IFAC

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

WARSZAWA 1969

Traffic and Vehicle Control

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

TECHNICAL SESSION
55

NOT

Organized by
Naczelna Organizacja Techniczna w Polsce

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

Traffic and Vehicle Control

TECHNICAL SESSION No 55

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

Organized by
Naczelna Organizacja Techniczna w Polsce

Biblioteka
Politechniki Białostockiej
1181070

BIBLIOTEKA SZKOŁY INŻYNIERSKIEJ w Białymstoku

K-1323

Contents

Wydawnictwa Czasopism Technicznych NOT - Polska

Zakład Poligraficzny WCT NOT. Zam. 119/69.

CONTRIBUTION TO ANALYSIS OF NONLINEARITIES IN THE GENERAL TRAFFIC LAW Q = Cv

by

Westeried Florian
General Automation Company,
Nozy (91), France

Macroscopic studies determining the dynamics of the unidirection free traffic were based as we know, on deterministic or probabilistic hypothesis, independently. Studies using deterministic hypothesis tried to justify traffic laws and traffic stability laws by means of:

- assimilation to compressivable fluid flow [1,2,3,4].
- generating the models of intervals between vehicles of rank (n,n+ 1) in a column [5,6,7,8,9].
- diffusion motivation.

Stochastic approaches yielded in analysis of the local spaces, cadences, delays and queues. It enabled also to determine mathematical functions defining colision or relaxation probabilities in the vehicles flow [10,11].

Deterministic models, using integration of the flow intensity, concentration and average velocity led to formal expressions, with their form and existence domain corresponding, more or less precisely, to the measurement results. In particular, these models furnished qualitative suggestions on the traffic propagation or collision waves as associated with properties of the local roads and with reflex of the average driver.

Stochastic models, when imposing probability properties by choice of the probability distribution (Poisson, Erlang, etc.), can describe the traffic's local conditions. Their application to quasi-stationary regimes is limited, however, by a nature of necessary computations.

In order to get advantages from two traffic interpretations mentioned above, the author presents a concept of a model of the general traffic law. The model takes into account point distributions of the vehicles' velocity and concentration, according to continous traffic law and to imposed security intervals.

The general law of the forward circilation in stationary regime without vehicles' bypassing, can be expressed for a unidirection displacement, as a function of the space "x" and the time "t", in form:

$$Q(x,t) = C(x,t) \cdot v(x,t) \qquad \text{.............} \quad (1)$$

where the traffic $Q(x,t)$, concentration $C(x,t)$ and velocity $v(x,t)$ are mathematical expectations related to time. Physical representation of this hypothesis is expressed by distribution of the traffic, varying along a road / referred to averaged measurement point /. The distribution can affect the traffic law described by the deterministic models mentioned before, if an auxilliary distribution is imposed.

Continuity equation for the intervals Δx, Δt admits a form:*

$$\frac{\partial C}{\partial t} + \frac{\partial (C \cdot v)}{\partial x} = 0 \qquad \text{.............} \quad (2)$$

or

$$\frac{\partial C}{\partial t} + v\frac{\partial C}{\partial x} + C\frac{\partial v}{\partial x} = 0 \qquad \text{.............} \quad (2')$$

Time differential of $C(x,t)$ when having $v = dx/dt$, after the terms' arrangement yields in

$$\frac{\partial C}{\partial t} = \frac{dC}{dt} - v\frac{\partial C}{\partial x} \qquad \text{.............} \quad (3)$$

From (2') and (3) it results

$$\frac{dC}{dt} + C\frac{\partial v}{\partial x} = 0 \qquad \text{.............} \quad (4)$$

On the other hand, recognizing the concentration as a ratio of the area occupied by k vehicles of length l_0, to the road segment Δx, around the measurement point when assuming probability density $P(t)$, it yields

$$C = \frac{k \cdot lo\, P(t)}{t \cdot v} \qquad \text{.............} \quad (5)$$

Assuming that $P(t)$ results from the general probability law e.g. the Erlang law, defined by a derivative

$$p(t) = \frac{dP}{dt} = \frac{\lambda^k}{\Gamma(k)}\, t^{k-1} \cdot e^{-\lambda t} \qquad \text{.............} \quad (6)$$

where $\Gamma(k)$ is a probability function for occurrence of k events / k arrived vehicles / at time t, and λ is averaged

*For simplicity, let us denote $Q(x,t) = Q$, $C(x,t) = C$, $v(x,t) = v$

vehicle flow at chosen period t, it yields

$$\frac{dC}{dt} = \frac{k\cdot lo\cdot p(t)v - k\ lo\ P(t)[v + t\ dv/dt]}{t^2 \cdot v^2} \quad \ldots. (7)$$

or, introducing (5)

$$\frac{dC}{dt} = \frac{k\cdot lo\cdot p(t)}{t\cdot v} - C\left[\frac{1}{t} + \frac{1}{v}\frac{dv}{dt}\right] \quad \ldots (7')$$

However

$$\frac{C}{v}\frac{dv}{dt} = \frac{C}{v}\left[\frac{\partial v}{\partial x}\frac{dx}{dt} + \frac{\partial v}{\partial t}\right] = C\frac{\partial v}{\partial x} + \frac{C}{v}\frac{\partial v}{\partial t} \quad \ldots (8)$$

Thus, considering /4/, /7/ and /8/ one can derive

$$\frac{k\cdot lo\cdot p(t)}{t\cdot v} - \frac{C}{t} + \frac{C}{v}\frac{\partial v}{\partial t} = 0 \quad \ldots (9)$$

If we assume existence of a functional relation in a form:

$$v = v(C) \quad \ldots\ldots\ldots\ldots\ldots (10)$$

which is theoretically justified by consideration of the vehicles position time changes, influenced by the concentration variations,

$$\frac{\partial v}{\partial t} = \frac{dv}{dC}\frac{\partial C}{\partial t} \quad \ldots\ldots\ldots\ldots\ldots (11)$$

then

$$\frac{\partial v}{\partial t} = u\frac{\partial C}{\partial t} \quad \ldots\ldots\ldots\ldots\ldots (11')$$

where u denotes a parameter describing the road quality. Substituting (11') to (9) and assuming for simplicity that the probability function is determined by the Poisson law (k = 1) while $c\cdot v = \lambda$ in the period t,

$$\frac{\partial C}{\partial t} = \frac{\lambda(1 - e^{-\lambda t})lo}{u\cdot C\cdot t} \quad \ldots\ldots\ldots\ldots\ldots (12)$$

Remembering relation (10), one may put [3]

$$\frac{dv}{dt} = -\alpha C^n \cdot v^m \frac{\partial C}{\partial x} \quad \ldots\ldots\ldots\ldots\ldots (13)$$

when assuming that drivers will adapt themselves to the conditions of temporary traffic, according to law of the n, m order proportionality, specific for the interval Δx.
Substituting (12) and (13) to (3) it yields

$$\frac{dC}{dt} = \frac{v}{\alpha\cdot C^n\cdot v^m}\cdot\frac{dv}{dt} + \frac{\lambda(1 - e^{-\lambda t})\cdot l_o}{u\cdot C\cdot t} \quad \ldots (14)$$

For the Poisson process the equation (5) results in

$$C = \frac{lo \cdot (1 - e^{-\lambda t})}{t \cdot v} \; ; \qquad (k = 1)$$

Hence

$$\frac{dC}{dt} = - \frac{v}{\alpha \cdot C^n \cdot v^m} \cdot \frac{dv}{dt} + \frac{1}{u} \cdot \lambda \cdot v \qquad \ldots . (14')$$

which, for $m = 1$, $n = -1$ and $\lambda = C \cdot v = \lambda(t)$ yields

$$\frac{dC}{dt} = - \frac{C}{\alpha} \cdot \frac{dv}{dt} + \frac{1}{u} C \cdot v^2 \qquad \ldots . (15)$$

Integrating this nonlinear differential equation one will obtain

$$v = \alpha \cdot \mathrm{Log} \frac{Csat}{C} + \frac{\alpha}{u} \int_{(t)} v^2 \, dt \qquad \ldots . (16)$$

where Csat denotes concentration of the saturation for which the circulation velocity is $v = 0$.

The derived equation (16), according to comprising of a logarithmic term, presents the Greenberg formula succesfully utilized for defining of the fundamental stable - traffic law.

The second term of eq. (16) is a square-integral of the velocity variations around a measurement point of the concentration /distributed in time t, in reference to employed measurement method: synchronous or asynchronous/.

Choice of adaptable formula resulting in different possibilities of defining p (t), m , n will permit for a better interpretation of the actual measurement /the experimental results will have been presented at the IVth IFAC Congress/.

The fundamental traffic law, expressed by relationship (1) can be transformed if assuming that security intervals between vehicles in a column are given by $l = lo(1 + a.v)$. Hence we shall obtain

$$Q = \frac{C}{\alpha} \mathrm{Log} \frac{Csat}{C} + \left[A - B \cdot \mathrm{arctg} \beta + \gamma \left(\frac{lo}{v} + \delta \right) \right] \ldots . (17)$$

where A, B, β, γ, δ, a are constants, l is a distance between vehicles' bumpers "head way" and lo is the avarege vehicle length.

Existence of the terms: arc tg or logarithm in the equation (17) can justify the a priori formation of the collision waves

in the velocity flow.
The author has the intention to continue the studies to develop new relations characterizing, in virtue of measurements, the properties of the parameteres u and α.
Extension of the presented theory will be done for determining of the lights' influence on local variations of the flow velocity and of their connection with automatic coordination of the traffic-control lights.

REFERENCES

1/ Lightill M.J., and Whitham GB : "On Kinematic Waves II.A Theory of Traffic Flow on Long Crowded Roads". Proc. Roy. Soc. Series A 229 /1952/

2/ Greenberg H. " An Analysis of Traffic Flow" Opns. Res. 7 / 1959 / p. 79-85

3/ Drew D.R. and Pinnel Ch. "Some Theoretical Considerations of Peak-Hour Control for Arterial Street Systems" in Traffic Control Theory and Instrumentation, edited by Horton Th. R., Plenum Press N-Y 1965 p. 91-123.

4/ Franklin R.E. "On the Flow-Concentration Relationship for Traffic" Communication Second International Symposium on the Theory of Road Traffic Flow London 1963.

5/ Tse-Sun Chow "Operational Analysis of a traffic Dynamic Problem" Opns. Res. 6, 1958, p. 827-834.

6/ Herman R. Montroll E.W., Potts R.B. and Rothery R. "Traffic Dynamics: Analysis of Stability in Car Fllowing" Opns. Res. 7, /1959/ p. 86-106.

7/ Chandler R.F., Herman R.E., Montroll E.W. "Traffic Dynamics: Studies in Car Following" Opns. Res. 6 /1958/ p. 165-181.

8/ Edie I.C. "Car Following and Steady State Theory for Non--Congested Traffic" Report VI, Port of N-Y Authority, N-Y, May 1960.

9/ Lee G. "A Generalisation of Linear Car-Following Theory", Opns. Res. 1965, p. 595-605.

10/ Harr M.E., Leonard G.A. "A Theory of Traffic Flow for Evaluation of Geometric Aspects of Highways". Hyghway Res. Bull n° 26 p. 87-91.

11/ Haight F.A., "Towards an Unified Theory of Road Traffic", Opns. Res. 1958 p. 813-823.

12/ Prigogine I. and Andrews FC "A Boltzmann-Like Approach for Traffic Flow" Opns. Res. 8, 1960, p. 789-797.

13/ Tournerie G. "The macroscopic theory of the unidirection circulation", Informational note, Ministry of Transport, Publics. SERC, 1964.

SETTING LINKED TRAFFIC SIGNALS TO MINIMISE DELAY

J.A. Hillier, B.Sc.,
Road Research Laboratory, Ministry of Transport
Crowthorne, United Kingdom

INTRODUCTION

Increasing traffic on the road networks of our city centres has made it important that the best use should be made of these expensive facilities. One step toward this is to co-ordinate the operation of traffic signals over the area.

Installations involving the central co-ordination of a number of signals have been operating in Europe and North America for some years. There was, however, no major evidence as to the benefits which might be expected from such installations and the British Ministry of Transport has invested £1 million in two experiments to assess what benefits could be achieved by existing or unconventional control systems. In West London, part of the western approaches to the City are controlled by a computer system which also includes very advanced television surveillance and facilities for manual over-ride. In Glasgow, signals controlling the whole of the city centre are controlled by a computer using only automatic techniques. This paper deals with methods of setting linked signals with particular reference to the Glasgow experiment, which is the direct responsibility of the Laboratory.

The Ministry's experimental programme is designed to assess existing and unconventional control schemes under standard conditions. To enable comparisons to be made between results obtained in the two projects, a basic control scheme is being used in each as a reference against which other control schemes are judged. The criteria on which an assessment may be made include journey time, frequency of stops, safety, capacity and the effect on environment. At the present state of knowledge, however, journey time is regarded as the most important of these factors and it has been adopted in both West London and Glasgow as the primary criterion for assessment. Each 'before' or 'after' study now consists of four instrumented cars carrying observers travelling for two weeks over carefully selected routes which include all major traffic movements on the main roads and many minor roads. The cars make a journey time assessment by recording journey distance in units of one-fivehundredth of a mile and time in seconds on a 5-hole punched paper tape. In addition the observer records the passage of the car past pre-determined timing points, which are usually signal stop-lines. Records are also being kept of the hours when particular control schemes are in operation to enable a detailed accident analysis to be made.

THE WEST LONDON EXPERIMENT

The area chosen for this project[1] lies to the west of Hyde Park Corner and covers about six square miles. About 70 existing signal controlled intersections, together with over 30 new signalled pedestrian crossings, have been connected to a digital computer at the control centre in the new Metropolitan Police Headquarters in Victoria Street, Westminster, some three-quarters of a mile east of the experimental area.

Additional equipment has been placed adjacent to each local controller in order to allow the computer to take over when required. The main functions of these outstations are similar to those of the interposing units described in greater detail in the section on the Glasgow experiment.

Detectors

The computer receives traffic information from the following sources:-

(i) all existing pneumatic detectors on approaches to signalled intersections;

(ii) a small number of additional passage detectors installed away from junctions to provide extra data on traffic flows;

(iii) a small number of speed-measuring detectors to provide information on speeds of vehicle platoons; and

(iv) short queue and long queue detectors on all critical approaches to intersections. Queue detectors consist of two small loops, each 6 feet wide, placed 8 feet apart. If vehicle speeds between the two loops fall below a pre-determined figure, 8 miles/h, this is taken to indicate the presence of a queue.

Queue detectors are also provided where the exit from an intersection in one direction is liable to be obstructed by congestion. These are used to control a special computer control facility known as "split-phase working", which holds signals facing traffic travelling towards congestion at red for part or all of the "green" period, while permitting traffic to move in the reverse direction in the normal manner.

Data Links

A solid-state time division multiplex data transmission system operating over a single telephone circuit is provided between the control centre and each data outstation. It employs a scan cycle of one second, and in that period provides 12 channels outward from the control centre to the outstation and 24 channels inward. There are additional facilities for five further inward channels for the transmission of information on vehicle speeds. Counters in the data outstations accumulate vehicle counts during the one-second scan period and one-second totals are transmitted to the computer.

Central control equipment

At the control centre a Plessey XL-9 computer is used as a data scanner, which rearranges the 24 incoming channels from each data outstation into a standardized 72-bit format for presentation to the central processor computer, which is another XL-9. The data scanner also receives the output of signal control pulses each second from the central processor.

The central processor has a 16K, 24-bit word core store with a drum backing store of 83K, 24-bit words. The time for simple transfer and arithmetical operation is 5 microseconds. Peripheral equipment includes a tape reader, tape punch and a flexowriter.

An important feature of the installation is the control/display console. This consists of the three-position desk from which any intersection in the area may be disconnected from computer control for isolated vehicle-actuated operation or for manual control on site. Six important intersections in the area can be monitored from the control centre by closed circuit television equipment, which provides remote control of pan, tilt and zoom.

Control Schemes

The traffic control schemes to be studied involve the "strategic" selection of linking plans to suit the general traffic conditions in an area and "tactical" local modifications based on traffic conditions at the individual intersections. Provision is also made for the diversion of traffic round bottlenecks at times of extreme congestion.

THE GLASGOW EXPERIMENT

The experimental area covers about one square mile of the Glasgow central business and shopping district, including four bridges over the Clyde and about 80 Plessey traffic signals. A simplified diagram of the road network controlled by signals in this area is given in Figure 1. All these signals, apart from three pedestrian crossings, are vehicle actuated and those on several major streets are connected by local linking systems.

Additional equipment has been placed adjacent to each local controller in order to allow the computer to take over when required. The main functions of this interposing unit are:-

(a) to transmit information about the state of the signals and the detectors to the central control;

(b) to receive instructions from the computer; and

(c) to send the necessary signals to the controller to cause it to show the required signal aspects when under computer control.

Detectors

All approaches to traffic signals in Glasgow have pneumatic detectors, placed between 90 and 130 feet before the stop-line, which detect the passage of all axles. The output from each road separately is brought back to the computer centre.

Two forms of inductive loop detector will be installed in addition. Small-loop vehicle presence detectors will be used to establish the presence of queues waiting to turn in the centre of intersections and they may possibly be used at other places along the approaches. Large-loop vehicle presence detectors will be used in certain circumstances to establish how much traffic occupies the whole of a section of road between major intersections. The loop is laid in slots parallel to the kerb to cover the effective width of the road and extends from one intersection to the next. It gives a linear output according to the amount of traffic which is travelling or stationary above it. This form of detector, still in the experimental stage, is intended for use in the more advanced, unconventional forms of control and might be applied to automatic assessment of control schemes.

Engineers from the Marconi Company working at the Laboratory have developed a tribo-electric detector which looks promising as a more reliable alternative to the existing pneumatic detectors at signals. It operates from the electrical signal produced in a co-axial cable under impact. The cable is enclosed beneath a stainless steel strip which is mounted flush with the road surface. The detector is about 2½" wide by 1½" deep and is considerably easier to install than the existing detectors.

Data Links

Seventy of the intersections are connected to the control centre by multi-core cables laid through existing ducts originally provided for the Glasgow tramway system. Each group of six controllers is connected to the Centre by 100 pair cable with the following allocation of pairs:-

a) 14 separate pairs to each controller

b) one pair common to all six controllers for a telephone

c) five separate pairs common to each pair of controllers

The remaining ten intersections not on the multicore cable network are controlled from the computer centre by Smiths voice frequency multiplex data transmission equipment working over telephone circuits. Two pairs are provided from each intersection to carry data transmission and speech. The maximum number of control channels required between any of these local controllers and the computer centre is 18. Six channels transmit information from the computer centre to the local controller at speeds up to 5 bauds. Of the 12 channels in the opposite direction, 8 are capable of signalling at speeds of 35 bauds.

Incoming information. The following data are sent back to the control centre at all times whether the computer is controlling the signals or not:

(a) a continuous indication of the traffic phase which is being given a green signal. This information is used, among other things, to operate the lamps on the map display;

(b) axle counts for control and assessment purpose and to assist in maintenance by showing up detector faults. The detectors on each approach are fitted with pulse lengtheners which are scanned ever 25 milliseconds.

(c) a speed-timer signal which is associated with one of the features of modern British vehicle actuated controllers.

Computer "take-over" of signals. The computer takes over operation of a local controller via the interposing unit by sending a continuous electrical signal, and the local controller will only respond to computer commands to change its signal aspect while this take-over signal is being received. The local controller returns a continuous confirmatory signal when it has responded to the take-over signal. The safety features incorporated in the local controller, i.e., minimum green and intergreen periods are maintained while under computer control and demands by the computer to change the traffic phase can only be obeyed subject to the operation of the safety features.

A d.c. signal is sent on a separate circuit simultaneously with the take-over signal to guard against stray interference being interpreted as a take-over signal. It lasts until a confirmation signal is received. Receipt by the interposing unit of the take-over signal together with the d.c. signal operates a delay circuit. If the take-over signal continues to be received the delay circuit will release the local controller from the computer after a delay of two to three minutes unless the timing is re-set by the release of all the signal aspect relays during the intergreen period as a result of a command from the computer. This feature ensures that the central equipment cannot take over the operation of the local controllers and then fail to switch the traffic signals.

When the take-over signal ceases the timing is accelerated so as to reduce the delay before release to 3 to 5 seconds. This prevents the occurrence of a long period without a signal change when the central equipment relinquishes its control. The retention of a short delay enables the computer to retain control of the local controller even if a burst of interference causes a short break in the take-over signal.

Central Control Equipment

The central computer is a Marconi Myriad I with a core store of 16K 24-bit words. The time for simple transfer and arithmetical operations is 3 micro-seconds. Input to the computer is through two Facit readers which can read 5 or 8-hole tape at 500 characters/s; the reading heads are electro-static. Outputs from the computer is produced by a Facit punch operating at 150 characters

Eight levels of interrupt are provided for dealing with transfers to and from peripheral equipment. This consists of a Sperry drum with a storage capacity of 80K 24-bit words, a console typewriter for the output of emergency messages to the operator and the input of small amounts of data, a Benson-Lehner graph-plotter and the single-bit input and output unit. Traffic and signal data from the interposing units are presented in the control centre on a bank of 1152 contacts. The computer can control up to 576 pairs of relay contacts for transmitting commands to the local controllers. The relays remain set until instructed to change by the computer. Data is transferred to and from the computer in groups of 24 bits at a time.

A very simple map display is being used with lamps indicating the conditions of the signals and whether or not they are under computer control. This display is intended to show that the control system is behaving generally in the way expected but it is not intended for use as an aid to manual control.

Provision has been made for up to 100 analogue input current signals of 0-5 milliamperes to be received from devices such as the large loop detectors for measuring the concentration of traffic on a signal approach. Each analogue input is converted into a 7-bit binary number suitable for the computer by an analogue/digital converter.

The computer is housed in a pre-fabricated air-conditioned building provided by the Glasgow Corporation. The building also includes office accommodation for the staff and facilities for servicing the instrumented cars used during assessments.

Control Schemes

The programme of control schemes which is being assessed in the Glasgow experiment is given below. It may of course be modified in the light of experience or new developments.

Fixed-time progressions selected by time of day. This is one of the simpler forms of co-ordinated control. The cycle times and splits for a given traffic condition, i.e., morning peak, evening peak, or mid-day, are based on observations made beforehand. The appropriate plan is then selected according to the time of day.

No detectors are used and the system depends entirely on historical traffic information.

Linked vehicle actuated operation. The linked vehicle-actuated flexible progressive system is that used commonly in British linked systems today. Details of its operation are given in reference 2. A master controller determines the common cycle and exerts overriding control of the local vehicle-actuated controllers at certain points in that cycle.

Pneumatic vehicle passage detectors are used at all signals.

Fixed cycle progressions maintaining equal degress of saturation. It has been shown that maintaining an equal degree of saturation on each traffic phase produces minimum delay with fixed-time signals and steady traffic[3]. The degree of saturation of an approach is the ratio between the flow actually arriving and the maximum flow which could be handled with the green time allocated. This control system will operate in conjunction with fixed-time progressions as a means of allocating to the main and side roads appropriate greens inside the fixed cycle length.

The normal pneumatic detectors are also used in this system.

Fixed-time progressions selected according to traffic conditions. This system is very similar to the first described above, having a library of fixed-time plans to suit differing traffic conditions. The appropriate signal plan is selected from the library on the basis of information from a small number of detectors which are used to carry out a simple form of traffic pattern recognition. A control scheme of this type commonly used in the United States is given the name "P.R." system.

BIBLIOTEKA
w
Białymstoku
SZKOŁY INŻYNIERSKIEJ

Signals changing on cessation of saturated flow. In this control scheme the signals will operate in isolation. At each, a traffic phase will be allowed to run for so long as at least one of the roads controlled by that phase has traffic entering the intersection at the saturation flow* rate. The existence or otherwise of saturation flow will be determined by a statistical examination of the gaps between the traffic passing over the pneumatic detectors.

Minimisation of delay by prediction and on-line simulation. The minimum delay control scheme is based on a limited prediction of traffic arrivals and is intended for application to those major intersections which become critically overloaded as traffic increases. The object of the scheme is to minimise delay by increasing capacity at these intersections, i.e., the primary bottlenecks. The less important junctions on roads leading to major intersections would operate on a fixed-time basis. The computer will simulate delays from estimated arrivals at the stop-line in each section. No change in traffic phase would be made if the total delay over the cycle would be reduced by waiting. If the calculation shows that it would be worse to wait another second, further calculations would be made of the effect of change in 2, 3, 4 or 5 seconds in the future. If all these indicate that it would be worse to wait then the signals would change immediately.

This control scheme will use both small-loop and large-loop vehicle presence detectors in addition to pneumatic detectors.

TECHNIQUES FOR SETTING LINKED FIXED-TIME TRAFFIC SIGNALS

It can be seen that most of these control schemes depend on the provision of fixed-time progressions. The conventional method of obtaining settings for linked fixed-time signals has been to use time distance diagrams, produced originally by hand but recently by off-line computer methods. This technique is workable for single roads carrying either one-way or two-way traffic; the aim is to maximise the bandwidth (the proportion of the cycle for which a vehicle unimpeded by other traffic and travelling at a pre-determined speed on each section of the main road could enter and pass through the system without meeting any of the lights at red). A serious disadvantage of this aim is that the bandwidth that can be obtained is almost always insufficient to deal with the amount of traffic that can, and does, pass through the system. A further disadvantage is that delay is not considered and no account is taken

---

* Saturation flow is that flow which crosses a stop-line from a queue when the signal is green. It is the maximum rate of flow possible on any given approach.

of vehicles which turn into the main road from the side roads.

The Combination Method for Setting Linked Traffic Signals

To overcome some of these disadvantages of the time/distance diagram the Combination method (proposed by Whiting and described in reference 4) was developed for setting signal progressions to give approximate minimisation of delay on certain types of network. In order to avoid a complicated simulation and considering all possible combinations of the settings of the signals, the method makes the following simplifying assumptions about the behaviour of traffic:-

(i) the settings of the signals do not effect the amount of traffic on the routes used;

(ii) all the signals have a common cycle (or have a cycle which is a sub-multiple of some master cycle);

(iii) at each signal the distribution of the effective green time among the phases is known;

(iv) the delay to traffic in one direction along any section (link) of the network depends solely upon the difference between the settings of the signals at each end of the section; it is not affected by any other adjacent signals in the network.

The last assumption is the critical one since it is not true for all conditions of traffic but when the network is fairly heavily loaded, i.e., when most traffic problems become apparent, the assumption is sufficiently accurate.

On the basis of these assumptions it is possible to derive a relation between:

(i) the delay to traffic travelling in one direction along a section, and

(ii) the relative timing of the start of the main road greens at the signals at the beginning and end of the section. This relative timing is known as the difference of offset.

If the cycle is divided into an equal number of steps (say 50, which is a common number in British linked signal systems) then the delay/difference-of-offset relation for a given section will be a histogram of 50 steps, each step representing the delay associated with a given difference of offset between the signals at each end of the section.

A typical delay/difference-of-offset histogram is shown in Fig.2. These relations are essential to the Combination method but the manner in which they are produced, i.e., by simulation or by calculation, is not important. When using the Combination method to obtain signal settings for Glasgow, very simple assumptions were made about the behaviour of traffic entering and travelling down the section. Later versions of the method have been improved by the inclusion of, for example, allowance for platoon dispersion as the traffic travels down the section.

