Aspects of automatic train control

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Additional Information:

- A Doctoral Thesis. Submitted in partial fulfillment of the requirements for the award of Doctor of Philosophy of Loughborough University.

Metadata Record: https://dspace.lboro.ac.uk/2134/10613

Publisher: © Ian Peter Milroy

Please cite the published version.
This item was submitted to Loughborough University as a PhD thesis by the author and is made available in the Institutional Repository (https://dspace.lboro.ac.uk/) under the following Creative Commons Licence conditions.

For the full text of this licence, please go to: http://creativecommons.org/licenses/by-nc-nd/2.5/
ASPECTS OF AUTOMATIC TRAIN CONTROL

by

Ian Peter Milroy, M.A., M.I.E. Australia

A doctoral thesis submitted in
partial fulfilment of the requirements
of the award of Doctor of Philosophy
of the Loughborough University of Technology

July 1980

Supervisor:--

W. Forsythe, M.Sc.
Department of Electronic and
Electrical Engineering

© by Ian Peter Milroy 1980
ABSTRACT

This thesis describes research and development work carried out by the author into the control of traction and braking systems on rail vehicles.

After a review of recent developments, the problem of driving a train under minimum-energy control subject to timetable and operational constraints is discussed. This is partitioned into two sections. Firstly, target time and velocities for key points on the journey are computed; these are communicated to or stored on the train, together with route and vehicle data. Secondly, an on-board digital system drives the train to each target according to control algorithms which incorporate a predictor-corrector module, whose function is to determine which of two criteria of performance is to be used (minimum-energy when running early or on-time, minimum-time when running late).

Most of the thesis is devoted to the analysis and design of the train-borne control system. The general form of the optimal control (of tractive or braking effort) is determined by the application of Pontryagin's Maximum Principle over each section of the journey. However, the moments of transition between the various modes of control are calculated by a method which involves a look-ahead model in the predictor module, rather than by iterative solution of the state and co-state equations.

An important aspect of the design is the dynamic response of the braking sub-system, which may include a substantial pneumatic transport lag within the control loop. S-plane and z-plane design procedures for the required discrete control algorithms to achieve a specified transient response are derived.

The thesis concludes with a chapter on the instrumentation required for the train-borne control system.

Key-words

Railway automation, train control, automatic train operation, energy-minimisation, Pontryagin's Maximum Principle, digital control.
## CONTENTS

### CHAPTER 1: THE PROJECT
- 1.1 Aims of Project ........................................ 1.1
- 1.2 Program of Work ........................................ 1.2
- 1.3 Structure of the Thesis ................................ 1.6

### CHAPTER 2: A REVIEW OF THE LITERATURE ON AUTOMATIC TRAIN CONTROL
- 2.1 Early Developments in A.T.O. ......................... 2.1
- 2.2 Why Automate Railways? .............................. 2.2
- 2.3 Summary of Reasons for Railway Automation .......... 2.3
- 2.4 Energy Consumption of Rail Vehicles ................ 2.4
- 2.5 Algorithms and Software .......................... 2.5
- 2.6 Conclusion ........................................... 2.6

### CHAPTER 3: TIME-KEEPING CONTROL OF AN AUTOMATICALLY DRIVEN TRAIN
- 3.1 Data Needed by Automatic Time-Keeping System .... 3.1
- 3.2 Reasons for Automatic Time-Keeping ............... 3.2
- 3.3 The Concept of a "Supervisory" System for Time-Keeping Control ........................................ 3.3
- 3.4 A Predictor-Corrector Strategy ....................... 3.4
- 3.5 The B.R. Autodrive System .......................... 3.5
- 3.6 Prediction Algorithm .............................. 3.6
- 3.7 Correction Algorithm .............................. 3.7
- 3.8 Sensitivity of Predicted Time-Keeping Error to Velocity Disturbances .......................... 3.8
- 3.9 Simulation of Time-Keeping Algorithm ........... 3.9
CHAPTER 4: PARTITION OF THE TRAIN CONTROL PROBLEM .. 4.1

