IFAC

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

WARSZAWA 1969

Systems Applications

Fourth Congress of the International Federation of Automatic Control
Warszawa 16–21 June 1969

TECHNICAL SESSION
60

Organized by
Naczelna Organizacja Techniczna w Polsce

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

Systems Applications

TECHNICAL SESSION No 60

FOURTH CONGRESS OF THE INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL
WARSZAWA 16 – 21 JUNE 1969

Organized by
Naczelna Organizacja Techniczna w Polsce

Contents

Wydawnictwa Czasopism Technicznych NOT - Polska

Zakład Poligraficzny WCT NOT. Zam. 88/69

3

THE RELEVANCY OF CONTROL THEORY TO EDUCATIONAL ENROLMENT

Paul Alper

UNESCO Project at the Instituto Politecnico Nacional

Mexico City

Mexico

and

Peter Armitage

Higher Education Research Unit, London School of Economics

London

England

1. Background

As is becoming increasingly obvious to everyone, education is a very big business indeed. Not only is the educational sector of the economy a large one, but moreover, education can be viewed as so powerful a stimulus to the economy such that we speak of "investing" rather than spending money on education. With this in mind, it is easy to see why it is desired to replace the qualitative, quaint and outmoded vague, handwaving of previous educational planning with qualitative, realistic and up-to-date, precise methodologies in order to obtain an optimum or at least a better return for our investment.

Perhaps the most important problem*, or at least the one which has produced the most amount of analytic discussion in the enrolment problem: "How many students will there be in the various sectors of education in the forthcoming years?" The enrolment problem has been the subject of investigation for both the industrial giants of the world and for the underdeveloped nations. There exist models for local communities as well as models drawn up by international organizations which are supposedly applicable to entire subcontinents. There even exists an enrolment model in one country for the white population and quite another for the non-whites.

However, these mathematical models have not proved to be very effective[1]. Quite often, the educational system was viewed as merely an exogenous input

---

* Many people would dispute that enrolment is the most important problem in educational planning. The very recent turmoil and unrest in universities around the world would seem to indicate that the most important problem, so important that it dwarfs all other considerations, is rather: "What is the purpose of education?" a far more difficult and penetrating question than the subject of this paper.

to the economic system as exemplified by either Rate of Return, Manpower Planning or Social Demand Models; little or no mention is made as to how the necessary numbers of educated people are to be produced. Econometric models have been made which display some interaction between the economic system and the educational system, but the stress is on the economy and little detail is paid to the educational system.

Other investigators have chosen mathematical models of enrolment which emphasize enrolment per se but too often these paediometric models have taken the form of what is called input-output analysis. In this method, enrolment is seen as a Markov process such that the number of students in sector i at year t+1, $n_i(t+1)$, is equal to the number of students in sector j at year t, $n_j(t)$, multiplied by the transition proportion of those students who want to go to sector i, ρ_{ij}, summed over j plus the new entrants to the system's sector i, $b_i(t)$:

$$n_i(t+1) = \sum_j \rho_{ij} \, n_j(t) + b_i(t) \qquad (1)$$

It is easy enough to see why input-output analysis is so tempting. The argument for it runs that it is a necessary first step and seems to give a cohesive, comprehensive picture of what is occurring in a neat tabular form and is very appealing to economists brought up on Leontief matrices because of the inherent similarities between partitioning the economy and partitioning the educational system. Moreover, input-output analysis seems to possess an objective rather than subjective basis, a very desirable feature for research in the social sciences.

However, this view of input-output analysis is rather naive. The main objection to input-output analysis is that it provides no insight into the educational system. Nor does it tell one what would occur if different decisions were made. Furthermore, the supposed objectivity doesn't actually exist because drawing up the input-output table tacitly assumes that the status quo will be maintained.

Although the ρ's supposedly represent the percentage social demand of those in sector j for sector i, in reality, the number of students who eventually go to sector i from j is greatly influenced by past decisions as to the number of places made available in many of the sectors of the educational system. Thus, when today's planner, using data from preceding years divides n_i by the number who came from j, he does not obtain the trend of ρ_{ij} on which to base predictions and optimum strategies but rather he obtains the complex result of the interaction of his previous decisions with the social demand.

Furthermore, even though it has been repeatedly stated that because of statistical uncertainty, it is necessary to have more aggregation[2], the proponents of input-output analysis seem to be heading in the opposite direction, namely increased disaggregation. Although we desire to know as much detail as possible about educational enrolment, it is quite frankly impossible in the conceivable future to have matrices of the order of 100 x 100; yet, there exist advocates of input-output analysis who propose matrices of the order of 1000 x 1000.

II. Introduction of Control Concepts

At first, almost exclusively, all mathematical approaches to educational enrolment were from the point of view of economists. Recently, even though economists still predominate in the field, some investigators have attempted to introduce control concepts. Perhaps the most important contribution of a control outlook is the recognition that models as exemplified by Equation (1) are structurally empty.

Some research workers have suggested that Equation (1) be reformulated as:

$$n_i\,(t+1) = \text{Min}\left[\left\{\sum_j^N \rho_{ij}\, n_j(t) + b_i(t)\right\},\ u_i(t)\right] \qquad (2)$$

where $u_i(t)$ is the number of places provided; in the language of the economist, $u_i(t)$ represents the supply side of the question while the rest of the other terms represents the demand side. This formulation makes explicit that the decision-maker can and does influence the resulting distribution of students and teachers. That is to say, there do exist some means of actively manipulating enrolment, rather than just passively observing preceding events.

Koenig[3] has attempted to set up a model on a control basis for enrolment, not in the nation, but at one particular university. This model attempts to formulate enrolment as a typical, modern control problem using state space concepts whereby the system is brought from one state of nature to another. The control variable is conceived of as the number of assistantships or the amount of money made available in order to steer students into the desired paths. In other words, the transition proportions can be viewed as a function of the money obtainable in the receiving sector

$$\rho_{ij} = f(\text{money in } i) \qquad (3)$$

In Section III is found a discussion which is pertinent to some possible shortcomings and deficiencies of this approach. Let it just be noted that no matter how well thought out a plan based on control theory for Michigan State University is, the plan is in deep trouble if no recognition is paid to actions independently undertaken at the University of Michigan (or indeed Ohio State or Illinois or Indiana). Clearly, when the outside factors are as important or more important than the factors within the system being studied, the plan must take this into account.

Another application of control concepts to educational enrolment may be found in Smith and Alper[4]. There it is considered how to provide the optimum number of places each year, the decision variable u, which will satisfy a fixed but unknown demand for those places, μ, in order that either the expected cost is minimized

$$\text{Min E}[W] = \text{Min E}[u - \lambda \min(u, \mu)] \tag{4}$$

or that the probability that the cost is less than or equal to a fixed quantity is maximized

$$\text{Max Prob}(W = S) = \text{Max Prob}[u - \lambda \min(u, \mu) = S] \tag{5}$$

The result obtained via dynamic programming for a sequence of decisions is quite surprising. In the first situation, Equation (4), the problem turns out to be a truly dual one, information is actively acquired via the decisions made while in the second case, Equation (5), this is not true and a n-stage decision process degenerates into n-single stage processes; this indicates that different criteria can give rise to markedly different results. Because these problems are analytically tractable, a comparison for various probability distributions of μ can be made between the optimum and nonoptimum but reasonable strategies (rolling planning). Figure 1 gives a block diagram version of the problem.

As interesting and revealing as this problem is, as soon as more realism is added - for example, if μ is a function of time rather than a fixed quantity or if μ is a vector rather than a scalar - the problem becomes intractable. Even if solutions were analytically possible, there remains the enormously difficult task of assigning a numerical value to λ, a parameter which relates the benefit attached to a student who occupies a place to the cost of providing that place.

Von Weizsacker[5] treats the problem of determining the proportion of a citizen's lifetime which should be spent in school in order that the nation's income be optimized. He next assumes that the nation's income is maximized when the individual's income is maximized and deals from then on with this problem. In Fig. 2 is the block diagram; the decision variable is the proportion of time spent learning, z, and the dynamics of acquiring knowledge is assumed to be

$$\dot{m} = z - h\,m \qquad (6)$$

where m is the level of skill or training and h is the rate of obsolescence of knowledge. The amount of money earned is a function of the skill level multiplied by the proportion of time spent not in school. Therefore, the total discounted life income for T years is given by

$$L \sim \int_0^T e^{kt} f(m)\,(1-z)\,dt \qquad (7)$$

where k is a discount factor.

Von Weizsacker finds by means of Pontryagin's maximal principle that for f(m) obeying reasonable economic assumptions (decreasing marginal return --$f(0) \geq 0$, $f'(m) > 0$, $f'(m)/f(m)$ is a decreasing function of m) that the solution is given by one of three cases depending on the form chosen for f(m) and the particular values of h and k:

Case I. m(t) is identically zero for all t and no training pays the most.

Case II. m(t) has a unique maximum. The life span is divided into two periods, all learning followed by all earning -- bang-bang solution.

Case III. m(t) has a unique maximum. The life span is divided into three periods, all learning, followed by part-time learning (but z = constant) followed by all earning.

Note that z is monotonic; once the proportion of time spent learning decreases, it never reverses direction.

Inasmuch as having a monotonically decreasing proportion of time spent learning goes against many modern theories of education, it is possible that the fault lies in either the dynamics of acquiring knowledge or the criterion or both. Certainly, there can be much criticism concerning the basic assumption that a nation's income is maximized when each individual's income is maximized; moreover, the optimization of discounted life income, while a fine principle of economics, may be far too limited an objective function for education.

III. A Closer Look at Some Basic Difficulties

At first glance, it would seem quite obvious that control concepts can be very usefully applied to educational enrolment. After all, the educational system surely bears some resemblance to an industrial process in the sense that the raw materials, the uneducated students, enter and are acted upon by teachers to produce educated people who are either removed from the system or return as teachers. Decision-makers exist who affect the educational process analogous to the way controllers affect an industrial process; analogous to criteria for judging an industrial process, similar criteria can be thought of for judging the educational process.

Unfortunately, an industrial process and the educational process, upon further investigation are not so isomorphic to each other. Even in a country such as England which is far more centralized than the United States, educational decisions made by the central decision-maker can be ignored or effectively modified. Thus, we no longer have a controller but a "suggester"; it is as if some of the knobs on the central control panel of an industrial process had shafts which either didn't engage the process or shafts which were connected to some unknown element somewhere in the process.

Equally serious from the point of view of control theory, is the process itself. At the best, the time constants for educational enrolment is measured in years, rather than hours or days as in an industrial process. Consequently, any optimization search technique, such as parameter perturbation for example, which depends on slight alteration of significant factors would be extremely slow.

Far worse than the time constant problem, is the fact that the educational process is non-stationary to the point where past response to previous stimuli may be completely misleading as to what present-day responses would be to the same stimuli. The educational processes of many countries have undergone changes that may be described as explosive and therefore, little faith can be had in the assumption that what was true ten or twenty years ago will be true today.

Moreover, the educational process has enormous data problems. With respect to England and Wales*, which have more comprehensive data than most other countries, the following has been noted[5]:

> "So far as stocks are concerned, ... a considerable volume of stock data is now published in Statistics of Education. Even so, any ambitious plans to introduce much more detailed models quickly run into gaps in our knowledge. For example, data on the age distribution of university students for each separate year of study are not collected and while the qualifications of teachers are known, it is not known what subjects they teach. Where gaps have only recently been plugged, there is also the problem that time series of past values may consist of only two or three points. There have also been changes in classifications which affect the published statistics and changes in the actual structure of the educational system which have profound and diffuse repercuss ons on stocks. For example, starting in 1960, courses in teacher training colleges were extended from two to three years. It is virtually impossible to isolate the effects of such a change, or estimate what would have happened had the change not taken place.

> "Flow data are much more sparse. The most reliable data at present available relate to teachers because of the existence of the central record of teachers (an individualized data system set up originally for superannuation purposes). Since until recently little attention has been devoted to flow data, educational returns have in the past concentrated upon statements of stocks at some particular time with no analysis of these stocks by the sources from which they immediately came. There are some cases of flow data, but they remain few. For example, the school leavers sample survey gives information on the destinations of leavers (to university, further education, colleges of education and employment). The volume of flow data can be expected to increase in the future but, for the present, where flow figures are lacking they have to be inferred from stock data .... Inferred flows are a temporary necessity but a poor substitute for data on the actual flows."

---

* Scotland and Northern Ireland have a different structure and manner of collecting data from England and Wales, illustrating yet another aspect of the difficulties of data even within one entity known collectively as the United Kingdom of Great Britain and Northern Ireland.

Concerning disaggregation and data requirements of England and Wales[5]:

> "Suppose that we have reason to believe that the performance of students at universities differs according to their secondary school backgrounds, and that this appears relevant to the purposes for which a particular model calculation is being carried out, then this would force us to introduce more detail into the classification embodied in the model structure. It would be necessary to distinguish, say, university science undergraduates who had come from one type of secondary school in one box and those who had come from another type of secondary school in another box in order to keep the boxes homogeneous. While our model language could cater for this kind of elaboration, it must be emphasized that <u>full</u> flow information is needed on each permitted movement. If we introduce further breakdowns in our structure, but have to infer most of the flow data required by the new breakdown, then the resulting exercises may do little more than display our imaginative ability."

But the data situation is even more serious because of the "bottleneck" problem. When the demand for places exceeds the supply in a sector, then it is said that a bottleneck exists in that sector and an overspill occurs. Clearly, if bottlenecks exist in the educational process, the model must therefore reflect this phenomenon in order to have any validity regardless of the temptation to ignore it on the grounds of analytic difficulties or democracy*. Since the recognition of bottleneck models is of much more recent origin and since they have only just begun to be studied[6], data for this purpose is much less adequate. We do not know for each sector of education what the "true" demand for places was, whether this demand exceeded the supply of places or posts and whether there were any consequent overspills. As before, we can only proceed by exercising our imagination in order to reconstruct feasible interpretations of what happened in the past. It will also be necessary to say in each case how the applicants from the various possible sources are to be chosen and it will also be necessary for

---

* An imaginative but spurious argument advanced against the existence of bottlenecks is that there would be a revolution in the country if students could not get their first choice. Inasmuch as no revolution has occurred, students must be satisfied and thus bottlenecks do not exist!

every group of unsuccessful applicants to describe its alternative behaviour. These two processes correspond to the selection procedures adopted by admission authorities and the alternative preference or reallocation of frustrated applicants.

The analytic difficulties introduced by bottlenecks are considerable and different selection and/or allocation procedures can yield different solutions[4]. Moreover, we no longer have a neat series of equations but rather now have a computational algorithm which is strongly dependent on which overspills take place and in which order.

Continuing a closer look at the process itself, it is seen that the "physics" of what is taking place is very poorly understood. The functional dependence of the transition proportion upon the student-teacher ratio or the scholarship money made available is highly debatable; some people claim that ρ_{ij} is proportional to the student-teacher ratio while others feel their experience indicates that ρ_{ij} is independent of the student-teacher ratio.

Many suggestions have been made about the forces and mechanisms which underlie demand but few seem to have been satisfactorily investigated and some are difficult, or even impossible, to quantify. In other cases it would be of no value in predicting educational developments if the postulated relationships were reliably known. For example, if educational demand is affected by heredity and family income, then future changes in heredity patterns and the distribution of family incomes would have to be known in order to estimate changes in educational demand, i.e. it would be necessary to extrapolate heredity and family income in order to estimate future demand*.

Looking more closely at possible indices of performance, unlike an industrial process for which reasonable criteria concerning maximum

---

* This important qualification should not be misconstrued: the fact that such relationships would be of little or no predictive value does not mean that studies of them are futile, for they may be of considerable value in understanding the behaviour of the real system <u>after</u> the event. If some unexpected occurrence can be adequately explained after the event, there is the consolation that no further explanation is required even though anticipation was not possible.

productivity, minimum fuel consumption or minimum time can be thought of, the educational process defies the setting up of a quantitative criterion unanimously agreed upon.

Even if we could say with precision what the economy must produce in the future, this could not be translated with comparable precision into the distribution of educational achievements of the labour force due, among other things, to our poor state of knowledge on labour substitution.

While it may be comparatively easy to set up an economic criterion such as maximizing discounted life income: "Are not increased literacy, responsible participation in society and in politics, identification with national goals, improvement in physical and mental health, and development of attitudes and values favourable to progress just as important as national income and industrial productivity? May not the ultimate economic return of such results be more significant than the immediate gains in productivity? If so, how do we really measure the economic benefit of education?"[7].

In the present state of knowledge, it does not seem that either man-power demanded by the state or social demand of the individual can be satisfactorily represented and yet an even greater difficulty lies in fusing the two together into an objective function. If we could adequately state the expected educational needs of the economy and demands of society we could reasonably expect to resolve the conflicts between them by defining their relative importance. In order to do this we would have to develop a definite sense of the value of meeting, or failing to meet, both kinds of needs. If we could say "one 'production unit' is worth x 'satisfaction units'", then the way would be open to saying what the educational system ought to do. However, we are unable to make statements of this kind and the vast body of educational literature and discussions offers little practical guidance and is surprisingly inadequate in making clear statements on educational objectives.

Naturally, when it is recognized that the process and the criterion must be stochastic and not deterministic, the mathematical analysis is still further greatly complicated since many parameters are needed to reflect the complexity of the real system. Optimizing a stochastic control process containing so many parameters, only vaguely known, would seem to be far beyond the state of the computational art.

Even if the analytic problems mentioned concerning the controller, the process and the criterion were circumvented, a very embarassing problem still remains. It is desirable to apply control theory to educational enrolment and the reverse must be avoided. Little purpose is served if educational enrolment is put into a control framework, just to make enrolment more comprehensible to the control engineer. The very framework of control theory is shaky with regard to educational enrolment. Not only is the controller reduced to a suggester, but the error detector and the idea of feedback, so central to all discussions of control concepts, can be questioned; the political and thus less than rational aspects of education, the frequent turnover of ministers of education, the long time between cause and effect, makes one wonder whether control theory is fairly irrelevant to educational enrolment.

IV. Conclusions

As was mentioned, perhaps the key contribution of control theory to educational enrolment is the recognition that it is necessary to embody into our models a manipulable variable in order to reflect the active influence the decision-maker has upon the distribution of students and teachers; without such a formulation, our models run the risk of being structurally inadequate and planning is reduced to passively observing what has taken place previously and unknowingly projecting the status quo.

However, as important as this new way of thinking about the educational process is, a point of view alone is not enough. There remains a multitude of very real conceptual problems which will not disappear by merely moulding the situation within a control framework. Much more effort must be invested into research in order to find out in a quantitative fashion, for example, why students prefer certain sectors of education to others. Pilot studies to find out how, for example, increases in stipends would affect social demand, though expensive, must be undertaken in order to ascertain the effects of decisions, regardless of the temptation to claim the pilot studies would be undemocratic and unenlightening because no minority should be singled out and this minority would behave differently anyway just because it is being studied.

A control framework can be very useful and rewarding but only if the real problems are not brushed aside in favour of a familiar methodology.

Notes and References

1 Alper, P., "Decision-Making and Optimization in Educational Enrolment", Proceedings of the IFAC Conference in Cleveland, Ohio, June 1968, pp.298-305.

2 Alper, P., "A Critical Appraisal of the Application of Systems Analysis to Educational Planning Models", IEEE Transactions on Education, June, 1968.

3 O.E.C.D., Mathematical Models in Educational Planning, Paris: 1967.

4 Smith, C.S. and P. Alper, "Control Strategies for an Educational Planning Problem", Canadian Operational Research Journal, Vol. 5, No. 3, November, 1967.

5 Armitage, P., C.S. Smith and P. Alper, Decision Models for Educational Planning, London: Allen Lane, The Penguin Press, forthcoming (1969).

6 Clough, D.J. and McReynolds, "State Transition Model of an Educational System Incorporating a Constraint Theory of Supply and Demand", Ontario Journal of Educational Research, 1966, Vol. 9, No. 1.

7 Dressel, P.L., "Comments on the Use of Mathematical Models in Educational Planning", Mathematical Models in Educational Planning, Paris: O.E.C.D., 1967, pp.275-290.

$$\text{Min } E \sum_{i=1}^{N} W_i = \text{Min } E \left[\sum_{i=1}^{N} u_i - \lambda \, \text{Min} \, (\mu, u_i) \right]$$

or

$$\text{Max Prob } \left(\left(\sum_{i=1}^{N} W_i \right) \leq S \right)$$

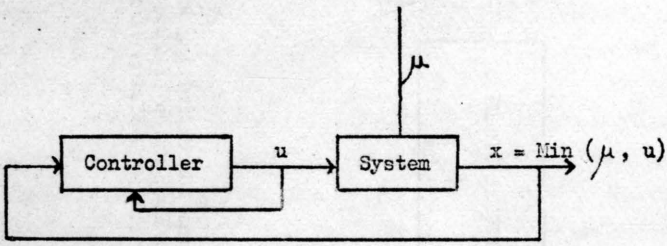


Figure 1. Block Diagram for Optimization Example of Smith and Alper for Determining Sequence of Decisions to Satisfy Unknown Demand for Places.

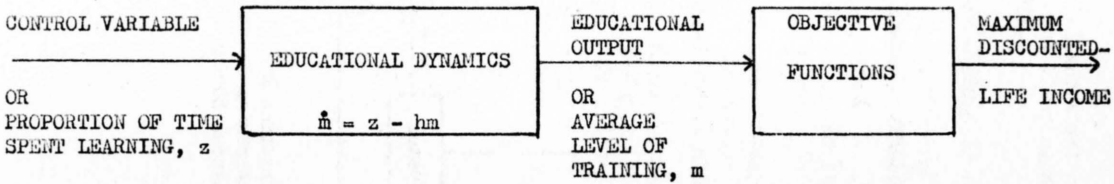


Figure 2. Block Diagram for Optimizing Discounted-Life Income with Proportion of Time Spent Learning as the Control or Decision Variable.