When all the individual delay/difference-of-offset relations for each section or link in the network are known it is possible to combine the relations of adjacent links, provided that they are connected in a simple series or parallel manner.

When combining links in series, the relation for link A → B combined to that for link B → C gives an overall relation for A → C. For each of the 50 offsets of A → C, the 50 possible settings of B are tested and the minimum recorded. This requires 2,500 calculations.

For combining links in parallel the relation for A → B can be combined directly with the reverse of the relation B → A giving an overall relation A ⇄ B. This requires 50 calculations.

By repeated operations of this type a suitable network can be completely reduced until only one delay/difference-of-offset relation remains. It represents the delay on the whole network in terms of the offsets between two points on it. From this the difference-of-offset between these two points which gives minimum delay over the whole network can be chosen. Going back over the process used to obtain the overall relation it is then possible to find the minimum-delay offset of each signal. The particular value of the technique is that it is systematic and therefore can be carried out by computer. The amount of work involved is approximately proportional to the number of links and is very considerably less than it would be if every possible permutation of offsets throughout the network were to be considered.

The requirement that individual links must be connected in a serial or parallel manner dictates the type of network to which this method could originally be applied. These networks which could be completely solved consisted of ladders or a tree* made up of ladders. The Glasgow street network shown in

---

* A tree is a network in which there are no closed loops; a ladder is considered as two parallel routes with cross links. For a tree made up of the ladders each link of the simple tree has been replaced by a ladder.

Figure 1 is more complex than this and the Combination method was applied by removing unimportant links to bring the network to the required form. Alsop[5] has subsequently proposed a method by which the Combination method can be extended to some more complex networks. Figure 3 shows examples of various types of network.

Modified Combination method. The basic method has been modified to provide a technique which will allow approximate minimisation of delay on a two-way road when the signals are not heavily loaded, and assumption (iv) above is not strictly valid. In this case the road is considered as a simple ladder network made up of two one-way roads connected at the intersections by imaginary links. The traffic in each direction is then assumed to be travelling in platoons, in such a manner that it passes through each intersection at saturation flow during a "dummy" green which is just sufficiently long to handle the traffic. An upper limit is placed on the total green time allowed to the main road at each intersection. It is then stipulated that the "dummy" greens in opposite directions at each intersection may not be separated at their extremes, i.e., from the start of the first to the end of the second, by more than the total main road green time available at that intersection.

Preliminary tests of the Combination method. The modified method was used to produce linking plans for 8 signals on the Cromwell Road, London. The "before and after" study showed that the plan produced by this method reduced average journey times in the mid-day period between peaks by a statistically significant 8 percent. These savings in journey time were estimated to be worth £10,000 per annum and it was felt that the results justified a full-scale trial.

Trial of Combination method in Glasgow. Signal settings produced by the un-modified Combination method were compared with the existing mixture of linked and isolated vehicle-actuated signals in Glasgow during Autumn 1967. It was decided that there should be three plans, for the morning peak, evening peak and between-peak periods respectively. For each plan, the calculation of cycle times and the preparation of input data for the computer took about 10 man days. Taking off the results and checking took a further 3 man days. The programme took about 6 hours to run on a slow computer; it is estimated it would take between 5 and 10 minutes on Myriad I. These times do not include the collection of traffic data or its reduction to the correct form, or decisions on the traffic phasings and safety features at each signal.

The "before" survey was carried out in October with the signals under local control and the "after" survey was carried out a fortnight later in November when the signals were under computer control. The signals were switched to computer control for one week before the 'after' survey to enable drivers to become used to the changed conditions.

In this first instance each survey covered the weekdays of two weeks using two cars. Flow measurements were obtained by a team of about 16 observers who took short sample counts throughout the network.

Preliminary examination of the results showed that the Combination method settings had reduced the number of vehicle hours per hour being spent in the network during the working day by about 12 per cent. However, journey times were very dependent on the level of flow in the network so graphs relating vehicle-hours per hour and vehicle-miles per hour were plotted, see Figure 4. These showed that the new settings had produced reductions in journey time of approximately 11 per cent during the morning peak, 8 per cent in the period between peaks and 20 percent during the evening peak.

The Combination method results shown for the evening peak were obtained during a second assessment in May 1968. Those obtained in November 1967 have not been shown because they were obtained during darkness. They were indistinguishable from the "existing" results obtained during daylight.

The results shown in Figure 4 are still being analysed statistically and the improvements quoted may be revised, but the present conclusion is that the Combination method settings produced an improvement of 12 per cent in the average journey times of vehicles in Glasgow during the period covered by the survey. It is estimated that, if this improvement were maintained, the savings in the time of vehicles and occupants* would be worth about £600,000 per annum. The capital cost of the equipment installed to date, which includes some features for experimental use only is about £330,000.

The period of the trial was too short to allow any analysis to be made of changes in accidents.

TRANSYT

The results quoted above show that a substantial improvement in journey time can be obtained by using the Combination method to select the best offset between linked fixed-time signals. Another method called TRANSYT, which optimises both offsets and splits and which can be applied to any type of network was developed by Robertson, a Plessey engineer, when working at the Road Research Laboratory. The traffic model used in TRANSYT includes allowance for the flow interaction between successive sections of roads, a simple but effective representation of platoon dispersion, and flow control by signals or by another traffic stream which has right-of-way. The overall impedance to traffic is

* Assuming an average vehicle hour in Glasgow during the day to be worth 17/6d.

measured by a performance index that can be chosen with any desired balance between journey time and number of stops. The optimisation process uses a "hill-climbing" technique to minimise the performance index by altering the points within the signal cycle at which each green signal starts. The cycle time of the system is not changed automatically by the optimisation procedure, although the effect of different cycle times can be considered by successive runs.

TRANSYT traffic model. TRANSYT makes the following assumptions about the traffic situation:-

(i) all major junctions in the network have signals (or are controlled by a priority rule);

(ii) all the signals in the network have a common cycle time or a cycle time of half this value;

(iii) traffic entering the network does so at a constant specified rate on each approach;

(iv) the proportion of traffic turning left or right at each signal remain constant throughout the cycle; it does not depend on where the traffic approaching the signal came from.

Traffic patterns. The common cycle of the signals is divided into 50 equal units of time. All TRANSYT's calculations are made on the basis of the average values of the flow rates and vehicle queues which are expected to occur during each of these units of time. No representation is made of individual vehicles.

The traffic flowing into a link is obtained by taking the appropriate fraction of the traffic leaving upstream links. The pattern of traffic entering a link will be displaced in time and modified during the journey along the link, due to the different speeds of the individual vehicles and platoons of vehicles which will be partly dispersed. The process of platoon dispersal can be expressed by the formula:

$$q'_{(i+t)} = F.q_i + (1-F).q'_{(i+t-1)}$$

where q_i is the flow in the ith time interval of theinitial platoon

q'_i is the flow in the ith time interval of the predicted platoon

t is 0.8 times the average journey time (measured in the time intervals used for q_i)

and F is a smoothing factor which was found to be related to the journey time by the expression

$$F = \frac{1}{1 + 0.5t}$$

Calculation of delay. The average delay per unit time is calculated in two parts which are added together; the first corresponding to a uniform pattern of arrivals at the cycle is obtained by simulating the traffic behaviour over the cycle; the second allows for random variations from cycle to cycle and depends on the average degree of saturation at the stop-line being considered.

TRANSYT optimisation procedure. The first step is to calculate the performance index of the network for an initial set of signal timings. The next stage is to alter the offset of one of the signals by a predetermined number of $\frac{1}{50}$ cycle units and to re-calculate the performance index of the network. If it is reduced the offset is altered successively in the same direction by the same number of units until a minimum value of the index is obtained. If the initial step increases the value of the index, the offset of the signal is altered in the opposite direction. The process of optimising the offsets of each signal in turn is repeated a number of times to obtain the final signal settings.

Re-allocation of green time. In the previous section the amount of green time allocated to each arm of a junction was assumed to be unchanged during the hill-climbing process. TRANSYT is also capable of optimising the start of each green on its own at a signal before proceeding to the next signal. This enables TRANSYT to re-allocate the green time between the various approaches to a signal to reduce the performance index, but it is not allowed to reduce a green time below a specified minimum value for that approach.

Data output. Two forms of data output are available from the Myriad I computer programme on which TRANSYT is run at present. The first consists of tabulated values of the signal settings and the expected traffic behaviour on each link; the second consists of the graphical presentation of flow patterns within the network. The graphs are a development of the conventional type of time-distance diagrams used by traffic engineers to set signal progressions.

Preliminary tests of TRANSYT. Settings produced by TRANSYT were tried on the 8 signals on Cromwell Road used for the earlier tests of the Combination method, which gave an 8 percent improvement in journey time by optimising offsets only. TRANSYT settings which optimised both offsets and splits gave a further 12 percent reduction on average journey time during themid-day period.

The new traffic data collected for TRANSYT was also used to up-date the splits calculated by hand for the Combination method. Using these new data the Combination method gave offsets for Cromwell Road almost identical to those given by TRANSYT. However, it was felt that TRANSYT might be a better method to use in a large network like Glasgow because of its ability to optimise both splits and offsets.

Trial of TRANSYT in Glasgow. A comparison of TRANSYT and the Combination method was carried out in Glasgow during May and June 1968. The preparation of a TRANSYT plan required about the same amount of manual effort as the Combination method: running time on Myriad I was about 1½ hours. The TRANSYT model predicted that journey times would be about 3 percent lower than those given by the Combination method. As before, observations of journey times and traffic flow were made by instrumented cars and road-side observers, but the number of instrumented cars was doubled from two to four.

The results of the trial are shown in Figure 5. Average journey times with TRANSYT were lower by about 5 percent in the morning peak, by about 2 percent between peaks and by about 5 percent during the evening peak. Over the whole day the reduction was about 4 percent. None of these differences was statistically significant, however.

The difference between the TRANSYT journey times and those on the existing system was statistically significant. If these reductions in journey time were maintained for a year it is estimated that the savings in time for vehicles and occupants would be worth £750,000 per annum.

CONCLUSIONS

Both the Combination method and TRANSYT are successful techniques for setting a network of linked fixed-time signals. There was no conclusive evidence that one was better than the other in Glasgow, but both produced substantially lower average journey times than the existing system set by hand. This is a significant result because the existing Glasgow system, a mixture of linked and isolated vehicle-actuated signals, was itself believe to be more advanced than those commonly in use in other parts of the world.

ACKNOWLEDGEMENTS

The author acknowledges the contribution made by his colleagues, particularly Dr. Joyce Holroyd, Mr. P.D. Whiting and Mr. D.I. Robertson. Published by permission of the Director of Road Research. Crown Copyright. Reproduced by permission of the Controller of H.M. Stationery Office.

REFERENCES

1. INSTITUTION OF CIVIL ENGINEERS. Proceedings of the Symposium on Area Traffic Control in February 1967. Paper 1 by B.M. Cobbe, also papers 4, 5, 10, 13, 14 and 15.

2. DEPARTMENT OF SCIENTIFIC AND INDUSTRIAL RESEARCH, ROAD RESEARCH LABORATORY. Research on road traffic p.340-341 London 1965 (H.M. Stationery Office).

3. WEBSTER F.V. and B.M. COBBE. Traffic signals. Ministry of Transport, Road Research Technical Paper No.56, London, 1966 (H.M. Stationery Office).

4. HILLIER J.A. The area traffic control experiment in Glasgow. Proceedings of the 35th Annual Meeting of the Institute of Traffic Engineers "A World Traffic Engineering Conference" held at Boston, Mass., October 1965. Reprinted as Glasgow's experiment in area traffic control. Traffic Engng Control 1965/66, 7 (8), 502-9, (9), 569-71.

5. ALLSOP, R.E. Choice of offsets in linking traffic signals. Traffic Engng Control. 1968, 10 (2), 73-5.

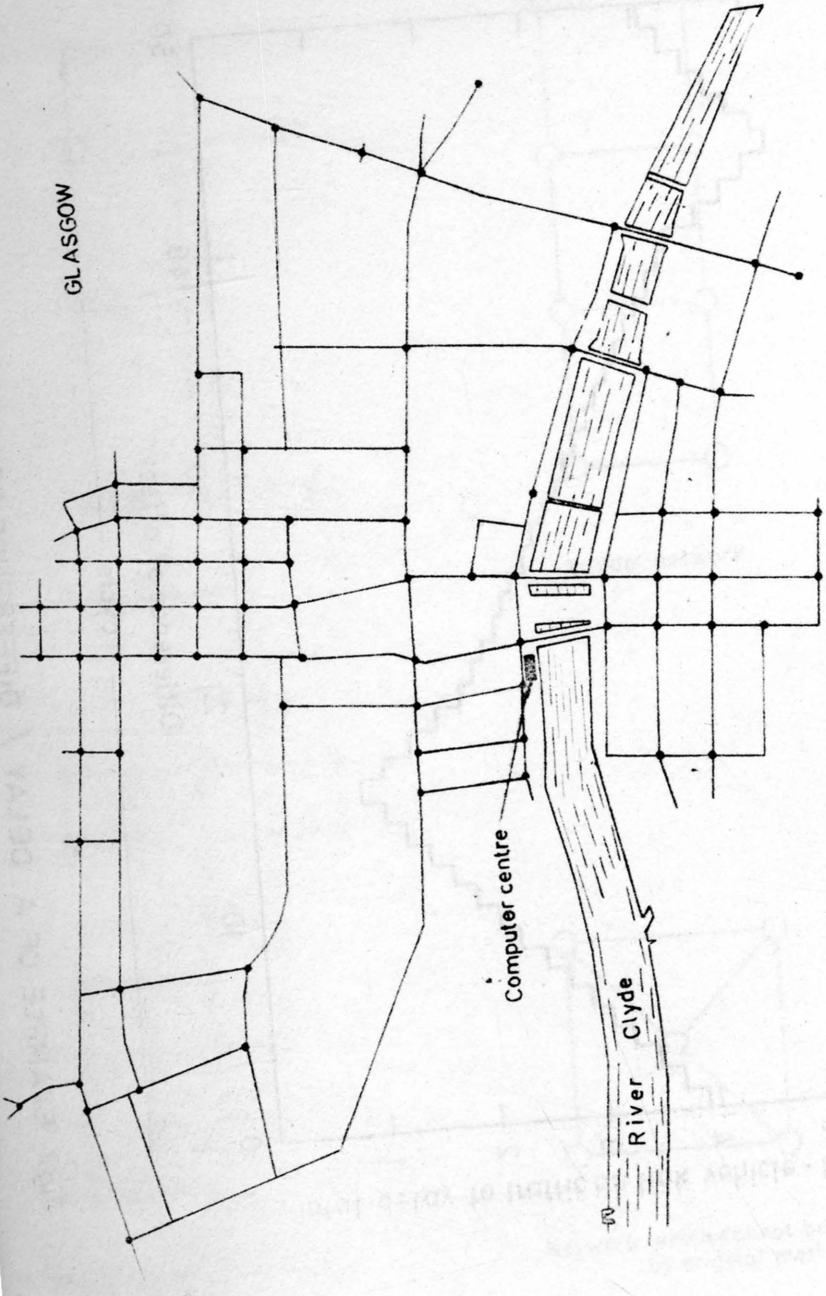


Fig. 1. DIAGRAM OF SIGNAL NETWORK CONTROLLED BY COMPUTER

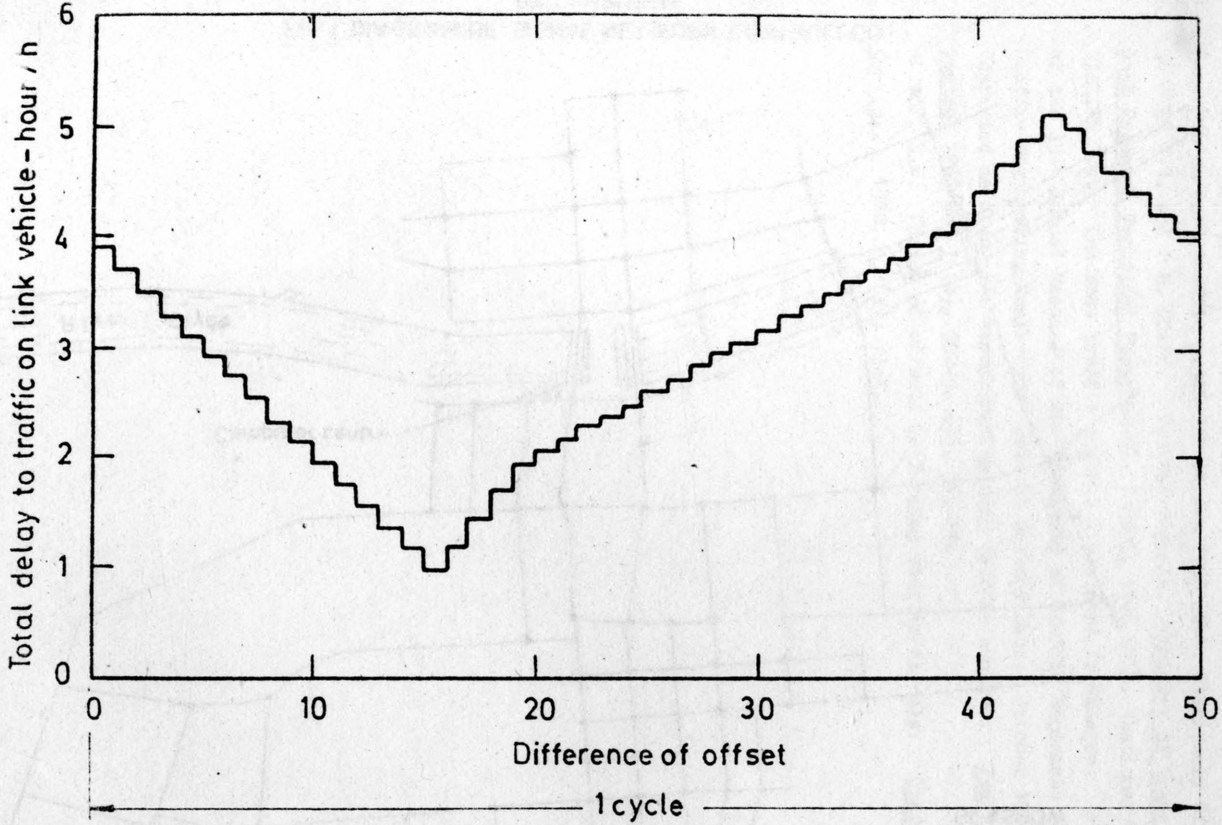


Fig.2. EXAMPLE OF A DELAY / DIFFERENCE-OF-OFFSET RELATION FOR LINK

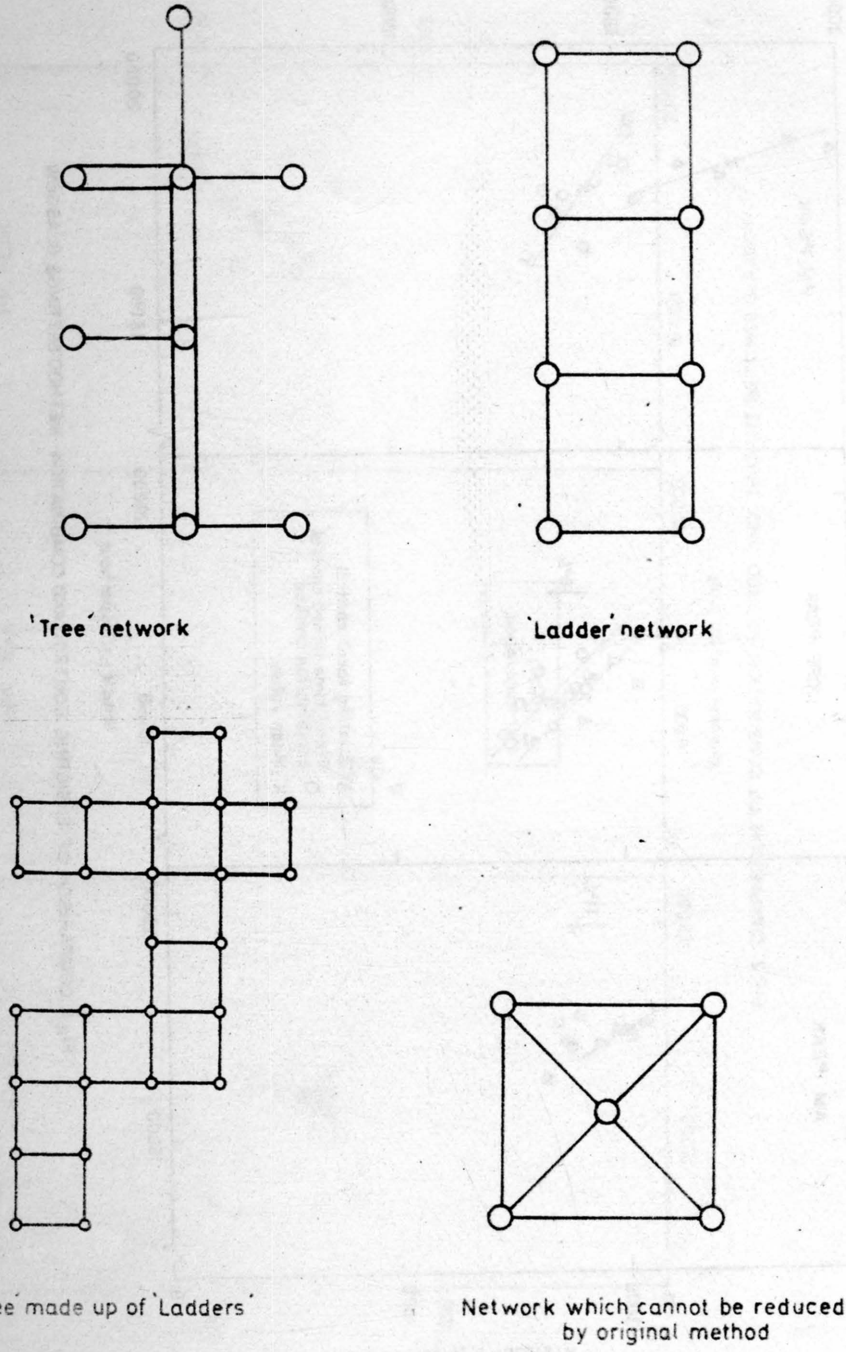


Fig.3 TYPES OF NETWORK

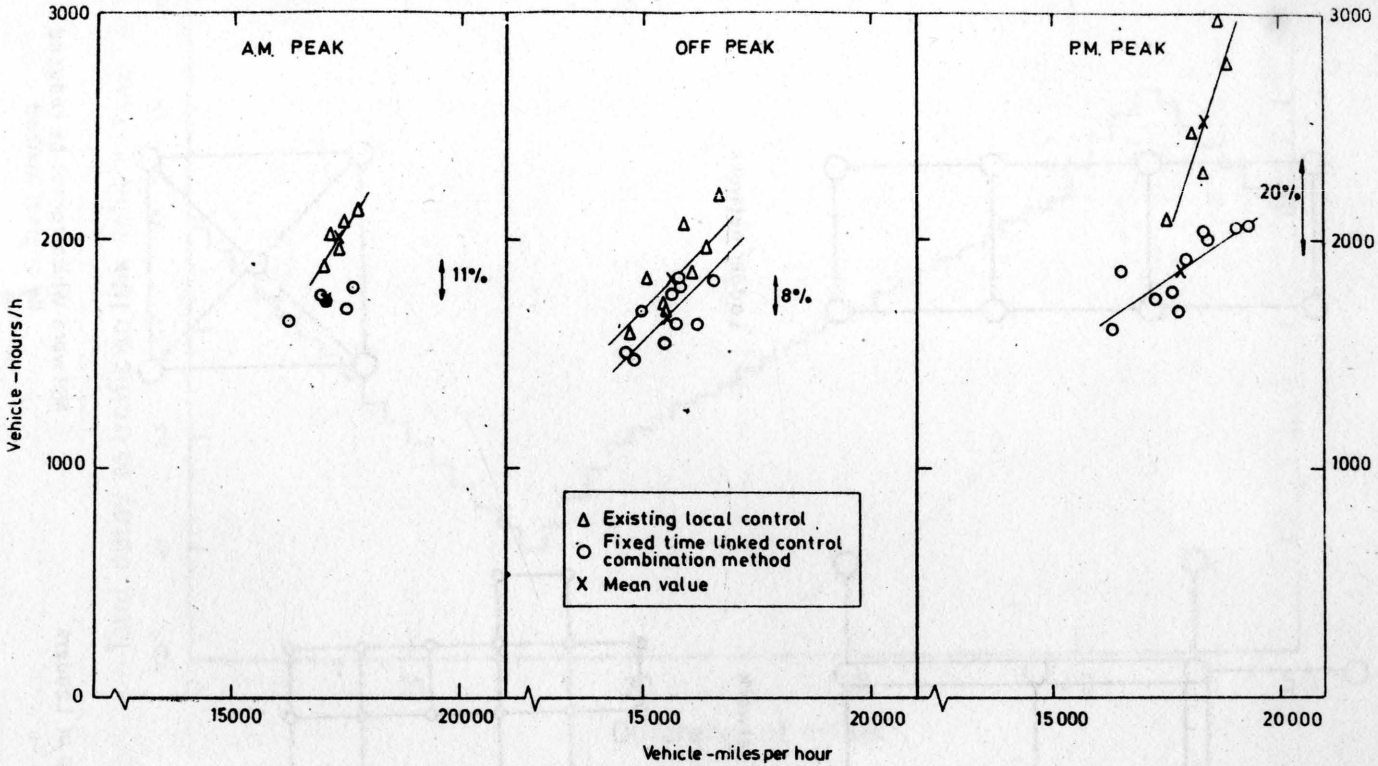


Fig.4. COMPARISON OF EXSISTING CONTROL AND COMBINATION METHOD SETTINGS, GLASGOW

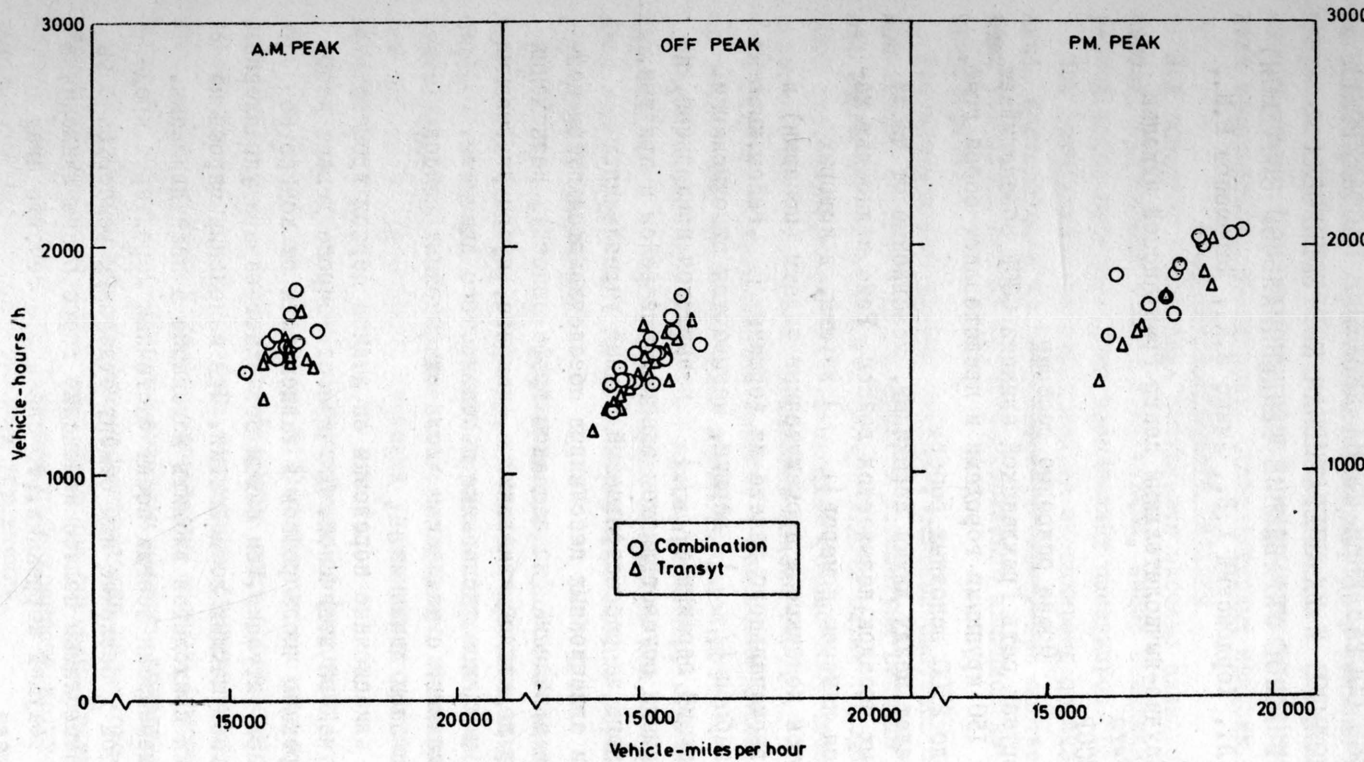


Fig. 5. COMPARISON OF COMBINATION METHOD AND TRANSYT SETTINGS, GLASGOW

АЛГОРИТМЫ СОСТАВЛЕНИЯ ПЛАНА ДВИЖЕНИЯ ПАССАЖИРСКИХ САМОЛЕТОВ И ИХ ОПЕРАТИВНАЯ КОРРЕКТИРОВКА (АВТОМАТИЧЕСКОЕ УПРАВЛЕНИЕ ТЕНДЕНЦИАЛЬНОЙ СИСТЕМОЙ)

Атабегов Л.Д., Кордонский Х.Б., Линис В.К., Парамонов Ю.М., Фролов С.Р.