4.1 Timetable Planning Program ............... 4.4

4.1.1 Inputs to the Timetable Planning Program ............... 4.4

4.1.2 Function of the Timetable Planning Program ............... 4.8

4.1.3 Outputs from the Timetable Planning Program ............... 4.12

4.2 Chapter Summary .............................. 4.15

CHAPTER 5: APPLICATION OF PONTRYAGIN'S MAXIMUM PRINCIPLE ...................... 5.1

5.1 General Procedure for Applying Pontryagin's Principle ...................... 5.3

5.1.1 Step 1 ............................... 5.3

5.1.2 Step 2 ............................... 5.3

5.1.3 Step 3 ............................... 5.4

5.1.4 Step 4 ............................... 5.4

5.1.5 Step 5 ............................... 5.5

5.1.6 Steps Taken ............................... 5.5

5.2 Pontryagin's Method Applied to Minimum-Energy Control of a Rail Vehicle over a Journey Section ...................... 5.5
5.2.1 Dynamic Model .................... 5.6
5.2.2 Criterion of Performance ........ 5.7
5.2.3 The Hamiltonian .................... 5.8
5.2.4 The Co-State Variable Dynamics ... 5.8
5.2.5 The Constraints on Control u .... 5.9
5.2.6 The General Form of the Optimal Control ................................. 5.11
5.2.7 Optimal Switching of Control Modes ................................. 5.15
  5.2.7.1 Second Switching Decision ................................. 5.16
  5.2.7.2 First Switching Decision ................................. 5.20
5.2.8 Summary of Minimum-Energy Control ................................. 5.26
5.3 Minimum-Time Control .................... 5.26
  5.3.1 Introduction ................................. 5.26
  5.3.2 Analysis ................................. 5.27
5.4 Discussion ................................. 5.29
5.5 Chapter Summary ................................. 5.31
  5.5.1 ................................. 5.31
  5.5.2 ................................. 5.31
  5.5.3 ................................. 5.31
  5.5.4 ................................. 5.31
  5.5.5 ................................. 5.32

CHAPTER 6: ON-TRAIN CONTROL SYSTEM ................................. 6.1
  6.1 Inputs to the On-Train Control System ... 6.2
  6.2 Functions of the On-Train Control System 6.3
  6.3 Outputs from the On-Train Control System 6.4
  6.4 Choice of Sampling Interval ................................. 6.5
  6.5 Choice of Data Word Length ................................. 6.6
  6.6 Intel 8085 Implementation ................................. 6.8
CHAPTER 7: ANALYSIS AND DESIGN OF THREE TERM CONTROL SYSTEM FOR BRAKING PHASE