THE USE OF MULTIACCESS COMPUTERS FOR THE MANAGEMENT AND CONTROL OF PROFESSIONAL LITERATURE

J. Francis Reintjes
Massachusetts Institute of Technology
Cambridge, Massachusetts, U.S.A.

The vast increase in recorded knowledge that has occurred during the twentieth century is complicating the traditional procedures we use to gain access to this knowledge. Uncertainty about what is already known and what has been recorded on a subject often leads to unnecessary duplications of effort and to frustration when duplications are later discovered.

Multiaccess computers, operated in an online mode, now offer an opportunity to establish orderly management and control of the professional literature, and thus they have potential for improving the efficiency of information transfer among members of the professional community. These computers can accommodate billions of bits of information in secondary storage; they enable many users to access and process the digitally-stored information simultaneously from remote terminals; and they can be organized so that users can work back and forth with the machine, in dialog fashion, as user and machine progress toward acquisition of the information being sought.

A machine-oriented library system in which a multiaccess computer is the central storage and processing element is illustrated in Fig. 1. A data base, consisting of a catalog of professional literature, is stored in the disk files of the computer, together with the computer programs which are used to operate upon the data base and to retrieve the information being sought. The catalog is accessed through a user console consisting, as a minimum, of a keyboard for inputting user requests and a cathode-ray-tube display for presenting machine replies. When a satisfactory piece of information is found, the full text of the document containing the information can be obtained automatically at the user's station from a microfiche storage, transmission, and receiver system to be described below.

The power of this kind of information-transfer system resides in two key elements which react continually with each other as searches are executed -- the machine and the user himself. The quantity of information that can be stored in the machine and the speed with which it can be processed add new dimensions to information-cataloging procedures. The fact that the user can negotiate for information smoothly in a real-time mode and continually exercise options as the dialog progresses, enables one to take full advantage of the user's intellectual powers in the searching process. Thus, in the terminology of the control engineer, the information-transfer system includes a human operator in the man-machine loop, with full advantage being taken of the operator's memory capacity, decision-making abilities, and response characteristics.

An example of how a user may engage an information-transfer system, illustrated in Fig. 1, follows:

User types the following request at his keyboard terminal: Search for literature on Optimal Control.

The system responds by displaying on the screen of the user terminal: Two hundred documents found. Do you wish to see their titles? Reply yes or no.

The dialog might then continue as follows:

USER:	No.
SYSTEM:	You may qualify your request if you wish to continue.
USER:	Display the number of books and the number of journal articles on Optimal Control.
SYSTEM:	20 books found. 180 journal articles found.
USER:	Search for journal articles published since 1967.
SYSTEM:	25 articles found. Do you wish their titles displayed?
USER:	No. Search for articles on Optimal Control in Nonlinear Systems published since 1967.
SYSTEM:	5 articles found. Do you wish their titles displayed?
USER:	Yes.
SYSTEM:	(Titles of 5 documents are displayed at the user's console.)
USER:	Display Abstract of document No. 18762.
SYSTEM:	(Abstract is displayed.)
USER:	Display full text of 18762.
SYSTEM:	Full text is displayed, page by page at a rate controlled by user.

Another powerful aspect of a computer-stored information system stems from the fact that in this kind of system information is stored in electrical-signal form. Conceptually, therefore, it is unnecessary to duplicate the information in many widely separated geographic locations, since the signals are easily transmittable over wire and radio communication links. Worldwide transfer of the stored information is possible through use of communication satellites.

An experimental information-transfer system based on the configuration shown in Fig. 1 has been designed under Project Intrex and is being made available to a selected group of users for their evaluations. The central computing system is an M.I.T. modified IBM-7094. In modified form the machine will accommodate approximately 80 million English words in secondary (disk-file) storage, and a maximum of 30 users can engage the machine at a time from remote typewriter or graphical (cathode-ray-tube) consoles. The machine is designed so that the computer program of each user who is online is executed in short, discrete steps. The user can also exercise some control over the rate at which his program proceeds and the course he wishes it to take. The system therefore is able to react rapidly to user commands and it responds in a flexible manner, in accordance with the wishes of the user as he sees his preliminary results evolving.

It should be noted that a multiaccess computer capable of serving a multiplicity of users simultaneously is ideal as an information-transfer system for the professional literature. The salient requirements for such a system are these: it should be capable of storing a large data base which can be easily and quickly accessed by a community of users through use of a single set of computer programs; and furthermore, the system should be able to take full advantage of the knowledge and decision-making capabilities of the human operator by permitting him to conduct a dialog with the machine as he negotiates for the precise item of information he is seeking. The manner in which the experimental system is organized to meet these requirements is described below.

Literature. The literature chosen for the experimental system has been selected from the field of Materials Science and Engineering (MS and E). This choice was made on the basis of several factors which turned out to be optimal for the Materials Science area at our location. We have access to a large community of researchers in MS and E, and several groups are interested in participating in the selection of literature and monitoring of our work as it proceeds. These groups will provide an experimental body of users when the system becomes operational. Active interest of professionals in the management and control of the literature of their field is a highly desirable attribute when one undertakes the development of a well-ordered system of the kind being described. Choice of material which constitutes the literature base, agreement on formatting, and establishment of guidelines and procedures for subject indexing by the authors, reviewers, professional societies or others, are important considerations which can contribute to the effectiveness of a machine-stored literature base. Although our literature is presently in MS and E, it should be emphasized that the automatic-control field appears as an excellent area for exploitation of modern information-transfer techniques because of the worldwide cohesiveness of its professional membership.

The Catalog. The catalog which describes the salient features of the documents in a collection serves as a "finding tool" for the literature itself. Because of the high storage capacity of modern computing machines and their rapid processing capabilities, the catalog can be enriched beyond present standards to make it a valuable entry point to the literature. If one is willing to devote a few hundred words to the characterization of a document, it should be possible to reveal to the user many document features before its full text is requested. In the experimental system we are attempting to determine the characterizations which are most helpful in the retrieval process. We are devoting approximately 500 words per document (journal article, report, book) to document descriptions and approximately 50 attributes per document are being included. Principal among these are the author's name, document title, abstract, author's purpose, professional flavor and level, and a set of words and phrases which describe the material in the document. The amount of the document to which the phrase applies is indicated by a number weight. Our objective in the experimental system is to ascertain which attributes among the many we are using are most helpful in retrieving relevant documents from the catalog. A catalog of at least 10^4 documents has been set as a goal for our experiments.

The Catalog File. The catalog file is constructed to give the user several ways for requesting information and maximum flexibility in making his request. Inverted files of authors and subject words are provided, as well as an inverted file of titles. The subject words are compiled from the list of subject phrases that are used to describe each document. In the file of subject words, the order of occurrence of each word in a phrase is noted as well as the phrase in which it is located and the number weight of the phrase containing the word. Thus a user may enter the catalog by word, a sequence of words, author's name, or document title. The file also contains, of course, the additional information indexed for each document (abstract, author's purpose, and so forth), and these items are available upon request, once the document number is known.

The User's Console. A user of the literature information-base accesses it by means of a display console which can be located remotely from the central computer. Requests for information are entered through a keyboard, shown in Fig. 2. The keyboard is composed of upper- and lower-case English-letter keys and a selection of special keys to enable quick execution of frequently used commands. Each user command appears on the cathode-ray-tube screen, together with machine responses. A light pen is available for designating items of displayed information which should be acted upon by the system. If, for example, a series of document titles is displayed, the user may save some and delete others by pointing the light pen at the document numbers to be saved and depressing the appropriate key.

A 128-track magnetic-drum storage device and a small buffer computer are interposed between the display console and the central multi-access computer. This configuration permits 10 display consoles clustered within a radius of 500 meters to be served efficiently and with minimum delay to the user. The data rate between central computer and magnetic drum is 2000 bits per sec; between drum and each display console, approximately 4.5 megabits per sec.

As a further convenience to the user, a series of programmable switches is provided beneath the cathode-ray-tube display. These switches designate a series of logical operations which must be followed, once an action is taken and a switch is activated. The possible logical operations appear on the CRT screen as displayed labels directly above the switches. For example, suppose the programmable switches are initially in a mode that labels them EDIT, DISPLAY, PRINT, SEARCH. Actuating the EDIT switch now relabels the switches to read ERASE, CLOSE-UP, CHANGE, UNDERLINE, thereby indicating the kinds of editing that may now be performed. Further actuation of a new operation will reveal the additional operations that may be performed as a result of making a choice among the displayed operations.

Full-Text-Access System. Once a user has decided he wishes to see the full text of a document, he may place his request through his keyboard. The document may then be obtained in either of two forms: either as a 35-mm film strip or as a page-by-page display on a CRT screen. At present the full-text CRT is separate from the screen he used to access the catalog. The full-text display equipment is located adjacent to the catalog display console.

Presently the technology of digital-computer storage does not permit economical storage of the full-text of documents within the computer. To store even a small library of 2×10^4 professional journal articles imposes storage requirements beyond current capabilities. If one assumes that the 2×10^4 articles average 5×10^3 words per article plus one half-page photograph per article, then approximately 1.5×10^{10} bits of information must be stored within the machine.

In looking for various possibilities for storing full text outside the computer, one is attracted immediately to photographic film as a highly reliable, inexpensive, high-density storage medium. On the basis of information-density capability, cost and compactness, full-text storage in image form on microfilm is the most attractive among such alternatives as film, magnetic tape, and magnetic disk. Our experimental system stores the full text of documents contained in our computer-stored catalog on microfiche, a 10-cm by 15-cm strip of film, with each microfiche containing approximately 60 pages of text at an 18-to-1 reduction from original size.

An overall block diagram of the full-text retrieval system is shown in Fig. 3. The microfiche are stored in an automatic storage and retrieval device which accesses documents automatically by means of a magnetic selector operating under computer control. Each microfiche has a ferrometallic clip attached to its longer edge, and this clip is notched according to the binary code representing the document numbers contained on the microfiche. When a document is requested by number, the appropriate microfiche is automatically withdrawn from storage and positioned so that the first page of the document may be optically scanned by means of a flying-spot scanner. The electrical output signal from the photomultiplier tube of the scanning system is then transmitted via coaxial cable to the receiving station. Since several receiving stations may be included in the system, each at a different geographic location, the video signals are encoded with an address at the transmitter. At the receiving station the video signals are displayed either on a cathode-ray storage tube from which the user may read the document a page at a time, or he may receive a 35-mm copy of the complete article, ready for viewing on a microfilm viewer. In the former case, the microfiche being scanned is stepped along to each succeeding frame upon receipt of a command from the user; in the latter case, the stepping process occurs automatically as the scanning of each page is completed.

The total time required to receive a film strip of a 5-page document is, in terms of component times:

Find and withdraw microfiche from store	5 sec, max
Position first frame	2 sec, max
Step to each succeeding 4 frames	2 sec, total
Cut off film strip at receiver, feed to processor	1 sec,
Develop and dry film	90 sec,
Total	100 sec, max

A highly important consideration in the design of the microfiche scanning system is the number of scan lines which should be used. We have made extensive analyses and tests to determine the influence of the number of scan lines on resolvability of the received images. From these tests we conclude at least 2,000 scan lines per page are needed in order to preserve legibility of the smallest size type commonly used for mathematical symbols, particularly for subscripts and superscripts. This large number of scan lines imposes stringent demands on the bandwidth of the transmission link if the images at the receiver are to be refreshed at a rate which minimizes annoyance due to flicker (at least 50 times per second). Bandwidths of the order of 100 megahertz are required, depending on the exact number of lines and refreshment rate employed. To circumvent need for such high bandwidths, and to improve performance speed of the text-access system, the screen-refreshment concept was abandoned altogether. In our system we scan a frame only once, at a half-second rate. As a result, transmission-bandwidth requirements are those for standard U. S. A. television, approximately 4.5 megahertz. This approach makes it necessary to keep a microfiche out-of-store only long enough to scan each page of an article once -- about 11 sec for a 5-page article. By attaching a binary-coded address to the video signals corresponding to each page, we are able to bridge several receiving stations at remote locations across the video transmission line. Only the station which senses its own address responds to an incoming signal.

The price one pays for the elimination of image refreshment is that a means for storing images must be provided at the receiving station in order to capture and retain the one-shot transmissions. In the system illustrated, this storage is provided at the film station by the film itself, and at the CRT station by the 11-inch storage tube being used as a viewing device. A sample of a full page of text which has been microfilmed, scanned, transmitted, reproduced on 35-mm film, enlarged, and printed as part of this paper is shown in Fig. 4. In evaluating the quality of this reproduction, one should bear in mind that it has been necessary to subject the original page to extra processing steps to reproduce it in illustrated form.

The full-text-access system described above meets very closely the basic requirements which should be set for a machine-oriented library system, namely, the system should provide guaranteed access to full text at remote locations. Guaranteed access to text is essential in order to avoid the frustrations that result from inability to obtain material from a library because it is already on loan, or misshelved, or at the bindery, or misplaced. Rapid access at remote locations is essential so that the whole concept of conserving the user's time throughout the entire retrieval process is preserved.

Status of our Research. The overall objective of our research program is to conduct experiments which will lead to specifications of operational machine-stored library systems of the future. By bringing our experimental library to a selected community of users who have a bona fide need for the information it contains, we shall obtain valuable insights about user requirements and system characteristics which are needed to meet these requirements.

We are making careful measurements of the time required to develop the literature base in our system. It is evident that manual indexing of the literature in depth after its publication is time-consuming. Other approaches to indexing must be examined. Our skilled catalogers are spending, on the average, seven minutes per page on the in-depth subject indexing of professional journal articles, and an additional 5 minutes per article on the development of additional entry points. When one includes the time required for descriptive cataloging and to place the catalog material in digitally encoded form, a substantial investment of time goes into the manual-cataloging process.

Several alternatives are open which can assist in reducing cataloging costs. One approach is to develop a standard subject-indexing procedure which can be agreed upon by the professional community and which can be performed by the author at the time of submission of his papers for publication. If definitive guide lines for subject indexing can be postulated, the author should be in the best position to subject index his paper. It is our purpose to develop guide lines for subject indexing which will gain acceptance for use in cataloging for machine-stored library systems.

A second alternative approach to subject indexing makes use of the digital computer for automatic indexing. We have experimented to a limited extent with automatic (machine) indexing and have obtained promising results. In one experiment we compared all words in the titles and abstracts of articles with the subject index words used by human catalogers who indexed the articles. Subject-index terms which were used by catalogers and which appeared in the title and abstract are placed into a computer-stored dictionary. This dictionary is then applied to the titles and abstracts of new articles as they come along. Our tests on a sample of 200 documents show that our completeness and relevance factors are approximately 60 percent.

The urgent need at present is for an abundant supply of easily-obtainable digitally-encoded literature so that intensive research and exhaustive testing of results on automatic indexing can go forward. Translating literature which already exists in printed form into digitally encoded text is very time-consuming and costly. Hence, experimentation with large samples of literature is likely to be prohibitive. A tabulation of possible points in the literature-generation process where digital encoding may occur, together with advantages and disadvantages, follow:

Digital Encoding Performed by	Advantage	Disadvantage
Author	Avoids need for retyping the manuscript	Author needs a special typewriter
Professional Society	Relieves author of requirement for special typing equipment	Adds to cost of publication because of retyping
Printer	Digital encoding is already taking place here if type is being set under digital control	Printers must convert to automatic type setting equipment
Automatic Character Readers	Enable automatic encoding of the literature which already exists in graphical form	The equipment is expensive and the encoding is not error-free

Obviously it is highly desirable that printing be accomplished by means of computer-controlled type setting processes. A digitally encoded version of the manuscript then becomes a natural by-product.

Conclusions. The multiaccess computer operated in an online mode offers excellent potential for improving the effectiveness of information transfer among members of a professional community. These machines are most effective for information transfer when the capabilities of the human operator are taken into account in the overall system design. In computer-oriented information-transfer systems, maximum value accrues to the various groups employing the systems when uniform standards of indexing and formatting are followed. Project Intrex is currently conducting research which should be helpful in the establishment of such standards. A major cost factor in the development of a machine-stored library system is the time required for professional indexers to generate the catalog of literature being stored. Availability of digitally encoded text material can be of benefit here, since it can serve as a source of experimental material in research on automatic-indexing techniques.

Acknowledgements. The research reported here was made possible through support to Project Intrex, Massachusetts Institute of Technology, by the Carnegie Corporation, the National Science Foundation and the Council on Library Resources.