Научно-вычислительный центр Гражданской авиации
Рига
СССР

Общее описание системы

Транспортная сеть Гражданской авиации СССР обеспечивает связь между 150 крупными городами и представляет собой граф, имеющий около 2500 основных ребер.

Один перелет между двумя вершинами, независимо от числа промежуточных посадок, называется рейсом. Рейс полностью характеризуется набором номеров $\{j_1, \dots j_k\}$ вершин, в которых осуществляются конечные и промежуточные взлеты (посадки) и временем первоначального взлета из вершины j_1. Рейсы, имеющие одинаковые наборы $\{j_1, \dots j_k\}$ вершин, называются одноименными. Путь соединяющий вершины $\{j_1, \dots j_k\}$ называется авиалинией. По данной авиалинии осуществляются одноименные рейсы и при том, не менее одного за рассматриваемый период управления.

Перевозки авиационных пассажиров обеспечиваются путем назначения сети авиалиний (на заданном графе связей), назначения числа рейсов по каждой авиалинии и, наконец, путем увязывания всей рейсов в единое расписание пассажирского движения.

Набор авиалиний с указанием числа одноименных рейсов называется планом движения.

Спрос на авиационные перевозки за данный период времени является случайной величиной. Случайность спроса делает необходимой коррекцию числа рейсов в зависимости от конкретно складывающейся ситуации. Тем самым возникает задача управления в условиях возмущающих воздействий. При изменении спроса на авиаперевозки приходится вводить изменения в план движения. Однако эти изменения всегда носят частичный характер и представляют собой добавление или отмену отдельных рейсов. Было бы совершенно немыслимым полное изменение всего плана движения в порядке его текущей коррекции. Это связано с тем, что план движения задает не только условия перевозки пассажиров, но и условия функционирования всей системы Гражданской авиации : загрузки аэродромов, работы аэродромных служб, вспомогательного транспортного обслуживания пассажиров, распределенных средств и т.д.

Формальное описание системы

Под объектом управления понимается система авиационных связей, включающая в себя граф связей, набор авиалиний, набор рейсов, набор потенциальных авиационных пассажиров (рис.I).

Под потенциальными авиационными пассажирами понимаются лица, претендующие на перемещение, которые, при соответствующих условиях, смогли бы воспользоваться авиационным транспортом.

Рассматриваемый объект является большой системой с огромным числом связей и параметров. К параметрам системы относятся: спрос на авиаперевозки между каждыми двумя городами (более 5000 существенных величин), авиалинии (около 700), числа одноименных рейсов по каждой авиалинии (около 700).

Управляющее воздействие состоит в изменении числа рейсов по линиям, или изменении вида линий (посадок, путей).

Управляющее воздействие по одному параметру осуществляется при получении одного управляющего импульса. Так, для управляющего воздействия по всем параметрам надо подать порядка I400 управляющих импульсов.

Изменение спроса на перевозки имеет систематические и случайные составляющие. К систематическим относятся, например, сезонные изменения. Изменение спроса во времени приводит к изменению состояний объекта во времени и это изменение мы будем называть траекторией движения управляемой системы в пространстве параметров.

Поведение системы задается переходной функцией $P\{X(T)<M|\overline{X}(t), t\in(\tau,s)\}$, где $X(T)$ состояние системы в момент времени T, $\overline{X}(t)$ – траектория (перечнь всех состояний) системы за время (τ,s). Как известно, переходная функция дает возможность прогнозировать будущее поведение системы (имеется ввиду вероятностный прогноз) при известном прошлом поведении.

Оценка поведения системы осуществляется на основе функционала, заданного на траектории движения системы. Таким функционалом может явиться прибыль или себестоимость перевозок за рассматриваемый период времени управления.

Задача управляющих воздействий состоит в оптимизации траектории движения системы, т.е. в выборе траектории, обращающей функционал в максимум (минимум). В вероятностном плане речь может идти либо о математическом ожидании функционала, либо о предельном значении при бесконечношаговом процессе управления.

В рассматриваемом виде объект управления (рис.I) является примером управляемого случайного процесса. Однако здесь имеются особенности, которые оказывают коренное влияние на метод управления этим объектом, и на построение управляющей системы.

Как уже отмечалось, объект управления является большой системой. В этой системе имеется огромное количество связей. Это создает определенные трудности в управлении.

Назовем объемом управления за время ΔT число управляющих импульсов, находящихся в объекте управления и требующих управляющих воздействий. Большая система называется тенденциальной, если запаздывание управляющего воздействия по отношению к соответствующему импульсу растет вместе с ростом объема управления.

Заметим, что запаздывание управляющего воздействия характерно для всех динамических систем. Особенность тенденциальной системы состоит в росте запаздывания при увеличении числа импульсов поданных в систему.

Легко понять, что при управлении тенденциальной системой имеется опасность, что управляющее воздействие будет осуществлено уже тогда, когда входы коренным образом изменились по сравнению с их состоянием при подаче управляющих импульсов, т.е. произойдет рассогласование входов и выходов.

При малом объеме управления, например, при нахождении в объекте одного управляющего импульса управляющее воздействие наступает достаточно быстро.

Применительно к транспортной системе это означает следующее Если подать в объект большое число импульсов, т.е. потребовать большого числа изменений в количестве рейсов и списке авиалиний, то это будет означать требование перестройки расписания пассажирского движения, работы аэродромных служб и перераспределения материальных средств. Время, потребное для осуществления управляющих воздействий будет велико. Если же подать один импульс, например, потребовать ввода дополнительного рейса на линии МОСКВА-ЛЕНИНГРАД, то это может быть осуществлено в течение нескольких часов.

Укажем теперь особенности управления тенденциальной системой:

I. Необходимость перспективного планирования. Учитывая наличие запаздывания управляющих воздействий, для успешности управления, необходимо прогнозировать траекторию движения системы и заранее указывать дискретные моменты времени управляющих воздействий. Перспективный план представляет собой набор управляющих импульсов с большим объемом управления, который заранее вводится в объект управления с учетом запаздывания его реакции.
2. Текущее управление должно быть малым по объему. Текущее управление имеет целью исправления рассогласования входов и выходов при непредвиденных (в прогнозе) изменениях входов. Поскольку реакция системы на текущее управление должна быть быстрой, то это управление должно быть малым по объему.
3. При нестационарном случайном процессе изменения входов (имеется ввиду случайная составляющая) максимум (минимум) функционала недостижим. Для достижения максимума (минимума) функционала, при наличии запаздывания управляющих воздействий, необходимо обеспечить абсолютно точное предвидение изменения входов. При нестационарном изменении входов такое предвидение невозможно.

В соответствии с изложенными особенностями строится система управления тенденциальным объектом. Эта система включает в себя следующие блоки :

I. Блок прогноза входов (прогноза спроса на авиационные перевозки).
2. Блок расчета плана движения (упреждающего набора управляющих импульсов) с учетом связи между числом рейсов и спросом на перевозки.
3. Блок выделения локальной ячейки текущего управления и расчета импульсов текущего управления.

Заметим, что нами не рассматриваются блоки, которые обеспечивают реализацию импульсов в управляющее воздействие. Это связано с тем, что такие блоки "приводов" не носят автоматического характера и "привод" осуществляется посредством исполнителей - людей. Перечисленные же выше блоки управления являются электронными и осуществляются автоматически в ЭВМ. Схема управления объектом, без учета блоков-"приводов", изображена на рис.2.

Ниже каждый из блоков управления будет описан более подробно.

Блок прогноза входов

Как показано на рис.2, прогноз входов основан на предыстории движения системы. Тем самым речь идет об экстраполяции поведения входов на будущее время. Экстраполяция носит основанный характер, если процесс изменения входов является процессом с сильным последействием.

Входы выступают как случайный спрос на авиаперевозки между городами. Здесь содержится типовая ситуация спроса-предложения, имеющая место при любом виде торговли.

Довольно большое распространение имеет ошибка, когда вход изучается изолированно, без учета его связи с выходом[4]. При этом для прогнозирования входов используются простейшие экстраполяционные формулы (полиномы и т.д.), без построения математической модели спроса-предложения. Ошибочность этого подхода состоит в непонимании, что точки изменения предложения образуют в процессе изменения входов (спроса) вложенную цепь Маркова и этим нарушают последействие. Как показали прямые расчеты,использование экстраполяционных формул приводит при прогнозировании спроса на авиаперевозки к ошибкам, иногда составляющим 1000 % от искомого числа.

Таким образом, для экстраполяции входов нужна математическая модель, учитывающая связь между входом и выходом. Довольно характерно, что речь идет о математическом описании деятельности человека, как субъекта перевозок. Нет нужды говорить, что на деятельность человека оказывают влияние огромное число факторов. Но все же среди них можно указать доминирующие, а влияние остальных учесть путем представления деятельности человека в виде случайных событий. При выборе вида транспорта и выборе решения о совершении поездки доминирующими факторами являются ресурсы денег и времени данного пассажира, стоимость и комфорт, предоставляемый тем или иным видом транспорта. Под комфортом понимается время нахождения в пути, частота курсирования общие удобства. Частота курсирования (число между городами А, Б рейсов), как видно из рисунка 2, есть выход объекта управления. Таким образом образуется органическая связь между входом и выходом. Заметим, что с ростом частоты курсирования увеличивается комфортабельность перевозок, т.к. пассажиру легко подобрать удобное для него время отправления (прибытия). Поэтому рост до некоторого предела число рейсов способствует повышению спроса на перевозки.

Задача построения модели состоит в указании вероятности того, что потенциальный пассажир воспользуется авиационным транспортом.

Не останавливаясь на подробностях, укажем, что если транспортная связь между городами обеспечивается только железной дорогой и авиацией, то в предположении логарифмически-нормального распределения денежных ресурсов пассажиров (этот вид распределения получил подтверждение при статистическом обследовании доходов населения социалистических стран[7] имеют место следующие формулы :

Вероятность того, что потенциальный пассажир согласится совершить поездку

$$P^{+} = 1-(1-q)\Phi\left(\frac{\ln \delta^{*} - m}{\sigma}\right), \quad \delta^{*} = \min(\delta^{a}, \delta^{ж}) \tag{I}$$

Если при этом $\delta^{a} > \delta^{ж}$, то вероятность того, что пассажир воспользуется авиационным транспортом есть

$$P^{a} = \frac{1-(1-q)\Phi\left(\frac{\ln \delta^{a} - m}{\sigma}\right)}{1-(1-q)\Phi\left(\frac{\ln \delta^{ж} - m}{\sigma}\right)} \cdot \left(1 - \frac{k^{ж}}{k^{ж}+k^{a}} \exp\left(-\frac{k^{a}}{168}(\alpha\tau^{ж} - \tau^{a})\right)\right) \tag{2}$$

если $\delta^{a} < \delta^{ж}$, то

$$P^{a} = 1 - \frac{1-(1-q)\Phi\left(\frac{\ln \delta^{ж} - m}{\sigma}\right)}{1-(1-q)\Phi\left(\frac{\ln \delta^{a} - m}{\sigma}\right)} \cdot \frac{k^{ж}}{k^{ж}+k^{a}} \exp\left(-\frac{k^{a}}{168}(\alpha\tau^{ж} - \tau^{a})\right) \tag{3}$$

здесь $\delta^{ж}, \delta^{a}, k^{ж}, k^{a}, \tau^{ж}, \tau^{a}$ – стоимости билетов, числа поездов (рейсов) в неделю, времена передвижения по железной дороге и в авиации ;

q, m, σ, α – некоторые константы

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-\frac{z}{2}\right) dz$$

Общее число пассажиров, желающих воспользоваться авиационным транспортом в отрезок времени $(t, t+\tau)$, представится как

$$\mathcal{N}^{a}(t, t+\tau) = \mathcal{N}(t, t+\tau) \cdot P^{+} \cdot P^{a} \qquad (4)$$

где $\mathcal{N}(t, t+\tau)$ общее число потенциальных пассажиров, между городами А и Б. Значение $\mathcal{N}(t, t+\tau)$ как и константы, входящие в формулы (I) – (3) получаются путем статистической обработки предистории. Заметим, что практически речь идет об определении величины $\mathcal{N}^{a}$ за август месяц каждого года.

С помощью формулы (4) рассчитываются не параметры процесса изменения входов, а случайная величина – суммарный авиационный поток за август месяц. Для расчета процесса изменения входов необходимо на основе истории движения объекта установить систематическую составляющую изменения входов. К такой составляющей относятся сезонные колебания потоков и рост потока вблизи праздничных дней. Процесс анализа систематической составляющей изменения входов является весьма сложным, многофакторным анализом и здесь, естественно, подробно не рассматривается. Заметим только, что каково бы ни было качество прогноза изменения входов, оно не может учесть неожиданное изменение обстановки. В качестве яркого примера, укажем на влияние землетрясения в Ташкенте на рост числа перевезенных пассажиров.

В конечном счете, прогноз входов дает общее описание их изменения (применительно к перевозкам между парой городов А – Б в виде

$$Y(t) = G(t)\rho(t) + C(t) \qquad (5)$$

где $G(t), C(t)$ – детерминированные функции, зависящие от выходов, а $\rho(t)$ – стационарный случайный процесс

Пуассоновского типа.

Расчеты прогноза входов весьма громоздки. Достаточно указать, что для изучения предистории движения объекта в ЭВМ вводится 450 000 первичных данных. В результате расчета выводится порядка 10 000 данных. Сказанное означает, что прогноз должен осуществляться на сравнительно далекую перспективу. В настоящее время прогноз делается за год до ввода плана в действие.

Блок расчета перспективного плана

В целом процесс изменения входов, заданный формулой (5), является нестационарным. Однако имеются сравнительно длительные отрезки времени, когда $G(t)=const$, $C(t)=const$, т.е. имеются участки стационарного спроса. Естественно, что моменты нарушения стационарности спроса являются моментами, когда требуется управляющее воздействие.

Расчет перспективного плана состоит из следующих этапов :

а) разбиение периода управления на участки стационарного спроса ;

б) расчет потребных авиалиний для участка наибольшего стационарного спроса ;

в) расчет потребного числа рейсов для каждого участка стационарного спроса. При этом учитывается связь входов и выходов.

Рассмотрим указанные этапы несколько подробнее. Разбиение на участки стационарного спроса осуществляется на основе изучения предистории движения объекта, при этом на следующий год целиком переносятся закономерности предыдущего года.

Обычно участок наибольшего стационарного спроса приходится на август месяц. Для этого месяца производится расчет авиалиний Учитывая тенденциальность системы и допустимость срока запаздывания управляющих воздействий по отношению к плану не более, чем за год (речь идет о плане краткосрочной перспективы) в основу расчета вкладывается список линий, действовавших в предыдущий год. На основе специальных алгоритмов производится поиск новых авиалиний, по парам городов, которые ранее не имели авиационной прямой связи. Скомпанованный таким образом список авиалиний берется за основу дальнейшего расчета.

Затем производится расчет необходимого числа одноименных рейсов. В основу расчета кладется требование перевозки всех пассажиров, желающих осуществить перелет. При указанном расчете имеется обратная связь со входами и со списком авиалиний. Связь со входами основана на использовании процесса последовательных итерраций расчета. Эти итеррации учитывают, что функция спроса является выпуклой по отношению к числу рейсов. На каждом шаге итеррации используется модель линейного программирования типа :

а) ограничения $\sum_{i \in \omega_j} z_i \geqslant b_j$, $j = 1, 2, \ldots m$;

$z_i \geqslant 0$, $i = 1, 2, \ldots n$;

б) функционал минимума себестоимости перевозок $\sum_{i=1}^{n} c_i z_i$

Здесь z_i - вспомогательные неизвестные, так называемые варианты загрузки авиалиний, ω_j - множество всех вариантов загрузки, обслуживающих j -тую пару городов.

Решение линейной задачи дает возможность указания числа необходимых рейсов. Далее осуществляется возврат к формулам прогноза (2), (3), (4) и вновь расчет числа необходимых рейсов и т.д. Процесс итерраций быстро сходится и практически достаточно двух итерраций.

Полученное решение анализируется в ЭВМ для уточнения списка авиалиний. С этой целью, по каждой авиалинии подсчитывается число перевезенных пустых кресел и производится уточнение состава промежуточных посадок и назначение новых авиалиний. В конечном счете обеспечивается, чтобы на каждом участке авиалинии процент использования кресел составлял порядка 80 %.

После уточнения состава авиалиний вновь производится процесс расчета числа рейсов с итеррациями.

Следующий этап вычисления плана связан с учетом наличного парка самолетов. А именно, по подсчитанному плану определяется потребное число самолетов всех типов. Эти величины сравниваются с располагаемыми. Затем методом выравнивающих коэффициентов производится уточнение плана, при котором по мере необходимости уменьшается число рейсов и рейсы перераспределяются между типами самолетов. Выравнивающие коэффициенты учитывают, что при уменьшении числа рейсов,в первую очередь необходимо обеспечить доставку дальних пассажиров и пассажиров, где нет железнодорожного транспорта.Общая схема расчета плана приводится на рис.3.

Блок выделения ячейки текущего управления и расчета импульсов

Как уже отмечалось, необходимость в текущем управлении возникает при непрогнозированном изменении входов. Заметим, что потребность в текущем управлении существенно зависит от качества прогноза поведения входов и качества разработки перспективного плана.

При небольших отличиях поведения входов от прогнозированного имеется некоторый резерв плана, который позволяет не прибегать к текущему управлению. Таким резервом является планируемый коэффициент заполнения кресел (80 %). Очевидно, что повышение спроса на перевозки по отношению к запланированному даже при 10 % не требует увеличения числа рейсов.

При больших отличиях поведения входов (15 % и выше) приходится прибегать к текущему управлению. Соответственно встает вопрос об необходимом объеме управления.

Общий граф авиационных связей весьма велик. При расчете текущего управления надо выделить подграф связей, внутри которого производится перерасчет числа потребных рейсов и изменение состава авиалиний.

Обозначим через Y_{j_1,j_2} пассажирский авиапоток между городами A_{j_1}, A_{j_2}. Обозначим через $y^{(k)}_{j_1,j_2}$ пассажирский поток, перевозимый по линии с номером k между городами A_{j_1}, A_{j_2}. Авиалинии с номерами $k_1, \ldots k_\ell$ называются связанными по потоку (j_1, j_2) если

$$\sum_{i=1}^{\ell} y^{(k_i)}_{j_1,j_2} = Y_{j_1,j_2}, \quad y^{(k_i)}_{j_1,j_2} > 0, \quad i = 1,2,\ldots \ell.$$

То авиалинии, для которых $y^{(k_i)}_{j_1 j_2} = 0$ являются на участке (j_1, j_2) изолированными.

Допустим, что поток $Y_{j_1 j_2}$ существенно изменился по отношению к спрогнозированному. Тогда извлекается ячейка текущего управления состоящая из авиалиний, для которых $y^{(k_i)}_{j_1 j_2} > 0$ и связанных с ними по потокам авиалиний. Как показывают расчеты, объем ячейки текущего управления, для правильно выполненного перспективного плана невелик и составляет обычно, не более пяти авиалиний.

В пределах ячейки текущего управления производится весь расчет, который уже был изложен при описании блока расчета перспективного плана. Пример ячейки тенденциального управления плана движения 1968 года приводится на рис.4. Объем расчета ячейки мал и требует малого времени (не более часа). Объем управления сравнительно не велик. Обычно он не превышает трех - пяти импульсов.

Заключение

Изложенная система управления используется в практике Аэрофлота с 1967 года. Опыт показывает, что многие части этой системы нуждаются в дальнейшем развитии. Это касается как блока прогноза, так и блока расчета перспективного плана. Однако, сами принципы управления себя полностью оправдали. Эти принципы видимо пригодны и для других тенденциальных систем, связанных со спросом - предложением.

СПИСОК ЛИТЕРАТУРЫ

1. ГЕРЦБАХ И.Б.,КОРДОНСКИЙ Х.Б.,НИКИТИНА А.Б. О статистической оценке величины неудовлетворенного спроса на пассажиркие авиаперевозки. Труды РИИ ГА, выпуск 73. "ПРИМЕНЕНИЕ МАТЕМАТИЧЕСКИХ МЕТОДОВ В ПЛАНИРОВАНИИ УПРАВЛЕНИЙ". Рига, 1966.
2. ЛИНИС В.К. Об оптимальном планировании транспортной деятельности. Труды РИИ ГА, выпуск 73. Рига, 1966.
3. ЛИНИС В.К. Организация авиационных пассажирских перевозок. Материалы к Всесоюзной конференции по применению экономико-математических методов в отраслевом планировании и управлении, секция №-1, М., 1966.
4. МОТ Ж. Статистические предвидения и решения на предприятиях, "ПРОГРЕСС", М.,1966.
5. ПАРАМОНОВ Ю.М., МОЛЧАНОВА Л.В. Анализ распределения числа пассажиров вблизи праздничных дней. Труды РИИ ГА, 73, Рига, 1966.
6. ПАРАМОНОВ Ю.М., ФРОЛОВ О.Р. Анализ и экстраполяция пассажирских перевозок. Материалы к Всесоюзной конференции по применению экономико-математических методов в отраслевом планировании и управлении, секция №-1, М., 1966.
7. ЭЛТЭТЕ Э. Исследование характера и свойств распределения доходов, сборник статей "ЖИЗНЕННЫЙ УРОВЕНЬ", Статистика, М., 1966.

А в т о р ы :

АТАБЕГОВ Леонид Дмитриевич, ведущий инженер Научно-вычислительного центра Гражданской авиации.

А д р е с : г.Рига-19, ул.Ломоносова, 1, Научно-вычислительный центр ГА.

КОРДОНСКИЙ Хаим Борисович, доктор технических наук, профессор, старший научный сотрудник НВЦ ГА.

ЛИНИС Валдемар Карлович, руководитель группы математиков-программистов НВЦ ГА.

ПАРАМОНОВ Юрий Михайлович, кандидат технических наук, начальник отдела НВЦ ГА.

ФРОЛОВ Ор Романович, ведущий инженер НВЦ ГА.

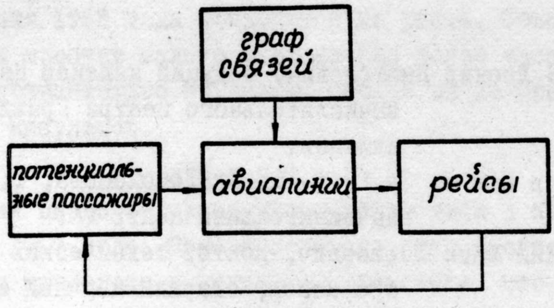

граф связей
потенциальные пассажиры
авиалинии
рейсы

Рис. 1

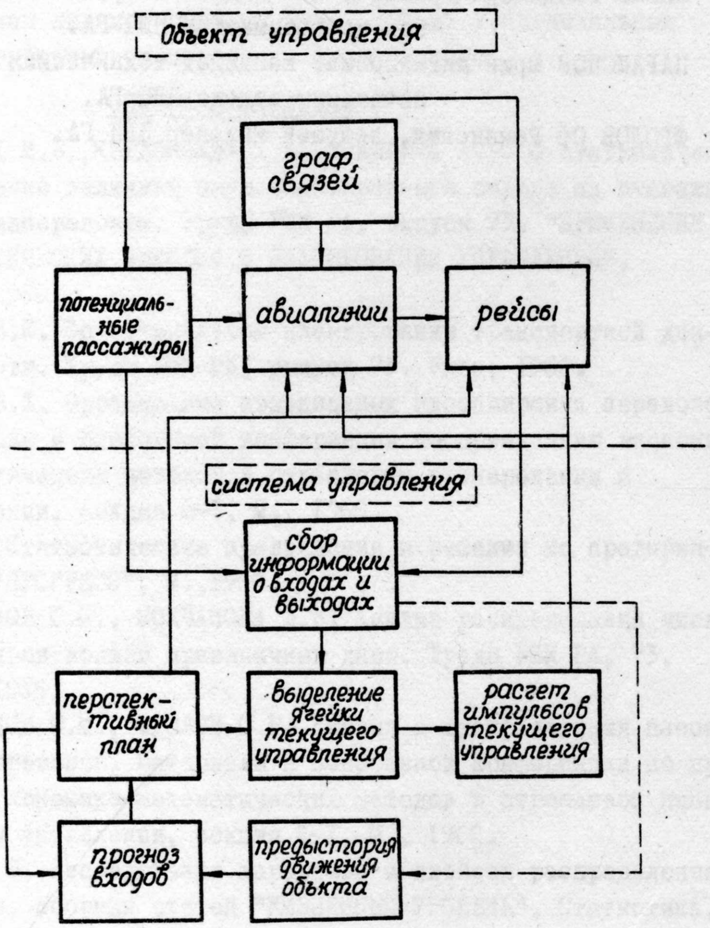

Объект управления
граф связей
потенциальные пассажиры
авиалинии
рейсы
система управления
сбор информации о входах и выходах
перспективный план
выделение ячейки текущего управления
расчет импульсов текущего управления
прогноз входов
предыстория движения объекта

Рис. 2

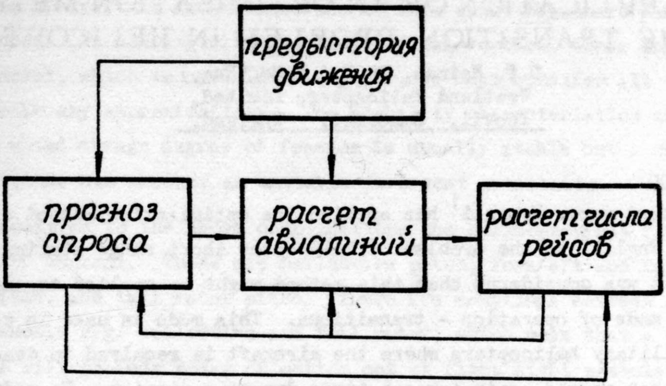

предыстория движения
прогноз спроса
расчет авиалиний
расчет числа рейсов

Рис. 3


Москва
Ростов
Грозный
Мин. Воды
Махачкала
Баку

Рис. 4

THE APPLICATION OF AN OPTIMISATION METHOD TO THE TRANSITION PROBLEM IN HELICOPTERS

D.F. Haines, Ph.D., B.Sc.(Eng.)
Westland Helicopters Limited,
Yeovil, Somerset, England.