7.1 S-Plane Model of Longitudinal Train Dynamics

7.2 Analysis and Design of the Braking Control System

7.2.1 Proportional Control Only

7.2.2 Proportional & Integral Control

7.2.3 Proportional, Derivative and Integral Controller

7.3 Application of Design Procedures

7.4 Chapter Summary

CHAPTER 8: Z-PLANE ANALYSIS OF TRAIN CONTROL SYSTEM

8.1 Derivation of z-plane Transfer Function

8.1.1 Block Diagram of Autodrive System

8.1.2 Transfer Functions of Analog Elements

8.1.3 Evaluation of $G_A(z)$ for a Specific Class of Train

8.2 Interpretation of Transient Response Specification in the z Domain

8.3 Root Loci in the z Domain

8.3.1 Proportional Control (Braking Phase)

8.3.2 Effect of Increased Transport Lag

8.3.3 Effect of Increased Sampling Interval

8.4 Design Procedure for Two-Term Controller

8.4.1 z-Plane Transfer Function

8.5 Chapter Summary
<table>
<thead>
<tr>
<th>Chapter 9: The Concept of Partial Poles in the z-Domain</th>
<th>Page Nos</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.1 Causes of Delay</td>
<td>9.1</td>
</tr>
<tr>
<td>9.2 Previous Methods of Analysis</td>
<td>9.2</td>
</tr>
<tr>
<td>9.3 The Concept of Partial Poles in the z-Domain</td>
<td>9.3</td>
</tr>
<tr>
<td>9.4 Theoretical Considerations</td>
<td>9.5</td>
</tr>
<tr>
<td>9.5 Experimental Validation</td>
<td>9.6</td>
</tr>
<tr>
<td>9.6 Design Example</td>
<td>9.6</td>
</tr>
<tr>
<td>9.6.1 Introduction</td>
<td>9.6</td>
</tr>
<tr>
<td>9.6.2 System Parameters</td>
<td>9.7</td>
</tr>
<tr>
<td>9.6.3 Specification</td>
<td>9.8</td>
</tr>
<tr>
<td>9.6.4 Design Procedure for Proportional and Integral Controller</td>
<td>9.9</td>
</tr>
<tr>
<td>9.7 Conclusion</td>
<td>9.12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 10: Instrumentation for Automatic Train Opera...</th>
<th>Page Nos</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.1 General Review of Tachometer Instrumentation</td>
<td>10.1</td>
</tr>
<tr>
<td>10.1.1 Analog or Digital?</td>
<td>10.2</td>
</tr>
<tr>
<td>10.1.2 Types of Digital Tachometer</td>
<td>10.5</td>
</tr>
<tr>
<td>10.1.2.1 Capacitive Tachometers</td>
<td>10.6</td>
</tr>
<tr>
<td>10.1.2.2 The Active Electromagnetic Transducer</td>
<td>10.6</td>
</tr>
<tr>
<td>10.1.2.3 The Passive Electromagnetic System</td>
<td>10.7</td>
</tr>
<tr>
<td>10.1.2.4 Incremental Optical Tachometers (or Encoders)</td>
<td>10.8</td>
</tr>
<tr>
<td>10.1.2.5 Absolute Optical Encoders</td>
<td>10.9</td>
</tr>
<tr>
<td>10.1.2.6 Recommended Tachometers</td>
<td>10.10</td>
</tr>
</tbody>
</table>
List of Symbols

The following symbols recur in the thesis. Other symbols used only within one chapter are defined when first used. M.K.S. units are used throughout the thesis.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Acceleration</td>
</tr>
<tr>
<td>a_r</td>
<td>Desired acceleration</td>
</tr>
<tr>
<td>B</td>
<td>Specified deceleration in braking phase</td>
</tr>
<tr>
<td>E</td>
<td>Mechanical energy</td>
</tr>
<tr>
<td>E_m</td>
<td>Minimum mechanical energy over a journey section</td>
</tr>
<tr>
<td>f_i</td>
<td>A general function</td>
</tr>
<tr>
<td>H</td>
<td>Hamiltonian function</td>
</tr>
<tr>
<td>J</td>
<td>Criterion of performance for optimal control</td>
</tr>
<tr>
<td>K_0, K_1, K_2</td>
<td>Coefficients in Davis running-resistance formula</td>
</tr>
<tr>
<td>K_p, K_i, K_d</td>
<td>Coefficients in three-term control algorithms</td>
</tr>
<tr>
<td>M</td>
<td>Train mass</td>
</tr>
<tr>
<td>M'</td>
<td>Effective train mass (including rotational inertia)</td>
</tr>
<tr>
<td>P_i</td>
<td>A co-state variable</td>
</tr>
<tr>
<td>R</td>
<td>Train running resistance</td>
</tr>
<tr>
<td>s</td>
<td>The general s-domain complex variable ((\sigma + j\omega))</td>
</tr>
</tbody>
</table>
\( t \)  
Time

\( T \)  
Tractive effort (except in Chapters 8 and 9)

\( T \)  
Sampling interval (Chapters 8 and 9)

\( t_a \)  
Allowable time-keeping errors

\( t_c \)  
Coast initiation time

\( t_e \)  
Time keeping error

\( t_j \)  
Journey time remaining

\( t_T \)  
Target time

\( t' \)  
Predicted time of an event

\( t'_b \)  
Predicted time at constant deceleration

\( t'_j \)  
Predicted journey-time remaining

\( t'_T \)  
Predicted time of arrival at target

\( u \)  
A control variable (usually acceleration)

\( v \)  
Velocity

\( v_h \)  
Speed-holding velocity reference for autodrive

\( v_r \)  
Velocity reference for autodrive

\( v' \)  
Predicted velocity

\( V_p \)  
Permitted maximum velocity

\( V_T \)  
Target velocity

\( x \)  
Position on track (distance)
\( x_b \quad \text{Braking point} \\
x_c \quad \text{Distance at which coasting is initiated} \\
x_d \quad \text{Distance to go, to target} \\
x_g \quad \text{Distance at which gradient changes} \\
x_i \quad \text{A general state variable} \\
x_p \quad \text{Distance at which } V_p \text{ changes} \\
x_s \quad \text{Distance at which speed-holding commences} \\
x_T \quad \text{Target position} \\
x' \quad \text{Predicted position} \\
z \quad \text{The general } z\text{-domain complex variable} \\
\theta \quad \text{Angle of gradient} \\
\zeta \quad \text{Damping ratio (of transient response)} \\
\sigma \quad \text{Reciprocal of transient response time constant, and horizontal axis of } s\text{-plane} \\
\tau \quad \text{Transport lag (time delay in pneumatic brake system)} \\
\omega \quad \text{Angular frequency, and vertical axis of } s\text{-plane} \\
\omega_d \quad \text{Damped neutral frequency (of transient response)} \\
\omega_n \quad \text{Undamped natural frequency (of transient response)}
FOREWORD