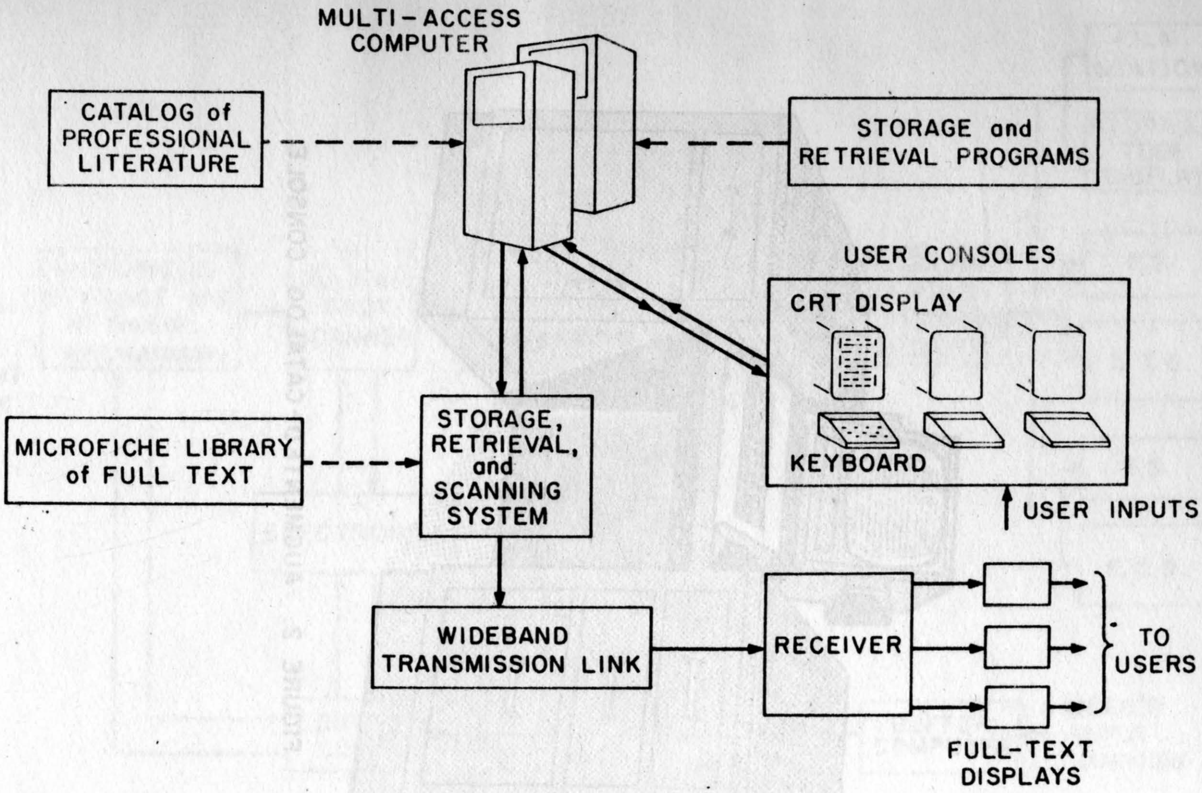


FIGURE I. FUNCTIONAL DIAGRAM OF A MACHINE-ORIENTED LIBRARY SYSTEM

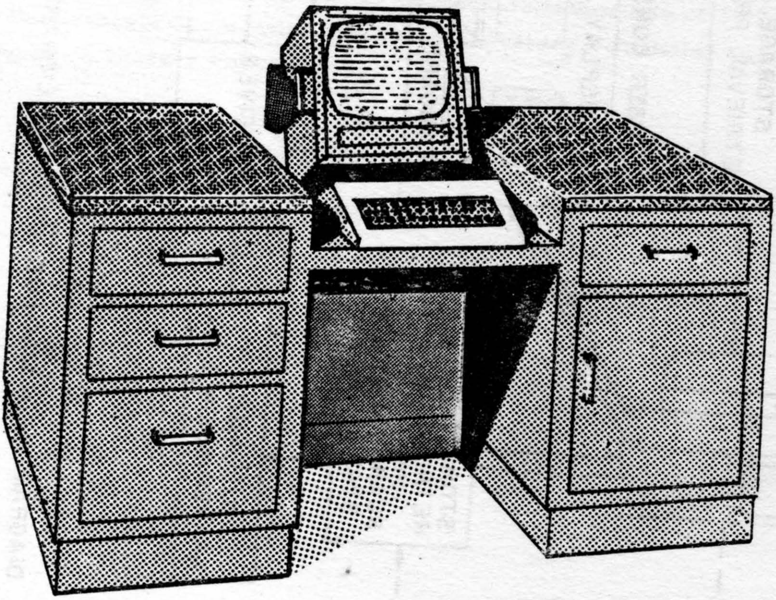


FIGURE 2. AUGMENTED-CATALOG CONSOLE

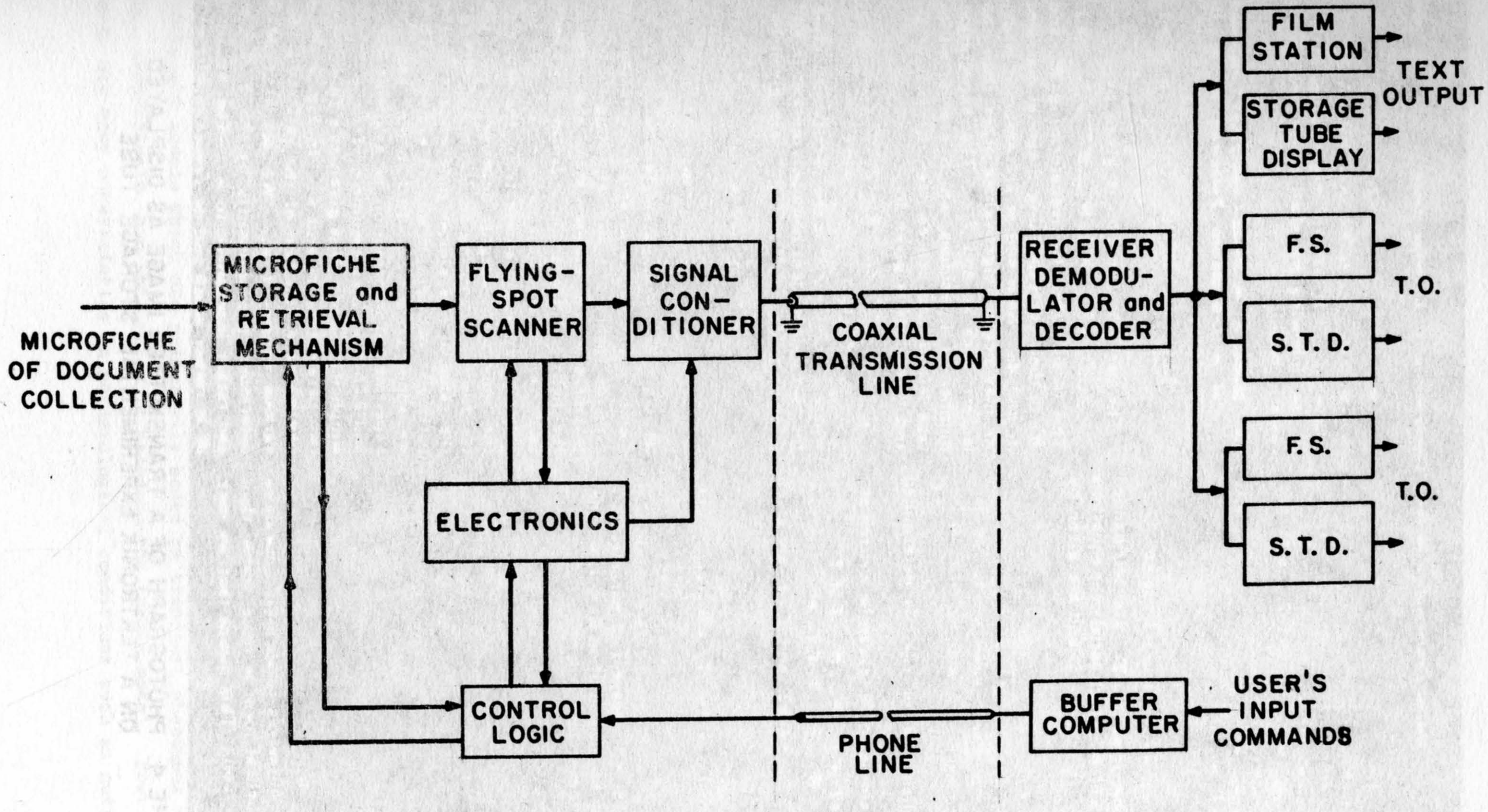
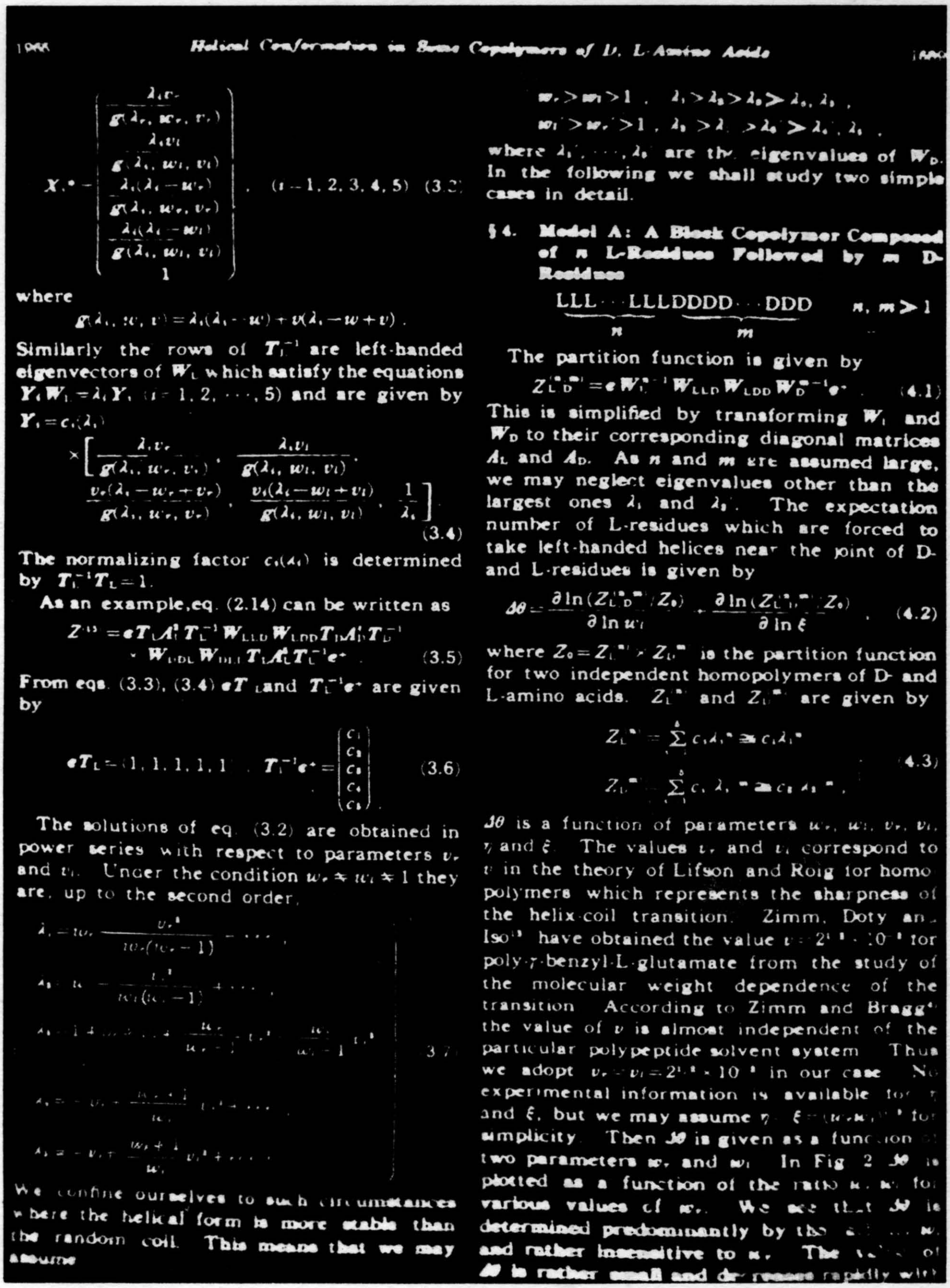


FIGURE 3. BLOCK DIAGRAM OF TEST-ACCESS SYSTEM



$$X_i^* = \begin{pmatrix} \dfrac{\lambda_i v_r}{g(\lambda_i, w_r, v_r)} \\ \dfrac{\lambda_i v_l}{g(\lambda_i, w_l, v_l)} \\ \dfrac{\lambda_i(\lambda_i - w_r)}{g(\lambda_i, w_r, v_r)} \\ \dfrac{\lambda_i(\lambda_i - w_l)}{g(\lambda_i, w_l, v_l)} \\ 1 \end{pmatrix}, \quad (i = 1, 2, 3, 4, 5) \quad (3.3)$$

where

$$g(\lambda_i, w, v) = \lambda_i(\lambda_i - w) + v(\lambda_i - w + v).$$

Similarly the rows of T_L^{-1} are left-handed eigenvectors of W_L which satisfy the equations $Y_i W_L = \lambda_i Y_i$ $(i = 1, 2, \cdots, 5)$ and are given by

$$Y_i = c_i(\lambda_i) \times \left[\frac{\lambda_i v_r}{g(\lambda_i, w_r, v_r)}, \frac{\lambda_i v_l}{g(\lambda_i, w_l, v_l)}, \frac{v_r(\lambda_i - w_r + v_r)}{g(\lambda_i, w_r, v_r)}, \frac{v_l(\lambda_i - w_l + v_l)}{g(\lambda_i, w_l, v_l)}, \frac{1}{\lambda_i}\right]. \quad (3.4)$$

The normalizing factor $c_i(\lambda_i)$ is determined by $T_L^{-1} T_L = 1$.

As an example, eq. (2.14) can be written as

$$Z^{(1)} = e T_L \Lambda_L^l T_L^{-1} W_{LLD} W_{LDD} T_D \Lambda_D^l T_D^{-1} \times W_{DDL} W_{DLL} T_L \Lambda_L^l T_L^{-1} e^+ . \quad (3.5)$$

From eqs. (3.3), (3.4) eT_L and $T_L^{-1} e^+$ are given by

$$eT_L = (1, 1, 1, 1, 1), \quad T_L^{-1} e^+ = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{pmatrix}. \quad (3.6)$$

The solutions of eq. (3.2) are obtained in power series with respect to parameters v_r and v_l. Under the condition $w_r \neq w_l \neq 1$ they are, up to the second order,

$$\begin{aligned} \lambda_1 &= w_r + \frac{v_r^2}{w_r(w_r - 1)} + \cdots, \\ \lambda_2 &= w_l + \frac{v_l^2}{w_l(w_l - 1)} + \cdots, \\ \lambda_3 &= 1 + \cdots - \frac{w_r}{w_r - 1} v_r^2 - \frac{w_l}{w_l - 1} v_l^2, \\ \lambda_4 &= -v_r + \frac{w_r + 1}{w_r} v_r^2 + \cdots, \\ \lambda_5 &= -v_l + \frac{w_l + 1}{w_l} v_l^2 + \cdots. \end{aligned} \quad (3.7)$$

We confine ourselves to such circumstances where the helical form is more stable than the random coil. This means that we may assume

$$w_r > w_l > 1, \quad \lambda_1 > \lambda_2 > \lambda_3 > \lambda_4, \lambda_5,$$

$$w_l' > w_r' > 1, \quad \lambda_1' > \lambda_2' > \lambda_3' > \lambda_4', \lambda_5',$$

where $\lambda_1', \cdots, \lambda_5'$ are the eigenvalues of W_D. In the following we shall study two simple cases in detail.

§4. Model A: A Block Copolymer Composed of n L-Residues Followed by m D-Residues

$$\underbrace{\mathrm{LLL} \cdots \mathrm{LLL}}_{n}\underbrace{\mathrm{DDDD} \cdots \mathrm{DDD}}_{m} \qquad n, m > 1$$

The partition function is given by

$$Z_{L^n D^m} = e W_L^{n-1} W_{LLD} W_{LDD} W_D^{m-1} e^+ . \quad (4.1)$$

This is simplified by transforming W_L and W_D to their corresponding diagonal matrices Λ_L and Λ_D. As n and m are assumed large, we may neglect eigenvalues other than the largest ones λ_1 and λ_1'. The expectation number of L-residues which are forced to take left-handed helices near the joint of D- and L-residues is given by

$$\Delta\theta = \frac{\partial \ln(Z_{L^n D^m}/Z_0)}{\partial \ln w_l} + \frac{\partial \ln(Z_{L^n D^m}/Z_0)}{\partial \ln \xi}, \quad (4.2)$$

where $Z_0 = Z_L^n \times Z_D^m$ is the partition function for two independent homopolymers of D- and L-amino acids. Z_L^n and Z_D^m are given by

$$Z_L^n = \sum_{i=1}^{5} c_i \lambda_i^n \cong c_1 \lambda_1^n$$

$$Z_D^m = \sum_{i=1}^{5} c_i' \lambda_i'^m \cong c_1' \lambda_1'^m . \quad (4.3)$$

$\Delta\theta$ is a function of parameters w_r, w_l, v_r, v_l, η and ξ. The values v_r and v_l correspond to v in the theory of Lifson and Roig for homopolymers which represents the sharpness of the helix-coil transition. Zimm, Doty and Iso[1] have obtained the value $v = 2^{1/2} \cdot 10^{-2}$ for poly-γ-benzyl-L-glutamate from the study of the molecular weight dependence of the transition. According to Zimm and Bragg[4] the value of v is almost independent of the particular polypeptide solvent system. Thus we adopt $v_r = v_l = 2^{1/2} \times 10^{-2}$ in our case. No experimental information is available for η and ξ, but we may assume $\eta = \xi = (w_r w_l)^{1/2}$ for simplicity. Then $\Delta\theta$ is given as a function of two parameters w_r and w_l. In Fig. 2 $\Delta\theta$ is plotted as a function of the ratio w_l/w_r for various values of w_r. We see that $\Delta\theta$ is determined predominantly by the ratio w_l/w_r and rather insensitive to w_r. The value of $\Delta\theta$ is rather small and decreases rapidly with

FIGURE 4. PHOTOGRAPH OF A TRANSMITTED IMAGE AS DISPLAYED ON A TEKTRONIX EXPERIMENTAL STORAGE TUBE

COMPUTER CONTROL SYSTEM FOR AIR POLLUTION

T. Takamatsu*, Y. Sawaragi*
M. Naito*, Y. Akagi*
I. Hashimoto*, Y. Ikeda*
K. Kawata*, T. Mizoguchi**

* Kyoto University, Kyoto
** Department of Public Nuisance
Control, Osaka Prefectural
Government, Osaka
Japan

1. INTRODUCTION

There have been several researches in the field of air pollution control, but many of those are based on the stochastic correlations between the generations of undesirable pollution and the local meteorological conditions[1,2,3,4], a few are related to the mathematical models on the generation of undesirable pollution[5,6].

In Osaka Prefecture, Japan, which has many industrial cities, problems on public nuisance caused by air pollution are getting acute, and the telemetering system has been installed to always catch the pollution and the meteorological conditions at several monitoring stations in the Prefecture. Using the informations from these stations, the following emission control system is now being planned.

The conceptual diagram of control principle is shown in Fig.1. That is, the mathematical model of transportation process of pollutant can be built based on the informations from the monitoring stations mentioned above, other meteorological conditions and the amount of exhaust pollutants, and then the future pollutant concentrations in monitoring stations may be estimated by using this model. If this estimated value is expected to continue over a critical value, the amount of exhaust pollutants should be controlled so effectively that future value does not be over the critical, and moreover it is desirable that social and industrial demerits or losses by control actions are minimized in the region under the consideration.

Mathematical models are generally classified to phenomenological and stochastic models which are principally based on transport phenomena and not, respectively, but thi paper is based on the former model. However, there are many uncertainties in meteorological conditions such as motion of air, atmospheric temperature, and also in transport conditions, and then it is easily recognized that some deterministic models can not be used. In this paper, at first, an adaptive control method in which the values of parameters are modified so that the differences between calculated and measured values of pollutants concentration in monitoring stations are minimized, is taken account. However, in general, local meteorological conditions are suddenly changed, and those phenomenological predictions are difficult, therefore it may be desirable to predict the future values of model parameters from stochastic view point.

Main problem of mathematical model building is related with diffusion of pollutants in air, and many steady state diffusion equations have been reported by C.H.Bosanquet & J.L.Pearson[7], O.G.Sutton[8] etc.[9]. It is easily recognized from all these equations that transport phenomena of pollutants are strongly affected by wind velocity and direction, variances of these values, stability of atmosphere and so on, that is, by local meteorological conditions. These equations, however, are based on data obtained from experiments by flat surface and one or a simple allocation stack, and the intact application of these equations to the atmosphere on a large city may not be reasonable. In this paper, such steady state transport equations are applied to only the horizontal transportation through rather upper zone of atmosphere, and in living zone which is in the scores of meters from the surface, a pseudo perfect mixing zone is assumed in which each monitoring station is included. Then, between the concentration in upper zone based on steady state transport equations and the concentration in a perfect mixing zone, the vertical transportation of pollutant is assumed. Fig.2 shows the schematic illustration of the model mentioned above.

2. VERTICAL MIXING OF AIR

It is a well known fact that diffusional characteristics of pollutant are strongly affected by so-called the stability of atmosphere. Generally, at high wind velocity, pollutant is immediately transported in the horizontal direction by bulk flow of air and is easily diffused in the vertical direction by turbulence. On the other hand, at low velocity when the accumulation of pollutant occures, the diffusivity of pollutant in the vertical direction influences the generation of undesirable pollution.

In general, one dimensional diffusion equation of mass is

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial z}\left(D_z \frac{\partial C}{\partial z}\right) \qquad (1)$$

Assuming the analogy of heat and mass diffusions,

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\left(K_z \frac{\partial T}{\partial z}\right) \tag{2}$$

and
$$D_z \propto K_z \tag{3}$$

are recognized. From the previous explanation it may be obvious that D_z or K_z is a parameter changed by the stability of atmosphere. That is, the vertical dispersion of pollutant increases as the value of D_z or K_z becomes large. In addition, it is easily recognized that the value of D_z or K_z increases as the wind velocity increases and as the wind velocity distribution in the vertical direction du/dz decreases[10]. Also, as the temperature gradient in the positive direction increases, the wind velocity distribution becomes steeper[10]. Therefore, it may be convenient to introduce the following Richardson's number Ri for predicting the value of D_z or K_z,

$$R_i = \frac{\frac{g}{T}\left(\frac{dT}{dz} + \Gamma\right)}{\left(\frac{du}{dz}\right)^2} \tag{4}$$

Using the simple rule of turbulent flow and Richardson's number, G. Yamamoto has reported the following relation[11].

$$D_z = k^2 z^2 \left(\frac{du}{dz}\right) \cdot \left(1 - \sigma R_i\right)^{\frac{1}{2}} \tag{5}$$

The similar relation may be formed for the thermal diffusivity K_z as follows,

$$K_z \propto z^2 \left(\frac{du}{dz}\right) \cdot f(R_i) \tag{6}$$

For obtaining the value of K_z, an expression for the wind velocity gradient du/dz must be derived from some measurable variables. Previous experiences have given rise to the power law of the velocity profile written as follows[12],

$$u(z) = u(z_h) \cdot (z/z_h)^{\beta} \tag{7}$$

The two factors, β and f(Ri), were estimated by using the data on temperature and velocity profiles measured at the Tokyo Tower as follows,

$$\left.\begin{aligned} \frac{K_z}{z^2\left(\frac{du}{dz}\right)} &= f(R_i) = \alpha_1 && \text{for } R_i \geq 0 \\ \frac{K_z}{z^2\left(\frac{du}{dz}\right)} &= f(R_i) = \alpha_1 + \alpha_2 R_i && \text{for } R_i < 0 \end{aligned}\right\} \tag{8}$$

$$\beta = \alpha_3 + \alpha_4 (dT/dz) \tag{9}$$

Eq.(9) is compared with the result obtained by R. Frost[13] in Fig.3.

Fig.4 shows the relation between $K_z/z^2(du/dz)$ and Ri.

From the above condideration the following procedure for obtaining the value of K_z is summarized.

i) If the temperature distribution can be always measured at a monitoring station and if the wind velocity at any height u_h can be always measured, the value of β is known by Eq.(9), the value of du/dz is by Eq.(7), the value of Ri is by Eq.(4), and then the value of K_z at the place having a monitoring station can be obtained from Eq.(8) as a function of z.

ii) In addition to the temperature distribution, if the wind velocity distribution is always measurable, the values of α_3 and α_4 can be estimated by comparing the calculated values with the observed value, and also the values of α_1, α_2 can be determined.

In the model previously presented, it is necessary to know the diffusivity in the vertical direction between a perfect mixing zone and the upper zone, and then it is convenient to evaluate an average value of K_z. Therefore, using the diffusion equation of

$$\frac{\partial T}{\partial t} = K_{av}\frac{\partial^2 T}{\partial z^2} \tag{10}$$

the value of $K_{av}(t)$ is determined from the data of temperature gradient by the least square estimation[15]. Knowing the values of K_{av} and dT/dz, the values of α_1, α_2 can be determined, therefore, it is desirable to obtain the temperature gradients at all monitoring stations. But this realization may be almost impossible from the view points of economical and social problems. Therefore, it may be inevitable that the values of α_1, α_2 are identified only at a few stations where the temperature gradients are measurable, and that the values of K_{av} at other stations are calculated by using the above values of α_1, α_2 and the wind velocity at each station. The prediction of α_1, α_2 may be carried out by the consideration of past data, and especially it might be more practical to introduce the principle of stochastic data processing.

3. FUNDAMENTAL CONCEPTS OF MODEL

3.1 Vertical Transportation

As illustrated in INTRODUCTION and Fig.2. the concentration in a pseudo perfect mixing zone, C may be expressed by the following first order differential equation,

$$V\frac{dc}{dt} = \sum_{in} Q_{in} + \gamma(C_B - C) - \phi \tag{11}$$

where, $\sum_{in} Q_{in}$ is the total amount of exhaust pollutants per unit time in this perfect mixing zone, C_B is the average concentration of the upper zone of the mixing zone, and it is assumed that the vertical transport proportional to $(C_B - C)$ is carried out and its proportional constant is expressed by γ. ϕ is the term showing the reduced amount of pollutants by adsorption, absorption by rain, chemical reaction and others. V is the capacity of a pseudo perfect mixing zone, but this value depends strongly upon the lay of the land or the constructive conditions of buildings at that place, and then this value must be estimated from the future experiences.

The value of γ should be interrelated to the value of K_z from the phenomenal view point. If this relation is represented by the simplest form, it may be

$$\gamma = \lambda_1 + \lambda_2 K_z \qquad (12)$$

Generally, it may be roughly assumed that the term of ϕ in which various rate processes should be included is proportional to C, that is,

$$\phi = k'VC \qquad (13)$$

3.2 Horizontal Transportation

Usual steady state diffusion equations are applicable to the case of a fixed bulk direction of wind. This condition may not be formed and the main wind direction changes from one to the other direction, for rather long time interval, and then pollutants exhausted from a stack may disperse over very wide angle. In general, the angle of spread of exhaust gas from a stack, that is, the angle in which pollutant sources influencing the pollutant concentration at a monitoring station are included, is depending on the wind velocity and on the variance of derection. In this paper, the value of this angle Θ is fixed as

$$\theta = 2\pi/16 \qquad (14)$$

and the wind direction is always measured by the division into 16 directions.[14] If the fraction of the period blowing from the j-th direction during τ hours is $\Delta\tau_j/\tau$, it may be reasonable that the weight of $\Delta\tau_j/\tau$ is multiplied to the usual value of pollutant concentration which is calculated by usual diffusion equation under the condition of the constant wind direction of the j-th. The fraction of wind direction and the average wind velocity during 1 hour are measurable at each station, in Osaka.