Introduction

In a recent paper Rynaski[1] has applied the optimisation method of Kalman and Englar[2] to the problem of helicopter short range station keeping. It was considered that this method might be applied to another helicopter mode of operation - transitions. This mode is used in certain types of military helicopters where the aircraft is required to descend, hover and then ascend again several times during a mission. To reduce pilot work-load the transition becomes an autopilot function and the pilot merely presses a button to initiate the transition down or transition up. This paper describes an investigation made to find a method that eases the task of finding suitable control laws for this particular mode of flight. The results could give the design engineer an appreciation of the characteristics, as a function of speed, of a helicopter with a new airframe.

The transition problem

The "transition down" term infers that a helicopter flying at cruise speed and an intermediate height is required to reduce speed and lose height at the same time, until the aircraft is hovering at some height, usually less than 50 ft. Conversely "transition up" means flying from the hover condition back to cruise speed while climbing. To achieve operational efficiency it is obviously necessary that a transition is performed in the minimum time subject to constraints of blade pitch angles allowable and aircraft attitude excursions. However the pilot will also want to know when he presses the transition-down button where the aircraft will eventually come to the hover. It is therefore advantageous that the aircraft rate of deceleration and rate of descent be constant during the manoeuvre. The problem therefore becomes one of controlling two variables e.g. forward deceleration and vertical speed with limits placed on the amount of control which may be used and the pitch attitude excursion of the aircraft.

A brief account of the dynamics of the helicopter are perhaps pertinent at this point so that the vehicle to be controlled may be appreciated.

The helicopter is a tightly-coupled dynamic system. Compared with fixed wing aircraft it is not possible to talk about the short period and phugoid mode. In dealing with the longitudinal-vertical three degree-of-freedom model, which is used in this paper, we must consider all the terms and not make any approximations. The stability characteristics are such that the speed change degree of freedom is usually stable but a change in pitch attitude can produce an unstable divergent oscillation.

In addition to the above complications the helicopter has four channels of control. These are collective pitch, fore/aft and lateral cyclic pitch, and tail rotor pitch. There are couplings between some of these channels e.g. collective and tail rotor, which mean that a change in a control will produce modes of motion not at first sight associated with that channel. All these facts then make the system designers problem more complex and use of modern control theory helps to ease the problem of finding suitable control laws.

Theory

Only an outline of the theory of optimisation used in this investigation will be required since many papers[2] have been written previously concerning the method.

The linearised equations of motion of the helicopter may be written as a set of first order differential equations in matrix form

$$\dot{x} = Fx + Gu \quad \text{.....................(1)}$$

$$y = Hx \quad \text{.....................(2)}$$

Where x - the vector of state variables describing the helicopter motion

u - the control vector

y - the vector of outputs which are to be minimised

F - the matrix of coefficients of the equations of motions describing the coupling among the state variables

G - the matrix of coefficients describing the effect of control inputs on the equations of motion

A quadratic performance index may be written:

$$2V = \int_0^{\infty} (y'Q\,y + u'R\,u)\,dt \quad \text{.....................(3)}$$

where Q - positive definite matrix whose elements weight contributions of each output in the index

R - positive definite matrix whose elements weight the contributions of each control motion in the index.

Using the calculus of variations the Euler-Lagrange partial differential equations of motion must be satisfied.

$$\frac{\partial L}{\partial x} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) = 0 \qquad \frac{\partial L}{\partial u} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{u}}\right) = 0 \quad \ldots\ldots\ldots\ldots(4)$$

Where

$$L = \tfrac{1}{2}\,(y'Qy + u'Ru) + \lambda'\,(-\dot{x} + Fx + Gu) \quad \ldots\ldots\ldots\ldots(5)$$

and λ is the adjoint state vector

Combining the equations we obtain

$$\dot{x} - Fx + Gu = 0 \quad \ldots\ldots\ldots\ldots(6)$$

$$Ru_o + G'\lambda = 0 \quad \ldots\ldots\ldots\ldots(7)$$

$$H'QHx + \dot{\lambda}' + F'\lambda = 0 \quad \ldots\ldots\ldots\ldots(8)$$

Therefore the optimal control law is

$$u_o = -R^{-1}\,G'\lambda \quad \ldots\ldots\ldots\ldots(9)$$

By substituting $\lambda = Px$, $\dot{\lambda} = P\dot{x}$ equations (6), (7) and (8) can be reduced to the matrix Riccati equation

$$0 = PF + F'P - PGR^{-1}G'P + H'QH \quad \ldots\ldots\ldots\ldots(10)$$

The gains for the control laws are determined by finding the positive definite symmetric solution to this equation and substituting P into (9) to give

$$u_o = -R^{-1}\,G'Px \quad \ldots\ldots\ldots\ldots(11)$$

Application to the transition problem

When a control system is designed that is required to control aircraft height designers have usually only considered achieving this control by applying collective pitch to the rotor blades. Similarly a speed control system would use cyclic pitch as its control. This convention is reasonable when the aircraft is flying at cruise speed in level flight but during a transition manoeuvre there is such interaction between the two channels aerodynamically that it is unlikely that any design produced by the above method will give optimum results. To illustrate this point control laws will be found with the performance indexes:

$$\left.\begin{aligned} 2V &= \int_0^{\infty} (q_u u^2 + r_{B_1}\,\Delta B_1^2)\,dt \\ 2V_1 &= \int_0^{\infty} (q_w w^2 + r_{\Theta_o}\,\Delta \Theta_o^2)\,dt \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots(12)$$

and

$$2V_2 = \int_0^\infty \left\{ [u \ w] \begin{bmatrix} q_u & 0 \\ 0 & q_w \end{bmatrix} \begin{bmatrix} u \\ w \end{bmatrix} + [\Delta B_1 \ \Delta\Theta_o] \begin{bmatrix} r_{B_1} & 0 \\ 0 & r_{\Theta_o} \end{bmatrix} \begin{bmatrix} \Delta B_1 \\ \Delta\Theta_o \end{bmatrix} \right\} dt \ldots (13)$$

The indexes given by (12) give two laws, one for cyclic pitch blade angle, the other for collective pitch blade angle, without any consideration being made for coupling between the channels. The control law given by (13) weights both output variables and the two control variables to give control laws which are geared together to produce the optimal responses.

Separated Channel Control Laws

The performance index used here was that of equation (12) with a range of $q_{u/r_{B1}}$ and $q_{w/r_{\Theta_o}}$ values. Further since a control law is required over a speed range where the aircraft characteristics are changing, laws were derived at hover, 30, 60 and 100 knots true airspeed. Solutions to the Riccati equation and subsequent transient responses were computed on an I.B.M. 360/40 machine with programmes written as suggested by Kalman, Englar and Bucy[2].

The linearised equations of motion are written with respect to a frame of reference fixed in the aircraft thus it is necessary to transform these to derive control laws whose signals are obtained from conventional forward speed and vertical speed sensors. All the laws given will be with respect to an earth-orientated frame of reference. The figures used are those for the SH.3D helicopter which is a heavy military machine currently being developed at Westland Helicopters, although the original airframe design is from Sikorsky of the U.S.A.

For the separate channel case two laws are obtained:

$$\Delta B_1 = k_u U + k_w W + k_r \dot{\Theta} + k_p \Theta$$

$$\Delta \Theta_o = \ell_u U + \ell_w W + \ell_r \dot{\Theta} + \ell_p \Theta$$

Where the k's and ℓ's are the derived coefficients.

Table 1 gives the fore/aft cyclic channel control laws for $q_{u/r_{B1}}$ equal to 1, .5, .25, .1 and .04 at HOVER, 30, 60 and 100 knots true air speed. Table 2 gives the equivalent figures for the collective channel.

For the control of forward speed it is evident from Table 1 that the signal from the vertical speed sensor has a very minor effect and could quite possibly be omitted. However, the pitch angle and pitch rate terms are quite large. This is not surprising since the normal automatic stabilisation of a helicopter is effected with a very similar control law. These pitch terms indicate that the aircraft is more stable in the middle speed range than at the extremes and this also bears out previous experience of helicopter system designers.

Table 2 shows that the collective channel control law changes very dramatically through the speed range. Except at hover the forward speed is negligible, but at hover this term is very large. A further point about the hover laws is that the pitch and pitch rate terms indicate that a positive change in collective angles pitch induces a nose-down attitude, whereas at other speeds a nose-up change occurs. This certainly illustrates the problems in designing control laws for large speed ranges in helicopters.

To indicate the responses obtained with some controls derived with channels separated Figs. 1 and 2 show the response to an error of -5 ft./sec. in forward speed and 3 ft./sec. in vertical speed. Fig. 1 is with $q/r = .1$ at 60 knots and Fig. 2 $q/r = .1$ at hover. The laws are:

$$\text{Fig. 1} \quad \begin{cases} \Delta B_1 = -.296U - .010W + .203\,\dot{\Theta} + .317\,\Theta \\ \Delta\Theta_o = -.009U + .206W - .109\,\dot{\Theta} - .196\,\Theta \end{cases}$$

$$\text{Fig. 2} \quad \begin{cases} B_1 = -.297U - .013W + .219\,\dot{\Theta} + .334\,\Theta \\ \Delta\Theta_o = -.462U + .237W + .226\,\dot{\Theta} + .531\,\Theta \end{cases}$$

The responses shown in Fig. 1 are not entirely acceptable. The vertical speed error W decays very slowly towards the end of the transient after a fast initial decrease. The collective pitch angle increment remains at a high level for a long time and this is not particularly desirable. Fig. 2 shows very much different responses. The vertical speed response includes a large overshoot and the rate of change would probably cause discomfort. The response for the collective pitch angle is quite unacceptable since the initial increment is some 3 degs. and this becomes -.5 degs. in about 1.5 secs.

It is evident that the problem must be approached with combined channels and the performance index of equation (13) should be used.

Combined channels control laws

The performance index, equation (13), contains 4 parameters, i.e. q_u, q_W, r_{B1}, $r_{\Theta o}$ so a very large number of cases could be used to find a control law with a suitable response. It was decided that the investigation here would be done in three stages.

(i) by experience choose q_u and q_W and with various r_{B1} and $r_{\Theta o}$

find control laws at one speed only.

(ii) pick the most suitable control law from those given by (i) and then find the laws at other speeds with the same r_{B1} and r_{Θ_0}.

(iii) apply each law found at each speed and compare the responses expecting that one law will be suitable throughout the speed range.

The choice made for the parameters in (i) was influenced by the experience gained from the separate channel investigation. q_u and q_w were both made equal to 1.0 which means an equal weighting on forward and vertical speed errors. It is usual in helicopters to use more cyclic than collective pitch changes since an increase in collective requires an increase in engine torque which in turn means a change in the tail rotor pitch angle. From (i) it was known that q/r should be .25 or less. The values for r_{B1} and r_{Θ_0} were therefore chosen as:

$$\frac{r_{B_1}}{r_{\Theta_0}} = \frac{5}{10} ; \frac{5}{15} ; \frac{5}{20} ; \frac{10}{15} ; \frac{10}{20} ; \frac{15}{20}$$

The speed chosen for the application of these r_{B1}/r_{Θ_0} values was 60 knots. Table 3 shows the coefficients as determined by solution of the Ricatti equation. As shown in Tables 1 and 2 the U term in the law and the W term in the B_1 law are small and might be neglected. A notable difference between these tables however is the small amount of pitch attitude rate and pitch attitude feedback required in the collective control laws. Although small it is probably not wise to say that the terms are negligible in all cases. For the same initial errors as in the separate channels cases the transient response with each law was found. Fig. 3 shows the response for the ratio r_{B1}/r_{Θ_0} finally chosen. This is when the ratio is 15/20. The amount of control used is small in both channels and the responses in pitch attitude and speeds are good. It may be argued that the collective pitch angle is unnecessarily large for a long time but in fact this was the best achieved in relation to the vertical speed error response. Thus the result obtained from (i) was that the parameters used in the performance index should be:

$$2V = \int_0^{\infty} \left(U + w + 15 \, (\Delta B_1)^2 + 20 \, (\Delta \Theta_0)^2 \right) \, dt$$

This index was now used to obtain control laws for the hover, 30, 60 and 100 kts. cases. These laws are given in Table 4. It can be seen that the major terms in all cases are similar throughout the speed range but differences do occur. Some signs associated with smaller terms change but this should not influence the overall effect. Each law was now applied to all speeds and to gauge the effect on vertical and forward speed responses two measures of these responses were made:

(i) amplitude ratio = $\dfrac{\text{1st overshoot amplitude}}{\text{initial amplitude}}$

(ii) 2% time - time for error to become less than 2% of initial error.

These results are shown in Figs. 4 and 5. Since in many cases there was no overshoot and the response was overdamped Fig. 4 cannot give a full picture. It does show however that the hover law gives a low amplitude ratio at all speeds in both channels. The 30, 60 and 100 knots laws give very unsatisfactory damping in vertical speed below about 20 knots and could not be used at hover conditions. Fig. 5 covers the whole speed range. It is undesirable that any 2% time should be greater than about 10 secs. so again only the hover law would seem to be satisfactory at all speeds. Some laws used at their own speed are not very good e.g. 100 knots law at 100 knots. Assuming that the hover law is good throughout the speed range it can be seen that the 2% time changes from less than 5 secs. for a vertical speed error at hover to 9.5 secs. at 30 knots and then to about 7 secs. at 60 knots after which it becomes just greater than 10 secs. at 100 knots. If it is acceptable that the response shall change this much then the hover law performance is fair at all speeds. Another law which is nearly acceptable throughout the range is the 60 knot law. Regarding the change in 2% time at low speeds this law is superior to the hover law but unfortunately the response above 90 knots becomes very slow.

Conclusions

An investigation has been carried out to find the usefulness of using the Kalman approach to optimisation in the problem of transitions in helicopters. Separate channel and combined channel laws have been formulated and it was hoped that a suitable law might emerge that could be used throughout the aircraft speed range. The results from the separate channel investigation were useful in obtaining a first estimate of the performance index parameters although actual laws given by this approach were unacceptable. The laws derived with the channels combined did not always give satisfactory responses but were very useful in finding the effect of a given law, which was an optimal law for one speed, at other speeds. It is considered that this method is useful in performing a preliminary study to get a feel for the transition problem associated with a particular helicopter with a new airframe and would shorten the time required to obtain a control law for use in a prototype aircraft. It is hoped that the results of this paper will be applied to a full non-linear simulation of the aircraft and their effect on transition profiles will be determined. These results would be presented at the congress.

Acknowledgements

I should like to thank Westland Helicopters Ltd. and the Ministry of Technology for permission to publish this paper. Also Mr. G. Dillaway of Westland for his advice and criticism on the manuscript.

References

1. Rynaski E.G. - Optimal Helicopter Station Keeping. 7th Int. Auto. Control Conf., Seattle 1966 pp. 374-81.

2. Kalman R.E., Englar T.S., Bucy R.S.., - Fundamental Study of Adoptive Control Systems, Vol.1 ASD-TR-61-27 March, 1961.

Notation

B_1 increment in fore/aft cyclic pitch blade angle

θ_o increment in collective pitch blade angle

U error in forward velocity relative to earth axes

W error in vertical velocity relative to earth axes

θ error in aircraft pitch attitude

q_u, q_w, r_{B1}, r_{θ_o} weighting factors

k_u, k_w, k_r, k_p fore/aft cyclic control law coefficients

l_u, l_w, l_r, l_p collective control law coefficients

True Air Speed kts.	q_u/r_{B_1}	k_U	k_W	k_r	k_P
Hover	1.0	-.98	-.016	.296	.654
30		-.98	-.017	.257	.608
60		-.97	-.020	.282	.613
100		-.95	-.023	.293	.713
Hover	.5	-.69	-.015	.275	.538
30		-.69	-.012	.237	.508
60		-.68	-.016	.260	.507
100		-.67	-.017	.269	.578
Hover	.25	-.480	-.014	.252	.440
30		-.484	-.008	.214	.422
60		-.478	-.013	.236	.416
100		-.463	-.013	.244	.466
Hover	.1	-.297	-.013	.219	.334
30		-.300	-.004	.183	.326
60		-.296	-.010	.203	.317
100		-.285	-.008	.209	.351
Hover	.04	-.179	-.011	.186	.251
30		-.184	.001	.153	.251
60		-.182	-.008	.170	.240
100		-.173	-.005	.176	.262

Cyclic control laws (separate channels)

$$\Delta B_1 = k_u U + k_w W + k_r \dot{\theta} + kp\theta$$

TABLE 1

True Air Speed kts.	q_u/r_{B_1}	ℓ_U	ℓ_W	ℓ_r	ℓ_p
Hover	1.0	- 1.08	.883	.212	.982
30		- .025	.888	- .317	- .125
60		- .011	.869	- .055	- .230
100		- .003	.866	- .083	- .400
Hover	.5	- .977	.593	.264	.952
30		- .020	.598	- .295	- .130
60		- .010	.580	- .070	- .222
100		- .003	.575	- .108	- .396
Hover	.25	- .726	.408	.226	.731
30		- .016	.394	- .270	- .139
60		- .010	.378	- .088	- .216
100		- .003	.373	- .134	- .380
Hover	.1	- .462	.237	.226	.531
30		- .009	.217	- .237	- .151
60		- .009	.206	- .109	- .196
100		- .009	.206	- .109	- .196
Hover	.04	- .300	.131	.226	.407
30		- .003	.110	- .203	- .140
60		- .008	.106	- .118	- .170
100					

Collective control laws (separate channels)

$$\Delta\theta_o = lu\,U + l_w W + l_r \dot{\theta} + l_p \theta$$

TABLE 2

r_{θ_0} $q_u = q_w = 1$	r_{B_1}	k_U	k_W	$k_{\dot{\theta}}$	k_{θ}
10	5	- .424	- .077	.231	.428
15	5	- .423	- .093	.237	.448
20	5	- .422	- .104	.242	.461
15	10	- .296	- .049	.206	.346
20	10	- .296	- .057	.211	.359
20	15	- .240	- .039	.192	.307

Cyclic control law at 60 knots (combined channels)

$$\Delta B_1 = k_u U + k_w W + k_{\dot{\theta}} \dot{\theta} + k_{\theta} \theta$$

r_{θ_0} $q_u = q_w = 1$	r_{B_1}	l_U	l_W	$l_{\dot{\theta}}$	l_{θ}
10	5	- .031	.203	-.024	-.103
15	5	- .033	.150	-.019	-.081
20	5	-.032	.120	-.016	-.066
15	10	-.013	.153	-.033	-.104
20	10	-.016	.123	-.028	-.086
20	15	-.005	.124	-.037	-.101

Collective control law at 60 knots (combined channels)

$$\Delta\theta_0 = l_u U + l_w W + l_{\dot{\theta}} \dot{\theta} + l_{\theta} \theta$$

TABLE 3

True Air Speed kts.	k_U	k_W	k_r	k_p
Hover	- .236	- .031	.203	.292
30	- .242	- .074	.234	.342
60	- .240	- .039	.192	.307
100	- .220	- .051	.193	.360

Cyclic control law (final weighting functions)

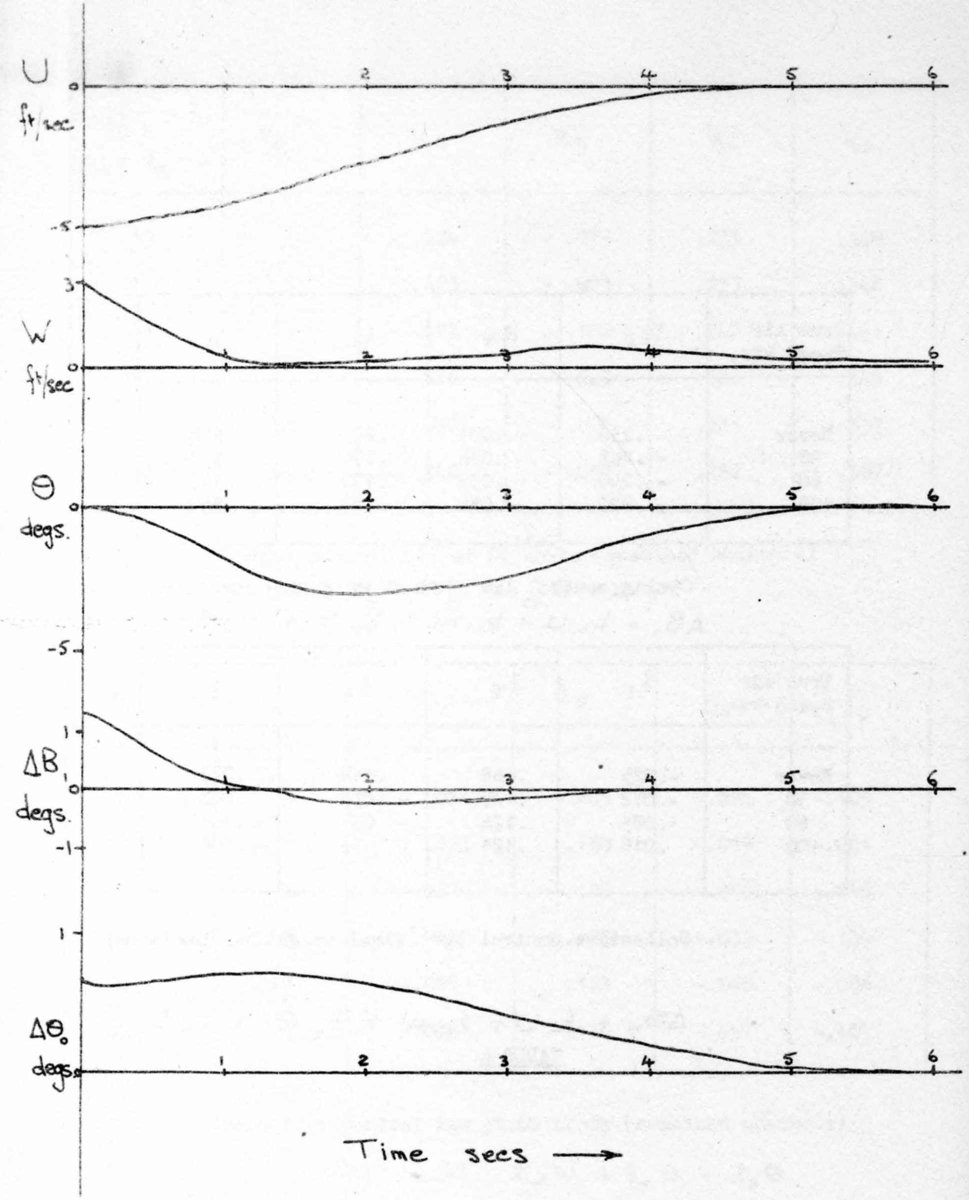
$$\Delta B_1 = k_u U + k_w W + k_r \theta + k_p \theta$$

True air Speed kts.	l_U	l_W	l_r	l_p
Hover	-.025	.165	.012	.022
30	-.072	.127	- .054	.048
60	-.005	.124	- .037	- .101
100	.016	.121	- . 63	- .188

Collective control law (final weighting functions)

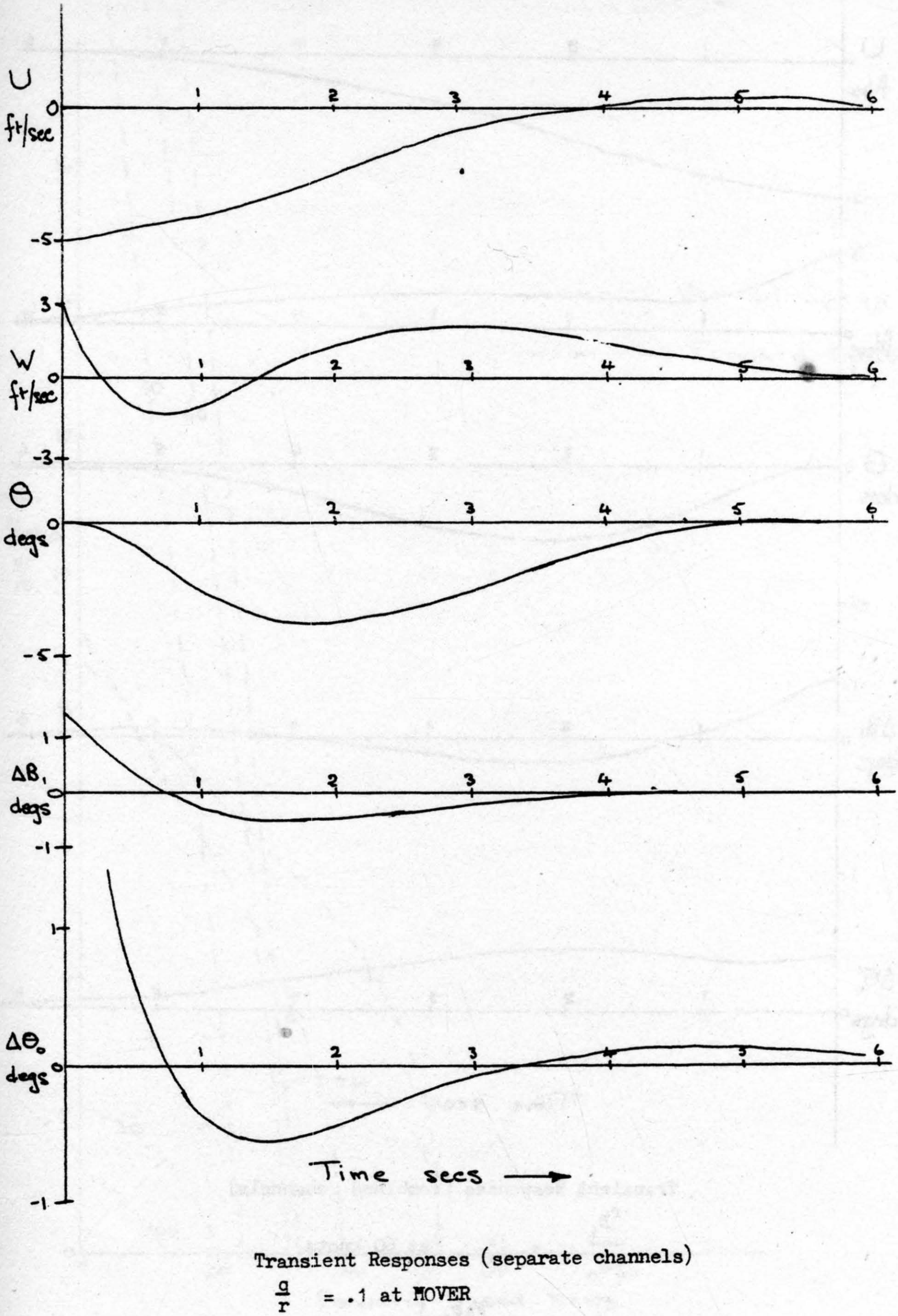
$$\Delta \theta_0 = l_u U + l_w W + l_r \dot{\theta} + l_p \theta$$

TABLE 4



Transient Responses (separate channels)
$\frac{q}{r}$ = .1 at 60 knots.

