economic benefits of automation are not limited to the replacement of labor, for it can improve quality, decrease spoilage, and increase machine and space productivity. But clearly its greatest economic benefit and most effective sales appeal is as a substitute for human operation and control. Let us not be evasive or timid on this point. Let us be frank, honest, and realistic. Let us not hide automation's greatest potential benefit—the elimination of labor.

I have heard numerous manufacturers state recently, "We want our equipment mechanized to take the hard work and drudgery out of the operation. But wherever possible, we also want it automated to eliminate the human element, to obtain consistent and predictable results which cannot be obtained when labor has control of the operation. We want to be able to know our costs in advance. We want the operation set up so the machine paces the worker, rather than the worker pacing the machine. That is as important as eliminating the worker."

Automation has given us devices which can see better, hear better, and measure better than human operators. They are more reliable, more powerful, and more precise. They think and move infinitely faster than humans. They never get tired, they willingly work around the clock, they do not make mistakes, they do not talk back, they are obedient, consistent, and fully predictable. They will not go out on strike, they do not ask annually for higher wages, and they have few personal problems.

From the standpoint of industry as a whole, one of the most important characteristics of electronic controls is their high degree of flexibility. This is a fact of basic economic importance in the potential of future automation. Mechanization as it has been applied in the automobile industry is feasible only in long-run assembly line, mass-production type of operations—a dramatic but relatively small segment of industry. We must bear in mind that *over 75 per cent of our manufacturing falls into lot sizes of less than fifty pieces.* Such short-run production simply cannot afford mechanization of any appreciable magnitude. If, however, mechanisms can be developed to carry out the necessary operations automatically, and at the same time permit ready conversion to other sizes and other shapes, or to completely different operations, we shall have gone a long way toward solving the problem of automating more than just a select segment of industry. Electronic controls appear to provide such flexibility. The bright new future, which they promise, lies in the possibility of mechanizing intermediate short-run operations, where any high degree of mechanization has thus far been impossible—in the aircraft industry for example. Those engaged in the production of automation equipment will do well to recognize that the green pastures are not all in Detroit. The intermediate and short-run market offers broad opportunities to the equipment manufacturer who can meet its special demands.

There are many automatic processes that are inherently flexible as far as product is concerned—die casting, heat treating, plating, painting, cleaning, etc. The automatic molding machine described earlier can be converted to casting other products simply by changing patterns. Now electronics seem to promise similar flexibility in other areas such as machining. The best-known example of such flexibility is MIT's numerically controlled milling machine in which a punched tape controls all action of the cutter in three dimensions. It will machine-out any simple or complex design dictated to it by the tape within the capabilities of the milling machine.

The Arma Corporation has developed a small lathe controlled by a punched tape much like a player piano roll. A technician working directly from blueprints punches the required information into the tape. Reportedly this lathe produced in 4 minutes a piece that required 30 minutes for a skilled machinist to make. Tape preparation required only 15 minutes; in other words, the machine was programmed and turned out a single piece in about two-thirds the time required by a skilled machinist and thereafter continued to produce at a rate seven times as fast! The estimated cost of the electronic control was $1500 or less than one-third the annual cost of a machinist. This machine performed only very simple work and is not yet commercially available, but its significance is apparent, especially its very high flexibility.

There are many others working on similar approaches. For example, development work is now being carried out in Palo Alto on an automatic machine tool controlled by means of a magnetic tape which provides up to fourteen channels of control. It is contemplated that companies which cannot afford the necessary taping equipment will send blueprints to a centralized taping service, which will then supply the tapes ready for use in the same manner as photographic films are developed and printed. Changing from one job to another will be just a matter of changing tapes and tooling.

When such machines as these are coupled together with flexible automatic loading, unloading, handling, inspection, and assembly devices with appropriate controls and interlocks, the automatic factory will come into its own. So far as I know, relatively little has thus far been done to obtain the same degree of flexibility in these other areas.

It must be emphasized that there are other important incentives to automation besides reduction in labor costs. The experimental Arma lathe, for example, through feedback control, held tolerances to 0.0003 inches, much closer than the skilled machinist can do; thus automation produces a better product with less scrap, and tolerances can become a matter of little concern. Further, automation can greatly increase operator safety. It can eliminate drudgery and improve working conditions. It can decrease space and equipment requirements by increasing machine productivity and permitting round-the-clock operation. It can make economically feasible operations and products which previously could not have been competitive. For example a new reusable shipping crate which can be quickly taken apart or put together without nails or tools is now being produced in Los Angeles. Thanks primarily to automation, this product has an extremely promising future, for the intricate and precise cutting and boring required are economically feasible only because a specially built automatic cutoff and boring machine has made production possible at competitive prices. This machine does more accurately in 6 seconds what it took 2 minutes to do on conventional tools. At an anticipated production rate of over 2000 cleats per day, the production savings are clearly apparent. The payoff is less than one and a half months—and the venture would have been impossible without the automatic machine.

18.4 DETERRENTS TO AUTOMATION

The foregoing examples indicate some of the technical possibilities. But to foretell the future of automation, we must ask ourselves some questions about automation from the standpoint of operating management. What use should we make of automatic equipment? How fast should we go? What are the basic considerations determining the applicability of automation?

To me, one of the striking things about automation is not what can be done in the future but rather how little has been done in the past. The past is indicative of what industry is likely to do in the future; hence, it warrants consideration here. There are thousands and thousands of operations now performed by human workers that could be performed more accurately, more efficiently, and at lower cost by automatic means, using devices which have long been available such as air and hydraulic cylinders, limited switches, solenoids, relays, photoelectric cells, etc.

Why has industry not taken greater advantages of the opportunities offered by these devices? What considerations have determined whether or not specific operations should be automated?

Availability of capital has of course been a prime consideration. Automation equipment is generally expensive. There has been little standardization of components and generally speaking each installation must be engineered for the job. Each proposed mechanization must compete for available dollars with other uses that might be made of the same funds—research, development of new products, increased inventory, improved tooling, and sales promotion. In small plants, funds are often lacking for even the simplest and most modest mechanization. No matter how great the potential payoff, there always seems to be some other place that offers even greater and more definite return.

Very often management does not know what is available, or how to use it, or what can be done. It is a full-time job to keep up with current developments in this rapidly growing field and to plan the application of such devices to the individual company's operation. Most managements simply have not had or have not taken the time to study the new developments in the light of their own company's needs.

Companies serving markets of low-volume, short-run, unpredictable-demand, and nonrepetitive operations are understandably cool toward automation. It has already been pointed out that mechanization on any appreciable scale is practicable only in mass-production operations. However, there is nothing sacred or necessarily desirable about 100 per cent automation. Like any good methods work, automation should be carried only as far as the situation warrants. In some very few situations, 100 per cent automation may be justified. In others, the operation may be made only 50 per cent automatic or even 5 per cent. The payoff is proportionately the same. It must be recognized that there are opportunities for automation in every plant.

One of the reasons why automation has not progressed more rapidly is that to obtain the full benefits requires more than just superimposing automatic mechanisms on existing production operations. Many existing products are simply not suited to automation in their present form. To redesign them of new materials and to re-equip the plant for their production may require extensive investigation and a major capital outlay, plus the scrapping of existing tooling and equipment. It may also require a major sales effort to gain customer acceptance of a radically new product. These expenditures constitute a hurdle not easily surmounted.

One of the greatest deterrents to automation is good, old-fashioned management inertia, especially when profits are substantial and competition is easy. Many companies do not think of new equipment until difficulties arise with the old. A new way of doing things is always suspect from the start for it involves changes which people make only with reluctance. There is a natural human tendency in industry to let well enough alone. A new process always brings uncertainties. I have spent a major part of the last few years in introducing a new and modern process to industry and have some knowledge of the difficulty of gaining acceptance for a new idea. In 1953 our salesmen were told, "Business is too good to change. Why upset things? We are making plenty of money. Why risk it?" In 1954 the same customers claimed, "Business is poor. We do not have the money, and cannot afford to modernize." This attitude reminds one of the Arkansas traveler who was told there was no point in fixing the roof when the sun was shining, and when it was raining, it was too wet up there to fix it. Unfortunately, there is a great deal of that kind of thinking in business too.

A friend of mine, a competent engineer and an excellent salesman, attempted unsuccessfully last year to sell automation equipment to industry in the San Francisco Bay Area. Within a year he was forced to give up. He likens his customers to the farmer who would not listen to the county agent's advice because, as he said, "Why should I take advice from him? I ain't farming half as good as I know how already!"

This situation is one known and recognized by every manufacturer of capital equipment. It is a most difficult job to introduce new and improved equipment to industry. The public will demand and pay for new gadgetry and emotional appeals on consumer products, but the buying motives of the industrial purchaser of industrial equipment are vastly different from those of the typical retail customer. Company after company introducing products to industrial consumers who it appeared "could not afford not to buy" have learned the hard way that only a small, advanced segment of industry is really actively interested in new processes, new equipment, and new materials. The electronics people have perhaps more to learn here than those who have been producing industrial equipment in the past, for they have generally been selling to aggressive, alert buyers in a field of rapidly advancing technology. When they begin selling control or production equipment to industry as a whole, they will find the situation considerably different.

One manufacturer of automatic inspection equipment has adopted a policy of attempting to sell an industry only when it is certain that the equipment can cut present costs in half. Even then it takes, on the average, a full year from the first call to the first sale. Another firm developed and put on the market a new and highly flexible machine tool capable of multiple simultaneous operations and rapid changeover from one job to another. Even in its own plant this company has had great difficulty in getting production people to visualize and make use of the new capabilities of the tool.

Labor has of course done its share to delay modernization through general resistance to change, insistence upon maintenance of former piece rates, and opposition to the assignment of workers to more than one machine. It has been stated that the United Automotive Workers is the controlling factor in the rate at which automation takes place in the automotive industry. A West Coast manufacturer reports that it has been forced to mount multiple tools on a common base in order to obtain union permission to utilize idle operator time.