In order to calculate the amount of exhaust pollutants, several concentric circles are drawn around each station, and many blocks are divided by these circles and the radial lines distinguishing 16 wind directions. For the convenience, pollution sources in each sector can be classified to point sources and plane source. Point sources represent strong sources such as power plants, and plane source consists of many weak sources such as vehicles or small stacks. The amount of pollutants from these sources should be determined or estimated in the preliminary investigations. Fig.5 illustrates the concepts of the horizontal dispersion model mentioned above.

For the simplicity, plane source can be replaced by line source as follows; the concentration of pollutant from a infinite line source at the distance of x_E, C_L, is given by

$$C_L = \frac{Q_L}{p u x_E} \tag{15}$$

and if the intensity of emission changes in the radial direction from x_1 to x_2, the concentration based on this region may be expressed by

$$C_L' = \int_{x_1}^{x_2} \frac{Q_{Li}}{p u x}\, dx \tag{16}$$

Since the total amount of exhaust pollutant in the sector ABCD, Q_A, should be equal for both plane and line sources, it must be

$$Q_A = x_E \theta Q_L = \int_{x_1}^{x_2} x \theta Q_{Li}\, dx \tag{17}$$

From the conditions of Eqs.(15)-(17), Q_{Li}=const., and $C_L = C_L'$, the following equivalent distance x_E for replacing place source by line source is given,

$$x_E = \left(\frac{x_2^2 - x_1^2}{2 \ln (x_2/x_1)} \right)^{\frac{1}{2}} \tag{18}$$

and then, the value of C_L can be calculated by

$$C_L = \frac{Q_A}{p u \left(\frac{x_2^2 - x_1^2}{2 \ln (x_2/x_1)} \right) \cdot \theta} \tag{19}$$

Usual steady state dispersion equations can be used to calculate the concentration caused by strong point sources, but Bosanquet & Pearson's equation[7] is used in this paper, because the number of parameters included in this equation is only two, p and q as follows;

$$C_P = \frac{Q_P}{\sqrt{2\pi}\, p q u x^2} \exp\left(-\frac{H}{p x}\right) \tag{20}$$

where, p and q are the parameters based on vertical and horizontal diffusion in perpendicular to the wind direction, respectively, and are defined by

$$p = D_z/(u z) \tag{21}$$

and

$$q^2 = D_Y/(u x) \tag{22}$$

Practically, since the wind velocity changes from time to time, it may be better to use the average wind velocity through past n hours, $\bar{u}$, for u in Eq.(20). In this case the value of n can be determined by the condition of

$$n-1 < \frac{x}{\bar{u}} \leq n \quad , \qquad (23)$$

and also the amount of exhaust pollutant, Q_p, must be the average value through n hours. Moreover, this correction can be applied also to the calculation of C_L by Eq.(19).

From the above considerations it is concluded as follows; the concentration at a station caused by the plane sources in the j-th direction can be expressed by

$$C_{Lj} = \sum_{\substack{\text{Radial} \\ \text{direction}}} \frac{(\Delta\tau_j/\tau)_n \bar{Q}_A}{\bar{p}\cdot\bar{u}\cdot x_E^2\cdot\theta} \qquad (24)$$

and for the point sources,

$$C_{Pj} = \sum_{\text{in j-th}} \frac{(\Delta\tau_j/\tau)_n \bar{Q}_P}{\sqrt{2\pi}\cdot\bar{p}\cdot\bar{q}\cdot\bar{u}\cdot x^2} \cdot \exp\left(-H/\bar{p}x\right) \qquad (25)$$

The total concentration at a station, C_B, can be represented by the summation to all direction, and it is

$$C_B = \sum_j (C_{Lj} + C_{Pj}) \qquad (25)$$

This value of C_B is related to the value of C by Eq.(11).

4. THE RESULTS OF SOME CALCULATIONS

Enough data to examine the above model have not yet been obtained, but one example of calculations was carried out by using the data of Table 1 and Table 2. Fig.6 shows the calculated values of C_B, but in this case the conditions of

$n=1\sim3$ hr, $H=50$m, $\bar{p}=0.2$, $\bar{q}=0.08$, $\tau=3$ hr, $H_c=50$m, $x_a=0.5$ km, $x_b=1$ km

$x_c=2$ km, $x_d=3.5$ km and $x_e=5.5$ km

were taken for simplicity(refer to Fig.5). The bold real line in Fig.7 shows the measured value of C at a monitoring station, and Fig.8 shows the reasonable assumed value of $\sum_{in} Q_{in}$.

Using the calculated values of C_B and the measured values of C at times, the optimum constant value of γ can be determined by the least square principle[15]. The objective function is

$$J = \sum_{\text{Sample points}} (\hat{C} - C)^2 \longrightarrow \text{Minimum} \qquad (27)$$

and the system equation is

$$V\frac{dC}{dt}=\sum_{in}Q_{in}+\gamma(C_B-\hat{C})-k'V\hat{C} \tag{28}$$

The narrow real lines in Fig.7 show the change of C satisfying the condition of Eq.(27), and Fig.9 gives the optimum values of γ at each time. These calculations were carried out by using the data of past three times every one hour.

It is necessary for air pollution control to predict the future value of C. If the future value of C_B can be predicted, the future value of C may be predicted by using Eq.(28) with the optimum γ given at present time. The future meteorological conditions must be predicted to calculate the future value of C_B, but this is very difficult. Therefore, it is forced to predict it by using only the past calculated values of C_B. In this paper, assuming that C_B can be expressed by the second order algebralic equation with time as follows,

$$\hat{C}_B=at^2+bt^2+c \quad , \tag{29}$$

the values of parameters a, b, c, are estimated under the condition of

$$\sum_{\text{Sample Points}}(C_B-\hat{C}_B)^2 \longrightarrow \text{Minimum} \tag{30}$$

The future value of C_B may be predicted from Eq.(29) with the calculated values of parameters. The broken lines in Fig.7 show the predicted values of C by using Eq.(28) in which C_B is substituted by the predicted value C_B of Eq.(29). It is recognized that the predicted values of C are well following the measured values.

5. CONCLUSION

There are many problems to be solved for air pollution control, but this paper is the first suggestion of model building for computer control system design. Especially, it is suggested to use the two kings of models for vertical and horizontal diffusions. However, the following approaches should be studied as quickly as possible to practical application of this control system. (1) The prediction of parameters γ, p, q etc. from the view points of stochastic data processing, (2) Obtaining the relation of K_z and γ. (3) Adaptive modification of the horizontal diffusion model by measuring C_B and so on. These problems are now being studied by our group.

At the final, the concept of this control system is given by Fig.10.

NOMENCLATURE

a, b, c : Parameters in Eq.(29) [-]

C : Pollutant concentration in pseudo perfect mixing zone [ppm]

C : Estimated value of concentration C [ppm]

C_B : Average pollutant concentration at the upper zone of the mixing zone [ppm]

C_B : Estimated value of concentration C_B [ppm]

C_L : Pollutant concentration at a monitoring station calculated by using equivalent line sources [ppm]

C'_L : Pollutant concentration at a monitoring station based on the weak plane source of a sector included in the j-th direction [ppm]

C_{Limit} : Upper limit of allowable concentration [ppm]

C_p : Pollutant concentration at a station from a point source [ppm]

D_y : Diffusivity based on horizontal diffusion in perpendicular to the wind direction [m^2/hr]

D_z : Diffusivity based on vertical diffusion [m^2/hr]

g : Gravitational acceleration [m/sec^2]

H : Effective emission height [m]

H_c : Height of pseudo perfect mixing zone [m]

K_{av} : Average thermal diffusivity based on vertical thermal diffusion [m^2/hr]

K_z : Thermal diffusivity based on vertical thermal diffusion [m^2/hr]

k : Kårmån constant [-]

k' : Coefficient related to the reduction rate of pollutant [1/hr]

p : Parameter based on vertical diffusion [-]

Q_A : Emission intensity in a sector [m^3/sec]

Q_{in} : Emission intensity in a pseudo perfect mixing zone [m^3/hr]

Q_L : Emission intensity of line source that is equivalent to plane source [m^2/sec]

Q_{Li} : Emission intensity of line source in the case that it changes in the radial direction [m^2/sec]

Q_o : Emission intensity outside a pseudo perfect mixing zone [m^3/hr]

Q_p : Emission intensity of point source [m^3/sec]

ΔQ : Reduced amount of emission intensity [m^3/hr]

q : Parameter based on horizontal diffusion in perpendicular to the wind direction [-]

Ri : Richardson's number [-]

T :	Temperature	[deg.]
t :	Time	[hr]
u :	Wind velocity	[m/sec]
u_h :	Wind velocity at z_h	[m/sec]
V :	Volume of a pseudo perfect mixing zone	[m^3]
x :	Distance from point source to monitoring station	[m]
x_E :	Equivalent distance from plane source to monitoring station	[m]
x_1,x_2:	Represent the radial length of adjoining circles, respectively, refer to Fig.6	[m]
z :	Height	[m]
z_h :	Nominal height	[m]
α_1, α_2, α_3, α_4 :	Parameters in Eq.(8) and (9)	[-]
β :	Parameter in Eq.(7)	[-]
Γ :	Dry adiavatic lapse rate=0.00986 °c/m	[°c/m]
γ :	Pseudo vertical transfer rate between C and C_B	[m^3/hr]
Θ :	Angle of sector	[radian]
λ_1 λ_2 :	Parameters in Eq.(12)	[-]
σ :	Parameter in Eq.(5)	
τ :	Sampling period for obtaining the average wind velocity	[hr]
$\Delta\tau_j$:	Period that wind is blowing from the j-th direction	[hr]
ϕ :	Term representing the reduced amount of pollutants in a perfect mixing zone	[m^3/hr]

REFERENCES

1. L. E. Niemeyer, Mon. Wea. Rev., 88, 88 (1960)
"Forecasting Air Pollution Potential"
2. C. M. Boettger & H. J. Smith, Mon. Wea. Rev., 89, 477 (1961)
"The Nashville Daily Air Pollution Forecast"
3. P. Williams, Jour. Appl. Meteor., 3, 92 (1964)
"Air Pollution Potential over the Salt Lake Valley of Utah as Related to Stability and Wind Speed"
4. G. C. Helzworth, Jour. Appl. Meteor., 1, 366 (1962)
"A Study of Air Pollution Potential for the Western United States"
5. D. B. Turner, Jour. Appl. Meteor., 3, 83 (1964)
"A Diffusion Model for an Urban Area"
6. M. R. Miller & G. C. Holzworth, An manuscript presented at the 59th annual meeting, Air Pollution Control Association, (1966)
"An Atmospheric Diffusion Model for Metropolitan Areas"

7. C. H. Bosanquet & J. L. Pearson, Trans. Faraday Soc., 32, 1294 (1936)
"The Spread of Smoke and Gases from Chimneys"

8. O. G. Sutton, Quart. J. Roy. Meteor. Soc., 73, 426 (1947)
"The Theoretical Distribution of Airborne Pollution from Factory Chimneys"

9. J. Sakagami, Natural Science Report of the Ochanomizu Univ., 11, 2 (1960)
"On the Relation between the Diffusion Parameters and Meteorological Conditions"

10. C. W. Thornthmaite & P. Kaser, Trans. Am. Geophys. Union, 1, 166 (1943)

11. K. Ito, ed., National Conference of Air Pollution Research, Corona (1965)
"Handbook of Air Pollution and Meteorology(Japan)"

12. E. L. Deacon, Quart. J. Roy. Meteor. Soc., 75, 89 (1949)

13. R. Frost, Meteor. Mag., 76, 14 (1947)

14. W. Ott, J. F. Clark, G. Ozdins, U. S. Department of Health, Education, and Welfare, Public Health Service Bureau of Disease Prevention and Environmental Control, National Center for Air Pollution Control Cincinnati, Ohio
"Calculation Future Carbon Monixide Emissions and Concentrations from Urban Traffic Data"

15. H. H. Rosenbrock & C. Storey, Pergamon Press, 7, (1966)
"Conputational Techniques for Chemical Engineers"

No. of point source		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Distance (Km)		9.7	9.0	8.7	6.1	6.2	8.1	4.1	9.9	7.2	3.9	2.9	8.0	3.8	3.5	2.2
No. of direction		1	1	1	1	1	1	1	2	2	2	2	16	16	16	16
Concentration (%)		2.67	2.8	2.8	3.0	3.0	2.9	3.0	3.25	3.0	3.0	2.6	2.8	2.95	2.9	2.9
Exhaust intensity (Heavy oil) (ℓ/hr) (Time)	1:00	810	4117	680	300	0	0	1740	1105	1200	0	1250	2000	1020	3200	120
	2:00	820	4117	675	400	0	0	1740	1105	1200	0	1250	2000	1040	3200	300
	3:00	830	4117	675	400	0	0	1740	1105	1200	0	1250	2000	1060	3200	300
	4:00	820	4117	665	300	0	0	1740	1105	1200	0	1250	2000	1060	3200	300
	5:00	1200	4117	665	300	0	140	1740	1105	1400	200	1250	2000	1010	3200	300
	6:00	1140	4117	680	1000	250	350	1740	1105	2750	600	1250	2000	1060	3200	300
	7:00	1000	4117	680	1250	250	910	1740	1105	2350	1050	1500	2000	1060	3200	800
	8:00	1890	4117	1030	1050	1250	930	1740	1105	1900	1100	1500	2000	1010	3200	2250
	9:00	1780	4117	1040	1250	1600	950	1770	1105	3200	1230	1500	2000	1110	3320	3150
	10:00	1750	4117	950	1250	1500	950	1770	1105	3400	1230	1500	2000	1150	3480	3150
	11:00	1670	4117	950	1150	1500	920	1770	1105	3400	1220	1500	2000	1200	3480	3150
	12:00	1520	4117	940	1150	1400	920	1770	1105	2350	1220	1500	2000	1210	3480	3150
	13:00	1610	4117	980	650	1000	579	1770	1105	3400	1220	1500	2000	1110	3320	1620
	14:00	2070	4117	980	1300	1500	1060	1770	1105	2750	1030	1500	2000	1060	3200	3150
	15:00	2110	4117	980	1300	1500	710	1770	1105	2750	1230	1500	2000	1130	3200	3150
	16:00	1940	4117	990	1150	1400	690	1770	1105	2750	1180	1500	2000	1090	3200	3150
	17:00	1590	4117	900	1100	1300	300	1770	1105	2750	1000	1250	2000	1080	3200	3150
	18:00	900	4117	910	1050	1200	463	1740	1105	1900	1000	1250	2000	990	3200	3150
	19:00	1110	4117	650	800	800	180	1740	1105	2750	1000	1250	2000	1040	3200	2350
	20:00	950	4117	720	800	700	0	1740	1105	2300	900	1250	2000	1060	3200	1100
	21:00	860	4117	700	650	500	0	1740	1105	2000	900	1250	2000	1080	3200	350
	22:00	1030	4117	730	600	300	0	1740	1105	1200	400	1250	2000	1070	3200	350
	23:00	860	4117	720	400	0	0	1740	1105	1200	0	1250	2000	1010	3200	300
	24:00	770	4117	680	400	0	0	1740	1105	1200	0	1250	2000	1000	3200	300

Table 1 - 1

	Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Meteorological conditions	Temp. (℃)	2.2	1.5	2.3	2.5	2.4	2.9	3.2	3.3	4.2	4.9	5.5	6.3	6.7	7.0	7.5	8.0	7.5	7.1	6.6
	Weather	○	○	◎	◎	◎	◎	◎	◎		∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
	Wind velocity	1.7	1.0	2.0	1.6	1.3	1.9	1.8	1.2	1.7	0.8	0.5	0.0	1.8	1.0	0.9	1.4	0.8	2.4	2.0
	Wind dir.	2	2	16	2	2	16	1	2	2	2	2	0	2	2	2	2	2	1	1

Table 2 ○ Fine ◎ Cloudy ∞ Smog

No. of plane source	Distance (Km)	Direction	SO_2 gas emission intensity	No. of plane source	Distance (Km)	Direction	SO_2 gas emission intensity
1	0.736	1	0.803 t/s	7	2.71	1	7.44 t/s
2	0.736	2	0.803	8	2.71	2	9.65
3	0.736	16	0.803	9	2.71	16	2.24
4	1.471	1	3.88	10	4.46	1	7.60
5	1.471	2	3.88	11	4.46	2	13.37
6	1.471	16	2.80	12	4.46	16	1.007