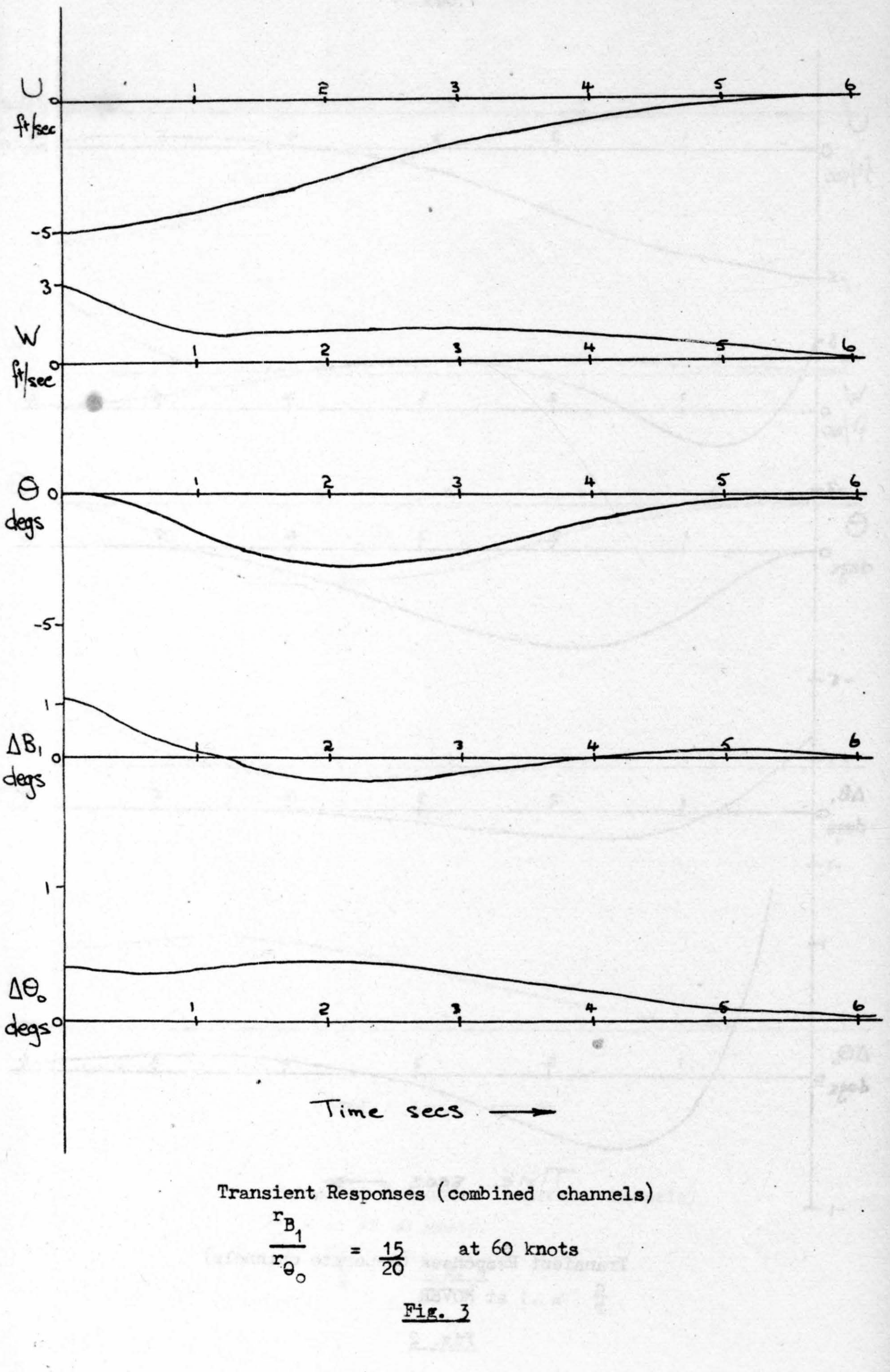
Fig. 1



Transient Responses (separate channels)

$\frac{q}{r}$ = .1 at MOVER

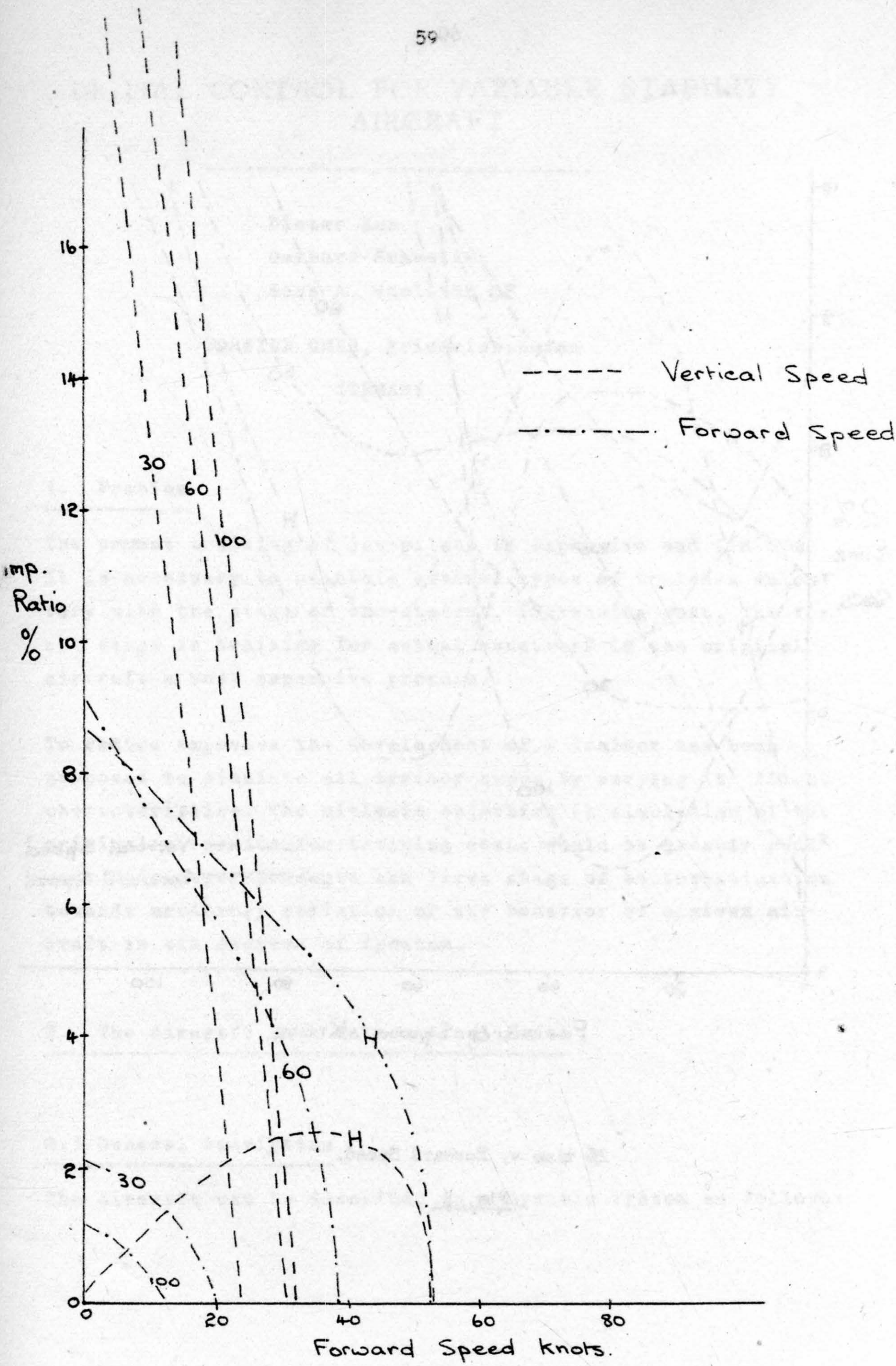
<u>Fig. 2</u>



Transient Responses (combined channels)

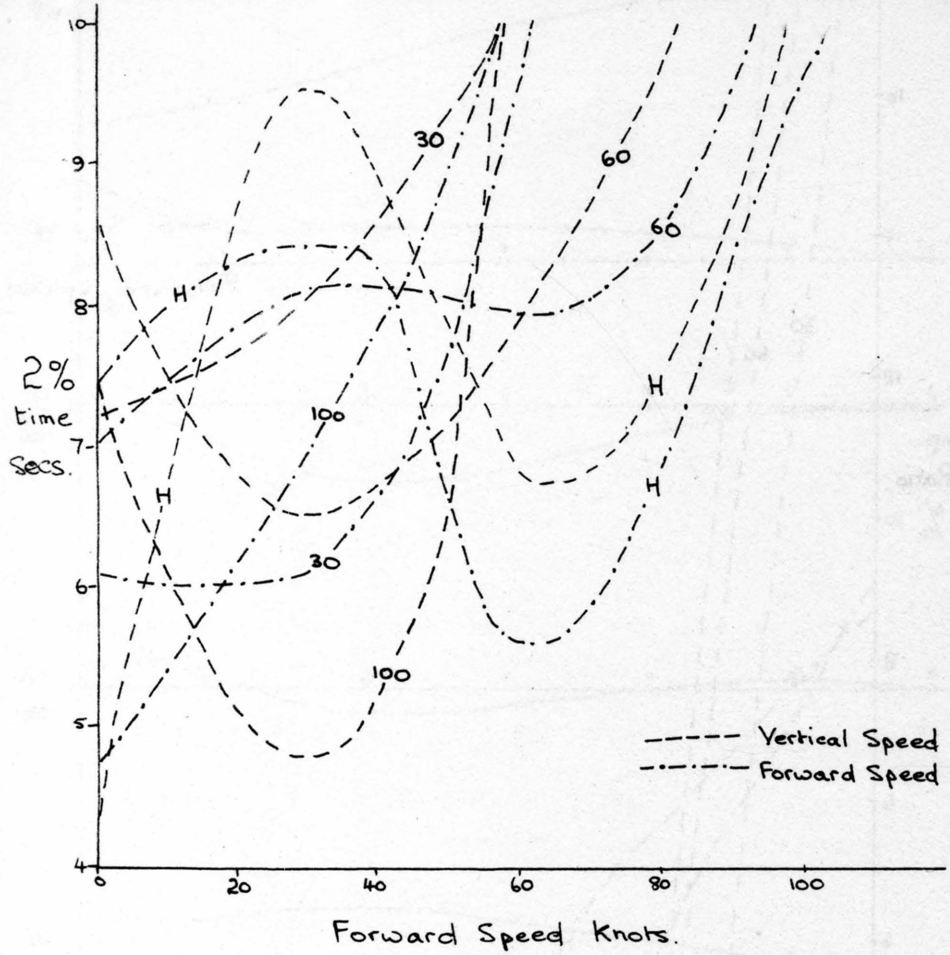
$$\frac{r_{B_1}}{r_{\theta_0}} = \frac{15}{20} \quad \text{at 60 knots}$$

Fig. 3



Amplitude Ratio v. Forward Speed

Fig. 4



2% time v. Forward Speed.

Fig. 5.

DIGITAL CONTROL FOR VARIABLE STABILITY AIRCRAFT

Dieter Bux
Gerhard Schweizer
Hans A. Seelmann

DORNIER GMBH, Friedrichshafen

GERMANY

1. Problem

The present training of jet-pilots is expensive and tedious. It is necessary to maintain several types of trainers which vary with the stage of the student, increasing cost. The final stage is training for actual maneuvers in the original aircraft a most expensive process.

To reduce expenses the development of a trainer has been proposed to simulate all trainer types by varying its flight characteristics. The ultimate objective is simulation of the original aircrafts for training costs would be greatly reduced. This paper presents the first stage of an investigation towards arbitrary variation of the behavior of a given aircraft in six degrees of freedom.

2. The Aircraft as a Mathematical Model

2.1 General Description

The aircraft can be described as a dynamic system as follows:

$$\dot{\bar{x}} = \bar{f}\,(\,\bar{x},\,\bar{u}\,) \qquad (2.1.1)$$

x is the state vector with the components:

$x_1 = v$	v_{x_f}	Velocity components in polar coordinates where α-angle of attack, β-slip-angle or aerplane-fixed cartesian coordinates.
$x_2 = \alpha$ or	v_{y_f}	
$x_3 = \beta$	v_{z_f}	
$x_4 = \omega_{x_f}$		Angle velocities in the airplane-fixed system.
$x_5 = \omega_{y_f}$		
$x_6 = \omega_{z_f}$		
$x_7 = \varphi$		Euler angles: φ - bank angle,
$x_8 = \vartheta$		ϑ - pitch angle.

The vector function $\bar{f}$ $(\bar{x},\bar{u})$ is obtained by applying Newton's equations in an airplane-fixed coordinate system. This process yields 6 equations for the the three components of linear and angular acceleration. The components of gravitational acceleration of the airplane-fixed system must be introduced as augmented state variables since they are dependent upon existing flight attitudes. These attitudes φ and ϑ, in turn are connected with the angular velocities via nonlinear differential equations.

Furthermore $\bar{f}$ $(\bar{x}, \bar{u})$ contain force and moment coefficients, which accordingly are nonlinear functions of the state variables. A system of analytically insoluble nonlinear differential equations is obtained.

2.2 Linearization

In the following investigations we examine the motion in the neighborhood of an arbitrary reference trajectory $\bar{x}_o$ (t), as a response to the control vector $\bar{u}_o(t)$.

One obtains:

$$\Delta\bar{x}(t) + \dot{\bar{x}}_o(t) = \bar{f}[\bar{x}_o(t), \bar{u}_o(t)] + \left.\frac{\partial\bar{f}}{\partial\bar{x}}\right|_{\substack{\bar{x}_o(t)\\ \bar{u}_o(t)}} \Delta\bar{x}(t) + \left.\frac{\partial\bar{f}}{\partial\bar{u}}\right|_{\substack{\bar{x}_o(t)\\ \bar{u}_o(t)}} \Delta\bar{u}(t) \qquad (2.2.1)$$

or

$$\Delta\dot{\bar{x}}(t) = \left.\frac{\partial\bar{f}}{\partial\bar{x}}\right|_{\substack{\bar{x}_o(t)\\ \bar{u}_o(t)}} \Delta\bar{x}(t) + \left.\frac{\partial\bar{f}}{\partial\bar{u}}\right|_{\substack{\bar{x}_o(t)\\ \bar{u}_o(t)}} \Delta u(t) \qquad (2.2.2)$$

since $\dot{\bar{x}}_o(t) = f[x_o(t), u_o(t)]$.

Equation (2.2.2) represents a system of linear time-variant differential equations of the following form:

$$\dot{\bar{q}} = \bar{\bar{A}}(t)\,\bar{q} + \bar{\bar{B}}(t)\,\bar{p}$$

2.3 Special Case

In conventional flight mechanics investigations an equilibrium stage is assumed instead of a general trajectory. In this case the equations assume a clear form, since:

$$\beta_o = \omega_{x_{fo}} = \omega_{y_{fo}} = \omega_{z_{fo}} = \varphi_o = 0 \qquad (2.3.1)$$

Neglecting the gyro moments of the jets and the intake impulses and verifying

$$c_L(\bar{x}_o) = c_M(\bar{x}_o) = c_N(\bar{x}_o) = 0$$

which is valid for the trimmed state, the system can be decoupled into two systems, each with four state variables for the longitudinal and lateral motions.

With equation (2.3.1) the longitudinal motion is obtained

$$
\begin{bmatrix} \Delta\dot{V} \\ \Delta\dot{\alpha} \\ \Delta\dot{\omega}_y \\ \Delta\dot{\vartheta} \end{bmatrix}
=
\begin{bmatrix}
-\frac{1}{m} C_{W_0}\rho v_0 F & -\frac{1}{m}\left(\frac{\partial C_W}{\partial \alpha}\frac{1}{2}\rho v_0^2 F + \sin\alpha_0 S_0 - G\right) & -\frac{1}{m}\frac{\rho}{2} v_0^2 F \frac{\partial C_W}{\partial \omega_{yf}} & -g \\
-\frac{1}{m} C_{A_0}\rho F & -\frac{1}{m}\left(\frac{S_0}{v_0}\cos\alpha_0 + \frac{\rho}{2} v_0 F \frac{\partial C_A}{\partial \alpha}\right) & 1 + \frac{1}{m}\frac{\rho}{2} v_0 F \frac{\partial C_A}{\partial \omega_{yf}} & 0 \\
0 & \frac{\rho}{2} v_0^2 F \frac{l_\mu}{J_{yy}} \cdot \frac{\partial C_M}{\partial \alpha} & \frac{\rho}{2} v_0^2 F \frac{l_\mu}{J_{yy}} \cdot \frac{\partial C_M}{\partial \omega_y} & 0 \\
0 & 0 & 1 & 0
\end{bmatrix}
\cdot
\begin{bmatrix} \Delta V \\ \Delta\alpha \\ \Delta\omega_y \\ \Delta\vartheta \end{bmatrix}
+
$$

$$
+
\begin{bmatrix}
\frac{\cos\alpha_0}{m} & 0 \\
\frac{\sin\alpha_0}{m v_0} & 0 \\
\frac{S_{zf}}{J_{yy}} & \frac{\rho}{2} v_0^2 F \frac{l_\mu}{J_{yy}} \frac{\partial C_M}{\partial \eta} \\
0 & 0
\end{bmatrix}
\cdot
\begin{bmatrix} \Delta S \\ \Delta\eta \end{bmatrix}
\tag{2.3.2}
$$

and the lateral motion

$$
\begin{bmatrix} \Delta\dot{\beta} \\ \Delta\dot{\omega}_x \\ \Delta\dot{\omega}_z \\ \Delta\dot{\varphi} \end{bmatrix}
=
\begin{bmatrix}
\frac{1}{m}\left(\frac{S_0}{v_0}\cos\alpha_0 - \frac{\rho}{2} v_0 F C_{W_0} - \frac{\rho}{2} v_0 F \frac{\partial C_Y}{\partial \beta}\right) & -\sin\alpha_0 - \frac{\rho}{2} v_0 \frac{F}{m}\frac{\partial C_Y}{\partial \omega_x} & \cos\alpha_0 - \frac{\rho}{2} v_0 \frac{F}{m}\frac{\partial C_Y}{\partial \omega_z} & \frac{g\cos\vartheta_0}{v_0} \\
\frac{\rho}{2} v_0^2 \frac{FS}{J_{xx}}\left(\frac{\partial C_L}{\partial \beta}\cos\alpha_0 - \frac{\partial C_N}{\partial \beta}\sin\alpha_0\right) & \frac{\rho}{2} v_0^2 \frac{FS}{J_{xx}}\left(\frac{\partial C_L}{\partial \omega_x}\cos\alpha_0 - \frac{\partial C_N}{\partial \omega_x}\sin\alpha_0\right) & \frac{\rho}{2} v_0^2 \frac{FS}{J_{xx}}\left(\frac{\partial C_L}{\partial \omega_z}\cos\alpha_0 - \frac{\partial C_N}{\partial \omega_z}\sin\alpha_0\right) & 0 \\
\frac{\rho}{2} v_0^2 \frac{FS}{J_{zz}}\left(\frac{\partial C_L}{\partial \beta}\sin\alpha_0 + \frac{\partial C_N}{\partial \beta}\cos\alpha_0\right) & \frac{\rho}{2} v_0^2 \frac{FS}{J_{zz}}\left(\frac{\partial C_L}{\partial \omega_x}\sin\alpha_0 + \frac{\partial C_N}{\partial \omega_x}\cos\alpha_0\right) & \frac{\rho}{2} v_0^2 \frac{FS}{J_{zz}}\left(\frac{\partial C_L}{\partial \omega_z}\sin\alpha_0 + \frac{\partial C_N}{\partial \omega_z}\cos\alpha_0\right) & 0 \\
0 & 1 & \tan\vartheta_0 & 0
\end{bmatrix}
\cdot
$$

$$
\cdot
\begin{bmatrix} \Delta\beta \\ \Delta\omega_x \\ \Delta\omega_z \\ \Delta\varphi \end{bmatrix}
+
\begin{bmatrix}
-\frac{\rho}{2} v_0 \frac{F}{m} \cdot \frac{\partial C_Y}{\partial \zeta} & -\frac{\rho}{2} v_0 \frac{F}{m} \cdot \frac{\partial C_Y}{\partial \xi} \\
\frac{\rho}{2} v_0^2 \frac{FS}{J_{xx}}\left(\frac{\partial C_L}{\partial \zeta}\cos\alpha_0 - \frac{\partial C_N}{\partial \zeta}\sin\alpha_0\right) & \frac{\rho}{2} v_0^2 \frac{FS}{J_{xx}}\left(\frac{\partial C_L}{\partial \xi}\cos\alpha_0 - \frac{\partial C_N}{\partial \xi}\sin\alpha_0\right) \\
\frac{\rho}{2} v_0^2 \frac{FS}{J_{zz}}\left(\frac{\partial C_L}{\partial \zeta}\sin\alpha_0 + \frac{\partial C_N}{\partial \zeta}\cos\alpha_0\right) & \frac{\rho}{2} v_0^2 \frac{FS}{J_{zz}}\left(\frac{\partial C_L}{\partial \xi}\sin\alpha_0 + \frac{\partial C_N}{\partial \xi}\cos\alpha_0\right) \\
0 & 0
\end{bmatrix}
\cdot
\begin{bmatrix} \Delta\zeta \\ \Delta\xi \end{bmatrix}
\tag{2.3.3}
$$

Here the presence of the common controls of an airplane is assumed.

ΔS = thrust displacement
$\Delta \eta$ = elevator displacement
$\Delta \zeta$ = rudder deflection
$\Delta \xi$ = aileron deflection

Therefore, if we assume small deviations from stationary horizontal flight the motion of an aircraft can be described utilizing a system of at least 8 state variables.

3. Discussion of present solutions

In reference [1] to [4] two different procedures are described for airplanes with variable stability. Until the present time these developments have been restricted to influencing the flight behavior for small deviations from a stationary horizontal flight.

1. The desired flight behavior is compelled by feedback: Response Feedback System (RFS), as illustrated in fig. 1.
2. The airplane follows the motions prescribed by a computer. Model Controlled System (MCS), as illustrated in fig. 2.

With only these developments it is impossible to generate a specific flight behavior because the customary four controls are not sufficient for the arbitrary variation of the state of an aerplane. Additional details are offered in Section 4. For these reasons only approximations could be obtained for the changes in flight behavior. We shall discuss this briefly.

For the RFS, α, $\dot{\alpha}$, $\dot{\omega}_y$, ω_y, v and ϑ are measured, each causes a displacement of the elevator via a given factor. Similarly, for the lateral motion β, ω_x, ω_z and their derivatives are measured causing displacements of the rudder and the aileron.

The gains can be chosen arbitrarily within a certain range and are adjusted through experimentation.

The MCS operates principally in the same way. However, the input signals for the displacements are the differences between the desired and the actual values.

Control possibilities as noted above do not allow complete arbitrary influence of the aircraft.

4. Discussion of New Solutions

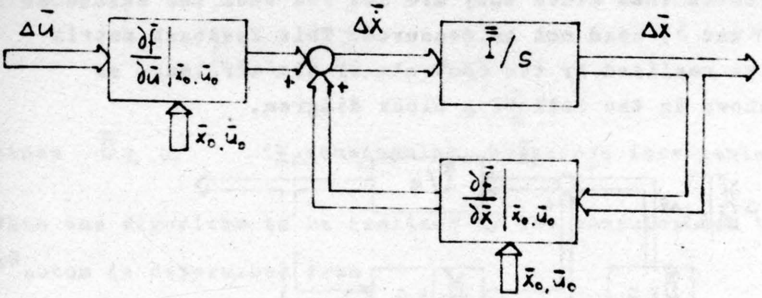
4.1 Continuous System

To simulate the dynamics of a given aircraft (index E) by a trainer (index S) the derivations are made to coincide. This can be done through proper feedback, as in the RFS. In the following discussion we shall develop such a system.

As derived above, the time-variant equation for small deviations from an arbitrary trajectory is

$$\Delta\dot{\bar{x}} = \left.\frac{\partial \bar{f}}{\partial \bar{x}}\right|_{\substack{\bar{x}_0\\ \bar{u}_0}} \cdot \Delta\bar{x} + \left.\frac{\partial \bar{f}}{\partial \bar{u}}\right|_{\substack{\bar{x}_0\\ \bar{u}_0}} \cdot \Delta u \qquad (4.1.1)$$

This equation applies to both the trainer and the aircraft to be simulated. Equation (4.1.1) is represented by the following block diagram:



For the airplanes to have the same behavior, they must have identical characteristic motions. These are described by

$$\Delta\dot{\bar{x}}_S = \frac{\partial\bar{f}}{\partial\bar{x}}\Big|_{\substack{\bar{x}_o,\bar{u}_o\\S}} \cdot \Delta\bar{x}_S \qquad (4.1.2.a)$$

and

$$\Delta\dot{\bar{x}}_E = \frac{\partial\bar{f}}{\partial\bar{x}}\Big|_{\substack{\bar{x}_o,\bar{u}_o\\E}} \cdot \Delta\bar{x}_E \qquad (4.1.2.b)$$

Index S characterizes the trainer and index E the aircraft to be simulated. Introducing artificial feedback for the trainer

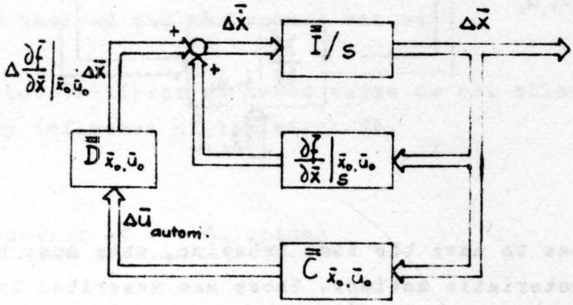
$$\Delta\dot{\bar{x}}_S = \frac{\partial\bar{f}}{\partial\bar{x}}\Big|_{\substack{\bar{x}_o,\bar{u}_o\\S}} \cdot \Delta\bar{x}_S + \Delta\frac{\partial\bar{f}}{\partial\bar{x}}\Big|_{\substack{\bar{x}_o,\bar{u}_o\\S}} \Delta\bar{x}_S \qquad (4.1.3)$$

one can generate the same behavior when

$$\Delta\frac{\partial\bar{f}}{\partial\bar{x}}\Big|_{\substack{\bar{x}_o,\bar{u}_o\\S}} = \frac{\partial\bar{f}}{\partial\bar{x}}\Big|_{\substack{\bar{x}_o,\bar{u}_o\\E}} - \frac{\partial\bar{f}}{\partial\bar{x}}\Big|_{\substack{\bar{x}_o,\bar{u}_o\\S}} \qquad (4.1.4)$$

In general the matrix $\Delta\frac{\partial\bar{f}}{\partial\bar{x}}\Big|_{\bar{x}_o,\bar{u}_o}$ **has rank 6. This matrix**

also indicates that since they are not fed back, the attitude angles (φ and ϑ) need not be measured. This feedback matrix can only be realized by the controls of the airplane, as clearly shown in the following block diagram.