High income taxes and unrealistically low depreciation rates allowable under the Internal Revenue Code have also retarded plant improvement. Plant managers logically ask, "Why invite labor trouble and tie up more capital when the Government takes away most of the savings in taxes?"

Another deterrent to modernization is archaic economic thinking. In most companies it is easier to obtain funds to keep an old machine working than to purchase a new unit, especially if the old machine has a substantial undepreciated value on the books. Before automation moves very far or very fast, businessmen must learn that book value has little or nothing to do with whether or not a machine should be replaced with a more modern machine. Obsolescence is a result of technological change and is not affected by accounting procedures. The question of whether or not a machine should be replaced by a more modern unit depends upon comparative performance in the future and not upon bookkeeping entries in the past.

Closely related to this type of thinking is the common hesitation to make a large capital outlay in a field of rapidly developing technology for fear a new development will make the expenditure obsolete before it can be written off. Rather than take this risk, some companies seem to prefer the purchase of an old standby model, tried and true, which is already obsolete.

These are some of the reasons why automation has moved slowly in the past, and they will continue to retard it in the future. Manufacturers of industrial equipment must recognize these deterrents and design and work around them. But there are powerful forces which are working in the other directions as well, strong incentives to automation.

18.5 INCENTIVES TO AUTOMATION

We have already described some of the improvements being made in automatic equipment. The machine-tool industry in general is not noted for its great imagination or progressiveness. But the leaders in the industry are becoming aware of the fact that the best way to sell new tools is to obsolete the old ones. Electronic controls are beginning to appear as significant improvements in industrial equipment. There are indications that competitive battle lines are forming for a struggle between the electronics and the machine-tool industries, a battle to determine which companies will build the new equipment—the conventional-equipment manufacturers with the assistance of the electronics people, or the electronic-control people themselves.

Increased competition is undoubtedly one of the greatest possible stimulants to modernization. We are today at the threshold of what appears to be the greatest capital boom in history. Widespread publicity on the benefits of automation has awakened management to the opportunities it offers. A major portion of the current capital spending is undoubtedly for automatic equipment. A leading manufacturer of automatic equipment reports that in 1957 it will produce each month twice as many units as were produced in the entire year 1947, and that its 1957 dollar sales will be eighty times as great as its 1947 sales. Factory sales of instruments for industrial control were reported to be \$3,000,000 in 1940, \$65,000,000 in 1953, and are expected to reach \$150,000,000 by 1960.

The increased cost of labor is among the most important incentives to automation. The higher labor pushes its wages and the less predictable labor becomes, the greater the likelihood of its replacement by automatic mechanisms. The firm whose rapid growth was just described reports that anticipated wage increases are one of the most important factors in their sales forecast. The guaranteed annual wage, which labor claims is designed to help protect the worker from technological unemployment, is much more likely to become a major stimulant to automation. Labor has on the whole been rather sensible in its approach to the problems created by automation. However, it has been completely "off-base" in linking the guaranteed annual wage to automation. Guaranteed employment can work only toward the replacement of the workers by the machine. Traditionally labor has been a variable cost; once labor costs are frozen by a guaranteed annual wage, labor thereby becomes a fixed cost in every sense of the word, and the incentive to its replacement is thereby increased.

Capital is more readily available for modernization than in the past. A new pay-as-you-go financing plan which allows the borrower to pay for new equipment with his depreciation allowance under the new tax laws has recently been announced. Various other equipment purchase plans attempting to gear payments to savings are available. Equipment leasing permits modernization with very little capital investment and often with substantial tax savings. *Steel Magazine* appraises leasing as follows: "Equipment leasing is something like free love—both have cost advantages over ownership or marriage; both involve relationships that may be terminated fairly easily; both are subject to legal difficulties not found in the more conventional arrangements; both are probably more talked about than practiced."

The overwhelming incentive to automation is national defense. The population of the NATO group of nations is outnumbered two to one by the Soviet bloc, and only by increasing man-hour productivity can this disparity be overcome. Defense production is typically large-volume, close-tolerance mass production, subject to manpower considerations and unrestricted by availability of capital—hence ideally suited to automation.

18.6 APPRAISAL OF AUTOMATION

The decision of when and whether to automate and how far to go in automation are primarily economic decisions. In some cases there may be no feasible alternative—processes in which radioactive materials, excessive pressures and temperatures, and tolerances and volumes not practicably attainable without automation are involved. But in most cases the decision to automate is based on the economic yardstick of dollars and cents, investment versus payoff.

The full benefit of automation can be obtained only by the complete rethinking of the entire product and process—complete redesign, new materials, new equipment, new plant—as opposed to piecemeal automation. One of the best examples of this is the use of printed circuits, automatic assembly, and dip soldering in the production of electronic equipment. Even this is a transitional step, using available components that were designed for conventional assembly. Hence it is only a foretaste of future electronic assembly.

Unfortunately, it appears that in most cases a complete conversion to automatic production is both economically and technologically impracticable. In general, therefore, automation must be carried out one operation, one process, one department at a time. A certain amount of the go-slow, wait-and-see approach is desirable, for automation is a young science and its most important developments are still in the future. But at the same time current progress must be made. We cannot wait for perfection. There are profit-making opportunities in the present, and competitive position may be lost if we drag our feet.

It is not practicable to set forth a universally applicable approach to automation, especially a program of piecemeal automation. But it is possible to indicate the type of questions which management must ask.

What do we expect to gain by automation? We do not automate to be stylish. Are we looking for economy? Labor saving? Reduced effort and fatigue? Safety? Freedom from labor troubles? Higher quality? Uniformity? Reduced material consumption? Increased production? Space savings? From such questions can be developed a master plan to insure that so far as possible all changes are directed toward a common integrated goal.

Where do we start? The criteria are the usual standards of industrial engineering—bottlenecks, operations offering the greatest cost savings per dollar expended, or trouble spots (safety, skill requirements, fatigue, tolerances, etc.).

On each proposed change we must ask, "What will it cost?" What can be recovered on old equipment? What indirect costs are involved—in engineering, product redesign, pilot line testing, market appraisal, sales promotion, tooling, training, relocation of labor, and interrupted production? Are adequate funds available? Are there better uses for those funds, uses that promise a more rapid or more certain return?

What risks are involved? Are there labor difficulties? Is the product ready for automation? Is the design thoroughly proved? Are there technological uncertainties? What is the risk of product obsolescence? Is the equipment convertible? What financial risks are involved?

What is the anticipated volume? Are there market uncertainties? What is the competitive situation? How long is the payoff period?

How do operating costs compare—for materials, labor, maintenance, power, space, overhead, spoilage, subsequent operations, and assembly? Are the necessary operating and maintenance know-how and facilities available?

What is gained (or lost) in terms of product value—attractiveness, quality, utility, durability? Are there intangible benefits—morale, safety, housekeeping, reduced supervision, competitive prestige?

There are many uncertainties, but we must be liberal in estimating the potential savings. Usually the payoff is much faster than expected. Rising material and labor costs have accelerated payoff in recent years. The indirect and intangible benefits typically add up to substantially more than the direct savings. In an expanding economy, a machine which was a substantial risk today may become an absolute necessity tomorrow. New and unforeseen uses and benefits often develop.

Carl Braun, a strong advocate of plant and equipment modernization, has expressed the economy of good equipment by comparing the benefits of a dynamic and progressive equipment policy with the hidden penalties of niggardly equipment policy. For every dollar saved by penny pinching on equipment, he estimated at least four dollars were lost in productivity, maintenance, materials, and overhead, with an even greater loss in intangibles.

We should ask ourselves such questions as "What is it worth to cut rejects in half? To cut tolerances in half? To have the operation paced consistently? What is greater safety worth? How much is it worth to get three times the production from the same floor area?" The first cost of automation is not important, and traditional ideals of utilization are not important. The important question is ultimate total unit cost.

Good automation does not cost; it pays. There is a challenge worth thinking about in the current advertisement advocating modernization, "Old equipment may cost a company nothing—nothing but its future."

Two very simple examples will serve to indicate the payoff of relatively simple automation. In the Berkeley plant of Magna Engineering Corporation a series of operations on the Shopsmith headstock (drill 9 holes, ream 1 hole, ream and counterbore 1 hole, tap 4 holes) were originally done on single-purpose tools—drill presses and tapping machines fitted with quick-change chucks, tumble jigs, and special fixtures. This tooling cost $5600 and machine time was 8 minutes per headstock.

A special six-station rotary table indexing machine to perform these operations was built for $12,150. The old equipment was sold for

$4000, leaving a net cost of $8150. Now one operator loads, unloads, and indexes the machine by pressing a button. Cycle time is 2 minutes per headstock. Labor saving, including overhead, is 47 cents per headstock. The equipment was completely amortized through labor saving alone on the first 17,400 units or eight and a half month's production. Thereafter annual savings have been $11,700. The new machine eliminated the need for a second shift, reduced space, scrap, and handling between machines, and produced a more uniformly machined headstock, thus improving product quality and facilitating assembly.

In a permanent mold operation making aluminum tables for the same product, two men were required to operate the mold, pour, pull cores, open, remove the casting, close, return cores to position, and lock. Only ten castings per hour could be made consistently. Under the new method the mold was automated for operation by one man who presses a button to start the automatic cycle. Output was increased to fifteen per hour. Mechanization cost $2300 and paid for itself in less than two months. Annual savings after payoff are $15,500, or over 600 per cent on the original investment. Mechanization of this operation reduced physical effort and decreased mold and core wear substantially. Castings were better and of more uniform quality because the automatic cycle was not subject to the discretion of the operator.