Table 1-2

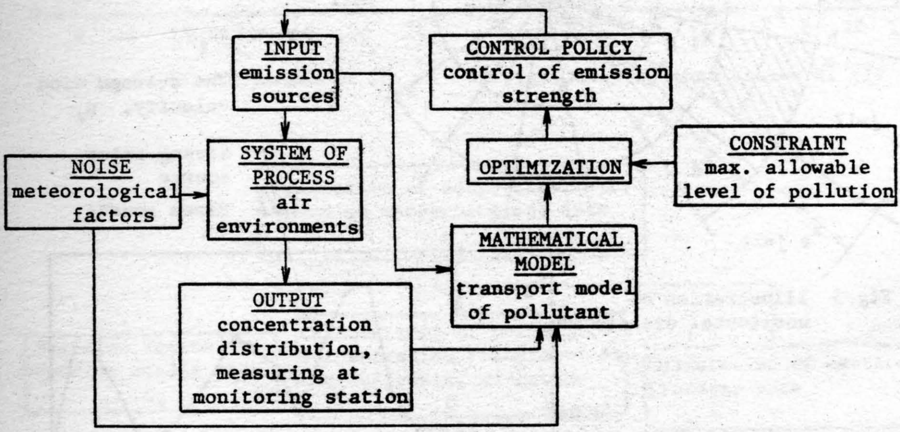


Fig.1 Conceptual diagram for air pollution control system

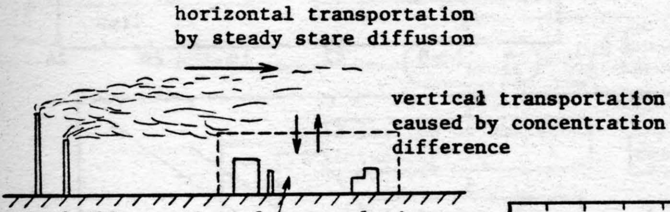


Fig.2 Illustration of accumulation transportation model

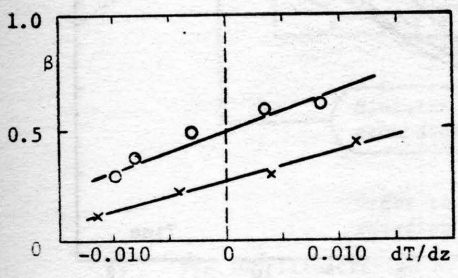


Fig.3 Value of β related to dT/dz

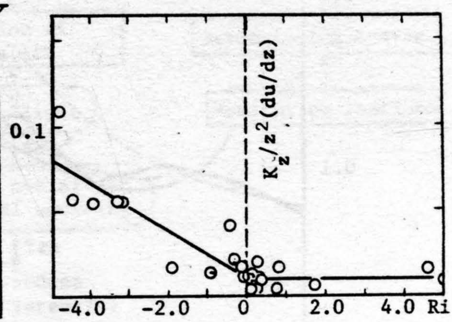


Fig.4 Relation between $K_z/z^2(du/dz)$ and Ri

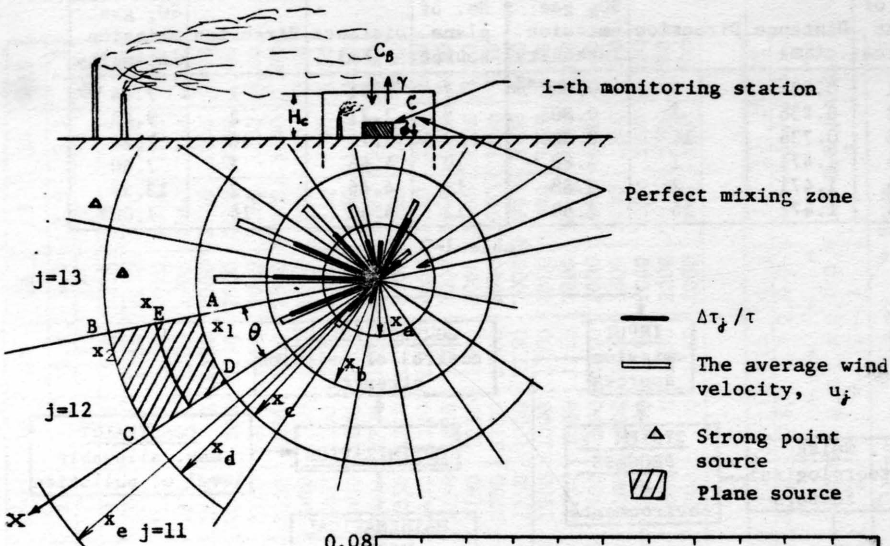


Fig.5 Illustration of horizontal dispersion

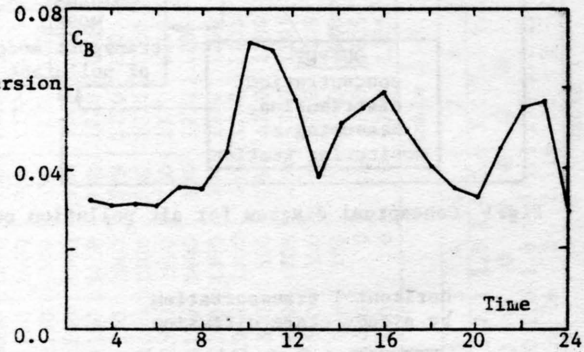


Fig.6 Calculated value of C_B

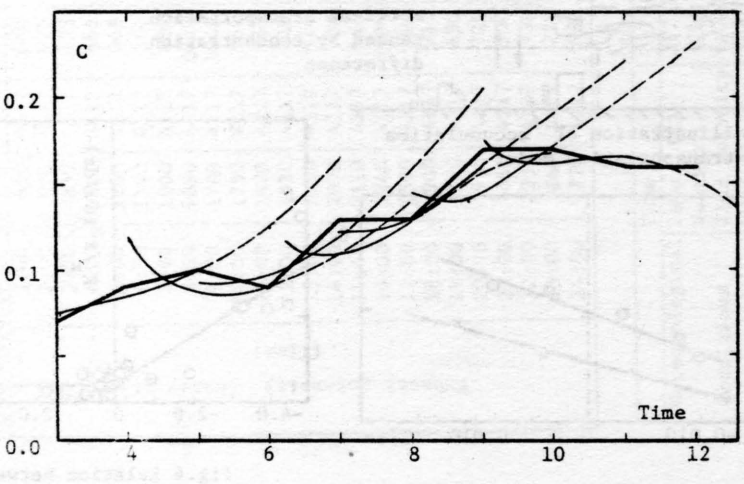


Fig.7 Measured, estimated and predicted values of C

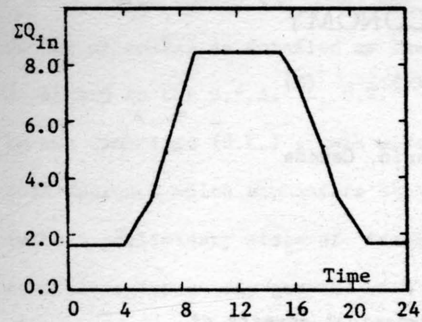


Fig.8 Assumed value of ΣQ_{in}

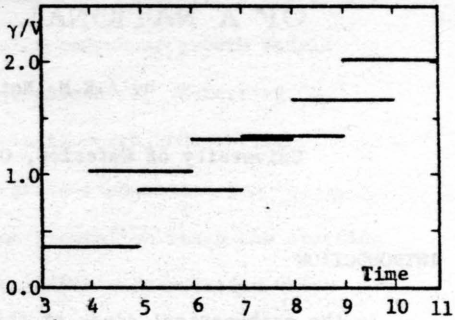


Fig.9 The optimum values of γ/V

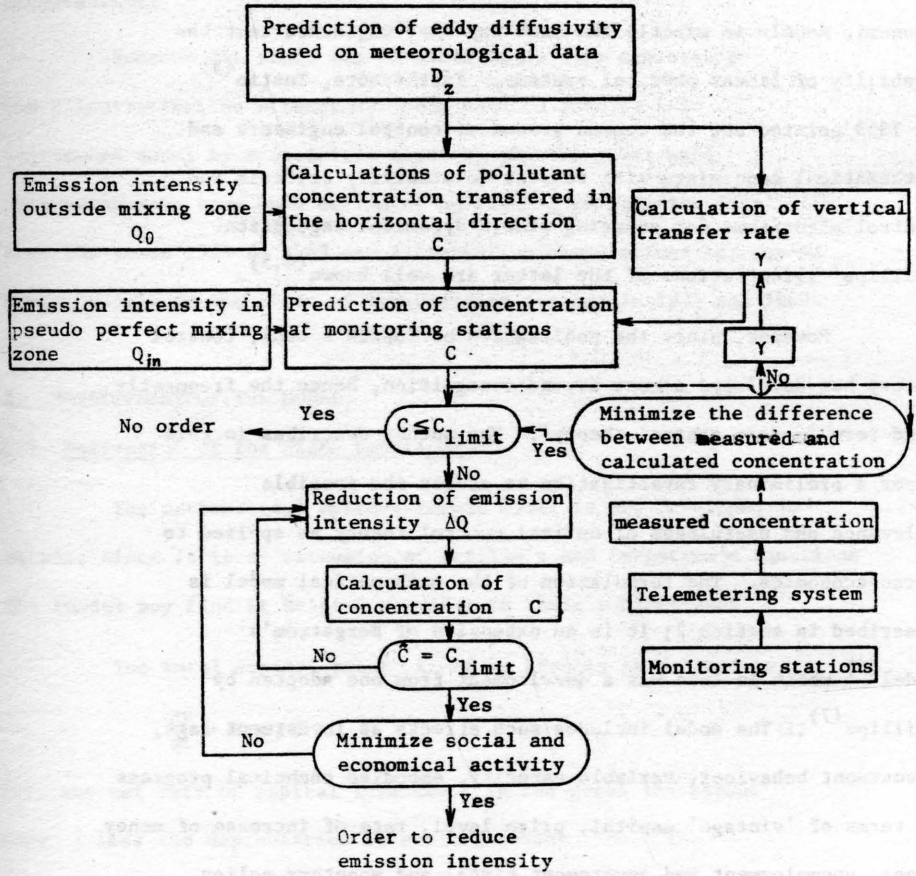


Fig.10 Illustration of calculation for control

DYNAMICALLY OPTIMIZED FISCAL AND MONETARY POLICIES FOR THE CONTROL OF A NATIONAL ECONOMY

by A.R.M. Noton

University of Waterloo, Ontario, Canada

1. INTRODUCTION

In the mathematical study of the dynamical aspects of macro-economics,[1] tests of stability have been commonplace for many years[2]. The tests have been applied to linear or linearized economic models in exactly the same way that engineers test the stability of linear physical systems. Furthermore, Tustin[3] in 1953 pointed out the common ground of control engineers and mathematical economists with respect to stability criteria and control algorithms for ensuring stable dynamical regulation. Phillips' illustrations of the latter are well known[4,5].

However, since the publication of Tustin's book, control theory has developed almost beyond recognition, hence the frequently used term 'modern control theory'. The author describes in this paper a preliminary investigation to assess the possible relevance and usefulness of optimal control theory as applied to marco-economics. The formulation of the mathematical model is described in section 2; it is an extension of Bergstrom's model[6] which in turn was a development from one adopted by Phillips[7]. The model includes such effects as investment lags, investment behaviour, variable capacity, embodied technical progress in terms of 'vintage' capital, price level, rate of increase of money wages, unemployment and government fiscal and monetary policy. It is however a 'closed' economic model i.e. foreign trade is not included.

Even though the model has a number of interesting features it is not of course as detailed as the multi-component growth models developed in the U.S.A.[8], U.K.[9] and elsewhere. Referring to the Cambridge (U.K.) growth model, it is worth contrasting that approach which emphasizes steady-state exponential growth (with a preliminary stage of 'transient growth' to reach the starting conditions for steady growth) with the author's formulation which imbeds all considerations of optimum growth, stability, transient policies and performance criteria into <u>one computation of dynamic optimization</u>.

Because the study was intended to be only exploratory and illustrative, no attempt has been made to justify this aggregated model by econometric studies. On the other hand most parameters have been estimated by fitting to Canadian data[10,11] over the years 1926 to 1967 and illustrative computations correspond approximately to the state of the Canadian economy in 1933 and 1963.

2. DEVELOPMENT OF THE MODEL

2.1 Derivation of the State Equations

The mathematical macro-economic model is now developed in detail; since it is an extension of Phillip's and Bergstrom's equations the reader may find it helpful to refer to their publications[6,7].

The <u>total</u> capital stock K of an economy is defined by

$$\frac{dK}{dt} = I - \delta K \tag{1}$$

i.e. the net rate of capital investment is the gross investment rate I less the depreciation or capital consumption. The latter is taken to be a constant fraction δ of the stock K. The integral form of (1) is

$$K(t) = \int_{-\infty}^{t} I(\tau)e^{-\delta(t-\tau)}d\tau \qquad (2)$$

where t denotes time.

The gross output, assuming no time lag between national output and national income, is taken to be ν multiplied by the capital in use and the required labour L per unit capital decreases exponentially with time at the proportional rate ρ due to technical progress. Thus, although the technical progress is embodied in the vintage of the capital, no substitution between capital and labour is permitted.

A feature of this model is variable capacity; specifically, capital stock (equipment or machinery) is not used unless all the 'younger' stock is in full use. However, unlike Bergstrom, an aggregation effect is introduced here (a) because it is more realistic and (b) because it avoids the equivalent of a transport lag in the mathematical model which would greatly complicate the numerical work. Let the probability of an entrepreneur having a machine between q and q + dq years old, as his oldest machine in use, be p(q). Then the expected value of the gross national output is

$$Y + \delta K = \nu\int_{0}^{\infty} p(q)\{\int_{t-q}^{t} I(\tau)e^{-\delta(t-\tau)}d\tau\}\, dq \qquad (3)$$

where Y is the net national income.

Changing the order of integration gives

$$Y + \delta K = \nu\int_{-\infty}^{t} I(\tau)e^{-\delta(t-\tau)}\{\int_{t-\tau}^{\infty} p(q)dq\}\, d\tau \qquad (4)$$

For illustrative purposes a zero order Poisson distribution was used for p(q). Although higher orders would probably be more realistic it would involve extra state equations. If

$$p(q) = \lambda e^{-\lambda q} \qquad (5)$$

where $1/\lambda$ is the average of the ages of the oldest machines in use, then (5) becomes

$$Y = \nu \int_{-\infty}^{t} I(\tau)e^{-(\delta+\lambda)(t-\tau)}d\tau - \delta K \tag{6}$$

and differentiating with respect to time yields

$$\frac{dY}{dt} = \nu I - \delta \frac{dK}{dt} - (\delta+\lambda)(Y + \delta K) + \frac{\partial Y}{\partial \lambda} \cdot \frac{d\lambda}{dt} \tag{7}$$

Following the previous statement about the required labour L, corresponding to equation (6), one writes

$$L = Lu \int_{-\infty}^{t} I(\tau)e^{-(\delta+\lambda)(t-\tau)-\rho\tau}d\tau \tag{8}$$

Lu being a factor depending merely on units of I and L. Equation (8) incorporates the same aggregation effect used in (6). Differentiating with respect to time

$$\frac{dL}{dt} = ILu\ e^{-\rho t} - (\delta+\lambda)L + \frac{\partial L}{\partial \lambda} \cdot \frac{d\lambda}{dt} \tag{9}$$

To obtain an expression for $\partial Y/\partial\lambda$, differentiate (6) partially with respect to λ.

$$\frac{\partial Y}{\partial \lambda} = \nu \int_{-\infty}^{t} I(\tau)\ (t-\tau)e^{-(\delta+\lambda)(t-\tau)}d\tau$$

$$= -t(Y + \delta K) + e^{-(\delta+\lambda)t}\ \nu \int_{-\infty}^{t} I(\tau)\tau\ e^{(\delta+\lambda)\tau}d\tau \tag{10}$$

Introduce a state variable X_8, defined by

$$\dot{X}_8 = I(t)t\ e^{(\delta+\lambda)t} \tag{11}$$

where $X_8(-\infty) = 0$, permits (10) to be rewritten as

$$\partial Y/\partial\lambda = -t(Y + \delta K) + \nu X_8 e^{-(\delta+\lambda)t} \tag{12}$$

Similarly, by introducing another state variable X_9, defined by

$$\dot{X}_9 = I(t)\ te^{(\delta+\lambda-\rho)t} \tag{13}$$

where $X_9(-\infty) = 0$, provides a corresponding expression for $\partial L/\partial\lambda$ to be derived from (8), viz:

$$\partial L/\partial\lambda = -tL + X_9 L_u e^{-(\delta+\lambda)t} \tag{14}$$

Consider now the autonomous variables T_I (indirect taxation), T_D (direct taxation) and G_c (government spending on goods and services). There is the identify

$$\frac{dK}{dt} = Y - C \tag{15}$$

where C is the total consumption less indirect taxation. The following linear law for domestic consumption was assumed (customary in such work).

$$C + T_I - G_c = C_1 + (1-s)(Y-T_D) \tag{16}$$

where s is the propensity to save and $C_1 = C_o \exp(\rho+\mu)t$. (where μ is the growth rate of the available force), hence the inclusion of such a factor with the autonomous term C_o. Equations (15) and (6) are combined as

$$\frac{dK}{dt} = sY - A \tag{17}$$

where

$$C = C_1 - X_{10}Y \tag{18}$$

if

$$X_{10} = \left(T_I + (1-s)T_D - G_c\right)/Y \tag{19}$$

X_{10} can therefore be regarded as a composite autonomous fiscal variable, henceforth referred to as the 'taxation variable'.

The lagged sum of government investment and private investment equals the gross investment I. Entrepreneurs determine the level of private investment and their behaviour has been taken to be described basically by the Phillips-Bergstrom function[5,6]. It includes one term for the anticipated growth rate (based on experience over the last T_2 units of time) and another term R_m corresponding to maximization of revenue by entrepreneurs. In fact the net proportional rate of increase of capital is assumed to be given by

$$\dot{K}/K = \frac{1}{(1+T_1D)^2}\left\{\frac{\alpha}{(1+T_2D)}\dot{Y}/Y + \gamma R_m + G_1/K\right\} \quad (20)$$

where $D = d/dt$ and G_1/K is planned government investment per unit capital stock.

The first term in brackets of (20) is the lagged rate of growth of national income, weighted by α. The term R_m is the marginal revenue from capital investment. Thus, if we use the classical revenue equation (aggregated for the whole economy)

$$R = Y - c(W/P)L - rK \quad (21)$$

where W is the money wage rate, P the price level, r the real interest rate (money rate less expected rate of increase of P) and c is a factor introduced dependent on the measure of wage rate W. Entrepreneurs seek to minimize R by adjustment of K and L, through the capacity variable λ. Thus, specifying that W/P and r cannot vary instantaneously

$$dR = \left[\partial Y/\partial\lambda - c(W/P)\partial L/\partial\lambda\right]d\lambda + \left[\partial Y/\partial K - c(W/P)\partial L/\partial K - r\right]dK \quad (21)$$

For a maximum, necessary conditions are

$$\partial Y/\partial\lambda = c(W/P)\,\partial L/\partial\lambda \quad (22)$$

$$R_m = \partial Y/\partial K - c(W/P)\partial L/\partial K - r = 0 \quad (23)$$

Consider a hypothetical burst of investment $dI = dKD(t)$ where $D(t)$ is the delta function; it follows from equations (6) and (8) that $\partial Y/\partial K = \nu - \delta$ and $\partial L/\partial K = L_u e^{-\rho t}$. Equations (23) becomes

$$R_m = \nu - \delta - c(W/P)L_u e^{-\rho t} - r \quad (24)$$

Entrepeneurs seek to satisfy equation (24) by investment, hence the inclusion of R_m in the investment function (20). This term differs from Bergstrom's expression because he derived an (approximate) production function to yield $\partial Y/\partial K$, whereas the author has not made such an approximation.

The money market was described along Keynesian lines[1] with

$$r = \kappa + \eta \log(PY/M) \quad (25)$$

where κ and η are constants. The equation is supposed to hold for small variations in price level P, national income Y and money supply M. It is more convenient to rewrite (25) as

$$r = X_{11} + \eta \log (PY) \quad (26)$$

where $$X_{11} = \kappa - \eta \log M \quad (27)$$

so that X_{11} can be regarded as a composite autonomous variable for the implementation of monetary policy, since it includes the effects of both changes on the interest rate and the money supply. In other words, this simple model does not distinguish between the various instruments of monetary policy.

Finally, for the labour market, Phillip's equation[10] was adopted. It was derived empirically from U.K. data and relates unemployment to the rate of increase of money wages

$$\dot{W}/W = - a_1 + a_2 U_n^{-a_3} \quad (28)$$

where a_1, a_2 and a_3 are empirical constants and the present unemployment is

$$U_n = 100 (1 - Le^{-\mu t}/L_1) \quad (29)$$

assuming that the available labour force grows as

$$L = L_1 e^{\mu t} \quad (30)$$

All the essential equations have now been introduced, in the process of which certain state variables have been named. Further manipulation is necessary to derive all the state equations (first order differential equations) but, since it is of a tedious nature it has been omitted. Instead the equations are summarized in appendix 1, with the following substitutions: $X_1 = \lambda$, $X_2 = K$, $X_5 = \dot{K}/K$, $X_6 = L$ and $X_7 = W$. Reference to that set indicates that the fiscal variable X_{10} (equation (19) has been written as $\dot{X}_{10} = u_1$ and the monetary variable as $\dot{X}_{11} = u_2$. In the terminology of optimal control theory u_1 and u_2 are the controlling variables and they correspond to changes in fiscal and monetary policy. This form is convenient since it permits the setting of initial values of X_{10} and X_{11} (e.g. taxation levels and the long term interest rate) according to the example under study. At the same time, from inspection of equations (20), (24) and (26) it will be observed that, due to the simplicity of this model, changes of planned government investment per unit capital stock have the same effect as changes of interest rate. In other words, only the combined term $(\gamma r - G_I/K)$ and not γr is significant, i.e. pure monetary policy cannot be distinguished.

The macro-economic model consists then of a set of 11 state equations with 2 controlling variables; they can be referred to generally in the form

$$\dot{\underline{X}} = \underline{\phi}(\underline{X}, \underline{u}, t) \tag{31}$$

where $\underline{\phi}$ is a vector valued nonlinear function. It is proposed to study discrete-time control. In other words $\underline{u}$ will be held constant in the kth sub-interval of duration h where $kh \leq t < (k+1)h$, because it is assumed that one might ultimately employ such techniques in real time when accounting data becomes available at say three month intervals. In that case it is appropriate to discuss the discrete-time form of (31), viz.,

$$\underline{X}(k+1) = \underline{f}\ (\underline{X}(k), \underline{u}(k), k)$$

However, it is important to note that, due to the complicated form of the model, no analytical expressions were available for the function f. The model was set up in continuous time and f was always computed numerically from (1), i.e. the treatment is not the 'period analysis' often used by economists.

2.2 The Performance Criterion

The performance criterion must now be formulated since the controlling variables are to be adjusted to minimize such a function, subject to the state equations (appendix 1), over a finite number of future sub-intervals of time (each of length h). It is proposed to maximize the national consumption C (a variable measured naturally in real and not monetary terms), while at the same time keeping unemployment and price increases in check. Hence one can expect to minimize a function of the form

$$\sum_{k=1}^{N}\{-\phi_1(C(k)) + \phi_2(U_n(k)) + \phi_3[(\frac{dP}{dt} / P)_k]\} \qquad (33)$$

where future time has been discretized as $(k-1)h$. ϕ_1, ϕ_2 and ϕ_3 are odd monotonic increasing functions. It is common, however, to modify such criteria.

Specifically, being realistic about modern political democracies although admitting to a controversial issue of the social sciences, a factor $\exp\ -(k-1)h/T_3$ is introduced to give more weight to the immediate future than the end of the optimization interval, which may be 20 to 30 years in the future.

Unfortunately, the choice of the functions ϕ_1, ϕ_2, ϕ_3, is somewhat arbitrary although perhaps not at all critical. For example, in studies of optimum growth, Cass[11] has shown that the form of a function corresponding to ϕ_1 does not actually affect the optimal growth path. In the absence of any specific forms for the three penalty functions the author has therefore chosen the following for the final version of (33):

$$\sum_{k=1}^{N} \left\{ -\log C(k) + S_2 \exp(U_n(k)/3) + S_1 \left(\frac{dP}{dt}/P\right) \right\} \exp(-(k-1)h/T_3) \tag{34}$$

The penalty term on unemployment is merely a choice of a function which increases rapidly when $U_n > 3$ per cent. The logarithmic function of C has been adopted to avoid that term growing exponentially compared to the others and hence causing difficulties in setting the weighting parameters. The latter are set by trial-and-error depending on computed variations of the variables in expression (34), for different typical starting conditions. Experience with other applications of optimal control theory has indicated that the choice of s_i is not critical.

2.3 Numerical Values of Parameters

Because only illustrative results were sought, no attempt has been made (at the time of writing) to confirm, modify or develop the model. On the other hand where possible parameters have been estimated by fitting to Canadian National Accounting Data[12,13]. Initial conditions on the state variables were adjusted to correspond approximately to the state of the Canadian economy (a) in 1933 and (b) in 1963, i.e. during a recession and a relative 'boom' period respectively. Numerical values are quoted with explanatory notes below, where the numbers in parentheses are the years over which fitting was implemented.

C_o = 2.60 (26-45) and 7.40 (48-67)

μ = 0.011 (26-45) and 0.022 (48-67)

η = 0.040 (36-67) estimated using equation (25) with the money supply M as the assets of Canadian banks and r as the average long term yield on Canadian securities. Yields were not however available before 1936.

δ = 0.051 (48-67). This figure was used for 1933 even though that 'non-exponential' period was avoided for fitting purposes.

ν = 0.40 (26-67) estimated by <u>comparison</u> with the condition for steady exponential growth $Y/K = \nu C_p - \delta$, which can be deduced from equations (1 and 7). C_p is the proportion of capacity in use.

ρ = 0.010 (1933) and 0.019 (48-67). The figure for the thirties is no more than a suggested value since rough fitting over 1926-45 gave 0.028, yet over 1926-35 the value was -0.039.

a_1 = 0.009, a_2 = 0.12, a_3 = 1.0 (26-67), slightly modified from Phillips' values(10) (in appropriate units) for Canadian data.

The law represents however no more than a trend and is probably one of the weakest points of the model.

s = 0.50 (26-67).

The unit of the real variables K, I, Y and C was taken to be billion (10^9) dollars at 1957 price levels. Total capital stock K was taken as the value given by capital consumption divided by depreciation rate δ, although this interpretation may be in doubt in the recession period. The inaccessible or unmeasured state variables X_1, X_8, and X_9 were estimated from estimates of the previous investment growth (equations 11 and 13) and the price equation (22). Other state variables were available from national accounting data.

It should perhaps be mentioned that in attempting to develop and use a more realistic aggregated model it is anticipated that recursive statistical nonlinear estimation procedures would be adopted for parameter estimation and treatment of inaccessible state variables.[19,20].