Obviously equation (4.1.5) must hold true:

$$\Delta \frac{\partial \bar{f}}{\partial \bar{x}}\Big|_{\substack{\bar{x}_o,\bar{u}_o \\ S}} = \bar{\bar{D}}_{\bar{x}_o,\bar{u}_o} \cdot \bar{\bar{C}}_{\bar{x}_o,\bar{u}_o} \qquad (4.1.5)$$

Matrix $\bar{\bar{C}}_{\bar{x}o,\bar{u}o}$ which determines the automatic control vector from the instantaneous state vector, is realized by a computer.

$\bar{\bar{D}}_{\bar{x}o,\bar{u}o}$ is the control matrix describing the effect of the control vector upon the aircraft.

Therefore $\bar{\bar{D}}_{\bar{x}o,\bar{u}o}$ must be at least of rank 6 and $\Delta\bar{u}_{autom}$ must be a 6-dimensional vector. Since usually an aircraft has only four controls with a defined control matrix, one must provide two additional controls, the necessary conditions for which are derived from the fact that the rank of the control matrix $\bar{\bar{D}}_{\bar{x}o,\bar{u}o}$ must be 6. Possible realizations of the control matrix are discussed in section 4.3.

Once $\bar{\bar{D}}_{\bar{x}o,\bar{u}o}$ is constructively fixed, $\bar{\bar{C}}_{\bar{x}o,\bar{u}o}$ can be determined from

$$\bar{\bar{C}}_{\bar{x}_o,\bar{u}_o} = \bar{\bar{D}}^{-1}_{\bar{x}_o,\bar{u}_o} \cdot \Delta \frac{\partial \bar{f}}{\partial \bar{x}}\bigg|_{\substack{\bar{x}_o,\bar{u}_o \\ S}} \qquad (4.1.6)$$

since $\bar{\bar{D}}_{\bar{x}_o,\bar{u}_o}$ is nonsingular, therefore invertable.

Thus the algorithm to be realized by the computer can be given. $\Delta\bar{u}_{autom}$ is determined from

$$\Delta\bar{u}_{autom} = \bar{\bar{C}}_{\bar{x}_o,\bar{u}_o} \Delta\bar{x}_S = \bar{\bar{D}}^{-1}_{\bar{x}_o,\bar{u}_o} \cdot \Delta \frac{\partial \bar{f}}{\partial \bar{x}}\bigg|_{\substack{\bar{x}_o,\bar{u}_o \\ S}} \Delta\bar{x}_S \qquad (4.1.7)$$

Since $\bar{\bar{D}}$ and $\frac{\partial \bar{f}}{\partial \bar{x}}$ are functions of the reference solution $\bar{x}_o(t)$ and $\bar{u}_o(t)$, they must be determined continuously from these values.

The simulation requires storage of many nonlinear functions of various variables and numerous arithmetic operations, even if only a low accuracy is desired. This can only be achieved by a digital computer. Thus the continuous equations must be discretized.

But we must first note that the control matrix $\frac{\partial \bar{f}}{\partial \bar{u}}\Big|_{\bar{x}_o,\bar{u}_o}$ must agree for both aircrafts so that they respond identically to the pilot's controls. This can be accomplished without difficulties. We have

$$\Delta\bar{u}_S = \bar{\bar{f}}^{-1}_{\bar{u}_S}(\bar{x}_o,\bar{u}_o) \cdot \bar{\bar{f}}_{\bar{u}_E}(\bar{x}_o,\bar{u}_o) \cdot \bar{\bar{S}}_{\substack{\bar{x}_o,\bar{u}_o \\ E}} \cdot \Delta\bar{u}_{Pilot} \qquad (4.1.8)$$

where the matrices $\bar{\bar{f}}_{\bar{u}_S}(\bar{x}_o,\bar{u}_o)$, $\bar{\bar{f}}_{\bar{u}_E}(\bar{x}_o,\bar{u}_o)$ are generated from the matrices $\frac{\partial \bar{f}}{\partial \bar{u}}\Big|_{S\,\bar{x}_o,\bar{u}_o}$ and $\frac{\partial \bar{f}}{\partial \bar{u}}\Big|_{E\,\bar{x}_o,\bar{u}_o}$ by omitting those rows consisting exclusively of zeros. Thus nonsingular 4 x 4 matries are obtained corresponding to the 4 conventional controls. $\bar{\bar{S}}_{\bar{x}_o,\bar{u}_o}$ is a diagonal matrix containing the gains between the pilot's control and the displacements of the operational aircraft.

4.2 Discrete System

The general solution of equation (4.1.1) is

$$\Delta\bar{x}(t-t_o) = \bar{\bar{\Phi}}(t-t_o)\Big|_{\bar{x}_o,\bar{u}_o} \Delta\bar{x}(t_o) + \int_{t_o}^{t} \bar{\bar{\Phi}}(t-\tau)\Big|_{\bar{x}_o,\bar{u}_o} \frac{\partial\bar{f}}{\partial\bar{u}}\Big|_{\bar{x}_o,\bar{u}_o} \cdot\Delta\bar{u}(\tau)\,d\tau \qquad (4.2.1)$$

where $\bar{\bar{\Phi}}(t-t_o)$ is the fundamental matrix as solution of the homogeneous system

$$\dot{\bar{\bar{\Phi}}}(t-t_o)\Big|_{\bar{x}_o,\bar{u}_o} = \frac{\partial\bar{f}}{\partial\bar{x}}\Big|_{\bar{x}_o,\bar{u}_o} \cdot \bar{\bar{\Phi}}(t-t_o)\Big|_{\bar{x}_o,\bar{u}_o} \qquad (4.2.2)$$

Here we assume, that $(t - t_o)$ is choosen so small that $\frac{\partial\bar{f}}{\partial\bar{x}}$ is constant in this period and the following equation holds true:

$$\frac{\partial\bar{f}}{\partial\bar{x}}(\tau) \approx \frac{\partial\bar{f}}{\partial\bar{x}}(t_o) = \frac{\partial\bar{f}}{\partial\bar{x}}\Big|_{\bar{x}_o,\bar{u}_o}, \quad t_o \leq \tau \leq t \qquad (4.2.3)$$

For the reference trajectory $\bar{x}_o(t)$ for $\bar{u}_o(t)$ is known

$$\frac{\partial\bar{f}}{\partial\bar{x}}\Big|_{\bar{x}_o,\bar{u}_o} = \frac{\partial\bar{f}}{\partial\bar{x}}(t)$$

is also known.

Expanding $\bar{\bar{\Phi}}(t - t_o)$ into a Taylor series at t_o yields:

$$\bar{\bar{\Phi}}(t-t_o) = \bar{\bar{I}} + \frac{\partial\bar{f}}{\partial\bar{x}}\Big|_{\bar{x}_o,\bar{u}_o} \cdot (t-t_o) + \left[\frac{\partial\bar{f}}{\partial\bar{x}}\Big|_{\bar{x}_o,\bar{u}_o} (t-t_o)\right]^2 \cdot \frac{1}{2!} + \ldots \qquad (4.2.4)$$

Neglecting all higher terms of $(t - t_o)$,

$$\bar{\bar{\Phi}}(t-t_o) \doteq \bar{\bar{I}} + \frac{\partial\bar{f}}{\partial\bar{x}}\Big|_{\bar{x}_o,\bar{u}_o} \cdot (t-t_o)\,. \qquad (4.2.5)$$

Assuming $t - t_o$ as the sampling time T of the digital system, we may write

$$\bar{\bar{\Phi}}(T) = \bar{\bar{I}} + \frac{\partial\bar{f}}{\partial\bar{x}}\Big|_{\bar{x}_o,\bar{u}_o} \cdot T \qquad (4.2.6)$$

The state transition matrix from time n T to time (n+1)T is then given by

$$\bar{\bar{\Phi}}[(n+1)T - nT] = \bar{\bar{\Phi}}(T)\Big|_{\substack{\bar{x}_0(nT)\\ \bar{u}_0(nT)}} = \bar{\bar{I}} + \frac{\partial \bar{f}}{\partial \bar{x}}\Big|_{\substack{\bar{x}_0(nT)\\ \bar{u}_0(nT)}} \cdot T \qquad (4.2.7)$$

Using equation (4.2.7) we have the general solution for t = (n+1) T, if Δu(t) = Δu(nT),

$$\Delta\bar{x}[(n+1)T] = \bar{\bar{\Phi}}(T)\Big|_{\substack{\bar{x}_0(nT)\\ \bar{u}_0(nT)}} \Delta\bar{x}(nT) + \bar{\bar{\Phi}}(T)\Big|_{\substack{\bar{x}_0(nT)\\ \bar{u}_0(nT)}} \int_0^T \bar{\bar{\Phi}}(-\tau)\Big|_{\substack{\bar{x}_0(nT)\\ \bar{u}_0(nT)}} \frac{\partial \bar{f}}{\partial \bar{u}}\Big|_{\substack{\bar{x}_0(nT)\\ \bar{u}_0(nT)}} \Delta\bar{u}(nT)\,d\tau \qquad (4.2.8)$$

Therefore, it is necessary to compute the values $\frac{\partial \bar{f}}{\partial \bar{x}}\Big|_{\bar{x}_0(nT),\bar{u}_0(nT)}$ and $\frac{\partial \bar{f}}{\partial \bar{u}}\Big|_{\bar{x}_0(nT),\bar{u}_0(nT)}$ from the instantaneous state before determing the control vector and the feedback in time n T. Then we obtain a very accurate simulation provided the sampling time T is sufficiently small.

4.3 Realization

The most important fact for an engineering realization is that the ranges of the trainer's control variables are sufficiently large for additional controls to simulate the original aircraft. This must be considered in the construction of the control system. The additional control essentially must vary the lift-versus-drag curve (using auxiliary flaps) and generate a lateral force (auxiliary jet, bleed from the main jets, or an additional control surface).

The computer must mainly store functions and compute the instantaneous state and control variables. The aerodynamic lift is an example:

$$A = \frac{\varrho}{2} v^2 F c_A \qquad \text{where} \quad c_A = f(v, \alpha, \omega_{y_f}, \eta, \eta_{KI})$$

The dependence of the airspeed v is correct only at constant flight level. Otherwise the function depends also on the Mach-number. Not only the state variables, but also the control variables are necessary for computing the aerodynamic functions.

A complete aerodynamic model needs a core of about 1,500 words with 24 bits each. Since two models must be stored, about 3,000 words are needed for the aerodynamic functions. To store the characteristics of the jet engines about the same amount of storage space is required. In addition the above control matrix must be stored. A total of about 8,000 words are needed for all this data.

Two additional storage blocks are necessary. One block serves as a buffer accumulator, for temporary storage of flight data which at certain instants are transferred to an external instrument, for example a tape recorder or a telemetry system. These data may already be partially analysed. The third storage block assimilates the required computer and control program comprising about 6,000 words with 24 bits each.

To summarize a digital airborne computer with 16 - 24 K (K=1,024) storage space à 24 bits is sufficient to solve the existing problem.

To retain minimum sampling time and computing time per sampling interval, the computer must have a cycle-time of 2 µsec. - The sampling time should be about 20 m sec. for this case. -

These proposed requirements for an airborne computer are fulfilled by several various types. The cost of such a computer is presently approximately 250,000.-- DM. Costs would be reduced by increasing technological experience and further production of the same types.

In Conclusion

The expenses of jet pilot training would be significantly lessened through further investigation of a flight simulator as proposed in this paper. The engineering and financial efforts would prove well justified and profitable.

BIBLIOGRAPHY

[1] John Kroll, Rudy H. Arendt and Francis E. Pritchard:
Development of a General Purpose Airborne Simulator.
NASA Contractor Report 641, Nov. 1966.

[2] Waldemar O. Breuhaus:
Recent Experience with In-Flight Simulation.
The AGARD Spezialists' Meeting on Stability and Control. Cambridge, England, Sept. 1966

[3] G. Klopfstein:
Avion à Stabilité Variable.
AGARD Bericht 537, Mai 1966

[4] W. Eldridge:
Use of a large Jet Transport as an In-Flight Dynamic Simulator.
AGARD Bericht 528, Mai 1966

[5] H. Roderer, W. Schnitzler, H. Seelmann:
Die Anwendung eines Hybridrechners bei der Simulation aerodynamischer Flugkörper.
Referat beim 5. AICA-Kongress, Lausanne, August 1961.

LIST OF SYMBOLS

Variable (ref. LN 9300)

φ - bank angle

ϑ - pitch angle

α - angle of attack

β - sideslip angle

v - airspeed

ω - angularvelocity

A - lift

c_A - lift coefficient

c_W - drag coefficient

c_Y - lateral force coefficient

c_L - rolling-moment coefficient

c_M - pitching-moment coefficient

c_N - yawing-moment coefficient

η - elevator displacement

η_{Kl} - flap displacement

ζ - rudder displacement

ξ - aileron displacement

F - wing area

m - mass of aircraft

I - moment of inertia

l_μ - reference wing chord

s - half wing span

s - Laplace - operator

s_{zf} - normal distance of thrust vector from C.G.

S - thrust

t, τ - time

T - sampling time

vectors

$\bar{p}$, $\bar{u}$ - control vectors

$\bar{q}$, $\bar{x}$ - state vectors

$\bar{f}$ - vector function

matrices

$\bar{\bar{A}}$ - coefficient matrix of the linear system

$\bar{\bar{B}}$ - control matrix of the linear system

$\left.\frac{\partial \bar{f}}{\partial \bar{x}}\right|_{\bar{x}_o, \bar{u}_o}$ - coefficient matrix of the system linearized with respect to the state $\bar{x}_o$, $\bar{u}_o$.

$\left.\frac{\partial \bar{f}}{\partial \bar{u}}\right|_{\bar{x}_o, \bar{u}_o}$ - control matrix of the system linearized with respect to the state $\bar{x}_o$, $\bar{u}_o$!

$\bar{\bar{C}}$, $\bar{\bar{D}}$ - coefficient matrices

$\bar{\bar{I}}$ - unity matrix

$\bar{\bar{S}}$ - diagonal matrix with the gains between the pilot's controls and the displacements.

$\bar{\bar{\Phi}}$ (t) - fundamental matrix of the linear system of differential equations.

indices

O - reference state

E - aircraft to be simulated

S - trainer

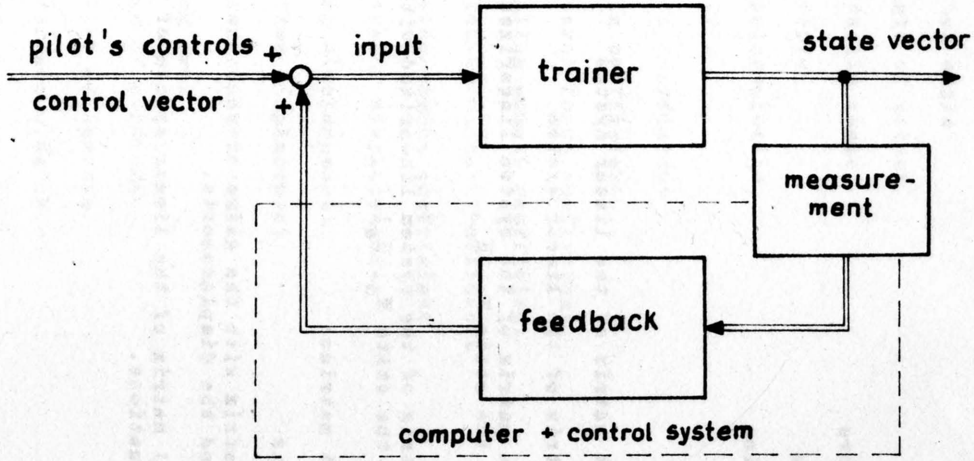


FIG. 1 :
RESPONSE FEEDBACK SYSTEM.

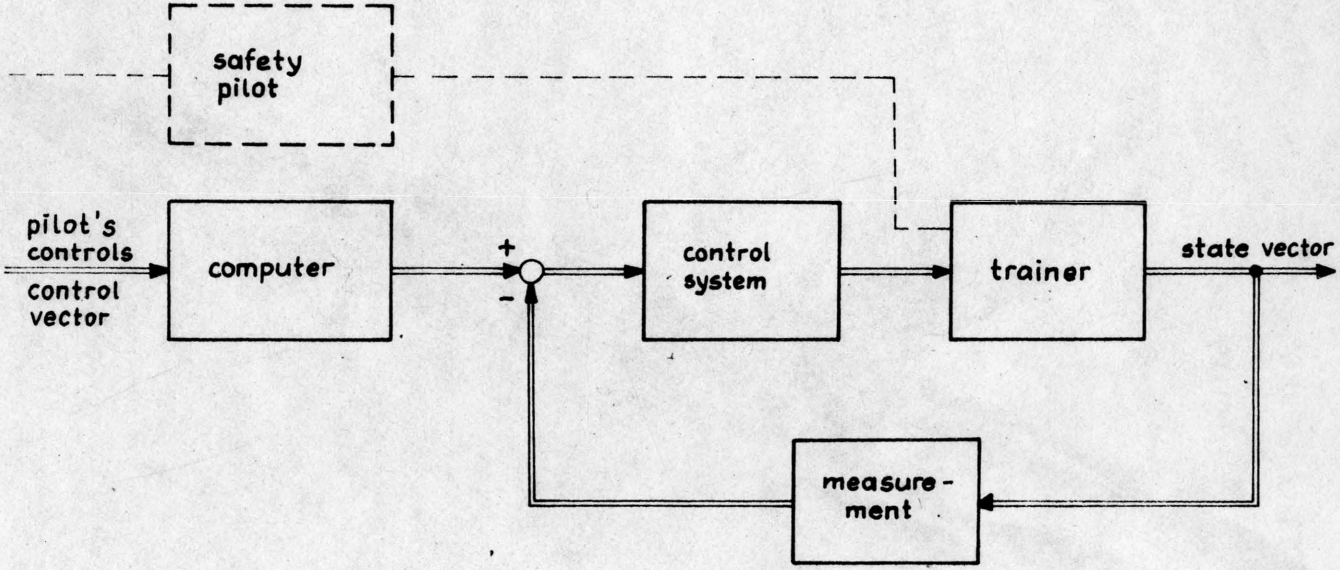
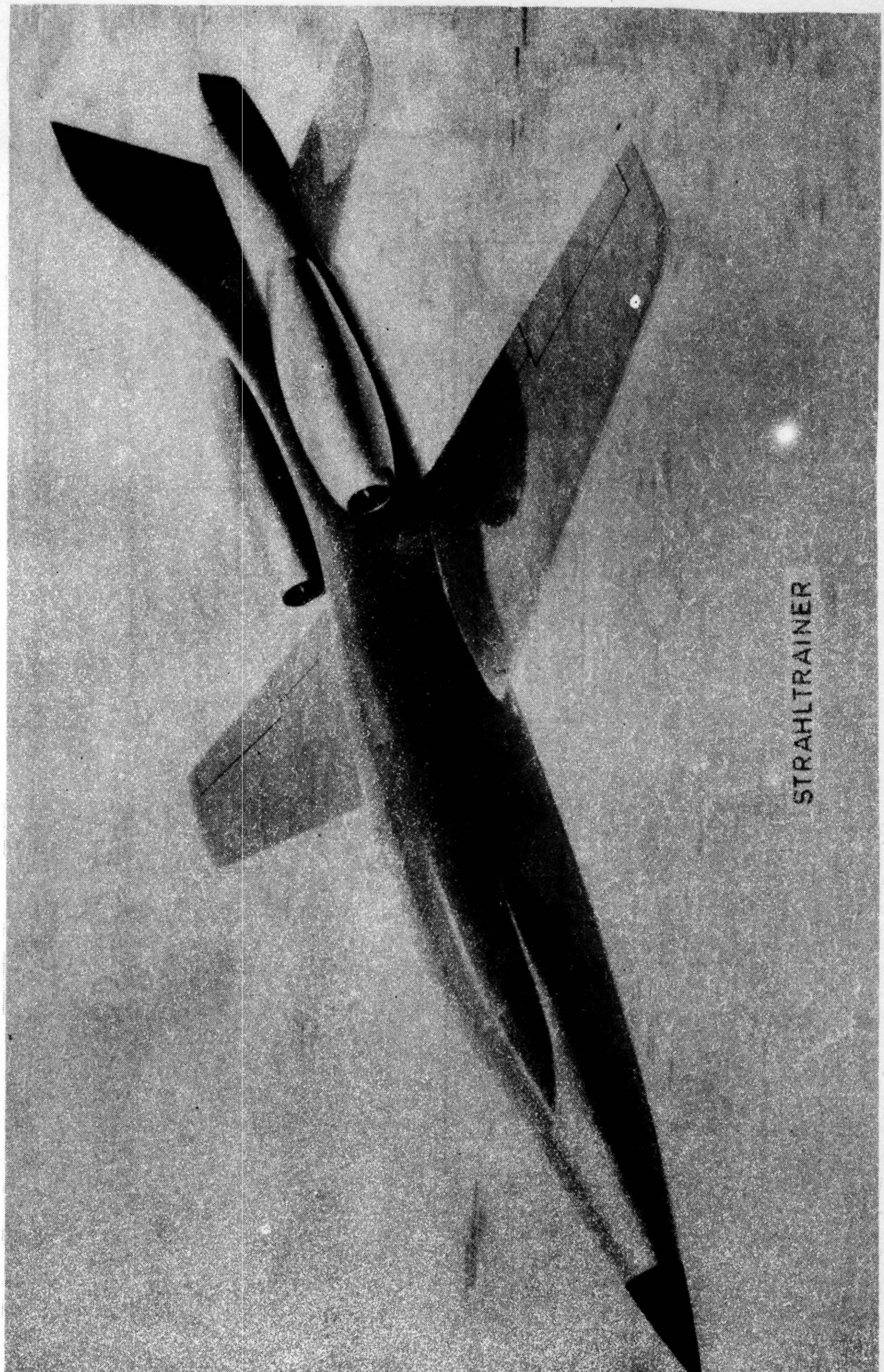


FIG. 2:

MODEL CONTROLLED SYSTEM.



STRAHLTRAINER

EXPERIMENTS ON A HYDROFOIL TEST CRAFT WITH A HYBRID FOIL SYSTEM AND AN AUTOPILOT

Y. Ohtsu*, T. Fujino*, M. Itoh*, H. Ohno* and K. Uchino**

Introduction

The hydrofoil craft is classified into three types according to the foil systems as shown in Figure 1:

(1) Surface-piercing foil system.

(2) Hybrid or combined foil system.

(3) Fully submerged foil system.

The surface-piercing foil system, having two surface-piercing foils, works at a substantially constant lift coefficient. The lift of the craft with this system is proportional to the foil area under water. This lift stabilizes the roll of the craft, and keeps the craft at the constant height and pitch angle from the surface of the water. Therefore, the craft rolls, heaves or pitches heavily owing to the wave.

In the fully submerged foil system, the lift is adjusted by the incidence angle or the flap angle of the foils. In order to control the altitude and the posture of the craft, some control system is required. The fully submerged foil system has a platforming capability, and is distinctly superior to the surface-piercing foil system. But, with this foil system the craft can not cruise in foil borne condition when the autopilot equipment fails.

In-between these two foil systems, there is a hybrid, or combined system, having surface-piercing bow foils and fully-submerged stern foils. This system is believed to possess more advantages than shortcomings of the aforementioned (1) or (3) system, the latter of which has deficiency in stability.

Up to the present, many papers have been presented concerning with the surface-piercing foil system or the fully submerged foil system. But there has been only few relating to the hydrofoil craft with the hybrid foil system. In this paper, an outline of longitudinal dynamics of the craft with hybrid foil system and its control system are given, and the foil borne cruising sea trials of the test craft in several autopilot modes are

The authors * are with the Mitsubishi Heavy Industries, Ltd., Tokyo, Japan
The author ** is with the College of Naval Architecture of Nagasaki, Nagasaki, Japan

described.

Test craft and autopilot equipment

The test craft has two split type surface-piercing fore foils having fully rotating submerged foils at their bottoms, and the fully submerged after foil which has a control flap at the trailing edge.

The test craft has an 8-meter overall length, 2.2-meter hull beam, 4.2-ton loaded weight, 280 hp (one gasoline engine), 30-knot foil borne cruising speed, 40-knot maximum speed. The general view of the test craft and the general arrangement of the autopilot equipment and also the measuring instruments are shown in Figures 2 and 3.

Controlled variables of the hydrofoil craft are altitude (or height), pitch angle, roll angle and yaw angle. In the test craft, the yaw angle is manually controlled by the rudder wheel, and other three variables are controlled by the autopilot system. The block diagram of the autopilot system is shown in Figure 4.

In the autopilot system, the height is controlled by changing additively each incidence angle of the fully rotating submerged fore foils. Each incidence angle is controlled by its own hydraulic servo mechanism. The roll angle is controlled by changing differentially each incidence angle of the fore foils. The pitch angle is controlled by changing the flap angle of the after foil. The flap angle is also controlled by the hydraulic servo mechanism.

The hydraulic servo mechanisms are composed of transistorized servo amplifiers, electro-hydraulic servo valves, hydraulic linear actuators and linear differential transformers to detect the stroke of the linear actuators. A plunger type variable displacement hydraulic pump is used as the hydraulic power source. Experiments on the hydraulic servo system under the dummy loads, (inertial load of 30 kgs and spring load of 600 kgs) showed that its characteristics could be approximated by a second order system in the following transfer function:

$$Gs(S) = \frac{\omega_{sn}^2}{S^2 + 2\zeta_{sn}\omega_{sn}S + \omega_{sn}^2} = \frac{(55.5)^2}{S^2 + 2(0.9)(55.5)S + (55.5)^2}$$

where

ω_{sn} : natural angular frequency (in rad/sec)

ζ_{sn} : damping coefficient of the hydraulic servo system.

As sensors of the controlled variables, a sonic height sensor and an electronic accelerometer are used to detect the heaving motion, and a vertical gyro and rate gyros are used to detect the pitching and rolling motions.

The specifications of these sensors are shown in Table 1. Electrical signals from these sensors become the controlling means of the autopilot.

The controlling means are composed of transistorized analogue computing amplifiers, multipliers, potentiometers and other elements. The maximum values of their inputs and outputs are ± 10 volts. Their functions are to adjust the level of signals from each sensor, to compute the height signal at the center of gravity, and to generate the necessary actuating signals.

Synthesis for the longitudinal plane control system

Generally, the hydrofoil craft dynamics could be treated separately as a longitudinal plane motion (i.e., heaving, surging and pitching motions) and a lateral plane motion (i.e., side-slip, rolling and yawing motions).