18.7 PROSPECTS FOR FUTURE AUTOMATION

What is the future of the automatic plant? We must try to be logical and realistic in appraising the prospects for automation and to avoid the temptation to overrate the possibilities. It is far easier to be a poor engineer than a good engineer, to find reasons why it cannot be done, rather than ways in which it can be done. We tend to think in terms of products and markets as they are today and find it difficult to see beyond to a greatly changed tomorrow. When we look around at Western plants, we see many where the volume, the type of product, the market, and the lack of finances make automation of any magnitude completely impracticable in the immediate future. For example, the president of a progressive company making highly standardized fiberglass-reinforced plastic products stated recently, "Why should we automate? Our labor is only 10 per cent of our costs. Show us how automation will reduce our materials consumption and we'll buy it."

On the other hand, when as a consumer we look at the articles we buy, we see many of them, such as pens, pencils, lamp bulbs, cigarettes, gasoline, and paint, which are already produced by processes that are almost completely automatic. We see others, such as hardware, clothing, toys, plumbing fixtures, plastics, radios, and television—now produced by semiautomatic processes—which offer great opportunities for further automation. When we look at these we lose any fear of overstating the possibilities.

Automation will not come overnight, for reasons we have already set forth. There will be many plants which will never enjoy extensive automation. Service and distributive industries do not lend themselves to automation. There will be many products which for various reasons will continue to be made by nonautomatic methods. And yet every plant will have some degree of automation, and we can expect that the vast majority of things we eat, use, and wear will within the next decade or so be made in plants where the only workers will be primarily technicians and not producers.

The subject of this discussion is the economics of plant automation. However, in making a general assessment of the economic prospects for automation, we must stretch our thinking far beyond manufacturing alone. The punched card, which may someday run our machine tools, has had its greatest development in the clerical field. In retailing, there is no reason why electronic equipment could not automatically record each sale, check the customer's credit, send out the bill, post the new stock balance, reorder if necessary, and give management a continuous up-to-the-second accounting of all operations.

Transportation offers tremendous opportunities for automation. The shortcomings of human control become critical with present-day air traffic congestion and supersonic speeds. Automatic electronic controls are not subject to such shortcomings, and it is to be expected that they will soon take over full control of aircraft in flight. The same may be true of rail, ocean, and even highway transportation. In freight yards, cars are today humped, braked, classified, and made up into trains by electronic controls. Computers handle plane reservations in large cities. Commuters can buy tickets in a four-second slot machine or may ride a moving sidewalk in railroad stations and airports. Telephone communications are largely automated today. A push-button warehouse has recently been announced.

The possibilities in agriculture are likewise great. We have witnessed an impressive mechanization in agriculture in the last thirty years. Many crops are already planted, cultivated, and harvested by machine. It is a rather simple step to automate these processes. It has been predicted that chemicals alone will displace more than one-third of our farm workers in the next twenty years. A West Coast electronic company has developed an electronic crop thinner which automatically detects and removes unwanted plants, four rows at a time.

The place of the small plant in a highly automated industrial society deserves serious consideration, for it would appear at first thought that the need for large capital investment and highly skilled management would of necessity lead to concentration of resources in large operations. In some industries this may be true, but there will be many others in which processes and equipment are simpler or demand is more limited, where small but well-integrated operations can better fill the need. We have already pointed out that automatic equipment need not necessarily be expensive or inflexible. Some of the automatic equipment described earlier, such as the Arma lathe, seem ideally suited to small business and will for the first time bring to small plants and intermittent production the opportunity of automation.

Every major mechanized industry—petroleum, automobiles, appliances, etc.—has a plethora of small companies serving the large. There will always be a need for small shops to install, service, and tool automatic equipment for larger concerns and to design and build special equipment. Small companies seem to thrive in a period of technological advance, and many of the significant steps in the development of automation controls are being taken by relatively small companies.

There will be local demands for foodstuffs, building materials, and other commodities which must be met by small local concerns. There will be situations where the imagination, vitality, flexibility, and lack of tradition and inhibition in small plants will offer an advantage over the rigidities and ponderous character of larger companies. In addition, there will be the service industries, those catering to the leisure market, and the distributive industries. On the whole, small business has little to fear and much to gain by automation.

18.8 IMPACT ON MANAGEMENT

To me the most exciting, the most challenging, and the least-explored aspect of automation is the impact of automation on management. We have already seen that not all plants will be automated to the same degree. But whether a plant becomes automated or not, it will face new management problems as a result of automation. Nonautomated plants must compete for the consumer's dollar and for labor with more automated industries. Companies whose direct competitors become automated will be forced to do likewise or fall by the wayside. And in companies where extensive automation takes place, many new management problems will be created.

It is not possible to discuss all the types of management problems that are likely to arise in nonautomated, partially automated, and fully automated plants. Therefore, some of the new problems that will be faced in the fully automated plant should be indicated. Each company may interpolate its position and problems, depending upon the degree to which automation may take place in its own industry.

The automatic plant will be relatively inflexible as to product and volume. Early writers envisioned factories equipped with tape-controlled automatons, replacing human workers and operating conventional general-purpose machines. It was suggested that if such a plant were producing vacuum cleaners and a shutdown threatened because of a shortage of electric motors, the plant could, with a little retooling and reshuffling of equipment, be converted to the production of motors until the shortage was remedied.

We have already discussed the flexibility of electronic control. Certainly many highly automated plants will be much more flexible than today's mechanized plants. Automation may even reach a rather high level in certain types of job shops.

But we can hardly expect the automatic plant to be an assortment of general-purpose automatic tools that can be changed over from one operation to another simply by changing tapes. In most automatic plants the special tooling—jigs, fixtures, dies—and single-purpose machines, inevitably expensive and suited only to the production of a single product, will severely limit the possibilities of change-over from one product to another. And in terms of layout, balance between operations, loading and unloading, and materials-handling equipment, the truly automatic plant will be strongly wedded to the single product for which it was originally designed. Thus design and product changes will be costly and flexibility with respect to product will be distinctly limited.

It appears likely that the automatic plant will also be relatively inflexible in volume output. Automation equipment is expensive. Because it is highly productive and can readily be operated around the clock, the output per dollar of investment may be very high and fixed costs per unit of output may be low. Nevertheless, fixed costs will always be high relative to variable costs. Labor costs will be fixed, because salaried maintenance, setup, engineering, and administrative personnel, in whom a large investment in training has been made, must be kept on the job as long as the plant is operating, regardless of the level of operations. And even materials cost will be relatively fixed, because long-term commitments for materials will be necessary to insure a continuous supply. The automatic plant need not accommodate itself to the convenience of a large work force, and competition will force it to operate around the clock. A plant operating twenty-four hours a day cannot expand its output in the event of an increase in demand; because of high fixed costs, it will be uneconomical to reduce operations in the event of a decrease in sales or price.

What does this mean for management? The function of long-range planning and the companion functions of coordination and control will be elevated to major importance. The automatic plant will require a highly sophisticated management—technically competent, decisive, capable of sound and realistic advance planning, able and willing to take substantial corporate risks based on carefully appraised facts, equipped with efficient administrative controls, and capable of better than normal coordination. The overall company point of view, the ability to coordinate operations as an integrated whole rather than as functional divisions of sales, production, and finance, will be emphasized, and functional lines will become more and more blurred. Company men, not functional specialists, will be needed on all levels.

The new responsibilities of the production manager are obvious. The automatic plant will be a highly integrated unit from incoming raw materials to finished product—vulnerable at any point. Emphasis will be on high utilization, with little time available for maintenance activities. Preventive maintenance will be a must. The normal pattern of plant operation may be reversed, with maintenance, tool changes, and service activities on the day shift and production operations on the night and swing shifts. Maintenance and standby personnel must be able to "double in brass," for the automatic plant must avoid the risks of human whims and frailties.

Administrative breakdowns, such as material shortages or faulty planning, will be as serious as mechanical breakdowns. A high premium will be placed on administrative competence, precise planning, effective control, and decisiveness. Relations with suppliers will take on added importance. Control over operations must be immediate and effective. Plant location will be less dependent on availability of labor, and we may expect to witness a continued decentralization to nonindustrial communities. Location need no longer cater to a pool of labor but to the living desires of trained technicians. A new type of engineer will be required, one whose training cuts across conventional lines—mechanical, hydraulic, electrical, and industrial—able to work as a part of a team, willing to take substantial responsibility, and able to make quick, sound, on-the-spot decisions. The production engineering function will also reach a new level. Methods study and problems of utilization of the human operator will tend to give way to problems of machine design, process improvement, equipment utilization, and maintenance.

Strangely, it may be the sales department and not the production function which will carry the major responsibility for the success of the automatic plant. The wheels will start to turn with the sales forecast and the overall operating plan—the responsibility of the top planning group. But on the sales department will fall the responsibility for developing and maintaining that market necessary to keep the automatic plant operating at an economical level. Full automation will be practicable only in plants producing standard products in large quantities. This constitutes another challenge to sales, gaining consumer acceptance of highly standardized products. In most cases products will have to be entirely reconceived in terms of the function they are to perform, and redesigned of new materials suited to automation. Once again this constitutes a burden on the sales department, which must overcome the consumer's natural resistance to radically new products. There are indications that seasonality of demand is gradually disappearing from the American business scene. This change must be hastened, for the automatic plant cannot be turned on and off with the seasons.

It is clear that the challenge to sales is as great as the challenge to production, and the competence with which the sales department can meet this challenge will determine to a large degree the success of automation in many companies and industries.

Automation will also bring many new problems for the personnel and industrial-relations group. It will be their responsibility to find, train, develop, maintain, and motivate a skilled work force made up primarily of technicians. Where will we locate such employees, what job satisfactions can we offer? What incentives? How will we measure the performance of such a group? What will automation do to the precious ingredient of leadership at the foreman level? What new industrial-relations problems will a shorter work week or a six-hour day bring? The problem of workers holding jobs and seniority rights in two plants is already a headache for management. A shorter work week will aggravate the problem. Will labor actively resist automation? Will increased pressure for the guaranteed annual wage come hand in hand with automation? What can be done to meet the problems of displaced workers? What seniority and mobility problems will arise? What new issues must be bargained during the transition period?