3. COMPUTATION OF THE CONTROL POLICIES

3.1 Numerical Techniques

The set of equations (31) correspond to those in appendix 1 and equation (32) denotes the equivalent of that set for discrete intervals of time. Optimal control is defined to be $\underline{u}(k)$ $(k = 1,2,\ldots N)$ which minimizes the performance criterion (34), now written as

$$V_r = \sum_{k=r}^{N-1} L\left[k,\ \underline{x}(k),\ \underline{u}(k)\right] + L\left[N,\ \underline{x}(N)\right] \qquad (35)$$

where $r = 1$. The last term is different because $\underline{u}(N)$ would be effective only outside the range of summation.

The problem is now in a form suitable for formal solution either by Dynamic Programming or the Discrete Maximum Principle. Since the state equations are nonlinear and the criterion (34) is non-quadratic a closed-form solution is not available and numerical iterative procedures must be employed. A simple first order gradient method in the function space $\underline{u}(k)$ was not used because experience has shown that final convergence is often so slow that it is difficult to know when to stop iterating and how closely the truncated sequence has approached the exact result. Methods which possess the property of final second order convergence are consequently preferable. One could proceed in terms of the Maximum Principle and use a discrete version of a second variation method[14] or, in terms of dynamic programming, using Mayne's formulation.[15] However, both these methods require second partial derivatives of f in equation (32) which, because analytical expressions for f do not exist, would incur excessive extra computing time. Noton's method[16] avoids using second partial derivatives of f but it does not give true second order convergence. The method of conjugate gradients[17], has therefore been applied to a first order expansion of dynamic programming; such a procedure requires only first derivatives yet it yields final second order convergence in a finite number of steps. Note that no constraints have been applied to the state or control vector.

The derivation of the numerical algorithms is not presented here since it follows easily, given a familiarity with the method of conjugate gradients for algebraic functions[18] and Mayne's formulation by dynamic programming[15]. The computing sequence is as follows, where superscripts denote iteration number:

1. Minimize $V_1(\underline{u}^i(k) + \alpha\,\underline{p}^i(k))$ by a linear search in α.
2. Set $\underline{u}^{i+1}(k) = \underline{u}^i(k) + \alpha\,\underline{p}^i(k)$
3. Use subroutine S to produce $G^{i+1}(j,k)$
4. Put

$$\beta^{i+1} = \frac{\sum_{k=1}^{N-1}\sum_{j=1}^{s} G^{i+1}(j,k)^2}{\sum_{k=1}^{N-1}\sum_{j=1}^{s} G^{i}(j,k)^2}$$

5. $\underline{p}^{i+1}(k) = -\underline{G}^{i+1}(k) + \beta^{i+1}\underline{p}^i(k)$ and return to #1.

for s controlling signals and (N-1) sub-intervals. The iterative process is started by entering subroutine S with some arbitrary $\underline{u}^1(k)$ and using $\beta^1 = 0$.

Referring to the linear search, $V_1(\alpha)$ is computed from equation (35) for different values of α. Increasing or decreasing steps were taken until $V_1(\alpha)$ starts to increase; the minimum was then located by quadratic interpolation through the last three points. Other authors have used the gradients $dV_1/d\alpha$ to permit cubic interpolation but, in the discrete-time case, computation of the gradient is very expensive in terms of computer time.

Subroutine S provides the gradient functions $G(j,k) = \partial V_1/\partial u(j,k)$. First $\underline{x}^i(t)$ and $\dot{\underline{x}}^i(t)$ (for cubic interpolation) are generated by integrating the state equations, given $\underline{u}(k)$, and stored every interval of integration. The gradient functions are then obtained by stepping the following two sets of equations backwards from k = N - 1 to k = 1.

$$W(j,k) = \frac{\partial L(h)}{\partial x(j)} + \sum_{i=1}^{m} W(i, k+1) \frac{\partial f_i}{\partial x(j)} \, j = 1,2, \ldots m \quad (36)$$

$$G(j,k) = \frac{\partial L(k)}{\partial u(j)} + \sum_{i=1}^{m} W(i,k+1) \frac{\partial f_i}{\partial u(j)}; \; j = 1,2, \ldots s \quad (37)$$

for m state equations, where $W(i,N) = \frac{\partial L(N)}{\partial x(i)}$. It will be observed that the partial derivatives $\partial f_i / \partial x(j)$ and $\partial f_i/\partial u(j)$ are required. Because no analytical expressions are available separate computations are required and it turns out that such operations account for the <u>majority of the computing time</u> in the whole process. Initially they were generated by perturbed numerical integrations of the state equations for each sub-interval. The system under study is however highly nonlinear and inevitably such a clumsy procedure gave rise to a conflict in the choice of the magnitude of the perturbations. Examples were found when the apparent 'optimum trajectory' depended on the magnitude of such perturbations. A more elegant procedure was therefore derived for <u>producing the summations</u> in equations (36) and (37) <u>directly</u> by the simultaneous integration of only (m + s) equations (per sub-interval). The equations are tedious to derive and complicated, since they involve the partial derivatives of the state equations, but there was a threefold saving in computer time. The analysis is given in appendix 2.

3.2 Illustrative Numerical Results

In order to appreciate the difficulties of obtaining numerical results with this macro-economic model is is important to realize that it is (a) a highly unstable system without control (c.f. Phillips' and Bergstrom's analytical studies[6,7] of a similar linearized version) and (b) it is significantly nonlinear. The major difficulties are associated with the sensitivity of the money wage rate to unemployment (equation 28). In the early stages of the numerical iterative process, unemployment levels can approach zero or even become negative. Equation (28) is clearly meaningless in such cases and it was therefore modified. Specifically, for U_n less than 1 per cent a linear law replaced function (28) which is continuous in magnitude and gradient with that function at $U_n = 1$.

A second modification was an approximation to an inequality state space constraint. $\lambda(=X_1)$ cannot be negative, otherwise it corresponds to using capacity that does not exist. Therefore a term $S_3\lambda(k)^2$ was added to criterion (34) where $S_3 = 0$ if $\lambda > 0$ and S_3 is set to a suitably large number (e.g. 10^5) for negative values of λ. In other words, negative excursions are penalized so heavily that only negligible overshoots occur in practice. In fact no such negative values occurred in the final trajectories presented in this paper. However, they did occur in other examples and invariably at early stages of the iterative process.

Referrring to those early stages, emphasis must be placed on the desirability of a good 'priming trajectory', i.e. first approximation to the control sequences $u_1(k)$ and $u_2(k)$. Due to the inherent instability of the system the computations often could not even be started without some adjustment of the controlling variables. Otherwise meaningless values, such as negative national income, might occur with a resultant stoppage in computer subroutines. One approach is to linearize the state equations and, with a second order expansion of function (34), use the easily calculated linear control law from dynamic programming. Alternatively one might try a sort of 'flooding technique' in which large numbers of trial trajectories are computed approximately for a family of functions for $\underline{u}(k)$ (say exponentials). The first proposal was rejected due to the nonlinearities of the system. The second was tried and rejected as being unworkable with an unstable set of equations. The following approach was found most reliable. Using a coarse interval of integration for fast approximate results, a small number of iterations are carried out for a number of sub-intervals, sufficiently small so that instability is not troublesome. The number of sub-intervals is then increased and more iterations carried out and so on. In other words, the optimization interval is increased in steps, the computer programe being written to step automatically of course. Once the full number of sub-intervals is in use a much larger number of iterations is employed, depending on the rate of convergence, and with a refined interval of integration.

Two examples of convergence are shown in figure 1 where the magnitude V_1 of the performance criterion is plotted* against number of iterations with the full number of subintervals. There were ten subintervals (corresponding to a period of five years in the model) so that, with two controlling signals, final convergence should

---

*Discontinuities correspond to re-started computations.

occur in 20 iterations[18], i.e. when the trajectory is near enough to the optimum for the validity of a second order expansion of V_1. There are two causes for concern in accepting such results as 'optimal', (a) only a necessary condition for relative minima is in use and (b) uncertainty that no further reduction in V_1 can be achieved by more iterations. As a form of check on relative versus absolute minimality, once one trajectory had been computed, the priming trajectory was changed considerably and yet, by manipulation of u_1 and u_2, kept within reasonable bounds. No examples were found of other local minima. The second difficulty of final convergence is probably more cause for concern due to the slow rate of convergence, e.g. fig. 1. Trajectories were therefore accepted finally only after V_1 had remained constant to a significant number of decimal digits for at least 5(N-1) iterations. Fletcher and Reeves[18] adopted the same criterion in their work on function minimization. Nevertheless some doubt still remains since the iterative process was found to be very dependent on the accuracy of the gradient functions and interpolation from the stored trajectory. On the other hand, only very slight changes occurred in the computed trajectories for the last fifty iterations.

All computing was carried out on the University of Waterloo IBM Computer, Systems 360/Model 75 using a WATFOR compiler; WATFOR is a subset of FORTRAN with slower execution time but much faster compile time and more elaborate error diagnostics. The compiled program required approximately 24,000 words of storage and typically 6,000 double words were assigned for number storage in double precision (14 decimal digits). As an example of computing time with 10 optimization sub-intervals spanned by 20 steps of fourth order Runge-Kutta numerical integration, the average time per iteration was 5.4 seconds. At a late stage of the work however the program was compiled in Fortran-G resulting in a threefold improvement in execution time.

Attention is now turned again to the exercise in a macro-economics. Reference to accounting data was made in section 2.3 for the estimation of parameters. However four parameters remain (T_1, T_2, α and γ) that can be estimated only by a major study of attempting to fit the interacting model equations to accounting data. Such a study is beyond the scope of this exercise and consequently one must accept merely illustrative values that have only the correct order of magnitude. For example, the time constant T_1 in double investment lag (the 'double' being little more than a guess) was set at 0.5 year from such qualitative hints as 'monetary controls since the war have required from six to nine months to produce a change in the direction of ease and a further six months for their maximum effect'.[21] T_2 is suggested as the order of 2 years, i.e. the memory time constant of entrepreneurs in anticipating the future growth rate. There is however, one condition which helps to fix α and γ, namely a steady state condition for a growth rate k from equations (20) and (24).

$$\alpha k + \gamma R_m + G_1/K = k \tag{38}$$

Insertion of numerical quantities for 1933 and 1963, with a typical <u>real</u> long-term interest rate of 3 per cent in R_m, suggest that $\alpha = 0.5$ and $\gamma = 0.1$ should be representative values. Such values have been adopted for the results presented below.

Finally, although they are not part of the model, the weighting parameters S_1 and S_2 of criterion (34) have to be adjusted and the social discounting time constant T_3 set. In principle these are to be chosen by government policy, or at least some federal agency, but in practice such decisions would not be taken without seeing the effects of different weightings. In these computations T_3 has been set at 10 years; since results only for an optimization interval of 5 years are presented, the discounting for later years is therefore only slight. Referring to the terms in (34), depending on price increases and unemployment, it turns out that the rate of increase of prices is approximated by the rate of increase of money wages. However, the latter is a function of the unemployment (through equation 28) and consequently in this model both terms are probably superflous. C(17) and C(18) have been set by trial-and-error to yield 'acceptable' behaviour of an economy.

It has been mentioned that parameters and initial conditions have been set to correspond roughly to the Canadian economy in 1933 and 1963 respectively. One should not expect however much correspondence and any comparison of computed and actual behaviour is not very meaningful for the following reasons: (a) The model has not been confirmed by econometric studies. (b) It has obvious simplifications such as no foreign trade. (c) Crucial growth parameters are ρ (technical progress or the growth rate of real output per unit labour) and μ (the rate of increase of the labour force). Long term average values have been used, yet these parameters vary considerably over a few years. For example, for the 1963-68 illustration $\rho = 0.019$ whereas the Canadian productivity growth rate was 2.7 per cent per annum between 1960 and 66 but 1.2 per cent per annum between 1963 and 67 (trailing alarmingly incidentally all the industrialized nations). Setting parameters for the 1933-38 interval is especially uncertain; for example the growth rate of real national net income Y between 1933 and 38 was 8.9 per cent per annum but such a value must be much higher than that for steady growth, i.e. there was a lot of unused capacity in 1933. In examining growth rates of Y and real total consumption C, it is therefore more meaningful to compare them to the values for sustained exponential growth as determined by the set numerical values in the model, viz.

$$\left.\begin{array}{ll}(\rho+\mu) = 0.021 & (1933\text{-}38) \\ (\rho+\mu) = 0.041 & (1963\text{-}68)\end{array}\right\} \quad (39)$$

Such rates are to be compared with the following initial values that were adopted for illustrative purposes:

growth rate expected by entrepreneurs, X_3 = 0.001 (1933)
" 0.061 (1963)

actual growth rate of capital, X_5 =-0.018 (1933)
" 0.041 (1963)

Four sets of results are presented in figures 2, 3, 4, and 5, two starting in 1933 and two in 1963. It is observed that a fivefold change in the weighting parameter on unemployment has been implemented to illustrate (a) how such parameters need be set only approximately and (b) how trial-and-error computations along these lines permit the setting of the weighting parameters dependent on permissible levels of unemployment. X_{10} shows changes of the composite taxation variable as a proportion of the national income (equation 19). In examining r, recall that it is the real interest rate (money rate less expected rate of increase of the price level). Nevertheless, the simplicity and limitations of the model become evident from the computed changes of r, according to monetary policy, since large variations are suggested. Following the observation in section 2.1 above equation (31), the convention has been adopted of interpreting real interest rates greater than 7 per cent as equivalent to a reduction in government investment per unit capital; they have been plotted (in broken lines) as such, as a proportion of the initial values. The huge changes suggest perhaps that γ was chosen too small.

Referring to figs. 2 and 3 for 1933 there is apparently little difficulty (with this model) in overcoming the recession; the under capacity of 78 per cent is rapidly taken up and initially the taxation is reduced. Subsequently the total planned investment is eased off (by high interest rates and reduced government investment) to keep prices in check. Significant regulation by means of taxation is also indicated. Figs. 3 and 4 correspond to the 1963 case in which approximately 91 per cent of the capacity is supposed to be in use and investment reflects over optimism of the entrepreneurs. Adjustment for this 'boom' situation is largely in terms of an early steep rise in the interest rate (and cut-back in government investment) with only slight changes in the taxation level.

4. CONCLUSIONS

It will be appreciated that the objective of this study was merely introductory, to examine how useful might be the tools of modern control theory in macro-economics. The feasibility of (at least deterministic) computations has been illustrated but it has served only to underline what work must be done before any policy decisions might be taken in conjunction with an aggregated economic model. So much work has to be done to confirm relationships and fit parameters; the recursive techniques for nonlinear state estimation certainly appear highly relevant for such work. Some additions to the model are obvious and present little difficulty, e.g. foreign trade. Penalties on rates of increase of the price level then become more meaningful and international effects on the interest rate are also essential, especially for Canada in terms of U.S. rates of interest. Stochastic inputs to the system seem then inevitable. For this reason, and because some parameters will vary in a more-or-less random manner, there seems almost unlimited scope for applying approximations to adaptive stochastic control. Strategies for implementing fiscal and monetary control, in terms of the current and past data, would then emerge, i.e. closed-loop control instead of the open-loop control of the deterministic situation in this paper. Only after such researches could one hope to throw light on some of the long-standing controversies of economists, such as the relative merits of fiscal versus monetary policy.

5. ACKNOWLEDGEMENTS

The early part of the research associated with this paper was initiated while the author was a Visiting Fellow of Clare Hall, Cambridge in the Fall 1967. It was during this period that the author derived benefit from numerous discussions with Mr. D. Livesey of the University Engineering Laboratory and helpful suggestions from Dr. J.A. Mirrlees and Professor J. Meade of the Department of Applied Economics at Cambridge.

6. REFERENCES

1. Allen, R.G.D., 'Macro-Economic Theory', MacMillan, London, 1967.

2. Allen, R.G.D., 'Mathematical Economics', London, 1959.

3. Tustin, A., 'The Mechanism of Economic Systems', Heinemann, London, 1953.

4. Phillips, A.W., 'Stabilization Policy in a Closed Economy', Economic Journal, June 1954, pp. 290-323.

5. Phillips, A.W. 'Stabilization Policy and the Time-Form of Lagged Responses', Economic Journal, June 1957, pp. 265-277.

6. Bergstrom, A.R., 'A Model of Technical Progress, the Production Function and Cyclical Growth', Economica, Nov. 1962, pp. 357-370.

7. Phillips, A.W., 'A Simple Model of Employment, Money and Prices in a Growing Economy', Economica, Nov. 1961, pp. 360-370.

8. Liebenberg, M., Hirsch, A.A. and Popkin, J., 'A Quarterly Econometric Model of the United States', Survey of Current Business, 4, 5, (1966), pp. 13-40.

9. Stone, R., A Programme for Growth' Chapman and Hall, London, 1962, onwards.

10. Phillips, A.W., 'The Relation between Unemployment and the Rate of Change of Money Wage Rates in the U.K., 1861-1957', Economica, 25, No. 100, Nov. 1958, pp. 283-299.

11. Cass, D., 'Optimum Growth in an Aggregative Model of Capital Accumulation', Review of Economic Studies, 32, (1965), pp. 233-240.

12. 'National Accounts, Income and Expenditure', Dominion Bureau of Statistics (Canada).

13. 'Statistical Summary Supplements', Bank of Canada.

14. McReynolds, S.R. and Bryson, A.E., 'Successive Sweep Method for Solving Optimal Programming Problems', Proc. JACC, 1965 pp. 551-555.

15. Mayne, D., 'Second Order Gradient Method for Determining Optimal Trajectories of Nonlinear Discrete Systems', International J. of Control, 3, 1, (1966), pp. 85-95.

16. Noton, A.R.M., Dyer, P. and Markland, C.A., 'Numerical Computation of Optimal Control', IEEE Trans. Automatic Control, AC-12, 1, (1967), pp. 59-66

17. Lasdon, L.S., Mitter, S.K. and Waren, A.D. 'Conjugate Gradient Method for Optimal Control Problems', IEEE Trans. Automatic Control, AC-12, 2, (1967) pp. 132-138.

18. Fletcher, R. and Reeves, C.M., 'Function Minimization by Conjugate Gradients', Computer Journal, 7, (1964), pp. 149-154.

19. Cox, H., 'On the Estimation of State Variables and Parameters for Noisy Dynamic Systems', IEEE Trans. Automatic Control, AC-9, 1, pp. 5-12, (1964).

20. Noton, A.R.M. and Choquette, P. 'Application of Modern Control Theory to the Operation of an Industrial Process', Proc. IFAC/IFIP Symposium, Toronto, June 1968.

21. Jacoby, N.H., 'U.S. Monetary Policy', p. 201, Praeger, New York, 1964.

7. APPENDICES

7.1 Equations of the Model

Auxiliary equations:

$$Y = (X_2X_5 + C_1)/(s + X_{10})$$

$$U_n = 100(1-X_6\, e^{-\mu t/L_1}$$

$$Y_A = -t(Y + \delta X_2) + \nu\, X_8 e^{-(\delta + X_1)t}$$

$$L_A = -t X_6 + X_9 L_u e^{-(\delta + X_1)t}$$

$$P = cX_7 L_A / Y_A$$

$$r = X_{12} + \eta \log (PY)$$

$$I = (\delta + X_5) X_2$$

$$A_a = X_2(X_4 - X_5)/T_1 + X_2 X_5^2 - Yu_1 + (\rho + \mu) C_1$$

$$A_b = \nu I - \delta X_2 X_5 - (\delta + X_1)(Y + \delta X_2)$$

$$A_c = \left[A_a/(s + X_{10}) - A_b \right] / Y_a$$

State equations:

$$\dot{X}_1 = A_c$$

$$\dot{X}_2 = X_2 X_5$$

$$\dot{X}_3 = \left[A_a/((s + X_{10})Y) - X_3 \right] / T_2$$

$$\dot{X}_4 = \left[\alpha X_3 + c' + \gamma(\nu - \delta - r - cX_7 L_u e^{-\rho t}/P) - X_4 \right] / T_1$$

$$\dot{X}_5 = (X_4 - X_5)/T_1$$

$$\dot{X}_6 = ILue^{-\rho t} - (\delta + X_1) X_6 + L_A A_c$$

$$\dot{X}_7 = X_7(-a_1 + a_2 U_n^{-a_3})$$

$$\dot{X}_8 = I\, te^{(\delta + X_1)t}$$

$$\dot{X}_9 = I\, t\, e^{(\delta + X_1 - \rho)t}$$

$$\dot{X}_{10} = u_1$$

$$\dot{X}_{11} = u_2$$

7.2 Generation of Partial Derivatives of the State Transition Equations

As mentioned above, partial derivatives of the nonlinear functions f_i $(i = 1,2,...m)$ are not readily available. Consider first order variations about a nominal trajectory and let all partial derivatives be evaluated on that trajectory. From equations (31)

$$\Delta\dot{x}_i = \sum_{j=1}^{m} \Delta x_j \partial\phi_i/\partial x_j + \sum_{j=1}^{s} \Delta u_j \partial\phi_i/\partial u_j \tag{40}$$

A set of adjoint variables z_i $(i = 1,2, \ldots m)$ are introduced defined by

$$\dot{z}_i = - \sum_{j=1}^{m} z_j \partial\phi_j/\partial x_i \tag{41}$$

Put $b_{ij} = \partial\phi_i/\partial x_j$ and $c_{ij} = \partial\phi_i/\partial u_j$, then manipulation leads to

$$\frac{d}{dt}\left(\sum_{i=1}^{m} \Delta x_i z_i\right) = \sum_{i=1}^{m} z_i \sum_{j=1}^{s} c_{ij} \Delta u_j \tag{42}$$

or

$$\sum_{i=1}^{m} z_i(k+1)\Delta x_i(k+1) = \sum_{i=1}^{m} z_i(k)\Delta x_i(k) + \int_{(k-1)h}^{kh} \sum_{i=1}^{m} z_i \sum_{j=1}^{s} c_{ij}\Delta u_j dt \tag{43}$$

where $t = (k-1)h$ and $\Delta x_i(t) = \Delta x_i(k)$.