(A cautionary note)

"We create models in the mind, to describe and sometimes to
govern the behaviour of aspects of the physical universe ....
These concepts are based on observation and measurement. But
since all our measurements and observations are nothing more
than approximations to the truth, the same must be true of all
concepts and calculations resting upon them".

K. F. Gauss

"Theoria Motus ..."

1795
CHAPTER 1

THE PROJECT

INTRODUCTION

This thesis describes research and development work carried out by the author in the field of railway automation. Various aspects of the automatic control of individual trains were studied, initially as part of the work of the train control group of the Research Division of British Rail. The project was supported in its later stages by the Directorate of Transport Planning and Research (South Australian Department of Transport), who proposed that the work be influenced by the possibility of a semi-automatic, rather than a fully automatic implementation on Australian railway trains.

1.1. Aims of Project

The project aims were to investigate the following specific aspects of the automatic control of trains.

(a) the development of control algorithms to enable an automatic time-keeping capability to be added to a train control system under development at the British Technical Centre, Derby, U.K. (If possible, the algorithms should be applicable to other AUTODRIVE systems).

(b) the application of optimum control theory to the problem of controlling trains so that, as well as having an automatic time-keeping capability, they consume a minimum amount of energy subject to operational constraints.
(c) the analysis of the longitudinal dynamics of the train during the braking phase, and the development of appropriate design procedures for the digital control algorithms to be used by on-train systems during that phase.

(d) the study of the on-train instrumentation required in order to implement the control strategies and algorithms mentioned above.

1.2. Program of Work

During the period September 1976 to July 1977 the author, while attending Loughborough University of Technology as a full-time research student, was seconded to the Research Division of British Rail on a part-time basis, and worked as a member of the Train Control Group.

During this period, work on the British Rail T.A.C.T. (Totally Automatic Control of Trains) was being carried out, and the author participated in numerous test runs of the laboratory train GEMINI, which was equipped with a Hewlett Packard 21 MX minicomputer interfaced to the electric traction and electro-pneumatic braking systems through an 8-bit Intel microprocessor, to a digital data transmission link through a second 8-bit processor, and to instrumentation for measuring the train's state. (See Figure 1.1.)
Figure 1.1. Digital System on GEMINI
The tests were made to examine the feasibility of driving a typical commuter train automatically through the modes of acceleration, holding a constant speed, and braking to a stopping target determined by a central control system. The relevant hardware had been designed and developed by full-time staff members of the B.R. Train Control Group, and train control algorithms based on digital two- and three-term control strategies had been developed and coded, (Thomas 1980, Ashworth 1977).

The author's contribution to the train control work during this period was:

(a) assisting with the conduct and evaluation of tests and the refinement of control algorithms.

(b) initiating a proposal for and demonstrating the feasibility of "supervisory" automatic time-keeping algorithms which would enable not only "target velocity" and "target position" to be controlled, but also "target arrival time".

The aim of the refinement was to conserve energy, and to improve traffic regulation in the system as a whole by allowing precise control of time of arrival at critical network nodes (e.g. merging junctions, or passing loops on single-line railways).

The algorithms were demonstrated on the Membrain digital differential analyser at Loughborough University, and prior to departure from Derby they were implemented on the test train.
(c) evaluating and comparing various methods of measuring train velocity and acceleration for use in the control system. In the installed system an acceleration signal was derived from a seismic accelerometer; the author investigated a number of digital differentiation algorithms for deriving acceleration from an incremental encoder mounted on an axle-end, and developed new procedures for determining optimum values of resolution (pulses per metre) and sampling interval for such a system.