Automation will bring with it a tremendous increase in the bargaining power of labor, both because the automatic plant cannot afford work stoppages, and because as labor costs decrease as a percentage of total costs, management's resistance to wage increases and fringe benefits likewise softens. There is in this a real threat for nonautomated industries employing many men, for they cannot isolate themselves from the wage patterns set in automatic plants.

There will also be new problems for the financial executive, who must provide capital funds for these new facilities, give up the luxury of variable costs, adapt budgetary and financial planning to an operation dominated by fixed costs, and finance inventories through off-season periods in plants that cannot afford to shut down or even to reduce operations. The controllership function will continue to grow in importance. The trend to responsibility accounting by major cost centers rather than by operations or job lots will be accelerated.

What effect will automation have on industrial research? Will the investment in specialized production equipment and tooling "put a damper on" product and process research? There is much evidence that product change in some highly mechanized industries is seriously restricted by sunk costs. But can any firm isolate itself from technological change in a competitive market without becoming highly vulnerable to inroads by competitors or newcomers? In industries where technology is volatile, companies will be forced to write off equipment expenditures at a very high rate and include this fast write-off in their deliberations on the economics of automation, in their pricing, etc.

It seems clear that management responsibilities in all areas will be complicated by automation, that the future will demand a much higher level of management competence than required in today's competitive economy. There will be more decisions to be made, less time to make them, and the consequence of ill-considered action will be infinitely greater. Does this mean that today's harassed, overworked, hypertensive, insomniac administrator will find himself completely overcome by tomorrow's new and more exacting responsibilities? Must we wait for the evolutionary process to produce a superrace of managers before the automatic plant becomes fully practicable?

Historically the advance of technology has been a balanced advance on many fronts, with developments in one area providing answers to problems created in others. It appears that the same interaction may take place here as well—that automation itself may provide the solution to the dilemma of the increased management burden it creates.

In the first place, when production is automated, operations are converted from intermittent to automatic continuous production, and, thereby, thousands and thousands of day-to-day middle-management decisions, decisions on schedules, loading, work allocation, work methods, etc., have been erased. The die is cast, the gears of the automatic plant of necessity mesh together, and there remains in essence only the high-level decisions of how fast and how long they run. Even emergency action must be largely preplanned in advance. This is of course oversimplification, but the principle remains.

Secondly, managers of the automatic plant will have the assistance of electronic data-processing equipment, first cousin to the controls that will guide and control the production machines. These devices will play an equally important part in the automatic plant. For computers, coupled with new techniques of statistics, operations research, and linear programming, can help provide the answers that will help management administer the delicately balanced operations of the automatic plant. The potential use of computers in research, engineering, mathematics, and the sciences has been discussed. But has the contribution they can make to sales planning and analysis, procurement, inventory, and quality control been considered? Future business management can be greatly influenced by the availability of continuous, up-to-the-second information on practically every phase of the company's operation. No longer need we wait until tomorrow to obtain today's operating figures, nor need we wait until next month to determine this month's performance. In the automatic plant we cannot afford to wait. At any instant the computer can provide a statement of performance up to the second and, if desired, an extension to tomorrow, next month, or next year, compensated for seasonals, trends, competition, and other factors.

Businessmen flatter themselves today when they talk of taking "calculated risks." The phrase means only that we do not know all the facts but have considered the risks and unknowns and attempted to take them into account. Actually it adds up to a few incomplete facts plus considerable old-fashioned hunch and judgment. Tomorrow with the aid of the computer we shall really be able to calculate our risks and carry on a planning and control job worthy of the name *enlightened management.* There is a blessing here, but there is also a risk. Management can have so much information that it may not see the forest for the trees, may demand and try to use too much information, and may lose sight of essentials in the mass of detail. Self-discipline, an orderly sense of value and importance, and a firm grasp of the overall picture will be necessary.

18.9 SOCIAL IMPACT

There are other aspects to this matter of economics of automation. Industry does not operate in a vacuum. It operates in an economic and social environment which provides capital, materials, workers, and markets, and which regulates industry's activities whenever those activities operate to the detriment of the community. Thus, business must look not only to the impact of automation on its own internal affairs but also to the impact of automation on society. Business has very definite responsibilities to that society and must have answers to the questions and problems that automation will create.

Almost automatically, when we think of technological change, we think of unemployment. Some labor leaders have already resurrected the bogeyman of technological unemployment. It has been cynically suggested that displaced workers may even suffer the indignity of receiving their severance checks from a computer. I am certain the labor leaders know better. Their remarks are as stupid as the actions of Belgian workers who threw their wooden shoes or "sabots" into textile machinery and gave birth to the word "sabotage." But though these fears may be unfounded, they cannot be ignored. Management must recognize and deal with them. Automation cannot be rammed down labor's throat.

It certainly is true that in many cases the primary objective of automation is the elimination of labor. But will automation lead to widespread unemployment? Certainly not in the long run. Technology has always created more jobs than it has destroyed and has made possible a continually expanding economy which has absorbed more than it has displaced. But these hit-and-run conversationalists confuse the facts further with dramatic remarks such as, "We are not interested in the long run; we live in the short run; in the long run we'll all be dead."

We may be sure that labor will play automation for all it is worth at the bargaining table and in public debate. Those who wish to have a free hand in automating their own plants, who wish to develop and sell automation equipment must make it clear to all that automation is a great potential blessing which promises tremendous benefits for all mankind, for the worker and the consumer much more than for capital. Management must emphasize the promise of increased leisure, reduced prices, and improved standard of living, greater employment security, an economy of abundance, and the importance of automation in defense. It must be pointed out that technology does not eliminate jobs, it creates them; that a major strike in the automobile industry, for example, would put more people out of work than automation ever will.

Automation will not and cannot come as a tidal wave. It will develop at a disappointingly slow pace, one job, one department at a time. The techniques of automation are still only partially developed. Automation of even a minor segment of a plant requires much planning, product redesign, consultation with equipment producers, coordination with sales, construction, installation, debugging, etc. No existing plant will automate overnight. And new plants do not destroy jobs; they create new ones. There will be time for readjustment, time to find new jobs, time for retraining (a responsibility of labor as well as management)., and opportunities for normal turnover to absorb displacement.

Automation of clerical work may well be much more rapid than automation of industrial jobs, for the number of office workers has grown 60 per cent in the last ten years. Management has an insatiable desire for information—if it is accurate, timely, and can be obtained at low cost. Data-processing equipment can provide such information and may actually expand rather than contract clerical employment, just as mechanization expanded employment in the automobile industry. Furthermore, office work is in an area of high natural turnover, and a large displacement can be readily absorbed.

The total impact of automation on industry may be large in dollars and cents, but in terms of industry as a whole it will be rather small, at least in the immediate future. It has been estimated that less than 8 per cent of the work force is employed in industries where automation is technologically or economically practicable in the foreseeable future, and these industries are already largely mechanized.

Even when installed and working, automation will not eliminate all jobs in any plant. Phrases such as "the workerless factory," "machines without men," are figures of speech, figments of overstimulated imagination. It is estimated that for every ten workers replaced by automation, one good maintenance man must be added. Even in the automatic plant there will be many workers, technicians, setup men, programmers, tool makers, maintenance men, supervisors, and systems engineers. Our industries that have attained the highest degree of automation—automobiles, appliances, rubber, steel, etc.—are also among our largest employers. It is perhaps significant, certainly it should be reassuring to labor, that some of the old-time glass blowers, displaced by modern high-speed automatic bottle-making equipment, are today operating these same automatic machines and are of course earning much higher wages.

Labor has clearly stated its objective of a thirty-hour week. If they are successful, any possible decrease in the number of jobs will be spread among the work force as a whole. Furthermore a new leisure market will be created, a market for home furnishings, gardening equipment, sporting goods, entertainment, and travel, creating major expansion in these industries.

All these factors are minor when measured in terms of our rapidly expanding economy. Automation means a better product at lower cost. In this country that combination has always meant greatly increased sales. When Walter Ruether was facetiously asked how he expected to collect dues from automatic machines in a new automobile plant, he reportedly replied with the question how management expected to sell automobiles to them. Industry will not install automatic equipment unless it knows it can sell the product. One of the characteristics which distinguishes American management from management anywhere else in the world is its aggressive, imaginative, hard-hitting ability to sell.

All facets of our economy promise a tremendously expanded demand in the future—our universal desire for an improved standard of living, the increasingly large portion of our population represented by non-working, elderly people, the tidal wave of youth bursting the seams of our school systems, the defense program, and our commitments abroad. It has been estimated that just to maintain our present rising standard of living we must increase our productivity 50 per cent during the next ten years. In other words, we must step up our annual increase in productivity from its traditional 2½ per cent to 5 per cent. This hardly bespeaks widespread unemployment! Only through rather extensive automation can such a demand be met.

We cannot deny that some temporary unemployment will be created. We are living in a period of change, and this is one of the prices we pay for the benefits that change brings. Management must do its share to cushion the impact of automation, and if it wishes to avoid undue labor or government restrictions on its right to automate, it will make certain that automation is accomplished with the least possible disruption.

There is another aspect of automation with perhaps even greater importance, one of which management should make the most in its effort to gain recognition of the benefits that automation promises.

Too little attention has been given to the contribution automation can make to a fuller, richer life. Clearly it can bring a higher standard of living and increased leisure. But it can also increase job satisfaction by replacing monotonous, dull jobs with challenging, meaningful assignments. Mass production has been criticized for degrading the worker, for replacing skilled craftsmen with machine tenders. Norbert Wiener in his book *The Human Use of Human Beings* has commented, "It is degradation to a human being to hitch him to a plow and use him as a source of power. But it is an even greater degradation to assign him to a repetitive task in a factory which requires less than a millionth of his brain power."