The definition of the adjoint variables is completed with the boundary conditions

$$z_i(k+1) = W(i, k+1) \tag{44}$$

Then

$$\int_{(k-1)h}^{kh} \sum_{i=1}^{m} z_i \sum_{j=1}^{s} c_{ij} \, dt \, \Delta u_j + \sum_{i=1}^{m} z_i(k)\Delta x_i(k) = \sum_{i=1}^{m} W(i,k+1)\Delta x_i(k+1) \tag{45}$$

Therefore

$$z_j(k) = \sum_{i=1}^{m} W(i,k+1)\partial f_i/\partial x_j \; ; \; j = 1,2, \ldots m \tag{46}$$

$$\int_{(k-1)}^{kh} \sum_{i=1}^{m} z_i(t)\, c_{ij}(t)dt = \sum_{i=1}^{m} W(i,k+1)\partial f_i/\partial u_j ; \; j = 1,2,\ldots s \tag{47}$$

Define an additional s variables by

$$\dot{z}_{m+j} = - \sum_{i=1}^{m} z_i c_{ij} ; \; j = 1,2, \ldots s \tag{48}$$

where $z_{m+j}(k+1) = 0$, then

$$z_{m+j}(k) = \sum_{i=1}^{m} W(i,k+1)\ \partial f_i/\partial u_j \ ; \quad j = 1,2, \ldots s \qquad (49)$$

Thus, for each sub-interval, simultaneous integration in reverse time of the variables z_i $(i = 1,2, \ldots m + s)$ provides the required summations (directly) of the partial derivatives of the state transition equations.

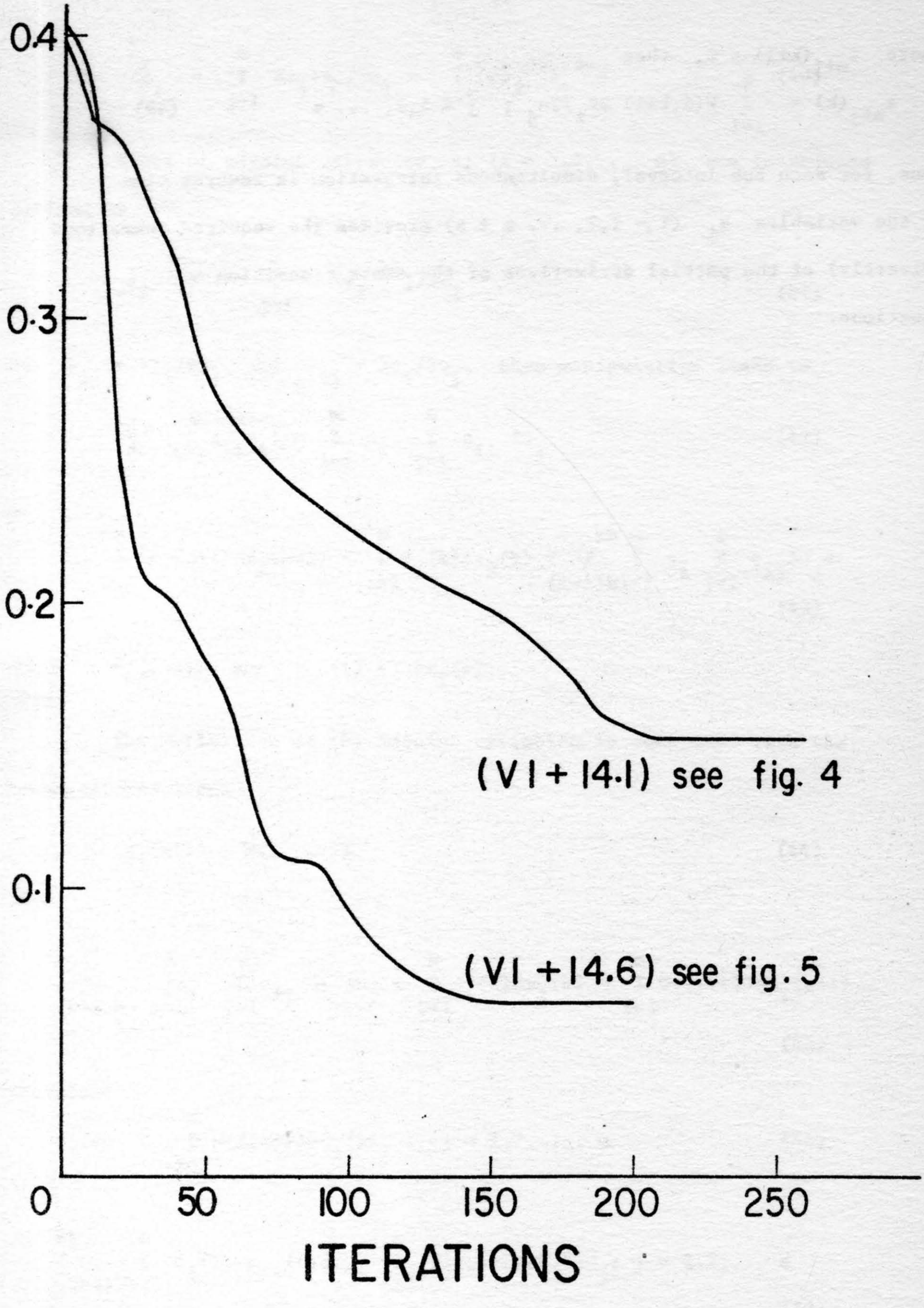
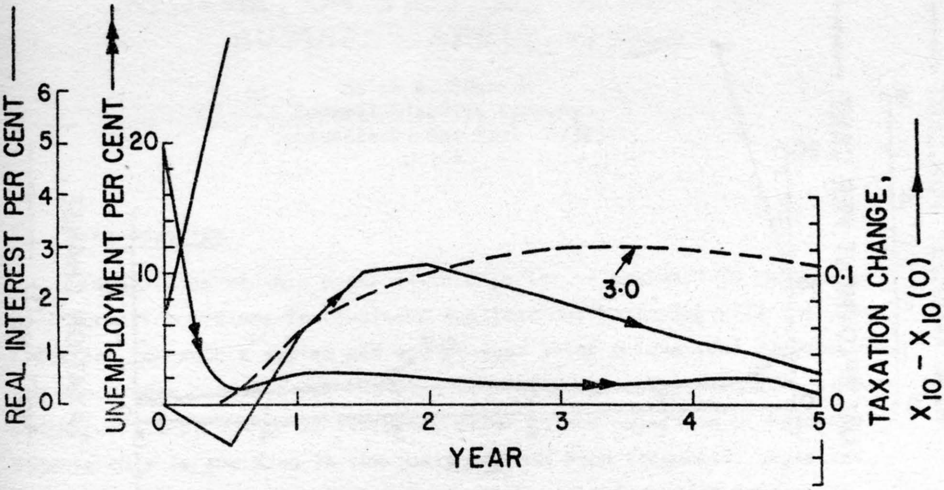
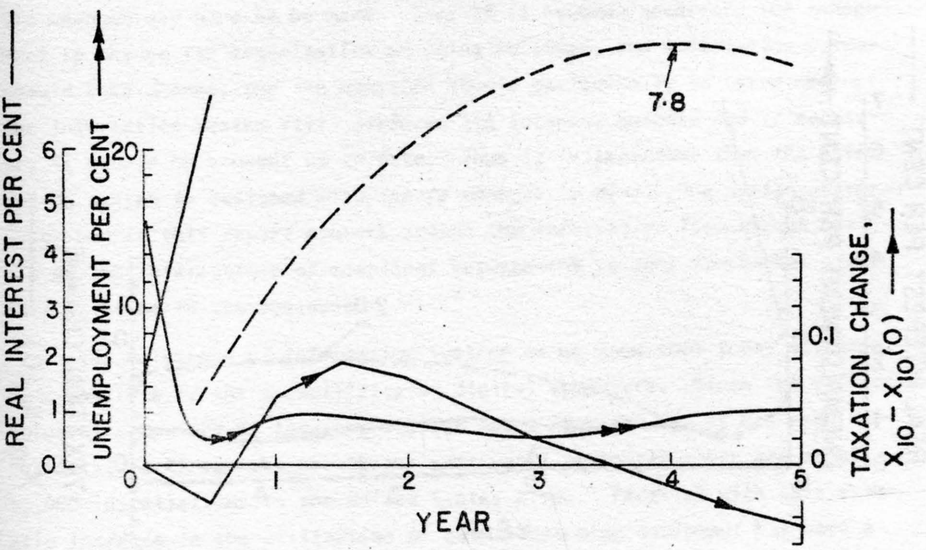


Fig. 1 CONVERGENCE AFTER A PRIMING TRAJECTORY
(10 Subintervals, 2 Control signals)



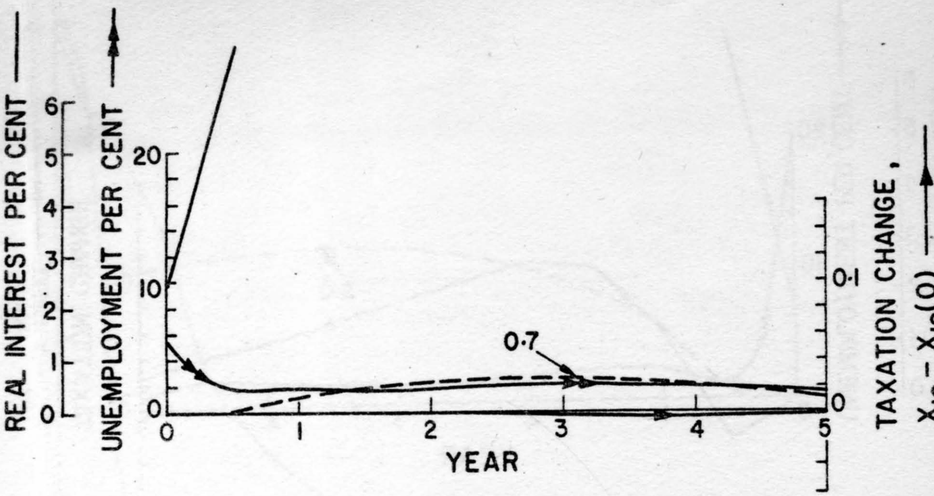
MEAN GROWTH RATES PER CENT = 8·3 (Y), 5·3 (C), 5·1 (P)

Fig. 2 CONTROL AFTER 1933, $S_1 = 1{\cdot}0$, $S_2 = 0{\cdot}05$



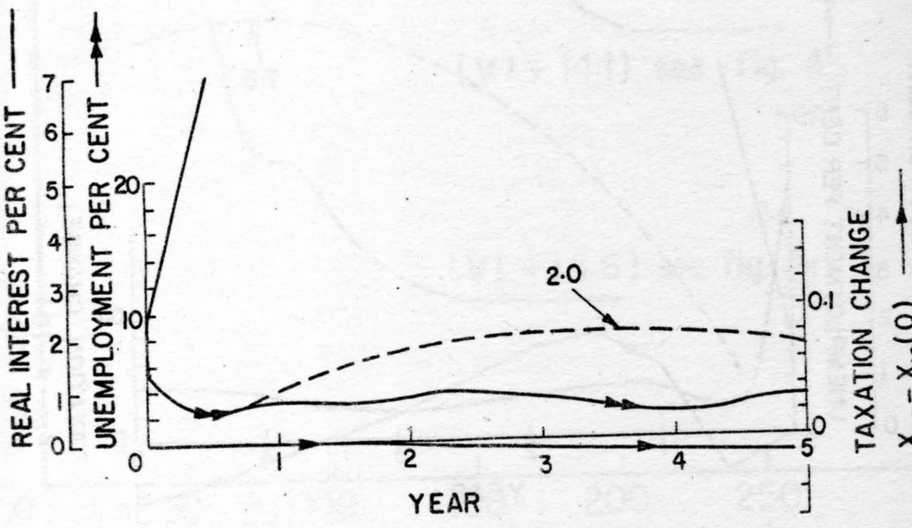
MEAN GROWTH RATES PER CENT = 7·1 (Y), 6·0 (C), 2·5 (P)

Fig. 3 CONTROL AFTER 1933, $S_1 = 1{\cdot}0$, $S_2 = 0{\cdot}01$



MEAN GROWTH RATES PER CENT = 4.6(Y), 5.3(C), 3.8(P)

Fig. 4 CONTROL AFTER 1963, $S_1 = 1.0$, $S_2 = 0.05$



MEAN GROWTH RATES PER CENT = 4.1 (Y), 6.4 (C), 1.6 (P)

Fig. 5 CONTROL AFTER 1963, $S_1 = 1.0$, $S_2 = 0.01$

INFORMATION SYSTEMS DESIGN FOR BUSINESS APPLICATIONS

Ralph R. Duersch
General Electric Company
Schenectady, New York 12305
U.S.A.

1) Introduction

The purpose of this paper is to describe an approach to the design of information systems for business applications beginning with the justification for such a system and ending just prior to the task of determining the proper hardware for the system. This report intentionally stresses the work that has to be performed prior to the selection of hardware because this is the area in the past that has been frequently neglected. Before proceeding to define what is meant by an information system, it should be stressed that such a system is never static. Minor changes occur continually in the process of managing a business and are executed within the organizational framework. However, if more significant changes occur because of competitive efforts or other factors, more drastic changes may have to be made. Thus if it becomes necessary for management to change its organization of doing business, the information system should also change, and the question should periodically be asked whether the information system still performs its intended purpose and if necessary it should be brought up to date. Thus it is important that the information system is designed with future changes in mind. The design approach described in this report centers around the information flow and an analysis of the interactions or couplings between the various functional elements existing in the business.

The existence of information systems as we know them today has been made possible by the capabilities of digital computers. Since the utilization of computers by industry started about 15 years ago it has seen approximately a 20 percent growth per year until presently there are about 40,000 installations in the United States alone. Parallel with this dramatic increase in the utilization of data processing equipment has been a progression of applications. Initially business applications were centered around payroll, invoicing and general accounting. This was followed by the development of personnel systems, inventory and production control systems and similar more complicated systems. Most recently programs have been installed to do business forecasting, investment planning and budgeting. Even more comprehensive information systems are likely to be designed in the future. The need for more powerful models, more comprehensive data bases, faster hardware, etc., to provide the additional capability for the simulation of planning and decision-making situations will become evident as management explores the capability of information systems to weigh alternatives, explore the effects of modified assumptions and to obtain the information needed for timely and accurate decisions.

II) Discussion Of Information Systems

Definition Of An Information System

In order to talk about information systems in general, it would be nice to have a definition for such a system. However, such a definition is not easy and there is no agreement as to which of the many possible definitions should be used. How all-encompassing such a definition is depends on the perspective or the goals and purposes of the user. One useful definition is as follows:

> An information system provides management with the information it requires to monitor progress, measure performance, detect trends, evaluate alternatives, make decisions, and to take corrective action.

This definition is very similar to the definitions one might select for management information systems. Certainly, a system could have as its purpose to provide information to top management for planning or decision making; it could furnish information to the functional division of a company, such as finance, manufacturing, marketing, etc., for the operation of the business; or it could be used to supply information to middle management to determine the detail status of the operation for purposes of daily control.

It might be helpful to give examples illustrating what is not considered to be an information system. It is not a report generator which, backed up by files of data, provides part of these data to various members of management. It is not a collection of programs which in aggregate produces all the reports requested. It is more than these. It is a closed loop system in which management closes the loop when acting on information received. There are conceivably many loops which are closed by management as indicated in Fig. 1 each of them capable of providing stability to the operation of the business. There is a structural similarity between the management functions of planning, organizing, integrating, measuring and control, and the conventional adaptive control system block diagram.

Characteristics Of An Information System

An information system generally consists of at least four parts as indicated in Fig. 2, showing the part of the system which might be computerized:

1. The supervisory program.
2. The data base.
3. The model base.
4. The reporting system.

The first of these parts contains the command and control programs. As the name implies these programs manage the operation of the system. That is, they control the input, update and retrieval of the data in the data base, and they execute the programs which are necessary to transform data into information and finally cause this to be the output of a particular report. The data base is the complete set of data in this system, arranged in a meaningful structure in files which often follow functional lines. This part of the system is a very fundamental part but it is only one of the four parts which make up the total system.

The model base of the information system contains the simulation models which are used by management or by the operating functions to find possible best solutions and obtain answers to the "what if" question. These models are management analogs in that they describe the real system as visualized by the manager who is responsible for that particular operation. This model base also contains the procedures and the equations to execute inventory control, project control, to determine personnel requirements, to do scheduling or forecasting, etc.

The reporting system represents the output function of the system. Its purpose is to present the information requested by the user. As the need is established for certain output, the form and format of the presentation should also be established. The most often used forms of output are paper reports (sent via mail or teletype), visual displays (CRT's) or oral communications. The formats are largely chosen by the recipient but the usefulness of graphs supplementing or even replacing tabular outputs should be considered.

Usefulness Of Information Systems

An information system is a system designed for use by the management and any justification should evaluate the benefits of a proposed system to management. Since the system is to be used by management, it is important for management to contribute its talents and support in all the design phases. The decision to establish an information system, and the support to design the system must come from all levels of the management structure. This cooperation between the management and the designers of the information system must continue through all phases of the development. While the information system analyst or designer can establish the necce-sary information flows and evaluate the system, they are not in a posit-ion to define future organizational goals and objectives. With this in-formation available, the system structure may be chosen to accommodate anticipated changes in the organization of the business. Thus management contributions in the early phases of design is not only what information will be required from the system, but an assessment of present and anti-cipated performance of the dynamic entity which we call "the business enterprise". The information system must be flexible enough to accommod-ate these changes in the organizational structure.

The usefulness of an information system can be described by consider-ing a number of areas:

A. Observation

1. Determining status.
2. Measuring progress.
3. Detecting trends.
4. Anticipating problems.

In these areas the issuance of reports makes it possible to determine the status of all operating functions of the business. If the performance measures have been set, and objectives or guidelines have been established, the performance monitored can be measured against these and the informa-tion conveyed to management. This will allow management to detect devel-opments in the early stages and prescribe action accordingly, or it may be used to make decisions automatically such as are made by an automatic in-ventory ordering system given criteria of performance and rules of action. It is also possible that trends may have been developed because of the particular way performances are measured and it is then possible to change the criteria in an adaptive sense. The anticipation of problems should be possible far enough in advance so that it is possible to take corrective action in time. This is one important goal of the information system. An example of the last point is the use of the PERT scheduling system where bottlenecks are detected in advance and can be constantly watched to guard against non-recoverable delays.

B. Control

An information system can be used to:

1. Provide accurate and timely information.
2. Provide for stable systems operation.
3. Make closed loop systems possible.

In each of these points the reporting function is one link of the total system which makes it possible to establish a closed loop operation. In this connection, it should be pointed out that just as the PERT system only points to the possible trouble spots and does not take actions to relieve these, an information system only provides the information requested but it is management that closes the loop of the system by acting on the information.

C. Analysis

Part of the information system may be used to:

1. Explore alternative approaches.
2. Find best solutions.
3. Test effects of strategy.

The functions served by the system in this area address themselves to the exploration of alternatives and to the long range planning responsibility. The system will make it possible to obtain answers to the question "what if" so often asked by management. As such analysis can provide predictive information so that quantitative answers with statistical information indicated, can be supplied for various alternatives that are worthy of consideration. This makes it possible to change factors, constraints, and other assumptions in a proposed course of action and to re-evaluate the consequences of this action so that the best course may be selected.

Justification to Establish An Information System

Competitive pressures today require a more rapid response to changes in the marketplace and have forced the trend to develop more sophisticated information systems. In the past, for applications which were essentially conversions from a manually operated system, it was possible to determine the benefits of a system by calculating the cost-savings in personnel. As more sophisticated information systems were designed justification became more difficult because not all benefits could be assigned monetary value, and the more sophisticated the system, the fewer the direct cost savings and the more numerous the indirect benefits appeared to become.

In order to get a reasonable answer to this question, a feasibility study is often undertaken. This in many cases means the examination of the existing system, an extrapolation of future needs and an analysis of these needs in terms of information systems requirements. Since so much of the answer to the feasibility study depends on the business's own position and its particular environment, no general answer is possible. It is necessary to be able to evaluate the consequences of planned management action in the future at the time the feasibility study is undertaken, since this will determine possible future changes or extensions of the system in existence today. The criteria involved in the justification of the information system is frequently a value/cost comparison in which value and cost, as a function of time, are determined for two or more alternatives. Thus it is important to have management's cooperation even at this early stage of the design.

In addition to future plans, a number of aspects may be suggested for your consideration:

1. Growth rates of the information to be handled.
2. Performance in terms of response time.
3. Speed of data available.
4. Customer service improvement.
5. Decreased cost of providing needed information.
6. Communications between source and recipient.
7. Uniformity of data available.
8. Quality of data entered.

Problems In System Development

The development of information systems began historically with the establishment of certain information being generated for a specific purpose. Let us assume this took the form of a computer program which solved a problem for a particular unit in the organization structure. This was the beginning of an avalanche of programs, each of which satisfied a particular need. As the applications became more complex, it became more difficult to forecast direct benefits of these applications and consequently, it became more difficult to justify establishment of such an application to management. Accompanying the progression of applications was a series of problems:

1. Lack of flexibilities to meet new demands.
2. Increased cost. Every change in hardware and most changes in organization brought about the necessities for creating larger programming staff and continual rework of programs.

3. Most programs were written in small packages and there was no overall planning.

As time went on the amount of data increased rapidly with time. The hardware capability grew, but management's use of data leveled off and even decreased. The result of this unplanned design was a generally inefficient system with much redundancy in data collection. Moreover, the usual practice of implementation by organizational units made it necessary for each department to collect, organize and store, large quantities of data which were often duplicated and inconsistent with the data collected by other organizational units.

Data existing in a system are isolated facts which do not have any meaning by themselves until they are related to other facts in a meaningful manner. There are a number of steps in going from raw data to information:

1. Data available.
2. Data recorded.
3. Data manipulated.
4. Data reported.
5. Data transformed into information requested.