(d) taking part in preliminary system design studies for a new project (B.R.A.T.O. - British Rail Automatic Train Operation) which is a successor to T.A.C.T. and involves a pilot scheme with actual commuter trains running on a section of track near Manchester which is shared with fast manually-driven intercity traffic and slow goods traffic. This project (in contrast to T.A.C.T. which was envisaged for dedicated all-automatic routes such as the Channel Tunnel) involves very considerable technical constraints because of the need for compatibility with the conventional signalling system.

Experience with these preliminary studies as part of the British Rail Train Control Group made the author aware of a wide range of railway operating procedures, techniques and conventions; this knowledge has been indispensable in keeping later research work practicable and realistic.

From the time of leaving U.K. in August 1977 to date, the author has been enrolled as a part-time external research student at Loughborough University, while carrying out normal academic duties as a Senior Lecturer in the School of Electrical Engineering
at the South Australian Institute of Technology. During this period the theoretical and simulation work concerned with the optimum control problem and the braking system dynamics has been carried out. Hardware and software for a laboratory rig to enable digital control algorithms to be evaluated have been developed and tested. The equipment is based on an Intel 8085 processor. (This work was carried out by a project student under the author's supervision, and is described in Appendix 1).

1.3. Structure of the thesis

The paragraphs above have presented a brief chronological account of the research carried out by the author. Due to the exigencies of his work the order of events has necessarily been a little abnormal - for example access to an equipped test train was available early in the piece, whereas the later research work has had to rely on simulation. It is hoped that this later work will be applied by the British Rail Group in a railway environment; the author will not be able to participate in this process.

Because of the above factors, the structure of the body of the thesis is not related to the chronological order of research, but is laid out in the context of the overall research aims. This structure is as follows:

Chapter 2 presents a review of recent work reported in the literature, and gives an overview of automation of a typical railway system's operation. The specific research work reported on here may be placed in the context of a distributed hierarchical control system. In this system some of the control functions (e.g. timetable planning) are carried out off-line, others by processors and staff
in central control rooms looking after many trains on hundreds of miles of route, and those functions which can be devolved out to processors and human drivers in individual trains are so devolved, in order to reduce the mass of data which must be transmitted to and handled by the central control facilities.

Chapter 3 is devoted to a description of the system developed by the author for predictive control of individual trains, to ensure that they arrive at their positional targets not only at the correct velocities, but also at the correct times as commanded by higher-level control staff or processors.

The system, proposed by the author, for transmitting only target positions, velocities and times from the control centres to the trains is offered as an elegant and simple solution to the vexed question of how to interface the higher-level "operations-research" packages which are becoming available (e.g. Junction Optimisation Techniques - a B.R. package) into an operational railway which requires for implementation the fine control of rapidly moving trains in real time.

It should be emphasised that the conventional "fail-safe" railway signalling system is retained and provides a mandatory safety envelope which the proposed system is not allowed to violate. (This is a fundamental constraint insisted on by British Rail, and allows scope for technical innovations to be tried out without the need for extensive modifications to already-installed vital safety systems). The reserved use of red and yellow signalling aspects for actual situations which are potentially dangerous (rather than for fine control of traffic) is seen as an enhancement of system safety rather than a degradation of it.
In Chapter 4 we commence the study of the optimum control of trains so as to minimise energy consumption; those parts of the hierarchy of command and control which influence energy-consumption are found to be the timetable planning function and the driving function. In this chapter algorithms are discussed which produce energy-optimal journey timetables subject to constraints fed in by the timetable planner. For these algorithms the journey is broken into a number of sections, and a sub-routine is required which can minimise the energy-consumption over a single journey-section.

In Chapter 5 such a routine is developed. Pontryagin's Maximum Principle is applied over a single journey section, and used to determine the general form of the optimal control so as to minimise mechanical energy at the wheel-rail interface, while achieving the correct target position, velocity and time.

In Chapter 6 the requirements of an on-train control system to enable the optimum control to be implemented are discussed, and it is shown that a conventional AUTODRIVE capable of following a commanded velocity reference may be used, with appropriate supervisory algorithms to alter the velocity reference as the journey proceeds.