It has also been said that "unskilled work is a mistake in engineering," that the most easily automated jobs are the repetitive, unskilled jobs. Automation could lead to "the human use of human beings." It could replace large numbers of unskilled workers on drab, monotonous jobs with highly trained technicians on challenging, responsible assignments of keeping the fabulous machines running. Thus the new technology may cure the evils of the old technology. In this sense, automation may be a reversal rather than an extension of the first industrial revolution.

This raises other pertinent questions to which management must find answers. We cannot deny that there will be some down-grading of labor whenever better jobs do not open or when workers cannot be trained for greater responsibility. What happens to those who are replaced and have neither the training nor the aptitude for higher-level work? Henry Ford I is reputed to have said that a whole stratum of humanity is unfit for anything but repetitive assembly line work. Even if Ford were right, there would be many semiskilled and unskilled jobs in nonautomatic industries.

The other side of the coin is equally unknown and perhaps more important to management. Will we be able to find the skilled technicians to supervise, service, tool, and maintain this highly complex equipment and adapt it to changing product demands? One large manufacturing company reportedly estimates that when its operations are automated it will have to step up its hiring of college graduates from 300 per year to 7000 per year! We already have a severe shortage of technical skill in industry. Certainly one of the great frontiers of psychology lies in improving our methods of determining human potentials. Thus far we have only scratched the surface. The experience of the military in selecting and training for highly technical jobs indicates that industry has a long way to go in this regard.

What will be the effect of automation on the business cycle? Will it reduce or aggravate the fluctuations in our economy? The evidence seems to indicate that automation will bring increased stability. It must bring stability, for the automatic plant cannot survive in a fluctuating economy. As we have seen, there will be few variable costs in the automatic plant and hence little incentive or opportunity to vary production in response to fluctuations of the market. Of course such cutbacks as may occur will be less likely to initiate a chain reaction of reduced purchasing power, for there will be few workers laid off. But the economic inflexibility of the automatic plant will necessitate investment and operating commitments which cannot be turned on or off in response to changing markets. We can today see evidence of the contribution that long-range business planning can make to a stable economy. Aided by the computer, we shall develop a level of planning skill and business predictability completely unattainable today. This all leads to the promise of an employment security much greater than that offered by unemployment compensation, make-work projects, guaranteed annual wage, or similar devices.

There has been much talk about the contribution automation can make to higher living standards and increased leisure. This presents business management and society with a question of real challenge. Will we have the intelligence, the character, and the incentive to use our added leisure and other benefits constructively, in ways that benefit and enrich rather than degrade and weaken? History teaches a sobering lesson: The more we have of creature comforts, the more we try to defend and protect what we have and the less we tend to pioneer and risk and innovate and strive for the ideals that led to our past great accomplishments.

19. The Future of Automation

DEAN E. WOOLDRIDGE

The Ramo-Wooldridge Corporation
Los Angeles, California

19.1 GENERAL PREDICTIONS

In a general way, it is not difficult to predict the future of automation. Ultimately, a major portion of the processing of data that now occupies the routine attention of many thousands of people will be accomplished automatically. Similarly, a large fraction of the routine repetitive activities of factory workers will be done by the use of the new automation techniques. Direct results of this spreading industrial revolution will certainly include a large increase in our national productivity. Indirect results will include the appearance of new products whose design and production are for the first time made economical because of automation. Hundreds of thousands of people will be employed by the companies engaged in the various ramifications of automation, with annual sales reckoned in billions of dollars.

Such predictions are easy to make, and they are certain to be true. Today the handwriting on the wall is so clear and unmistakable that it allows for no misinterpretation so far as the ultimate future of automation is concerned. What this handwriting on the wall does not spell out so clearly, however, is how long it will be before this bright future of automation is actually realized. Will most of American industry be well on its way to realizing the benefits that can arise out of automation techniques ten years from now, or will it be twenty years before the full effect is felt?

Certainly we will all agree that predicting the time scale for the development of a new technological or scientific trend is a different kind of business from predicting what the trend may ultimately lead to. We are all familiar with the concept of the dispensability of individuals in the development of new ideas. We know from history that when a field develops to a point where it is in effect ripe for new ideas and techniques to come out of it, frequently the new ideas and techniques appear to be discovered, almost simultaneously, by entirely unrelated workers in the field. However, from the perspective of the historian, the almost simultaneous discovery of some new idea may mean only that the two independent discoverers thought of the same thing within an interval of ten years or so. Although this interval is often insignificant when history is being written a few hundred years later, it can be of great concern to the participants in the making of history. We can be certain that if Einstein had not turned up with the theory of relativity, including the energy-mass equivalence, when he did, someone else would have developed an approximately similar theory before now, but it is quite possible that there might have been a delay of a dozen years in the process. An incidental result of such delay, if it had occurred, could have been that the atom bomb would not yet have been invented. Although this might or might not prove significant to the historian who writes up the activities of these times two hundred years hence, it may be that, from the narrow points of view that motivate most of us, such a time lag would be not at all insignificant.

And so it may be with the field of automation. Although the ultimate results are predictable, the rapidity with which we approach those results are likely to be materially affected by the activities, the decisions, the mistakes, and the successes of individuals, groups, and companies that are today in the field. If they choose improper methods of approach to the very great problems that exist in this field, it could well mean a delay of fifteen to twenty years before a given level of automation effectiveness is reached.

There may or may not be general agreement on this philosophy, but as long as this remains just philosophy, it is not likely to be very useful in helping us plan our course for the future. What we must do to turn this theorizing into useful channels is to see if we have any way of differentiating between approaches to automation that will lead to rapid progress and approaches that will result in delaying the attainment of the desired goals.

This, then, is my objective—it is not to discuss the future of automation as such, but rather it is to explore whether there are any general methods of approach that can be taken by individuals and organizations in the field that will promote the growth of automation more effectively than other methods of approach.

19.2 CHARACTERISTICS OF AUTOMATION SYSTEMS

We should note that the most effective performance in the development of a new field is usually exhibited by those who best employ lessons learned from past activities of a similar character. Therefore, let us start by inquiring whether there is any field similar to automation for which a past history of development exists from which we might hope to learn some useful lessons. What do we mean by a similar field? What are the distinctive characteristics of automation? It is apparent that one important feature of the field is that it is based upon the development and use of very complex equipment which, although not entirely electronic, is nevertheless centered upon electronic devices and electronic techniques. A second vital aspect of automation is that the equipment is rarely, if ever, an end-item in itself, such as a radio or an automobile, but usually appears as a tool or component that is integrated into a complex system involving men, machines, methods, and procedures—all organized to perform some basically nontechnical operation.

19.3 SIMILARITY OF MILITARY ELECTRONICS AND AUTOMATION SYSTEMS

Is there any field having similar characteristics that has already undergone extensive development? Indeed, there is one—it is the field of military electronics. There, too, the equipment, although by no means entirely electronic, is based upon electronic apparatus and electronic techniques. Like automation systems, military electronic systems are exceedingly complex. In fact, many of the specific techniques and apparatus components being used today in automation were developed in military electronic activities. Not only with respect to the nature and complexity of equipment, but also with respect to the type of operations in which it is used, there exists a great similarity between military electronics and automation apparatus. In each instance, the complex equipment must be operated by basically nontechnical people to perform as a part of a very intricate and extensive system of human organization, machines, methods, and procedures, to accomplish an overall result. In each case, the design of the equipment is intimately associated with and affected by the surrounding complex of men and procedures.

Let us analyze the lessons that have been learned in recent years in the development of military electronic systems, often at great expense. Automation may well be in the unique position, for a new field, of being able to draw on a reservoir of immediate past experience comprising the activities of thousands of people and the expenditure of billions of dollars.

19.4 MILITARY ELECTRONICS

Let us therefore now direct our attention to military electronics and how it has grown since the beginning of World War II. There were, of course, many applications of electronics in the last war. However, it is well known that the area of greatest electronics activity, and therefore of greatest long-range influence on the field of electronics, was that of radar. Project activities were carried out in a number of universities, and a great many industrial organizations quickly put large crews to work to develop the techniques that ultimately led to modern radar. And this equipment was, of course, used quite extensively by all branches of the Armed Services—on the ground, on the sea, and in the air. As we all know now, it proved to be sufficiently useful that it played, if not a key part, at least an important part in a wide range of military activities. Therefore, radar would appear to be a good subject to study in looking for lessons learned in military electronics that might be applicable to automation.

As a vehicle for developing the points of interest, let us consider the field of airborne radar. The first airborne applications of radar were in bomber planes to provide the navigator with means for obtaining position fixes at night or above the overcast. There were also early applications of radar equipment to night fighter aircraft to make it possible for the pilot of an interceptor plane to locate an enemy bomber at night and to approach close enough to have a chance of seeing the bomber plane outlined against the night sky, thereby permitting a gunfire attack.

19.4.1 The Black-Box Approach

Once radar equipment had been put into aircraft to facilitate navigation to the general vicinity of the target, it was natural that someone should propose a combination of radar with bombsight or fire control equipment to permit actual blind attacks to be made against enemy targets, rather than just the accomplishment of navigation to the vicinity of the target. The military organizations approached this matter in what was to them a normal way. They employed the black-box method. For each weapons project, they assigned to a suitable officer and staff of the military organization the systems management responsibility. In turn, this military project team made arrangements for a suitable contractor to provide what was to them simply a black box—a radar designed to be suitable for use with the bombsight or fire control equipment currently in use by the service in question. The theory was that the radar equipment would be delivered, designed to military specifications supplied by the customer, into the hands of the military project organization, and would then be combined with the usual bombsight or fire control computer, installed in the necessary airplanes, and connected together by military people to provide the hoped-for capacity for all-weather operations.