These five steps imply selectivity and transformation from the data to information. As these transformations are applied to the data, improvements are obtained. As shown in Fig. 3 where we suggest the transformation of data into information, and in Fig. 4 which shows the iterative loop in the design of information systems, one can see that there is no dependence on hardware in the design process until the last stage of that process is reached. Only then is it important to draw the conclusions relating to hardware. Unfortunately, the design of a system, if carried out at all, in the past has been largely determined by the hardware considerations that were often suggested by the manufacturer's sales and applications personnel. Until the implementation stage is reached, therefore, it may be immaterial whether the system will be computer oriented or turn out to be a manual system.

III) Steps In The Information Systems Design

The development of an information system should start by considering what is already available. This generally seems to make more sense than to try to start without taking into account the already available system. Because of the flexibility of the organization structure and its proneness to periodic changes, one of the recommendations in this report is the tie-in of any information system with the functional elements in the system instead of the organizational elements. These elements, manufacturing, marketing, engineering, etc., do not usually change drastically in an organization except that they may expand. Organization of the information system around other lines, such as a product line structure or geographic divisions, does not offer security from system obsolescence because of change in the organizational structure.

Definition Of An Information System

The initial analysis should try to establish all goals to be achieved with this information system so that it is possible to outline the long-range objectives of the entire system, although the total system may not be implemented immediately. In order to establish this broad outline for the information system it is necessary that a team from all the functions involved and someone who has broad corporate views is present. This insures the support from the corporate level as well as the functional level and makes it possible to incorporate the long term view into the design of the system.

The development of an information system as shown in Fig. 4 may be broken down into the two initial steps:

1. Documentation of existing systems.
2. Determination of information requirements.

Following the initial steps is the development of specifications of performance required and the corresponding system performance necessary. Once this is accomplished the design of the system can begin, followed by the actual implementation. In this procedure, each step logically follows the one preceding and can be accomplished one step at a time. This allows timely spacing of these steps in recognition of possible resource limitations. It therefore permits the planning of the resources allocated to the design of the information system according to the constraints present. It also allows interruptions and delays which may be necessary to satisfy priorities demanded by business conditions.

Analyzing and documenting an existing system can be accomplished by means of three aids:

1. The interview.
2. The flow diagram.
3. Analysis of document.

An accurate flow diagram reflects the actual sequence of steps in which the data and information flows through the system. A helpful addition to this diagram is the name of the function (department) involved, and the time (and cost) required to do each of these steps. This then establishes a good road map along which it is possible to trace the information flow. Fig. 5 shows part of such an information flow diagram for a Receiving/ Shipping Operation.

Along this road map of information flow are the positions of individuals who are acting on the system. It is important to find the interaction of each of these people with the system. That is, what are their responsibilities and objectives? What information do they receive to carry out their tasks? What requirements for information do these people feel they have? And, how do these requirements relate to what they actually receive? From this, the analyst must decide what are the actual information requirements necessary to carry out the tasks assigned to them.

Finally, the mechanism for transmitting the information along the road map must be explored. To do this, reports, forms and other documents are examined as are any oral communications that take place on an informal or formal basis. It is also important to try to discover any reports or other information, which exist but which are not a part of the existing formal information system.

In order to outline some of the information which has to be uncovered during this stage of the development, Fig. 6 is given.

Determination Of Information Flow Matrix

In order to effectively analyze the existing situation and for use in the system to be designed, the information flow can be depicted in a matrix which has as its rows the different functions in the organization which play a role in the flow of information and has as its columns the documents or reports and other communications which are used to transmit the information between the different functions listed. This matrix may be called the information flow matrix since it identifies and aids in the understanding of the information flow. At this point in the analysis, it is possible to propose changes in the existing system to take care of existing bottlenecks which may have been the cause of the consideration of a new information system. If such bottlenecks can be uncovered, this would be the first tangible evidence of benefit from the new information systems design effort. A skeleton of such a flow matrix is shown on Fig. 7. This matrix may be expanded in terms of additional columns or by replacing each document by all the data elements which it contains, if more than one functional area contributes to a document.

The next step in the development of an information system is the analysis of the information flow matrix in terms of the contribution of the individual document to the goals and the objectives of the information system. In order to proceed from the study of the old information system to the new system proposed, it is necessary to examine the functions represented by the rows and determine the information requirements for each of these functions according to their responsibilities. The requirements for data or information can then be entered into the appropriate column. An analysis of the resulting information flow matrix will then lead to groupings along functional lines and also to groupings of the data elements into reports. A benefit of this representation which is even more significant is the determination of existing couplings between the various functional elements through the flow of information.

Determination of Coupling

The information flow couples the various functional areas but the coupling and its degree is often not obvious when a large information flow matrix exist. In the implementation stages of the design it is important to know whether the information system can be separated into several subsystems and what the corresponding gain of such subdivision might be. Such a coupling matrix can be created by first separating the information flow matrix into a source matrix (S entries only) and a recipient matrix (R entries only). To get the coupling matrix elements, we compare each data element in a source matrix row with the corresponding element in the recipient matrix row. We count the number of coincident entries and the result will be an entry in the coupling matrix. Thus, for the matrix of Fig. 7, the coupling matrix is given in Fig. 8.

The entries in this matrix show that, by rearranging, it is possible to separate areas 3, 5 and 7 from 1, 2, 4 and 6, and that these two groups are independent from each other. While this subdivision could easily be seen in this example, partitioning algorithms will normally be required to accomplish this task.

If partitioning is not completely possible, the number of entries which are not part of the new sub-matrices will indicate where coupling exists. Consequently the functional area which is represented by the entry in the matrix will have to be a part of both systems. Of course, if the quantity of data is not large and the hardware to store all data in one single data base can be committed, and if the running times to generate the information from a single data base is not excessive, then there is no reason to subdivide the data base. However, if these conditions are not met it could be advantageous to separate the data base in two parts and to accept the redundancy. In addition to the redundancy, added costs and possible inconsistency should be watched.

Completion Of Design

Iteration is one of the standard procedures in arriving at a suitable system design. Fig. 5 indicates the basic iteration which has to be performed showing the complete loop. As a part of the iteration procedure, there are subloops. Some of the subloops are shown in Fig. 9. This figure shows that for each feasible arrangement of data, reports and functional areas, an analysis can be carried out as described previously with the results of each analysis compared against the others. This comparison is done with respect to the activities expected, the response time obtainable, and the size of the data base and other objectives. Given the activities and the requirements and objectives set by management, the performance specifications which must be met by the system (hardware and software if computerized) can be developed and these specifications in turn will place requirements and constraints on the implementation of the system.

At this point in the development of the system, it is entirely possible that the specifications cannot be realized despite the fact that the information requests have been reasonable. This may be caused by inadequate existing hardware or because insufficient funds are available for implementation, etc. It may also be the case that no hardware which meets the requirements is in existence. In any case, there may be a difference between what is needed and what can be obtained. Thus, modifications will have to be introduced in the objectives or the specifications or the resources by iteration. During these iterations, a systematic way to document the required activities is needed. The structure for such an activity record is shown in Fig. 10.

If a satisfactory system design has finally evolved and the specifications arrived at have resulted in the selection of proper hardware and software, the in-house system implementation can be begun. There are a number of aspects to this implementation:

1. Staff Selection.
2. Program coding and documentation.
3. Training.
4. Testing.

Some consideration of the coding should be given:

a) To establish priorities for implementation of subprograms.

b) For standardization.

c) Providing a mechanism for review and for authorization of change.

With respect to documentation, it is advisable to adopt formal procedures for maintenance and to enforce adherence, and to give special emphasis to the maintenance of data base documentation.

Since the environment of the information system is constantly changing, an information system designed today is probably obsolete within a few years unless constant effort is expended to keep the system as dynamic as the business itself. Since the business is continually subject to the pressures of its environment and adapting its mode of operation, the information system must also adapt. To prevent deterioration, the system should be re-evaluated as follows:

1) Use predetermined performance measures which are keyed to system objectives.

2) Include as benefits both cost reduction and increased performance potential.

3) Conduct periodic re-evaluations of systems costs and benefits.

In addition to keeping pace with the business, the role of the information system is expected to expand especially in the area of long range planning and decision making. This will mean more comprehensive simulation of management analogs and the use of historical data. As experience is gained with these techniques, their logical place will be in the information system.

IV) Conclusions

This paper discussed the nature of information systems and described the steps which must be taken to design an information system. To be a viable system, careful consideration should be given to all aspects so that the final design of the information system has the characteristics discussed. As you can observe, the efforts to establish such a system are considerable. The expense has been found to be of the same order of magnitude as that of acquiring the necessary hardware. The role of the control engineer in establishing such a system has not been sufficiently recognized, but future developments are likely to show that he will make valuable contributions.

The value of an information system is expected to become still greater, and its role more vital to management as the usefulness of models for the information system is exploited. Using his experience in modeling of physical systems, the engineer is in a good position to contribute to the modeling and simulation of economic systems. This development will put management in a far better position to make sound decisions than it has ever been in the past.

References

1) Ansoff, H.S. and Slevin, D.P. "An Appreciation of Industrial Dynamics" Management Science, 14, 7, March 68, pp 383-397.

2) Forrester, J.W., "Industrial Dynamics", MIT Press, 1961.

3) Forrester, J.W., "Industrial Dynamics After The First Decade", Management Science, 14, 7, March 68, pp 398-415.

4) Head, R.V., "Management Information Systems: A Critical Appraisal" Datamation, May 67, pp 22-27.

5) Homer, E.D., "A Generalized Model For Analyzing Management Information Systems", Management Science 8, 1961/62, pp 500-515.

6) Lieberman, I.J., "A Mathematical Model For Integrated Business Systems", Management Science 2, 1955/56, pp 327-336.

7) Lovenvirth, D.L., "Successful Information Management", 2nd Edition, C-E-I-R, Inc., Washington, D.C., 1968.

8) Roots, W.K., "Corporate Synthesis", Works Management, Nov. 66.

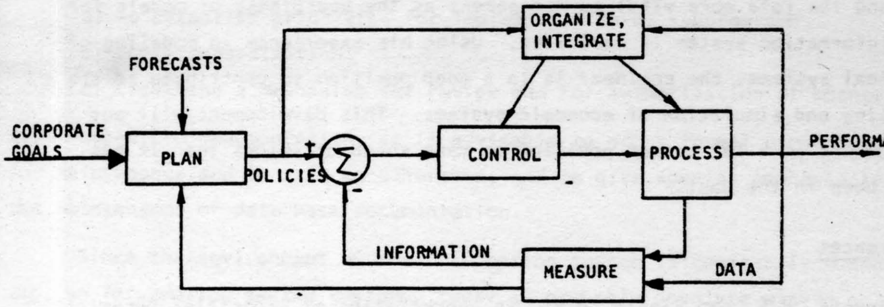


FIG. 1. LOOPS IN MANAGEMENT OF PROCESS

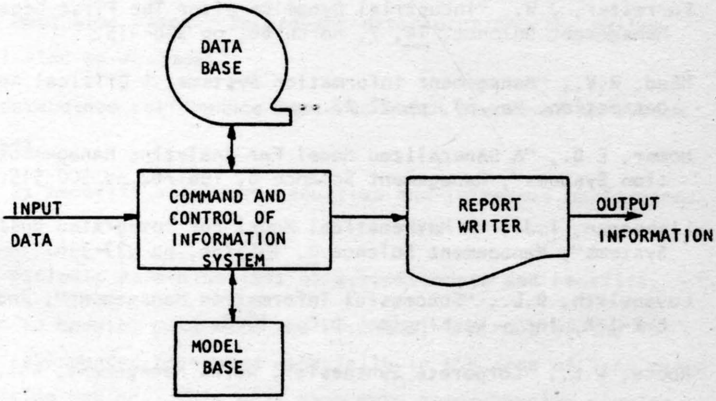


FIG. 2

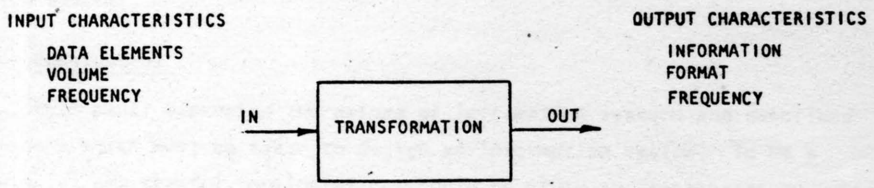


FIG. 3

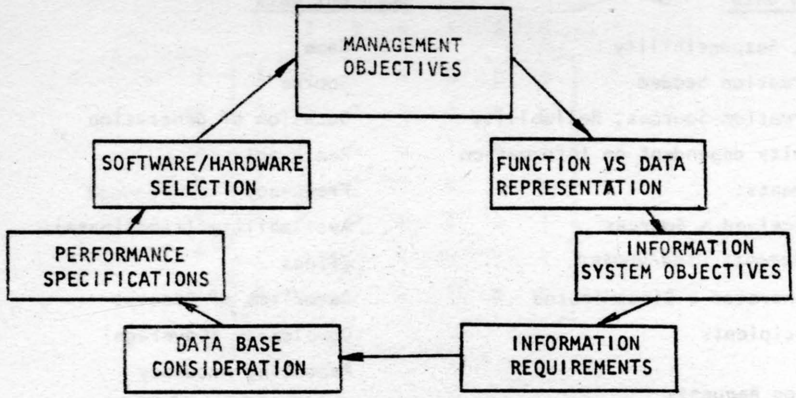

MANAGEMENT OBJECTIVES
FUNCTION VS DATA REPRESENTATION
INFORMATION SYSTEM OBJECTIVES
INFORMATION REQUIREMENTS
DATA BASE CONSIDERATION
PERFORMANCE SPECIFICATIONS
SOFTWARE/HARDWARE SELECTION

FIG. 4

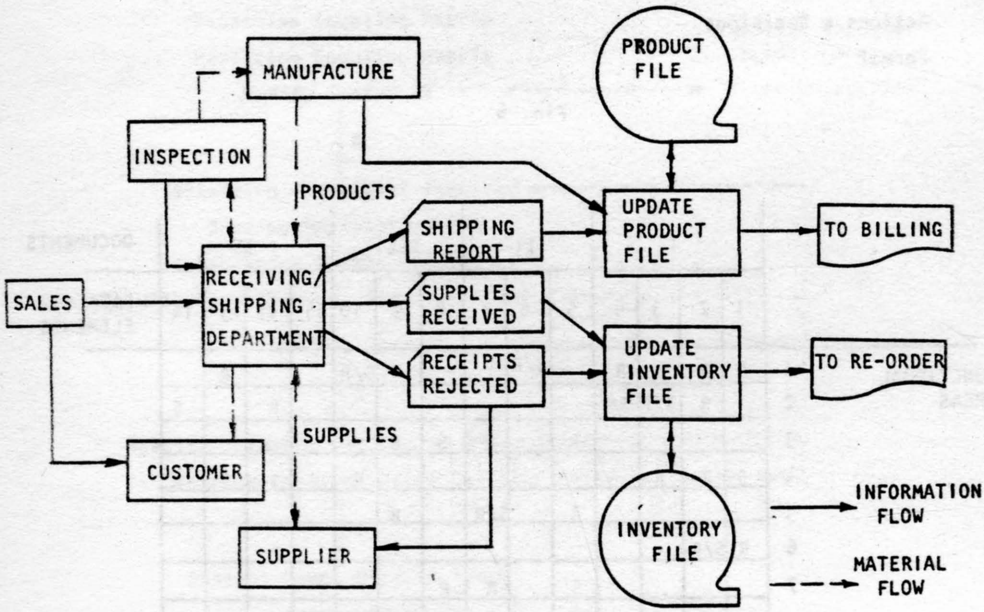

MANUFACTURE
PRODUCT FILE
INSPECTION
PRODUCTS
SHIPPING REPORT
UPDATE PRODUCT FILE
TO BILLING
SALES
RECEIVING/ SHIPPING DEPARTMENT
SUPPLIES RECEIVED
RECEIPTS REJECTED
UPDATE INVENTORY FILE
TO RE-ORDER
SUPPLIES
CUSTOMER
SUPPLIER
INVENTORY FILE
INFORMATION FLOW
MATERIAL FLOW

FIG. 5

Interview Data

Name, Responsibility
Information Needed
Information Sources, Reliability
Activity dependent on Information
Documents:
- Received & Sources
- Elements Used/Unused
- Generated & Disseminated
- Recipients

Information Requests

Name
Source
Frequency
Availability (timeliness)
Complexity (coverage)
Actions & Decisions
Format

Document Data

Name
Source
Occasion of generation
Recipient
Frequency
Availability (timeliness)
Effect
Date/Time of Receipt
Complexity (Coverage)
Reporting Accuracy
Availability of Redundancy
Action & Decisions

Fig. 6

DOCUMENTS	I				II			III			IV			
DATA ELEMENTS	1	2	3	4	5	6	7	8	9	10	11	12	13	14
FUNCTIONAL AREAS 1	R	R		R		S/R				S/R			R	
2		R	S/R	S/R								R		S
3					R			S	S					
4	S	R	R							R		S	S	R
5					R		S/R		R					
6	R	S/R										R		
7					S		R	R						

Where: R - recipient
S - source

FIG. 7: INFORMATION FLOW MATRIX

Source \ Recipient	1	2	3	4	5	6	7
1	2	0	0	1	0	0	0
2	1	2	0	2	0	0	0
3	0	0	0	0	1	0	1
4	2	1	0	0	0	2	0
5	0	0	0	0	1	0	1
6	1	1	0	1	0	1	0
7	0	0	1	0	1	0	0

FIG. 8: COUPLING MATRIX

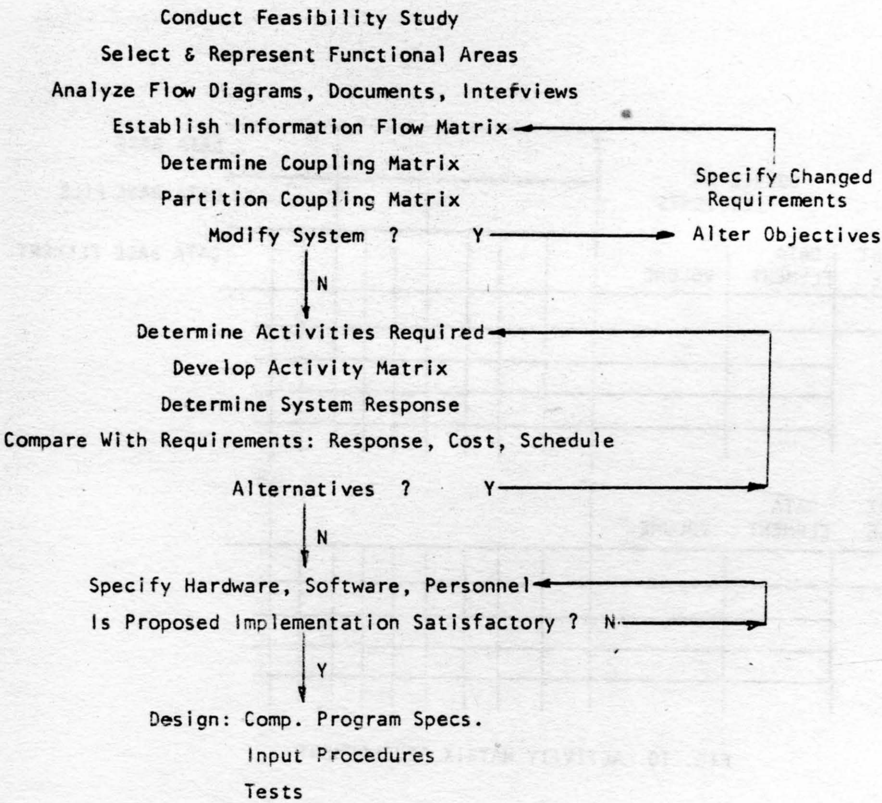


FIG. 9

IDENTIFY: SOURCE DOCUMENTS
RECIPIENT REPORTS

DEFINE: DATA ELEMENTS
RECORDS
FORMATS
VOLUME
FREQUENCY

DETERMINE METHOD TO ESTIMATE ACTIVITY:

FILE ACCESSES
RUN TIMES
I/O TIME
TYPE AND AMOUNT OF STORAGE

FORMAT:

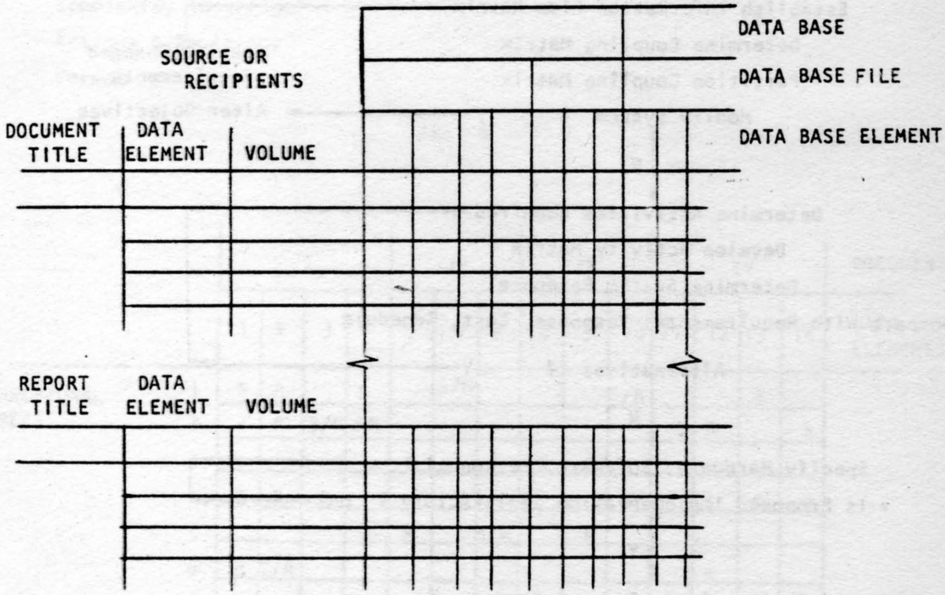


FIG. 10 ACTIVITY MATRIX DEVELOPMENT