In the test craft, only the roll angle is taken as controlled variables in the lateral plane motion, and the lateral plane control system is trea as the system of single input and one output. On the other hand, in the longitudinal plane motion the height and the pitch angle are taken as controlled variables. Then the longitudinal plane control system is treated as the system with mutual interaction between them. A sketch of the test craft and a co-ordinate system in the longitudinal plane are shown in Figure 5. In a equilibrium condition of the craft with a hybrid foil system, the height and the pitch angle are a function of the incidence angles of the rotating fore foils and the flap angle of the after foil. Then, the longitudinal dynamics of the test craft may be regarded as an interacting control system and studied by the same method as that in the fully submerged foil system[1].

Neglecting the surging term in the longitudinal dynamics of the craft, we obtain the block diagram of the autopilot control system shown in Figure 6.

In Figure 6,

Zg^* : Desired value (or set point) of the height at the c.g. of the craft (in m)

Zg : Height of the craft at the c.g. (in m)

θ^* : Desired value of the pitch angle (in rad)

θ : Pitch angle of the craft (in rad)

δ_1 : Incidence angle of the fully rotating fore foil (in rad)

δ_2 : Flap angle of the after foil (in rad)

$M_{11}(S)$: Transfer function of the craft (in m/rad)

$M_{31}(S)$: Ditto (in rad/rad)

$M_{12}(S)$: Ditto (in m/rad)

$M_{32}(S)$: Ditto (in rad/rad)

$G_1(S)$: Transfer function of the feed back element of the height loop

$G_1(S) = k_1 + g_1 S + h_1 S^2$

k_1 : Feed back constant (in volt/m)

g_1 : Feed back constant (in volt·sec/m)

h_1 : Feed back constant (in volt·sec²/m)

$G_2(S)$: Transfer function of the feed back element of the pitch loop

k_2 : Feed back constant (in volt/rad)

g_2 : Feed back constant (in volt·sec/rad)

$H_1(S)$: Transfer function of the hydraulic servo mechanism times K_1

$$H_1(S) = \frac{K_1 \omega_{sn}^2}{S^2 + 2\zeta_{sn}\omega_{sn} S + \omega_{sn}^2} = K_1 \cdot Gs(S)$$

K_1 : Master gain of the height loop (in rad/volt)

$H_2(S)$: Transfer function of the hydraulic servo mechanism times K_2

$$H_2(S) = \frac{K_2 \omega_{sn}^2}{S^2 + 2\zeta_{sn}\omega_{sn} S + \omega_{sn}^2} = K_2 \cdot Gs(S)$$

K_2 : Master gain of the pitch loop (in rad/volt)

We have obtained $Mij(S)$ as the following formula

$$Mij(S) = \frac{Kij\,(S-Z_{ij1})(S-Z_{ij2})(S-Z_{ij3})}{(S-P_1)(S-P_2)(S-P_3)(S-P_4)(S-P_5)}$$

The values of each parameter in $M_{ij}(S)$ calculated for the test craft are shown in Table 2. In Table 2, $Moij(S)$ means $Moij = \lim_{S \to 0} M_{ij}(S)$, that is static gain.

The transfer function of height and pitch loop in Figure 6 are respectively obtained as the following equations.

$$F_{11}(S) = \frac{\partial Zg}{\partial Zg^*} = \frac{M_{11}(S) + G_2(S)H_2(S)\left\{M_{11}(S)M_{32}(S) - M_{12}(S)M_{31}(S)\right\} H_1(S)}{\Delta}$$

$$F_{32}(S) = \frac{\partial \theta}{\partial \theta^*} = \frac{M_{32}(S) + G_1(S)H_1(S)\left\{M_{11}(S)M_{32}(S) - M_{12}(S)M_{31}(S)\right\} H_2(S)}{\Delta}$$

where $\Delta = 1 + M_{11}(S)G_1(S)H_1(S) + M_{32}(S)G_2(S)H_2(S) + G_1(S)G_2(S)H_1(S)H_2(S)\left\{M_{11}(S)M_{32}(S) - M_{12}(S)M_{31}(S)\right\}$

Substituting each value of $M_{ij}(S)$ shown in Table 2 and pre-determined values of k_1, g_1, h_1, k_2, g_2 in the characteristic equation of the system, and increasing the value of K_1, K_2 gradually in calculation, we obtain the roots of the characteristic equation. Root locus diagrams obtained from the above-mentioned calculations are shown in Figures 7 and 8, respectively. In these figures, dotted lines starting from the origin are drawn at an angle of 30 degrees to the imaginary axis. Each value of K_1 and K_2 at the point, where the dotted lines intersect the root locus starting from the root of the transfer function of the hydraulic servo system, is 0.5 and 2.5, respectively.

On the other hand, the values of Moij were measured experimentally without the autopilot in the foil borne cruising. Step response tests were carried out with the values of k_1, g_1, h_1, k_2, g_2 in Table 3. The optimum value of K_1 and K_2 obtained by the experiments and those by the mathematical synthesis are shown in Table 3. Though the values obtained by the experiments and the synthesis do not exactly coincide with each other, the values of each parameter obtained by the experiments do not differ so much from those by the synthesis, respectively. Therefore, it seems that the optimum open loop gain of the autopilot for the hydrofoil craft with the hybrid foil system could be obtained by the mathematical synthesis.

Foil borne cruising tests in several autopilot modes

The hydrofoil craft with the hybrid foil system does not require any autopilot for the foil borne cruising because of its self stabilizing characteristics. The autopilot is for improving its riding ability and comfortability in the foil borne cruising.

The foil borne cruising tests were carried out in several autopilot modes of the autopilot circuit shown in Figure 4. The several autopilot modes used in these experiments are tabulated in Table 4. In Table 4, the mode expressed in A.P.($\dot{\phi}$) means that the rolling rate of the craft dynamics is fed back by a rate gyro. The mode expressed in A.P.($\phi + \dot{\phi}$) means that the roll angle and the rolling rate of the craft dynamics are fed back by a vertical gyro and a rate gyro, respectively. Similarly, the mode A.P.($\dot{\theta}$) means that the pitch rate is fed back by a rate gyro and the mode A.P.($\theta + \dot{\theta}$) means that the pitch angle and pitch rate are both fed back by the above-mentioned vertical gyro and a rate gyro, respectively. The mode A.P.($\dot{h} + \ddot{h}$) means that the heaving velocity and the heaving acceleration are both fed back by the integrated signal of the accelerometer's output and by the accelerometer at c.g. of the craft, respectively. The mode A.P.($h + \dot{h} + \ddot{h}$) means that the height signal is also fed back by the sonic height sensor together with the mode A.P.($\dot{h} + \ddot{h}$). The rest is omitted.

Figure 9 shows one of the test results carried out under the same sea state in the foil borne cruising. In this figure, the abscissa shows the autopilot modes used and the ordinate shows the peak to peak amplitude of each quantity measured in foil borne cruising at the speed of 30 kts. The upper level indicated in wide bars shows the mean value of amplitude, and the upper level indicated in narrow bars shows the maximum amplitude. The data shown in Figure 9 is not a statistical result but one of the sea trials in foil borne cruising.

One of the raw data of oscillographs is shown in Figures 10 and 11. Figure 10 shows the result of the foil borne cruising tests in several autopilot modes: Without the autopilot (no control), and also A.P.($\dot{\phi}$), A.P.($\phi + \dot{\phi}$), A.P.($\dot{\theta}$) and A.P.($\theta + \dot{\theta}$). Similarly, Figure 11 shows the results in the other autopilot modes: A.P.($\dot{h} + \ddot{h}$), A.P.($h + \dot{h} + \ddot{h}$), A.P.($\dot{\phi} + \dot{\theta} + \dot{h} + \ddot{h}$) and A.P.($\phi + \dot{\phi} + \theta + \dot{\theta} + h + \dot{h} + \ddot{h}$). These figures suggest that the height, pitch angle and roll angle can not completely be controlled by the rate feed back only, but the deviations of them are considerably smaller than that of without the autopilot. Both their position and rate are controlled if the feed back signal of the height, pitch angle and roll angle is added to the rate feed back signal of them.

The hydrofoil craft with hybrid foil system has essentially a trim of the height, pitch angle and roll angle at a constant foil borne cruising without any autopilot. But, experiments was confirmed the feasibility of setting the trim of the craft to that of the desired values of the autopilot. Raw data of step response test to the height, pitch angle, roll angle inputs are shown in Figures 12, 13 and 14 to support the above-mentioned behavior. These figures show that each controlled variable i.e., the height, pitch angle, roll angle, was attained to the desired value when a step input is added.

Conclusion

The foil borne cruising tests for the hydrofoil craft with the hybrid foil system were conducted in several autopilot modes and the performances of the dynamic motion in each autopilot mode were compared qualitatively. The experiments showed that the deviation of the controlled variables from the trim of the craft was considerably suppressed under the rate feed back only, and that the deviation from the set point of the autopilot were decreased more under both position and rate feed back.

Acknowledgement

The authours express their thanks to those concerned at the Shimonoseki Shipyard & Engine Works, Mitsubishi Heavy Industries, Ltd. for the construction and the sea-trials of the hydrofoil test craft and their thanks also go to the parties concerned at the Nagoya Aircraft Works of the same company for the making of the autopilot equipment.

Reference

1 Fujino, T., Yumoto, T., Ohno, H., Ohtsu, Y., Uchino, K. and Itoh, M. Synthesis of Longitudinal control system of MH-3 fully submerged hydrofoil craft. Technical Review, Mitsubishi Heavy Industries, Ltd., Tokyo, Japan. Janu., 1966

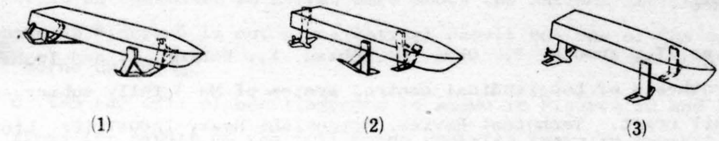


Figure 1. Three different types of foil configuration

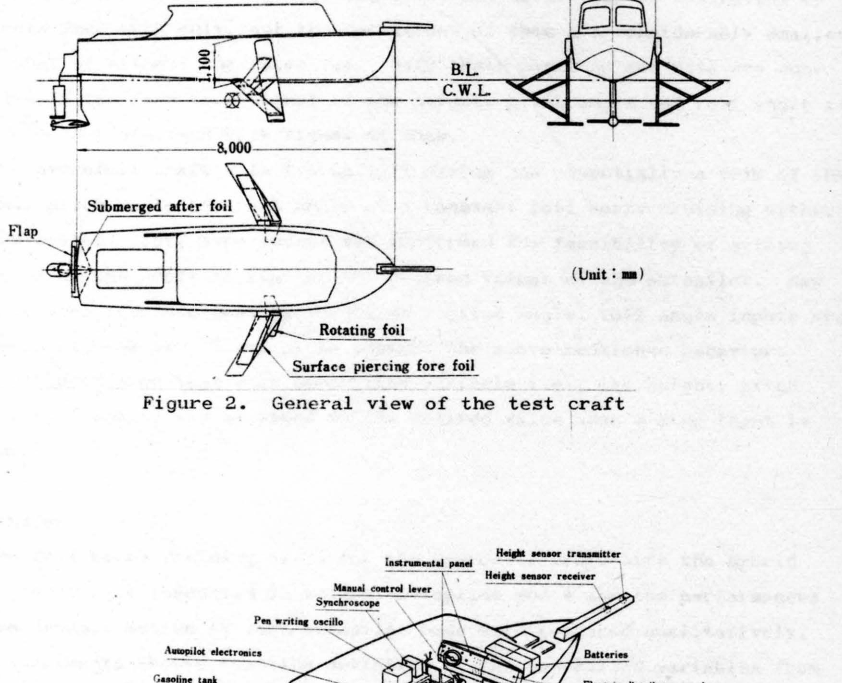


Figure 2. General view of the test craft

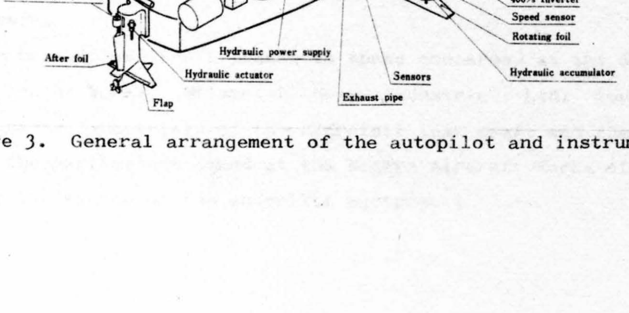


Figure 3. General arrangement of the autopilot and instruments

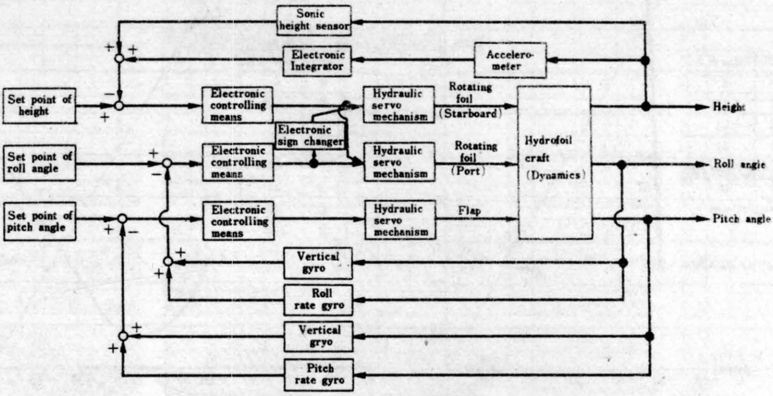


Figure 4. Block diagram of hydrofoil electronic autopilot

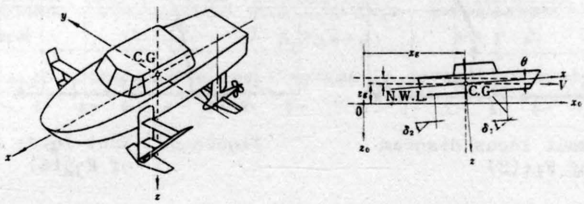


Figure 5. Sketch of the test craft and co-ordinated system in the longitudinal plane

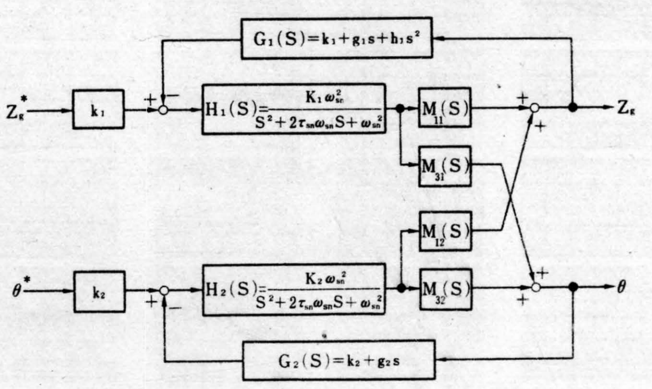


Figure 6. Block diagram of the longitudinal plane control system

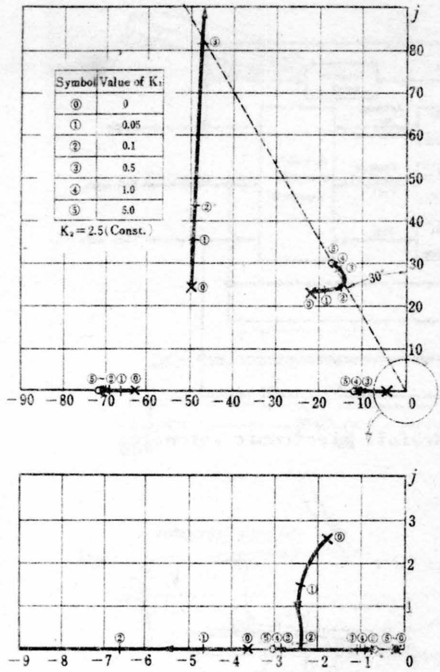


Figure 7. Root locus diagram of $F_{11}(S)$

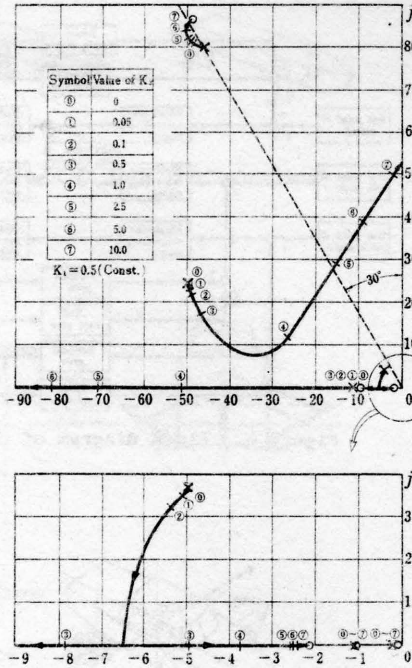


Figure 8. Root locus diagram of $F_{32}(S)$

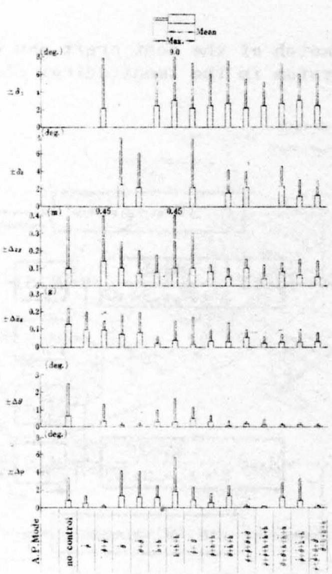


Figure 9. One of the test results in several autopilot modes

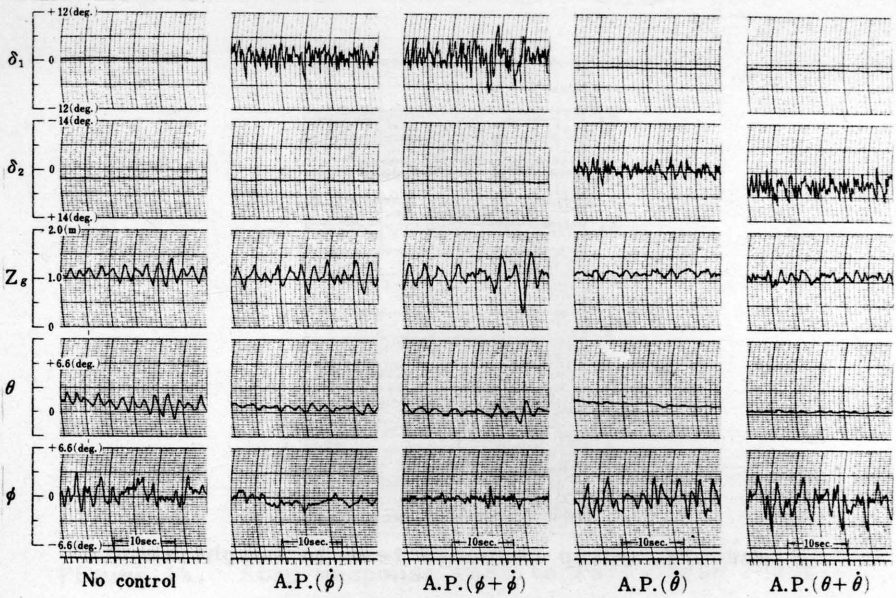


Figure 10. One of the oscillograms in several autopilot modes

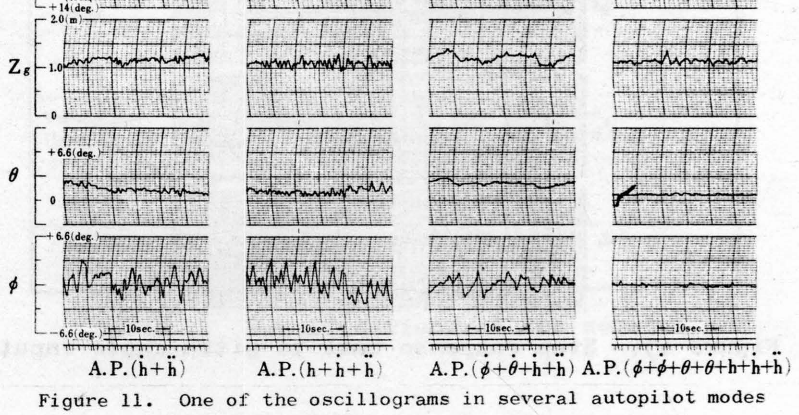


Figure 11. One of the oscillograms in several autopilot modes

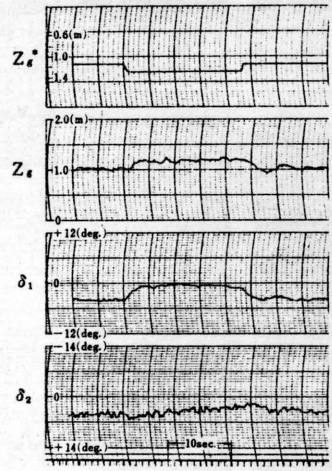


Figure 12. Step response test to height input

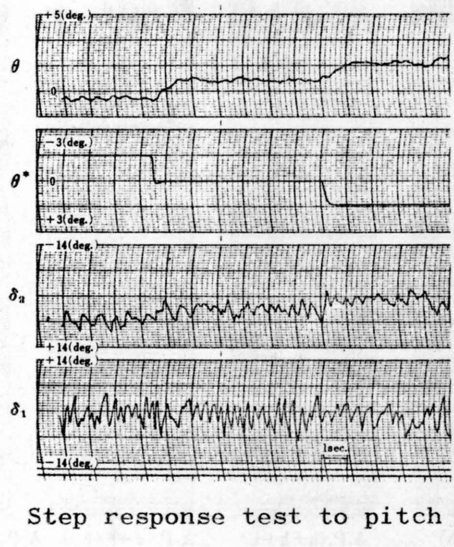


Figure 13. Step response test to pitch angle input

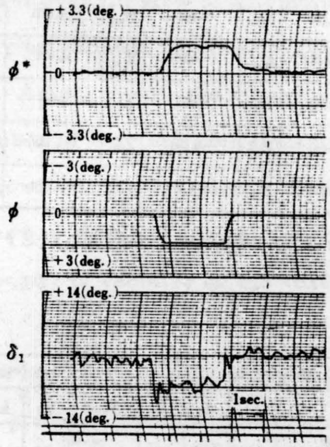


Figure 14. Step response test to roll angle input

Sensor	Type (Maker)	Essential specifications
Sonic height sensor	Type: Pulse (Japan Radio Co., Ltd.)	Range: 0.6～3.0meter Out put voltage gradient: 2V/meter Pulse: 20kc/sec.
Accelerometer	Type C−702407 (Kearfott)	In put range: ±1.5g Out put voltage: ±10V
Vertical gyro (Pitch and roll)	Type T−2105−1B (Kearfott)	Degree of Freedom: 360 deg. Roll, ±85 deg. Pitch. Out put voltage gradient: 0.206 V/deg.
Rate gyro (Pitch)	Type−T−2008−1A (Kearfott)	Maximum rate: 29 deg/sec Sensitivity: 0.465 v·sec/deg
Rate gyro (Roll)	Ditto	Ditto

Table 1. Specifications of the sensors

V = 30 (kt), Zg = 1.0 (m), $\theta = +2.0$ (deg)									
p_1	−0.165	M_{011}	−4.270	M_{012}	2.104	M_{031}	−0.379	M_{032}	−0.1645
p_2	−0.631	K_{11}	−33.450	K_{12}	−14.050	K_{31}	15.628	K_{32}	−12.069
p_3	±2.663j	Z_{111}	−0.185	Z_{121}	4.563	Z_{311}	0.244	Z_{321}	−0.052
p_4	−7.015	Z_{112}	−5.461	Z_{122}	−0.189	Z_{312}	−1.149	Z_{322}	−2.838
p_5	±1.787j	Z_{113}	±3.865j	Z_{123}	−11.210	Z_{313}	−5.624	Z_{323}	±2.986j

Table 2. Values of Mij(S)

	Height loop					Pitch loop			
	M_{011}	K_1	k_1	g_1	h_1	M_{032}	K_2	k_2	g_2
Result of calculation	m/rad −4.27	0.5	1	1.50	0.12	rad/rad −0.16	2.5	1	0.46
Result of experiment	m/rad −1.26	2.1	Ditto	Ditto	Ditto	rad/rad −0.29	9.5	Ditto	Ditto

Table 3. Comparison between calculated and experimental values of each parameter

Sensor / A. P. Mode	Sonic height sensor	Accelerometer	Vertical gyro (Pitch and roll)	Rate gyro (Pitch)	Rate gyro (Roll)
no control					
$\dot\phi$					○
$\phi+\dot\phi$			○		○
$\dot\theta$				○	
$\theta+\dot\theta$			○	○	
$\dot h+\ddot h$		○			
$h+\dot h+\ddot h$	○	○			
$\dot\phi+\dot\theta$				○	○
$\dot\phi+\dot h+\ddot h$		○		○	○
$\dot\theta+\dot h+\ddot h$		○		○	
$\phi+\dot\phi+\theta+\dot\theta$			○	○	○
$\phi+\dot\phi+h+\dot h+\ddot h$	○	○	○		
$\theta+\dot\theta+h+\dot h+\ddot h$	○	○	○	○	
$\dot\phi+\dot\theta+\dot h+\ddot h$		○		○	○
$\phi+\dot\phi+\theta+\dot\theta+h+\dot h+\ddot h$	○	○	○	○	○

Table 4. Autopilot modes and their sensors used

THE DYNAMIC CONTROL OF AUTOMOTIVE TRAFFIC AT A FREEWAY ENTRANCE RAMP

The need for control over the flow of automotive vehicles onto a high speed freeway is easily established by consideration of any one of several performance criteria. Among these, the more commonly used examples include: the large number of rear end collisions on freeway entrance ramps; the frequent occurance of rush-period "bottlenecks" just downstream from urban freeway entrance ramps; and, the reduced production of freeway when the density of vehicles on the roadway passes a "critical" limit.

To complement the need for control, early experiments in ramp metering indicated the potential benefits that were available when the flow of vehicles onto a freeway, was judically throttled.[1, 2] Subsequently, it was also shown that significant improvement in system behavior was obtained when the movement of vehicles onto the freeway was coordinated with the movement of vehicles in the upstream (right) outside-lane of the freeway.[3-16] This process, termed freeway-ramp merge control, coordinates the projected entrance onto the freeway of vehicles on a ramp with the projected arrival at the merge zone of vehicles that are in the upstream, outside-lane of the freeway.

There are presently two basic approaches to the freeway-ramp control problem. The first employs a single traffic control signal and is predicated on the use of a controller that is designed to release stopped vehicles from a given location on the entrance ramp. Thus only the initial release time and position of the merge vehicle are controlled, and no attempt is made to influence the vehicle after release. This approach is being developed at the Texas Transportation Institute (T. T. I.) at Texas A and M University under contract to the United States Bureau of Public Roads.

It is relatively unsophisticated in its "action" and is presently being employed in an experimental project on the Gulf Freeway in Houston, Texas. The second approach to the design of a controller does not attempt to control the initial time or place of the merge vehicle. Instead it attempts to modify the trajectory of merge vehicles on the ramp so as to guide them into available holes in the stream. This approach is presently being developed by Raytheon under a contract with the Bureau of Public Roads and is substantially more sophisticated in the scope of controller action.*

The work described in this paper concerns the analysis and design of a dynamic controller for the T. T. I. system during operation in the single-vehicle-merge mode. For this case a single vehicle is released to merge into the stream when a suitable target location is detected in the stream, and provided that the previously released vehicle has completed a successful merge.