But the black-box theory did not work. It was not that the radar equipment did not arrive more or less on schedule, for it did. Also, the size and shape of the black box worked out pretty well so that it fitted together physically with the computer. Furthermore, the plugs and interconnecting cables required to hook the two pieces of apparatus together mated pretty well. The only trouble was that, when put together, the radar and computing equipment generally interacted upon one another in obscure ways that had not been anticipated. There were a number of reasons for this. As only one example, there were important differences in the character of the target data supplied by a tracking radar and the target data supplied by optical tracking, with which bombsights and fire control computers had previously been used. The angular accuracy with which early radar equipment determined the position of a target was somewhat less than the accuracy with which a human observer could set cross errors on a visual telescope image of the target. This, of course, surprised no one. However, what was not so well anticipated was a set of subtle effects on the performance of computers that arose out of the "jitter" or random variation in target data, owing to the basic characteristics of radar. These properties of the radar data in some instances rendered completely inoperable computers that had performed very satisfactorily in their past optical applications. In other instances, the computers worked, but with a disappointingly low level of accuracy and effectiveness. Such matters as these are well understood now, but in the early days of military electronics they often came as dramatic and disheartening surprises to the military project people who sometimes abruptly discovered that in practice their hoped-for superior, advanced weapon operated in a very disappointing fashion.

19.4.2 Systems Integration

But our military people are not, and were not then, slow to learn. Therefore, it did not take them long to recognize that their basic mistake had been a failure to appreciate the fact that a radar bombsight is not obtained by adding a radar black box and a bombsight black box but rather must arise out of a single integrated development in which the design of the radar is intimately affected by the bombsight into which it is to work, and the design of the bombsight is strongly characteristic of the radar that is to supply it with data. In other words, a highly technical systems integration activity was required that had been missing in the early attempts by the military to provide themselves with advanced weapons systems. The solution to the problem required that the well-meaning, but technically unprepared, military project people get out of the coordinating business and assign to technical people the overall responsibility for the design of a single, integrated radar-computer combination. Here were the ingredients for an interesting industrial competition. Which type of organization was to be given systems responsibility—the computer company or the radar company?

At this stage, early in the last war, computers for military application were primarily mechanical in nature, were designed by people trained largely as mechanical engineers and mechanical designers, and were built by companies specializing in mechanical work. The newer radar techniques, however, were electronic in character, and the participants in the development of these new techniques were physicists and communication engineers. It might have been thought that there would ensue a long and possibly even bitter tug of war between the mechanical and the electrical groups for supremacy. Therefore, it was particularly interesting, and certainly of significance for the future of automation, that almost before the lines of battle could be drawn, the competition was over, with the physicists and electrical engineers left in clear, undisputed possession of the field. As is usually true of such decisive engagements, there was a very good reason for the one-sidedness of the victory. It arose out of the fact that, in this country at least, the educational training and subsequent professional experience of physicists and communication electrical engineers is much richer in scientific fundamentals than is the training given to mechanical engineers. As a result, electrical engineers and physicists are better suited to the solution of new types of previously unencountered problems than are mechanical engineers. This is particularly true of those physicists and electrical engineers who have had graduate training from our better institutions. It came much more naturally to these people to deal concurrently with the combination of radar and computer problems. Hence, no competition; the newer crop of electronics people took over systems responsibility.

It was in this way that most electronics people obtained their early indoctrination in the field of computing in the 1940–1945 era. In the process, not only did they learn to be ingenious in the invention of techniques by means of which electrical components could be caused to perform mathematical operations, but also, working with mechanical engineers as part of the team, they learned to understand the mechanical techniques that had been employed for this purpose in the past and to take them over and use them in conjunction with electrical techniques whenever the mechanical techniques were superior. To some extent, therefore, they ultimately absorbed what might have been their competition.

The impression should not be given that the companies that prior to the last war had been successfully engaged in the development and manufacture of mechanical computers for military applications all folded up and went out of business. In general, this did not occur. Frequently the established organizations saw the handwriting on the wall and began to add to their technical staffs considerable numbers of physicists and electrically trained men. It is a tribute to the management of these organizations that some of them were sufficiently fast on their feet to emerge after the war in a position almost as strong in the now electronically dominated field of weapons systems as the position they had held earlier on the basis of superior mechanical design competence.

To illustrate the point, for one military application the government let contracts for the development of a certain kind of airborne computer to two different companies, one traditionally known as a mechanical company, the other as an electronics organization. Both companies developed satisfactory computers performing roughly the same functions in the same types of airplanes. Both companies were given production contracts against which considerable numbers of computers were delivered and ultimately installed in military aircraft. When a comparison of the two kinds of equipment was made, it was discovered that the computers built by the electronics firm contained more gears than the computers built by the mechanical firm, and the computers built by the mechanical firm contained more vacuum tubes than those supplied by the electronics organization. The real point of the story is that, by this time, teams of physicists and electronics and mechanical engineers were in control of the systems development activities in both companies.

But the addition of a computer to a radar was only the beginning of a trend toward a general broadening of the scope of the technical assignments given to the designers of military electronic equipment. Another early step in this broadening process, in the instance of the bomber, was to employ the output of the computer to manipulate the control surface of the aircraft and thereby fly it automatically on a correct bombing course, rather than simply to display to the pilot steering instructions, as had been done in the past. In the instance of the interceptor plane used to attack enemy bombers, the analogous development was to use the output of the fire control computer to steer the plane on a proper gunfire attack against the enemy bomber. This type of problem led electronic engineers into such areas as the aerodynamic response characteristics of aircraft and the dynamics of flight. And so, as time went on, the trend in military electronics was to assign to the electronics organization more and more responsibility for the design or at least the integration of the growing numbers of increasingly complex equipment subsystems that were required to work together to produce some desired operational result.

19.4.3 Weapons Systems Concept

Along with the trend toward the broadening of the responsibility of the electronics development organization to cover other technical areas not previously associated with electronics, a concurrent trend developed toward expanding the responsibilities into nontechnical areas. Just as technical considerations led to the necessity of treating the radar and the computer that went with it as a single piece of equipment rather than as two separable black boxes, so, as the technical responsibility of the electronics development organization expanded to include entire weapons subsystems, such as radar navigating and bombing equipment or interceptor fire control apparatus, it was also necessary that the organization become more deeply involved in military operational considerations of a basically nontechnical nature.

It is important to understand how far this trend toward broadening responsibility has progressed in the field of military electronics. To illustrate, let me give some examples of problems that must be dealt with today by the members of a team charged with the responsibility of developing an electronics system for use in interceptor aircraft to assist in the shooting down of enemy bombers. Such problems include the following:

(1) How many bombing planes are likely to come in at one time? At what altitude? What kind of maneuvers will they perform? What countermaneuvers will be required of the interceptor for successful attack? Such matters as these can have a major bearing on the electronic computer designer's choice of the type of attack trajectory that is to be specified for use by the interceptor plane.

(2) What kinds of manipulations can the occupant of a supersonic aircraft really perform? One of the worst mistakes the designer of equipment can make is to impose upon the occupant of an aircraft operations he cannot really accomplish successfully, particularly under the stress of battle.

(3) How long does it take the pilot of a defending interceptor to get into his position in the plane, take it off, and get to altitude once an alarm has come in? This determines the time available for target kill and influences the choice of trajectory of attack to be designed into the computers.

(4) What will be the numbers and the training of the maintenance men the customer will be able to provide under battle conditions for keeping the electronic equipment operating? Equipment that may exhibit unusually high performance when maintained by men with doctorates in physics or engineering may in practice be quite inferior to simpler equipment of much lower ultimate performance if the maintenance staff is to consist of one high-school graduate.

And so it is that an almost indefinitely long list could be provided to illustrate the extent to which organizations today, successfully engaged in military electronic systems development, have been required to broaden their scope, first to include a wider variety of technical assignments than used to be considered appropriate for electronics organizations, and then to accept the responsibility for a wide range of concurrent, nontechnical matters. This development of the military weapons systems concept, important as it is, is not very widely understood today. Oddly enough, although this development has come about as the direct consequence of World War II, the principal strides made by the military in establishing suitable organizational and contractual arrangements for handling their difficult weapons systems problems have been made since the end of the last war. It is no accident whatever that the really giant strides in the applications of the newer electronic techniques to the science of warfare have also been made since the war and have been almost invariably associated with projects in which organizations, primarily electronic in character, have been given strong positions of systems leadership.

19.4.4 Summary

All this recent history may be summarized in the conclusion that the key to success in military electronics has been established to be the employment of a proper systems approach in the development work. The basic difference between improper and proper systems approaches has been illustrated as follows:

(1) The black-box approach to military systems is the improper approach and does not work. In this approach the customer attempts to specify to outside contractors the characteristics of equipment subsystems that it is hoped can be brought together with other subsystems by the customer to perform some overall operational function. In the complex weapons systems of today, this approach does not work because of the almost inevitable existence of a wide range of subtle, technical interactions among the various subsystems. As a result, the nontechnical military customer does not have an adequate degree of scientific or engineering competence to perform the systems integration function.

(2) The proper approach, which *does* work, is to assign systems responsibility for the design and integration of the various subsystems required to compose a complete operating weapon system to a single group of competent, versatile scientists and engineers, including physicists and electronics engineers, because of the superior training these men receive in the fundamentals of science and in the solution of new problems. Proper exercise of the systems management responsibility results in the development by such groups of an unusual ability to deal concurrently and quantitatively with a variety of nontechnical operations, as well as technical matters. This is the way in which a new and advanced system, which may employ the most complex equipment and technical methods, can best be made to fulfill the actual objectives of the development, which are generally operational or nontechnical in basic character.

To summarize, our government, in the last few years, has supported a gigantic experiment, costing hundreds of millions of dollars and involving many thousands of man-years of effort, testing these two approaches to weapons systems development. It would be tactless, of course, to enumerate the inefficiencies and failures that have characterized the first approach, and to show in detail how the outstandingly successful programs have always been characterized by the second approach. However, the record exists and is incapable of misinterpretation. In the military systems field, one approach is poor, the other is good!