Chapter 7 looks specifically at design procedures for the velocity-reference control loops which are at the heart of the proposed system, (and which have in fact been successfully implemented on Japan's Tokaido line and the B.R. T.A.C.T. system, as well as in Germany and France). The braking-phase is chosen for this part of the study, because it is well known that
this phase is the most difficult. Whereas the above project teams arrived at their control parameters by a valid iterative process of trial and correction on test trains, this thesis presents (for the first time to the author's knowledge) a rigorous design procedure utilising the root locus technique. Being based on a linear model of the braking system, the procedure provides first approximations for the control parameters, which can be adjusted to take account of system non-linearities. In Chapter 7 s-plane methods are used, and transport lag in the braking system is not taken into account.

In Chapters 8 and 9 we tackle the problem of transport lag, which is present to a significant degree in the braking systems of some rail vehicles. A new method of analysis is developed, in which a transport lag of less than one sampling interval is treated as a partial pole (or as a real pole and a partial zero) at the origin of the z-domain. Design procedures for z-domain control algorithms are developed, and again turned into algorithms suitable for implementation on any general purpose processor.

In Chapter 10 the thesis considers the problem of obtaining sufficiently accurate measurements of the train's state variables, (namely position, velocity and acceleration) for control purposes. A solution using a digital encoder on an axle is developed, and a novel design procedure is described, for determining the "best" resolution and sampling interval in such a system.

In the final chapter conclusions are drawn, and suggestions for further work are put forward.
We now turn our attention to the work published by others in the field of automatic train control.
CHAPTER 2
A REVIEW OF THE LITERATURE ON AUTOMATIC
TRAIN CONTROL

INTRODUCTION
In this chapter we place the author's research in the context of work carried out in various parts of the world in the field of automatic train control.

The shades of meaning attached to the phrases automatic train control (A.T.C.), automatic train protection (A.T.P.) and automatic train operation (A.T.O.) differ from one country to another, and even from one railway organisation to another; in this thesis we adopt the most common usage - A.T.C. is the general phrase, and embraces the two major subdivisions of A.T.P. (the safety system) and A.T.O. (the automation of the driving function). In basic terms, the only functions of the A.T.P. system are to reduce speed or bring the train to a halt in the event of a safety hazard or a failure of the A.T.O. system; much of the challenge of A.T.P. system design lies in ensuring that the hardware and software associated with the design are demonstrably 'fail-safe'. On the other hand the A.T.O. system involves the more refined control of traction and braking on each train so as to achieve a smooth flow of traffic in the network as a whole, as well as a smooth and satisfactory journey profile for each individual train, so as to satisfy the constraints imposed by the A.T.P., traction and braking systems, the adhesion limitations, the condition of the track and the requirements of passenger comfort (if any). In academic circles the A.T.O. function is sometimes referred to as "longitudinal control of rail vehicles"; there is a tendency for published work in this area to revolve around imagined or speculative types of novel transportation system whose
2.2

performance criteria can conveniently be built into a quadratic performance index. On the other hand some of the more mundane realities which beset the life of a control systems engineer working with a real (if conventional) train are conveniently neglected; these include such matters as the coefficient of friction of cast iron or composition brake blocks on steel wheels (the coefficient varies non-linearly but reasonably deterministically as a function of velocity and temperature), the running resistance of the train (which is a non-linear function of velocity, and also tends to vary stochastically with side-winds and curvature) and the dynamics of electro-pneumatic braking systems (which can be modelled reasonably satisfactorily as a conventional s-plane transfer function with an associated 'transport lag' or time delay).

For the above reasons, and because railway engineers are an esoteric group who tend not to report their findings in the academic journals, many of the citations in this chapter are of papers and articles published in specialist railway journals, or in the 'house' journals of the dozen or so major companies around the world who specialise in automatic train control equipment and software.

2.1 Early developments in A.T.O

The author's experience with the Train Control Group of British Rail was mainly in connection with A.T.O. rather than A.T.P., and for this reason most of the thesis is devoted to aspects of the former.

The first published indication that major development projects were taking place, leading to automation of the train-driving function on public railways, came at an international symposium on railway cybernetics in 1963, when Shinohara, Schmitz and Lagershausen...
(1963, 79)\textsuperscript{1} clearly spelt out the factors which had convinced railway managements in Japan and West Germany that such projects should be funded. At the same time a major British project (the construction of the London Transport Victoria Line, on which all trains were to be automatically controlled) was in an advanced stage but details had not yet been reported in the literature.