The System Model

The state of the merge control system in the single-vehicle mode is described by the number of vehicles waiting behind the control signal and the position on the ramp of the last merge vehicle released. (see figure 1) Thus, when there are N vehicles awaiting service and the last released vehicle is on the ramp at a distance X from the end, the system is described by a point in two dimensional space, (N, X). For this case: the variable X decreases from L to zero as the released vehicle moves to complete the merge; and the variable N increases by one for every arrival into the queue or decreases by one for every vehicle released from the queue.

---

* This system will include the capability to operate in the manner of the present T. T. I. if and when that is desirable.

This state model of the system is easily extended to the case of multiple-merge-vehicle operation by extension of the dimensionality of the state space to provide for up to m vehicles located on the ramp. However, little is gained by this extension at present because of the compound difficulty in analysis. Instead the model is first simplified by consideration of the system at the instants at which merges are completed. At these "renewal" instants the value of X is always zero and the system is completely described by the single integer value N . Thus, the system jumps from state to state at the completion of the merges, and N increases by K-1 at each renewal instant. (Here, K is a random variable that is equal to the number of arrival into the queue during the interval required for the last merge.) For this case, the state may decrease by at most 1 per merge (i. e., no arrivals). This model has been carefully analysed and is presented elsewhere.[17]

When the merge process is subdivided to differentiate between those vehicles that successfully complete a moving-merge into the stream within the merge zone and those vehicles that stop at the end of the acceleration strip before the execution of a successful merge, the state model is again two dimensional. However, in this case the variable X is defined to be A when there is a trapped vehicle stopped at the end of the acceleration strip and zero otherwise. Hence there are two states corresponding to each value of N , as shown in the state diagram in figure 2 . For this model, the system transitions occur as follows: the system moves from state (N, 0) to state (N-1, 0) when a moving merge is completed; it moves from (N, 0) to (N-1, A) when an attempted moving merge fails and the merge vehicle becomes trapped; it moves from (N-1, A) to (N-1, 0) when the trapped vehicle executes a successful merge; it moves from (N, A) to (N+1, A) when an arrival occurs while a vehicle is trapped; and, it moves from (N, 0) to (N+1, 0) when an arrival occurs while no vehicle is trapped.

For the system model described above there are arrivals, successful moving-merges, unsuccessful moving-merges and successful stopped-merges. When attention is focused upon the probability that the system is in a given state at the instant t, and the assumption is made that the probability of an arrival during the interval of time between t and t+dt is both dependent upon t and proportional to dt, then the system model represents a queuing system that is subject to a time-dependent Poisson arrival process. Released vehicles are then served in one of two ways; that is, with either a successful or an unsuccessful moving-merge. When the latter type of service occurs the vehicle then undergoes additional delay correspond to the completion of a stopped-merge.

A simplifying assumption is now made that the probability of a successful (or an unsuccessful) moving-merge, or the probability of a successful stopped-merge, during the interval between t and t+dt is proportional to dt. In addition all three quantities are assumed to be functions of the traffic parameter q, while the first two are also dependent upon the controller parameter T. Hence, these are only four transitional probabilities in the model; these are:

1) $\lambda(t)\,dt$ - the probability of an arrival during dt
2) $\mu(q, T)\,dt$ - the probability of a successful moving-merge during dt
3) $\mu_1(q, T)\,dt$ - the probability of an unsuccessful moving-merge during dt
4) $\mu_2(q)dt$ - the probability of a successful stopped-merge during dt

The system equations are then

$$[\dot{P}(t)] = [A(q, T, t)]\,[P(t)] \qquad (1)$$

where

$$[P(t)] = \begin{bmatrix} P_{0,0}(t) \\ P_{0,A}(t) \\ P_{1,0}(t) \\ P_{1,A}(t) \\ \vdots \\ P_{N,0}(t) \\ P_{N,A}(t) \end{bmatrix} \tag{2}$$

and $[A(q, T, t)]$ is shown with the arguments omitted.

$$[A] = \begin{bmatrix} -\lambda & \mu_2 & \mu & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & -\beta & \mu_1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \lambda & 0 & -\alpha & \mu_2 & \mu & \cdots & 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & -\beta & \mu_1 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & -\alpha & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & 0 & \lambda & 0 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & 0 & 0 & \lambda & \cdots & 0 & -0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & -\beta & \mu_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & \lambda & 0 & -(\mu+\mu_1) & \mu_2 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \lambda & 0 & -\mu_2 \end{bmatrix} \tag{3}$$

Here $\beta = \lambda + \mu_2$ and $\alpha = \lambda + \mu + \mu_1$.

The system is then considered when the transitional probability rates are constants, and the steady state probabilities are evaluated. For this purpose $\lim_{t \to \infty} [\dot{P}(t)]$ is set equal to zero and the solution to $[A]\,[P] = 0$ is obtained. Here $\lim_{t \to \infty} [P(t)] \overset{\Delta}{=} [P]$

$$\text{where } [P] = \begin{bmatrix} P_{0,0} \\ P_{0,A} \\ P_{1,0} \\ P_{1,A} \\ \vdots \\ P_{N,0} \\ P_{N,A} \end{bmatrix} \tag{4}$$

When the rows of the product [A][P] are considered in pairs, the first two yield the equations

$$\begin{bmatrix} \lambda \\ 0 \end{bmatrix} \begin{bmatrix} P_{0,0} \end{bmatrix} + \begin{bmatrix} -\mu_2 & \mu \\ \beta & -\mu_1 \end{bmatrix} \begin{bmatrix} P_{0,A} \\ P_{1,0} \end{bmatrix} = 0 \tag{5}$$

from which

$$P_{0,A} = \frac{\lambda \mu_1}{\beta\mu + \mu_1\mu_2} P_{0,0} \tag{6}$$

and

$$P_{1,0} = \frac{\lambda\beta}{\beta\mu + \mu_1\mu_2} P_{0,0} \tag{7}$$

Next the K^{th} pair of equations is given by

$$\begin{bmatrix} 0 & -\lambda \\ 0 & 0 \end{bmatrix} \begin{bmatrix} P_{K-1,A} \\ P_{K,0} \end{bmatrix} + \begin{bmatrix} 0 & \alpha \\ -\lambda & 0 \end{bmatrix} \begin{bmatrix} P_{K,A} \\ P_{K+1,0} \end{bmatrix} + \begin{bmatrix} -\mu_2 & -\mu \\ \beta & -\mu_1 \end{bmatrix} \begin{bmatrix} P_{K+1,A} \\ P_{K+2,0} \end{bmatrix} = 0$$

Z transforming this pair of equations yields, after simplification,

$$\begin{bmatrix} P_A(Z) \\ P_0(Z) \end{bmatrix} = \frac{P_0}{\Delta(Z)} \begin{bmatrix} Z^3 \lambda\mu_1 - Z^2 \lambda\mu_1 \\ Z^4(\mu_1\mu_2 + \beta\mu) - Z^2(\mu\lambda + \beta\mu + \beta\mu_1) + Z^2(\lambda\mu + \lambda\mu_1) \end{bmatrix} \tag{9}$$

where

$$P_A(Z) = \sum_{n=0}^{\infty} P_{n,A} Z^{-n} \tag{10}$$

$$P_0(Z) = \sum_{n=0}^{\infty} P_{n,0} \, Z^{-n} \tag{11}$$

and

$$\Delta_1(Z) = Z^4 (\mu_1 \mu_2 + \beta \mu) - Z^3(\alpha \beta + \lambda \mu) + Z^2(\lambda \alpha + \lambda \beta) - \lambda^2 Z \tag{12}$$

But

$$\lim_{Z \to 1} [P_A(Z) + P_0(Z)] = 1 \tag{13}$$

from which, by use of L' Hospitals Rule,

$$P_{0,0} = \frac{\mu_2(\mu + \mu_1) - \lambda(\mu_1 + \mu_2)}{\mu_2(\mu + \mu_1)} = 1 - \frac{\lambda}{\frac{\mu_2(\mu + \mu_1)}{\mu_1 + \mu_2}} = 1 - \frac{\lambda}{\mu'} \tag{14}$$

with N infinite. In addition the expected number of vehicles in the queue, E(n), is given by

$$E(n) = \lim_{Z \to 1} \frac{d}{dZ} [P_A(Z) + P_0(Z)] \tag{15}$$

from which, after appropriate manipulations and use of L' Hospitals Rule

$$E(n) = \frac{\lambda [\lambda \mu_1(\mu - \mu_2) + \mu_2(\mu + \mu_1)(\mu_1 + \mu_2)]}{\mu_2(\mu + \mu_1)[\mu_2(\mu + \mu_1) - \lambda(\mu_1 + \mu_2)]} = \frac{\lambda}{\mu_1} \frac{\left[1 + \frac{\lambda}{\mu'} \frac{\mu_1(\mu - \mu_2)}{(\mu_1 + \mu_2)^2}\right]}{\left[1 - \frac{\lambda}{\mu'}\right]} \tag{16}$$

Since the steady state probability of being empty, $P_{0,0}$, is zero when $\lambda = \mu' \triangleq \frac{\mu_2(\mu+\mu_1)}{\mu_1+\mu_2}$ and the expected steady state queue length is infinite at that arrival rate, $\lambda = \mu'$ corresponds to a borderline stability condition and μ' is the steady state service rate of the merge control system. Furthermore, when the system is modelled as a single queue with two dissimilar servers (where one server provides the user with moving merges at a rate of μ with probability $\mu/(\mu + \mu_1)$ while the second server provides stopped-merges at a rate of μ_2 with probability $\mu_1/(\mu + \mu_1)$ then the expected service time is

$$T_E = \frac{\mu}{\mu + \mu_1}\frac{1}{\mu} + \frac{\mu_1}{\mu + \mu_1}\;\frac{1}{\mu_2}$$
$$= \frac{\mu_1 + \mu_2}{\mu_2(\mu + \mu_1)} \tag{17}$$

provided that the choices of servers is random with proper associated probabilities. For that case the effective service rate is also μ' since

$$\frac{1}{T_E} = \frac{\mu_2\,(\mu + \mu_1)}{\mu_1 + \mu_2} = \mu' \tag{18}$$

The Design of a Controller

One useful and fairly common model for an urban freeway employs stochastic processes with slowly varying parameters to account for both ramp arrivals and highway flows. In particular, the peak-period ramp-arrival process is often described as a time-dependent Poisson process[18] with λ (t) as shown in figure 3. In addition, the intervehicle spacing for vehicles on the highway are described as independent samples from an Erlang distribution[19]

$$f(t) = \frac{(aq)^a\, t^{a-1}\, e^{-aqt}}{(a-1)\,!} \tag{19}$$

where a is an integer and q is a slowly varying function of time corresponding to average volume (see figure 4). Based upon these descriptions, and subject to the assumption that the process parameter vary slowly, the controller is designed to maximize the service rate μ' subject to a imitation of downstream freeway capacity.

The service rate μ' is a function of μ and μ_1 which are both dependent upon the controller parameter T. Hence by setting $\frac{d\mu'}{dT} = 0$, the values of T corresponding to the relative minima of μ' are located.

Particularly

$$\frac{d\mu'}{dT} = \frac{\mu_2(\mu_1 + \mu_2)(\frac{d\mu}{dT} + \frac{d\mu_1}{dT}) - \mu_2(\mu + \mu_1)\frac{d\mu_1}{dT}}{(\mu_1 + \mu_2)^2} = 0 \quad (20)$$

reduces to

$$(\mu_1 + \mu_2)(\frac{d\mu}{dT} + \frac{d\mu_1}{dT}) = (\mu + \mu_1)\frac{d\mu_1}{dT} \quad (21)$$

since $(\mu_1 + \mu_2)^2$ is positive and finite, and $\mu_2 \neq 0$

From eq. (21)

$$\frac{d}{dT}[\ln(\mu + \mu_1) - \ln(\mu_1 + \mu_2)] = 0 \quad (22)$$

When T corresponds to the threshold that is set on the minimum spacing between vehicles on the freeway into which a moving merge will be attempted, then all spacings in excess of this threshold may be termed " gaps." With this threshold limit set, the rate at which gaps appear in a stream with volume equal to q vehicles per second is:

$$\mu + \mu_1 = q \int_T^{\infty} \frac{(aq)^a t^{a-1} e^{-aqt}}{(a-1)!} dt$$

$$= q e^{-aqT} \sum_{i=0}^{a-1} \frac{(aqT)^i}{i!} \quad (23)$$

When the probability that a presented gap of t is accepted by a driver for a moving merge, is described by[19]

$$P_a(t) = 1 - e^{-Kt} \quad (24)$$

then the rate of successful moving merges is

$$\mu = q \int_T^{\infty} (1 - e^{-Kt}) \left[\frac{(aq)^a t^{a-1} e^{-aqt}}{(a-1)!}\right] dt$$

$$= q \left[e^{-aqT} \sum_{i=0}^{a-1} \frac{(aqT)^i}{i!} - \frac{(aq)^a}{(aq+K)^a} e^{-(aq+K)T} \sum_{i=0}^{a-1} \frac{[(aq+K)T]^i}{i!} \right] \quad (25)$$

and the rate of unsuccessful moving merges is

$$\mu_1 = q \frac{(aq)^a}{(aq + K)^a} e^{-(aq+K)T} \sum_{n=0}^{a-1} \frac{[(aq+K)T]^n}{n!} \qquad (26)$$

When the indicated derivatives are evaluated and employed in Eq. (21) then algebraic manipulations yields

$$\mu_2 = \frac{q\, e^{-(aq+K)T}}{(aq + K)^a} \left[\sum_{n=0}^{a-1} \frac{[(aq+K)^a (aq)^n - (aq+K)^n (aq)^a]\, T^n}{n!} \right] \qquad (27)$$

as the equation from which the optimum threshold settings are obtained, subject to the constraints $T \geq 0$. Since this equation is of the form

$$\mu_2 = A\, e^{-BT} \sum_{n=0}^{a-1} C_n T^n \qquad (28)$$

and

$$C_n > 0 \text{ for every } n \qquad (29)$$

the sum

$$\sum_{n=0}^{a-1} C_n T^n > 0 \text{ for } T > 0 \qquad (30)$$

In addition, the derivative of the expression

$$f(T) = A\, e^{-BT} \sum_{n=0}^{a-1} C_n T^n \qquad (31)$$

is

$$f'(T) = -K (aq + K)^a \left[e^{-BT} \sum_{n=0}^{a-1} \frac{(aqT)^n}{n!} \right] \qquad (32)$$

Inspection of this quantity, with $B = aq + K$ reveals that $f'(T)$ is negative for all positive T. Therefore, f(T) is monotonic decreasing for $T \geq 0$. Based upon this, it is concluded that a unique optimum solution for the controller threshold T exists. That value is zero when

$$\mu_2 \geq q \frac{1}{(aq + K)^a} [(aq + K)^a - (aq)^a] \tag{33}$$

and the positive number T_{opt}, obtained as the solution to Eq. (27), otherwise.

With T_{opt} evaluated as described above, the control policy is specified next. In particular, when the sum of the volume on the freeway and the demand on the ramp does not exceed the volume limitation for freeway, a threshold of T_{opt} is used for the ramp. This insures the highest possible service rate for demand on the ramp and results in minimum expected delay, minimum expected queue length, etc. When the sum of the freeway volume and ramp demand exceed the limitation, a threshold of T_{opt} is not acceptable. Instead a threshold T_c must be employed such that the sum of the freeway volume and the served portion of the demand equals the volume limitation.

At this point it is necessary to consider more exactly the heretofore undefined quantities of freeway volume, ramp service and volumetric limitation. For this purpose it is noted that experience (and traffic flow theory)[20] indicate that the short term production of any given point on a freeway cannot exceed some upper bound Q, (of approximately 2000 vehicles per hour) without significant risk of breakdown in the flow of traffic. Thus the T_A minute running average of flow past a critical point must be limited to approximately QT_A vehicles.* When this short term average volume is below the specified capacity limit, it is possible to admit additional vehicles from the ramp into the stream, provided that the running average of the sum remains below the critical value. Thus the average ramp service is limited to at most the difference between the limiting and actual average volumes.

---

* depending upon the facility

Many solutions are possible to the controller design problem that satisfy the above restriction. One such possibility allows the threshold to be set at T_{opt}, provided the restriction is met, and inhibits all merging otherwise. By this process all available roadway capacity is used as quickly as possible, and additional waiting vehicles are inserted into gaps as additional room for single vehicles arises. This solution, when associated with the limit condition in which T_{opt} equals zero, becomes the capacity adjusted metering system.[21, 22]

A second approach to the controller design problem involves the gradual adjustment of the threshold as a function of freeway volume. This technique has the usual advantage associated with smooth variation in controller policy and likewise, it has the added limitation associated with smoothing - extra delay. In particular, the first policy carries the risk of inserting an acceptable averaged number of vehicles into the stream too quickly (thus causing an avoidable breakdown) while the second policy includes the risk of adjusting too slowly (thus overloading the stream).

Summary

The analysis presented above is based upon use of a Markov system model with slowly varying parameters. This assumption is predicated on experimental results that indicate the variational time for the model parameters is between one and two orders of magnitude longer than the response times of the system (i. e., a control system response time on the order of 30 seconds[19] vs. a variational demand time on the order of one hour)[18]. In addition, the optimal control policy is derived by relating the stationary optimum to an experimentally derived characterization of the freeway flow process, with the added implied assumption that the time between the release of a vehicle and the time that vehicle reaches the merge zone is negligible. When this last assumption is not valid, the expected time to complete a merge is increased by t_a and the expected hourly service capacity is reduced from μ' to $\mu'/(1 + \mu' t_a)$. However, the evaluation of T_{opt} is independent of this condition, and the controller designs consideration remain unchanged.

Although T_{opt} is a function of q which in turn varies with time, it is possible to implement a controller in which a single value is selected for the threshold, (by assuming T_{opt} is a constant); or it is possible to let the threshold vary as a predetermined function of time (by assuming q is a known function of time); or it is possible to vary the threshold to reflect the measured value of q. Experiments with various values of threshold setting T have been performed on several of the ramps on the Gulf Freeway and empirical fixed threshold settings of approximately 3-5 seconds have proven to be acceptable. This mode of operation has come to be known as the Gap Acceptance Mode. In addition, with the threshold set to zero, the short term averaged volume restriction has been employed for merge control. This mode is known as the Demand-Capacity mode.

In both of the above modes significant improvements have been observed on the Gulf Freeway. These include[23] a 10% increase in rush hour volume; a speed increase of 30% during the rush hour; a reduction in the average travel time from 16 to 11 minutes for the 5 mile test section, a reduction from 145 to 75 in the yearly number of accidents; and, not a single ramp accident has occured during the two years that control has been in effect on the Telegraph Road Interchange inbound entrance ramp.

References

1. Foote, Robert S.: Installation of a Tunnel Traffic Surveillance and Control System, in "Traffic Control Theory and Instrumentation," Plenum Press, New York 1965

2. May, A. D., P. Athol, W. Parker, and J. B. Rudden: Development and Evaluation of Congress Street Expressway Pilot Detection System, "Highway Res. Board Rec. 21," 1963

3. Drew, Donald R.: Theoretical Approaches to the Study and Control of Freeway Congestion, "Texas Transportation Inst. Res. Rept. 24-1, 1964

4. Drew, Donald R.; and Charles J. Keese: Freeway Level of Service as Influenced by Volume Capacity Characteristics, Texas Transportation Inst. Res. Rept. 24-3, 1964

5. Drew, Donald R.: Stochastic Considerations in Freeway Operations and Control, Texas Transportation Inst. Res. Rept. 24-5, 1964

6. McCasland, William R.: Traffic Characteristics of the Westbound Freeway Interchange Traffic on the Gulf Freeway, Texas Transportation Inst. Res. Rept. 24-7, 1964

7. Wattleworth, Joseph A.: System Demand-Capacity Analysis on the Inbound Gulf Freeway, Texas Transportation Inst. Res. 24-8, 1964

8. McCasland, William R.: Capacity-Demand Analysis of the Wayside Inter Interchange on the Gulf Freeway, Texas Transportation Inst. Res. Rept. 24-9, Texas A and M University, College Station, 1965

9. Pinnell, Charles, Donald R. Drew, William R. MCasland, and Joseph A. Wattleworth: Inbound Gulf Freeway Ramp Control Study I, Texas Transportation Inst. Res. Rept. 24-10, Texas A and M University, College Station, 1965

10. Pinnell, Charles, Donald R. Drew, William R. McCasland, and Joseph A. Wattleworth: Inbound Gulf Freeway Ramp Control Study II, Texas Transportation Inst. Res. Rept. 24-13, Texas A and M University, College Station, 1965

11. Drew, Donald R.: Gap Acceptance Characteristics for Ramp-Freeway Surveillance and Control, Texas Transportation Inst. Res. Rept. 24-12, Texas A and M University, College Station, 1965

References, contd.

12. Wattleworth, Joseph A.: Peak-period Analysis and Control of a Freeway System, Texas Transportation Inst. Res. Rept. 24-15, Texas A&M University, College Station, 1966

13. Drew, D. R., L. R. LaMotte, J. H. Buhr, and J. A. Wattleworth: Gap Acceptance in the Freeway Merging Process, Texas Transportation Inst. Res. Rept. RF 430-2, Texas A&M University, College Station, 1966

14. Wattleworth, J. A., J. H. Buhr, D. R. Drew, and F. Gerig: Operational Effects of Some Entrance Ramp Geometrics on Freeway Merging, Texas Transportation Inst. Res. Rept. RF 430-3, Texas A&M University, College Station, 1966

15. Drew, D. R., J. A. Wattleworth, and J H. Buhr: Gap Acceptance and Traffic Interaction in the Freeway Merging Process, Proc. Traffic Systems Div., U. S. Bureau of Public Roads, 1966

16. Drew, D. R., J. H. Buhr, and R. H. Whitson: The Determination of Merging Capacity and Its Application to Freeway Design and Control, Texas Transportation Inst. Rept. 430-4, Texas A&M University, College Station, 1967

17. Wiener, Richard, "Some Theoretical Considerations in the Design and Operation of Gap Acceptance Ramp Control Systems for Urban Freeway," doctoral dissertation, Polytechnic Institute of Brooklyn, Brooklyn, N. Y., 1968.

18. Highway Capacity Manual, Highway Research Board, Special Report 87, 1965

19. Drew, Donald R.: "Traffic Flow Theory and Control," Chapter 9, McGraw Hill, New York, 1968

20. Haight, Frank A.: "Mathematical Models for Traffic Flow," Academic Press, New York, 1963

21. Wattleworth, Joseph A.: "Peak-Period Control of Freeway Systems; Some Theoretical Considerations," doctoral dissertation, Northwestern University, Evanston, Ill. 1963

22. Hodgkins, Edmund A.: Effects of Volume Controls on Freeway Traffic Flow - A Theoretical Analysis, Public Roads, February 1968

23. Buhr, Johann H.: Freeway Ramp Control, Civil Engineering, ASCE, June 1968.

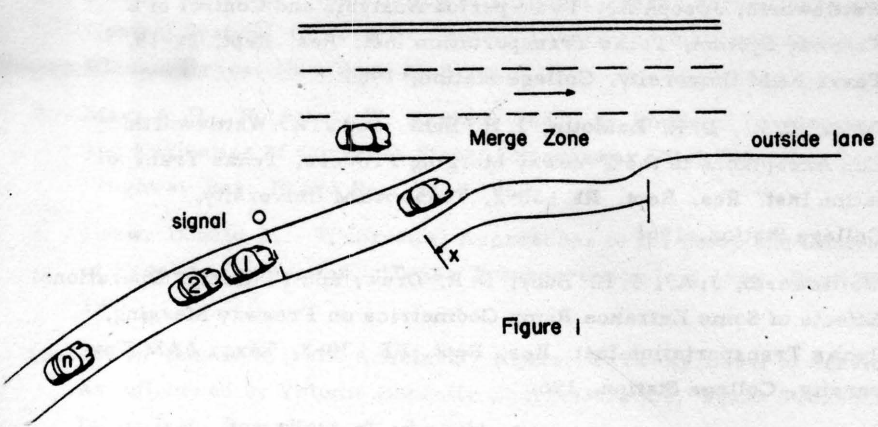


Figure 1 A Schematic Representation of a Freeway Merge Zone for for an Entrance Ramp.

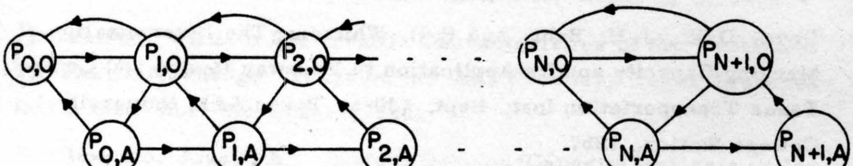


Figure 2 A Markov Model for the Merge Process

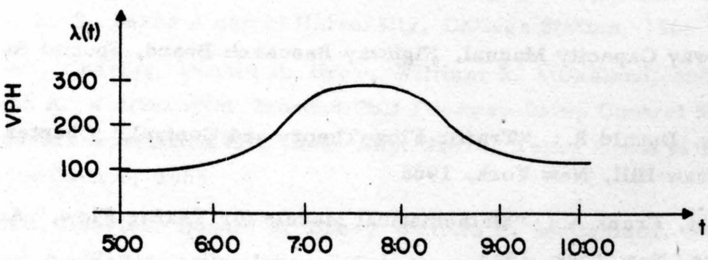


Figure 3 The Mean of a Time Dependent Piosson Process Used to Characterize the Ramp Arrival Process.

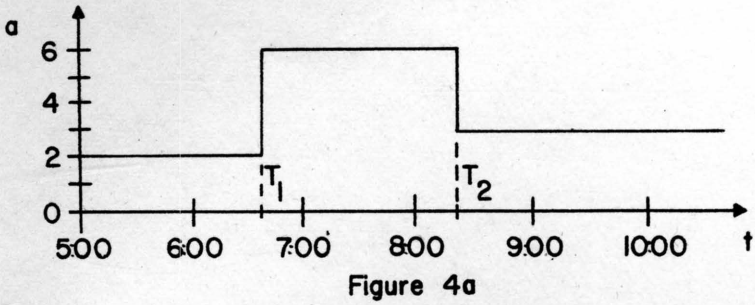


Figure 4a The Time-Dependent Erlany Parameter "a"

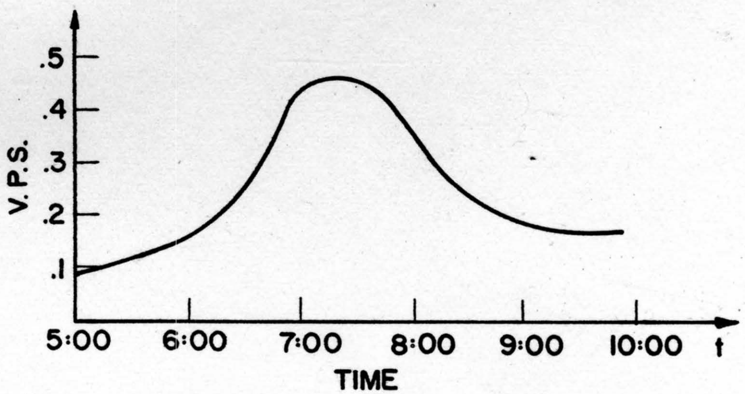


Figure 4b The Time-Dependent Freeway Volume

BIBLIOTEKA w Białymstoku SZKOŁY INŻYNIERSKIEJ