19.5 CARRY-OVER FROM MILITARY ELECTRONICS TO AUTOMATION IN BUSINESS AND INDUSTRY

In showing the unsuitability of the black-box approach, examples from airborne bombing and fire control were used to illustrate one of the kinds of interaction problems that so frequently arise. Attention was called to the very extensive, and at one time unsuspected, ramifications in the design of the computer equipment that are required because the statistically fluctuating character of radar data is so different from the character of optical data for which computers had formerly been designed. There is a good analogy to this kind of problem in the field of automation.

An excellent example is in the area of factory production control. It has to do with a problem to which automation experts sometimes address themselves—that of devising a combination of methods and electronic calculating equipment to derive machine-loading information that will permit the scheduling and use of the various machines in the shop so as to obtain the maximum utilization of men and machines compatible with meeting customer delivery dates. As all who have had occasion to tackle this problem know, optimum machine loading is by no means easy to obtain in a factory that makes a wide variety of items. However, any set of methods and equipment devised for "automating" the scheduling and assignment of machines must use various types of input data related to the availability of inventory supplies, the number, types, and delivery dates of customers' orders, and the like. Usually, some of this material is provided by the tabulating department of the company. This tabular information comes in the form of lists, reports, and summaries of various types, derived from inventory or customer order data.

The black-box approach as applied here would be one in which company management would stipulate to the group engaged to develop automation methods of machine scheduling a scope of responsibility that did not include the source and accuracy of the tabulating data, on the theory that this is a separate problem to be dealt with, if at all, by others. This can be recognized as similar to the black-box approach described in connection with the development of radar for combination with computing equipment to solve bombing and fire control problems. Just as in the military case, with such an approach to machine utilization, automation in a factory is likely to run into serious difficulties, and for much the same reason as that which lay behind the problems of the black-box approach to radar-computer combinations. Here is the trouble—or here at least is one kind of trouble that can arise. The tabulating data bearing upon such matters as the number of items in stock, purchase orders to be filled, and the like, are frequently based upon counts and checks that have been made considerably in advance of the time the final data arrived in the hands of production control for machine-scheduling use. It is not at all uncommon in a big factory to find that some types of tabulating data, by the time they reach the production-control people; are so badly out of date as to be quite unsuited for certain kinds of control purposes. How serious such inaccuracies may be depends entirely upon the use to which the data are put. Frequently factory people work out their own practical solution of this problem by trial and error methods that ultimately result in control procedures which, consciously or unconsciously, place a minimum of dependence on the large time-lag data. Under such circumstances, the introduction of an entirely new set of methods and procedures accompanied by sophisticated electronic equipment for bringing automation into the machine-loading problem might result in such a different use of the tabulating data that time-lag effects which previously gave no trouble would cause the new system to fail completely to meet its objective.

There is an interesting similarity between the weakness of the black-box approach in factory production control and the weakness of the black-box approach in a radar-computer combination. In each case, problems arise out of the unsuspected existence of certain inaccuracies in the basic input data. In the radar-computer combination, the inaccuracy stems from the characteristics of tracking radar. In the factory example just used, inaccuracies arose out of the methods of collection and processing of the tabular data. The conclusion is the same in both cases: it is risky to design a black box that is to be used as a portion of a larger, complex system of machinery and procedures, unless the design of the black box is accompanied and governed by a thoroughgoing analysis of the entire system within which the black box is to function. In other words, the black-box approach is as unsuited to the needs of automation development as it is to military weapons systems.

The apparent simplicity of this conclusion should not be allowed to conceal its very practical implications. For example, this principle asserts that rapid, efficient progress toward factory automation will not result from the narrow approach of developing automatic equipment to convey material from one machine to another unless the design of the material-handling equipment is associated with an analysis that shows it to be compatible with the requirements of the newer machines and methods of fabrication that may shortly be required as a result of an overall systems approach to the entire problem of the automatic production line.

As another example, rapid progress is not likely to result from a program to develop equipment and methods that will permit a factory to maintain its inventory at the lowest possible level, unless this work is integrated into a well thought-out program that takes proper account of the other closely related factors, such as machine utilization and delays in customer deliveries.

There is an unspecified implication in the examples just advanced to establish the unsuitability of the black-box approach. As in the military experiences, it has been assumed that the customer himself is unable to do the systems integration task and thereby properly supervise and control the work of the development team. Is this a sound general implication? That conclusion was reached in the military situation because of the technical complexities that were mixed up with the operational problems. In most cases the business and industrial customer is equally unsuited to these very analytical and technical tasks, if by the customer we mean the operational people within the company in question for whose use the automation methods and equipments are designed. More specifically, the man who has grown up as an operating expert in the field of production control in a large and modern factory is not likely to be equipped to determine the combination of procedures and electronic techniques that should be used to bring to his factory the fruits of automation. The field is much too technical and too complicated for that. The production control expert of the factory he is associated with, unfortunately, needs also to be one of the world's best-informed men with respect to the characteristics of modern electronics computing and data-handling equipment and methods. This he is not likely to be or to become. Just as in the case of military electronics, it is necessary for some one person or some one small group to encompass both the operational and the technical factors involved. Experience purchased by our government at the cost of many millions of dollars has shown that the approach that works is to give the systems design responsibility to the technical man and be patient while he learns what he needs to know of operational nontechnical matters. The other approach is rarely successful in fields such as military electronics or automation where the scientific and technical content is so very high.

The factory expert must, of course, work closely with the automation expert. If the problem in question is the design of a new data-processing system for automation of the production control functions of the factory, the factory experts should not be shoved aside and ignored by the team of physicists and engineers who set to work to devise a new system. This would lead to exactly the same kind of useless result as would be secured in a military systems development if the technical team were to ignore the military operational people in choosing their technical method of approach to the development of the military system. Although the customer must not himself attempt to perform the highly complex systems design job involving technical and operational factors, it is, of course, the customer who must supply to the scientific systems team the basic operational data that serves to define the problem. And if the scientifically dominated systems development team is really to be effective, its key members must have a breadth that makes it possible for them to talk to production control experts and develop an adequate understanding of their problems. Each automation systems development project, just like each military weapons systems development project, requires the establishment of extensive and detailed communication between the basically technical project team and the basically nontechnical operational people. Military experience has shown that such interrelations *can* be successfully established, and that basically technical organizations *can* develop project teams that, although scientifically dominated, demonstrate great aptitude for absorbing and properly integrating the operational problems into their considerations. The factory expert should have something to say about the development of a new production control system; but the factory man must consider himself to be primarily a source of input data to the project team, and he must recognize that as long as the field of automation involves as much technical complexity as it surely will for the next ten years, the people who do the constructive and inventive work will not be production control experts but instead must be men trained as scientists and development engineers.

And so, with abundant use of analogies from military experience these are the principal points of this discussion—the importance of a properly integrated systems approach to the automation of business and industry, and the key role that must be played by scientifically dominated teams in the control and management of such systems development projects.

These points have been emphasized heavily because they are vital in the future of automation, and also because it is not at all clear today that business and industry are aware either of their importance or of the great pertinence of the military electronics experience. There are many examples today in the business and industrial approach to automation of the employment of methods that have been more than amply tested, unsuccessfully, in weapons systems work. If this approach becomes prevalent in the automation activities of the next few years, then American progress in achieving the benefits of automation will be markedly slower than it needs to be. Certainly a ten-year difference in the achievement of a high level of automation can easily be the result if these matters are not properly understood and attended to by business and industry.

However, let us not view the future with alarm. In fact, there are some trends today which, if properly nourished and extended, could result in an adequate level of systems competence in the development of automation during the next few years. The rest of this chapter will be devoted to an appraisal of these trends.

19.6 APPRAISAL OF SYSTEMS APPROACH

If my arguments to this point have been convincing, then we should all be agreed that rapid development of automation requires that the solution of each factory control or data-processing problem in business or industrial establishments be governed by the results of a systems analysis treating all company operations capable of exerting a major influence on the choice of methods, procedures, and equipment best suited for the solution of the problem under attack. Furthermore, we should also be agreed that the extensive technical content of such systems work requires that it be accomplished by teams dominated by professionally trained scientists and development engineers, who, however, have also acquired a capacity to deal effectively with concurrent nontechnical operational matters. Now these conclusions appear to specify pretty completely the manner in which any company interested in the introduction of automation into its operations must go about accomplishing such a result. The procedure indicated is for the company to give a systems development contract to some organization with a high content of scientific and development engineering competence, and with an established record of performance in the design and development of electronic systems. If such arrangements are supplemented by the establishment of a proper degree of intimacy between the systems development group and appropriate operational people within the customer company, favorable conditions should exist for the ultimate determination of a combination of methods, procedures, and apparatus to bring to the customer the maximum quantity and quality of practical automation consistent with the state of the art.

However, there is a difficulty sometimes associated with the implementation of this approach to automation. This difficulty arises out of the fact that the thoroughgoing scientific systems analysis and equipment design procedures upon which the approach is based can frequently be a very expensive undertaking. The proper solution of a major systems development problem may well require a technical team of a hundred or more people working for two or more years. The price tag for such a program may run to several millions of dollars. So, although the procedure described is undoubtedly the proper approach for General Motors to take in bringing automation into its production lines, or for the Bank of America to employ in the development of mechanized data-processing procedures—where the size of the business and the ultimate benefits arising out of a properly designed system will recover many times over the cost of the development effort—such an approach may be out of the question for small or medium-size businesses. Not only do we not want to ignore small or medium-size businesses, but as a practical matter automation certainly will not really have arrived until it is suitable for this scale of industrial activity.

Certainly there is a practical answer to this problem. To develop it, we must start with the observation that there are two different ways in which a systems team can go at the problem of developing a combination of methods, procedures, and equipment for the solution of an operational problem. The method already discussed is one in which the development team undertakes to design new equipment to accomplish the required purposes. In this type of program, the development team usually limits itself to the employment of electronic and mechanical components, vacuum tubes, servo motors, semiconductor devices, memory elements, and the like, that are currently available, but attempts to devise ways of combining these components in circuits and electromechanical arrangements that may result in devices and subsystems quite different from any that have appeared before. When this kind of invention and design process is carried out with proper regard for the nontechnical operational objectives of the program, the most effective results are obtained. It is this kind of work that the major business and industrial establishments should support in order to obtain the maximum benefits from automation.