The first fully automatic train control systems (incorporating both A.T.P. and A.T.O.) came into service in Japan and the United Kingdom in the late sixties, and were predominantly analog in nature. During the seventies there has been a trend towards the use of digital technology in the command, communication, control and instrumentation systems associated with automatic train control, and in recent years the availability of cheap robust microprocessors and peripherals has led to a devolution of much of the control decision-making from central computers in control centres to individual or multiple microprocessors in each train. These trends, which reflect a major movement in the automation industry generally, are discussed in more detail by Forsythe Milroy and Thomas (1979, 42) in a paper "Towards Automatic Train Control" which is attached as Appendix 2. Eighteen references are cited.

2.2 Why automate railways?

The following books, papers and articles draw attention to various factors which have led railway managements to consider automation of the driving function on trains.

Barwell, Coales and Barton (1964, 18) wrote an important seminal paper on automation in railways.

\textsuperscript{1} The first number in parentheses indicates the year of publication; the second number indicates position in the alphabetical list of references at the end of the thesis.
In the United Kingdom, Barwell (1973, 17) in his book "Automation and Control in Transport" discusses both the reasons for and the technology of such developments as the A.T.O. system on London Transport's Victoria Line. This technology (developed by the Westinghouse Brake and Signal Company in co-operation with London Transport) has been transferred to railway organisations in many parts of the world through licensees and affiliates of the Westinghouse organisation. Gelbstein and Parkman (1974, 47) discuss important aspects of British Rail's A.T.C. technical philosophy in the specific content of in-cab signalling for the British Rail Advanced Passenger Train, then under development. Because of its higher speed limits on curves (a result of tilting coaches), and an improved hydrokinetic braking system, the wayside signals for conventional rolling stock are inappropriate for the A.P.T., and some ergonomists question whether human drivers can reasonably be expected to detect all wayside signals at speeds in excess of two hundred kilometres per hour (Gebherd, 1974, 87). Maxwell (1979, 89) spells out the developments which London Transport envisage for the years up to 1990; he covers the areas of automatic train control, automatic revenue collection, remotely-controlled passenger-information and passenger regulation systems, and remote control of tractive electrical power distribution. It is significant that the major factor in Maxwell's opinion is the need to reduce staff costs, which comprise 75% of London Transport's operating budget.

In Japan, Takoake (1976, 122) and Oshima et al (1976, 99) concur with Maxwell's predictions for the future development of automatic systems on urban rapid transit systems. Kitoake et al (1977, 75) and Kimara (1978, 74) discuss the ways in which the chopper-controlled commuter trains now in service on the fringes of Japanese cities are ideally
suited to full A.T.O., and Tabata (1978, 14) mentions the factors which led to full automation of the fast inter-city trains on the Shinkansen network.

In the United States of America the early failure of San Francisco's much vaunted Bay Area Rapid Transit System (Open University, 1975, 15) led to a phase of technical heart-searching about the directions in which railway technology was heading in that country. Allen (1976, 10) made a plea for relevance in transportation research, citing as a particular example the possibility of improving the capacity of existing railway routes by expenditure on improved signalling and automation equipment, rather than by expensive civil engineering works on "gee-whiz" (sic) new projects.

In making these comments Allen was echoing Cannon (1973, 31) who while U.S. Undersecretary of Transportation had delivered a reflective and philosophical Keynote address to the Joint Automatic Control Conference. "We have poured enough concrete now. The digital processor allows us to realise a vastly improved service from our existing transportation systems - at a relatively small investment in money and time - for the digital processor has opened a new era of automation in transportation".

These were fine words, but the present author, who rode on the B.A.R.T. system in San Francisco and held discussions with its engineers in May 1980, must report that (at least in the case of that system) Cannon's words must be regarded as a statement of intent rather than of accomplishment. The system still runs at less than half the planned capacity, despite the expenditure of many hundreds of thousands of dollars on software development, in an effort to demonstrate to the regulatory authorities of the State of California that the train protection system is fail-safe.
In 1973 neither Cannon, nor any engineer in the burgeoning microprocessor and microcomputer industries, realised the extent to which software development costs would run away with funds in the years to come, in comparison with low and falling computer hardware costs. Despite an improvement in the position since 1977, when the major microprocessor manufacturers started to support well-structured high level languages suitable for real-time instrumentation and control applications (for example Pascal and PL/M), the high cost of software development is consistently underestimated by technical management and remains the most pressing problem in applying microprocessor technology, in railways as in other industries.