19.7 BUILDING-BLOCK APPROACH TO AUTOMATION

There is, however, another approach to the same kind of problem that can be employed by the systems development team. In this approach, the team limits itself to the employment of complete devices or subsystems that are currently available, rather than to the employment of electrical and mechanical components currently available. For this kind of approach, the "building blocks" with which the development team works, instead of being transistors, vacuum tubes, motors, potentiometers, and the like, are electrical printers, arithmetic and control units, card readers, magnetic-tape files, and the like—assuming that units with such titles are currently available for purchase or lease from industrial suppliers. The "building blocks" with which the development team works are numerous and elemental, in the one instance, and few and complex in the other. However, the analytical and logical problems involved in the design of a suitable system of automation for a business or industrial establishment are much the same in the two instances, although physically the results might appear to be quite different, because in the first case new and extensive apparatus results, whereas in the second case the output of the systems approach is a combination of existing pieces of hardware.

Now, obviously, the first method of approach to an automation problem is a more flexible one and allows the development of equipment and procedures that are more precisely suited to the needs of the business or industrial establishment under study. In fact, the second method of approach is not practical at all until there has been enough development of the first type to provide on the market a suitably wide range of devices and subsystems of adequate flexibility to permit combinaton in various patterns to perform a variety of automation tasks. Ultimately, however, it is to be expected that this second condition will exist as more and more organizations get in the business and a wider and wider range of automation problems achieve solution.

Should a banking institution, for example, sponsor the development of automation for its own uses, it is quite likely that some of the resulting methods and equipment would be at least partially applicable to the activities of other banks enough smaller to be unable to sponsor the pioneering work themselves, but large enough to have similar operational problems. Since American industry never works in such a way as to exclude for very long the participation by smaller members of an industry in the advanced developments of their larger colleagues, we may safely expect that ultimately the pioneering work done by the large organizations will extend benefits in this way to other members of the same industry. This process will undoubtedly be helped along by the companies that design and manufacture the equipment. Their own interests, of course, will be best served if they can ultimately obtain a wide market for their products. Therefore, whenever they can do so without compromising the automation system that they are initially employed to develop for the major customer, they will choose methods of packaging the subsystems that go into their final system, and of standardizing the methods and procedures in different forms to meet the approximately similar requirements of other customers.

And so we can expect to find, as time goes on, that there will be developed a number of basic lines of automation, each line composed of equipment and methods and procedures in combinations that will make it possible to handle the automation problems of a class of business and industrial customers. It is not possible today to predict with any certainty how wide a range of customers may ultimately be served by any of these major lines of automation. Perhaps there will be a set of machinery, methods, and procedures that will be applicable only to the banking business, another set of machinery, methods, and equipment applicable only to the insurance business, another for department stores, another for automobile factories, and the like. Of course, companies in the automation business will prefer and will continue to work toward lines of equipment at least, if not methods and procedures, that can be employed across the board by all business and industrial establishments. Undoubtedly the final result will lie somewhere between these two extremes. There will be certain kinds of machinery that are applicable to a variety of data-processing and factory automation problems wherever they arise. There will be other types of machinery that will be more highly specialized to one industry. The same will be true of the methods and procedures that in each instance must work intimately with the machinery in order to accomplish the overall operational result.

But, whatever the final pattern, it is clear that the growth of this complex of methods and machines must stem from the large-scale systems study and equipment development programs, sponsored by major establishments, that will determine the basic character of the available equipment subsystems and methods. Solution of the automation problems of a large number of smaller customers will then follow, guided by the second kind of systems development approach, in which the design materials employed by the development scientists and engineers will consist of the available subsystems, rather than elemental electrical and mechanical components.

From the point of view of the smaller business and industrial establishment, this means of extending the applications of automation equipment is of great economic importance. The controlling factor is that by all odds the major part of the development expense is associated with the actual invention, testing, and reduction to practical form of the design of the devices and subsystems that make up the final equipment. The approach that employs existing equipment, therefore, although depending just as much on a high quality of scientific systems analysis, is much less costly and, if the problem under investigation is fairly similar to other problems for which major and expensive hardware development programs have been accomplished, then it is not unlikely that there will already exist devices and subsystems that can be combined to give a fairly satisfactory solution to the automation problem of the smaller business or industrial establishment.

19.8 OPERATIONS RESEARCH IN AUTOMATION

Finally, it may be interesting to note that the second kind of system development activity, although essentially not very different from the first kind, goes by an entirely different name. The term that has come to be applied to this kind of work is "operations research." More generally, operations research is a term used to describe the activities of scientifically trained teams who undertake the assignment of analyzing broad areas of operation of business or industrial establishments, to determine procedures and methods of making decisions that are most effectively geared to the basic objectives of the business. When, as is so frequently the case nowadays, the scope of assignment of the operations research team includes consideration of the possible employment of the newer techniques of automation to improve the efficiency of the enterprise, then the operations research assignment is indistinguishable from the integrated systems development operation, except that such an assignment is usually of the second type, where the team does not design new apparatus but instead limits itself to designing combinations of available equipment that can best meet the needs of the business in question.

It is no accident that operations research activities are beginning to appear and grow rapidly in various establishments throughout the country at about the same time that development of automation equipment is really getting under way. In each instance the principal work is being done by scientists and engineers who have learned their trade during and since the war in military electronic systems activities of one form or another. Some of these men have gone to organizations where they can use the detailed knowledge of electronic and electromechanical techniques, which they acquired while working on weapons systems, in the development of new kinds of automation apparatus. The others have made their way to operations research groups, where they find a direct carry-over to the business and industrial world of the analytical methods and the systems approach that they found so productive in military work. The rapid development of automation requires and will make use of both types of activity. As time goes on, we shall see a great expansion in the scientific and engineering staffs of the companies that engage in the development of the equipment of automation. In the companies that are successful, we shall also see an increasing dependence upon thoroughgoing systems analysis of the problems of major business and industrial establishments as the basis for equipment design. Concurrent with these developments, we shall see a rapid increase in the numbers and quality of operations research teams throughout the country, which, as time goes on, will carry a larger and larger fraction of the load of bringing to smaller business and industry in particular the combinations of methods, procedures, and equipment that will bring about the benefits of automation in the most economical and efficient manner. In each type of activity, principal dependence will be placed upon teams of scientists and development engineers, led by men who, in addition to superior professional attainments, will exhibit unusual breadth and capacity for comprehending and dealing simultaneously with nontechnical as well as technical problems.

19.9 CONCLUSIONS

The reason for this emphasis on the systems point of view is that experience has shown there is a certain subtlety about it that is likely to be overlooked. There is an illusory sense of simplicity about systems development work. There is something very logical and appealing about the idea of assigning to the production control organization of a factory the responsibility for working out and specifying the kind of automation equipment and techniques that should be employed to improve the efficiency of the production control operations. There was, of course, a similarly incorrect sense of fitness to the early military decision to assign to bombing and navigation experts the responsibility for working out and specifying the kind of equipment and procedures needed for the performance of improved bombing and navigation functions. The significance of the military electronics experience and its direct applicability to the field of automation are not generally appreciated by most businessmen. As a result, there is a great danger that business and industry will repeat a large fraction of the mistakes that have by now led many of our military people to a considerably more sophisticated and informed approach to major systems problems than characterizes the approach of many business and industrial executives.

If business and industry comprehend well and early enough the basic facts of automation, these trends will be supported and strengthened with the result that the national ability to move rapidly ahead in the field of automation will grow, and the bright new industrial revolution just now starting will bring major benefits in new products and in increased national output of old products within the next eight to ten years.

Index

ELECTRONIC DATA PROCESSING FOR BUSINESS AND INDUSTRY

By Richard G. Canning, *Partner, Canning, Sisson and Associates.* "This book makes a strong plea for a properly done and timely systems engineering approach in the hope that people will be guided to the *efficient* use of electronics. It presents some very easily understood examples of what electronics can do and follows in succeeding chapters with a summary of E.D.P. machines; programming of clerical operations; the systems study; the electronic system; the role of operations research; equipment characteristics; and management's program for obtaining a reliable system.

". . . this is recommended reading for (1) those who want an idea of what electronics is about, (2) top management, (3) those engaged in a survey, and (4) the equipment manufacturers." — *Cost and Management.*

1956. 332 pages.

OFFICE WORK AND AUTOMATION

By Howard S. Levin, Ebasco Services, Inc. A simple, quick review of how automation changes raw material into knowledge useful to management. ". . . an introductory text that does the job exceedingly well. Perhaps it is Mr. Levin's mathematics background that gives his writing such clarity, such logical organization. The book shows evidence of thoughful, detailed outlining. Even the format of the contents reveals the kind of logical structure upon which every good teacher builds his subject matter. The book is divided into five parts: New tools for the office, A common language, Electrons for the office, The Businessman and the scientist, and A changing office." — *Data Processing Digest.*

1956. 203 pages.

INTRODUCTION TO OPERATIONS RESEARCH

By C. West Churchman, Russell L. Ackoff, *and* E. Leonard Arnoff, all of the Case Institute of Technology. This new book answers a growing need in the many areas that are affected by operations research today. It fulfills two major objectives: it gives prospective consumers of operations research a sound basis for evaluating O.R. and for understanding its potentialities; and it provides future practitioners with a survey of the field and a basis upon which they can plan their further training.

Written by fourteen experts in the field, the book emphasizes the scientific approach to research. It stresses the importance of defining management problems in terms of objectives. Cogent discussions are given of inventory, linear programming, waiting line, replacement, and competitive and other mathematical models useful in operations research.

1957. 645 pages.

JOHN WILEY & SONS, Inc.

440 Fourth Avenue New York 16, N. Y.