Remaining with U.S. publications for the moment, Kaplan (1978, 72) devotes a complete paper to the applications of microprocessors in transport automation, and Kalra (1979, 69) provides an excellent review of A.T.O. practice in a number of urban rapid transit projects for U.S. cities; he also discusses practice in several European countries, but not in Japan. A recent U.S. Transportation Research Board publication (1979, 126) also provides an overview of United States practice in the field of A.T.O.

Automated train control systems are also coming into service or envisaged in many other countries, but the work is not extensively reported in the English speaking journals. The author has, however, found enough material to indicate that the technical, economic and social factors leading to railway automation are broadly similar throughout the world.

Relevant work in Germany is reported by Lagershausen (op.cit.),
Friedlander (1974, 43) and Kalusa (1978, 70). In France, a metropolitan rapid transit A.T.O. system at Lille is described anonymously in (1974, 2) and the S.N.C.F. inter-city semi-automatic A.T.O. system is discussed by Rey (1977, 106). A proposed semi-automatic system for the long stretches of rural track in Australia is discussed by Forsythe, Milroy and Thomas (op. cit.), and the use of digital processors in railway automation generally is discussed by Milroy (1980, 91). A paper by Suderburg (1977, 118) reports on a fine system developed in Sweden for maintaining tractive effort right on the limit of adhesion for each axle under snow conditions; this paper has particular relevance in the context of the optimal driving strategy discussed in Chapter 5 of this thesis. Developments in Canada, Italy, Switzerland and the Union of Soviet Socialist Republics are reported respectively by Pelletier (1977, 102), Solimini (1977, 119), Bogdan et al. (1976, 21) and Koulayev (1978, 77).

Having reviewed the literature discussed above, and a number of less substantial technical articles which touch on the subject of A.T.O. and are listed in References, it is now possible to draw up a list of the factors which influence the answer to the question "Why automate railways?"

2.3 Summary of reasons for railway automation

No special significance should be attached to the order of the list which follows, except to note that in the minds of railway staff (and probably passengers) there is a broad ranking of priorities. Safety factors come first, then the need for good timekeeping, then the need to keep costs down, and finally the need to satisfy what a railwayman regards as 'operational niceties', such as the desirability of conserving energy.
The following factors have influenced railways in their progress towards automation of the train driving function:

a) Over 50% of fatalities in railway accidents result from drivers passing through signals at red (Andrews, op cit.). Only about 20% of the total number of accidents result from this cause, but they are the accidents which kill most people.

b) A further 23% (approximately) of fatalities in railway accidents result from signalman error. The most common human factor is a lapse of short-term memory, after allowing unusual train movements during "permissive working".

c) Especially on busy railways (i.e. on routes running close to maximum capacity), there is a need to control groups of trains; for example at a merging junction it is necessary to create a gap in a convoy of trains on the main line, to enable a train from the branch line to enter without any trains being stopped (which is disruptive of traffic). Pearson (1973, 14) has clearly demonstrated the inadequacy of the fixed-block signalling system for this type of fine control.

On less busy railways such as the single-line mineral runs in Australia, it is equally important to control time of arrival at passing loops, to allow heavily-laden trains to keep moving. The extra cost in fuel alone of having to restart one typical Australian National Railways coal train is now in excess of $A 160, or £75.

d) In order to be able to control groups of trains, it is necessary to be able to control the speeds and positions of individual trains on the network.

e) Drivers, especially on high speed trains and in bad weather, are placed under undue stress trying to detect conventional wayside signals.
f) When disturbances to the working timetable occur, it is desirable to be able to correct these disturbances in such a way as to minimise their propagation through the network. To achieve this, it is helpful if railway staff in control centres are able to rapidly issue a corrected timetable for each train; this correction requires to be translated into appropriate velocity or acceleration commands for the traction system - a problem which it is difficult for a human driver to solve intuitively in an energy-efficient way.

g) Since 1973 governments, and to a much lesser extent railway managements, have become increasingly concerned with the need to conserve energy, especially in the case of diesel locomotion or the use of electricity derived from oil-fired power stations.

In the author's experience, energy conservation is not taken seriously by railway staff if it results in reduced operational capability; this thesis therefore considers ways in which energy savings may be made without that disadvantage (for example by using timetable slack more efficiently).

h) Staff costs account for up to 75% of total costs in a typical railway (Maxwell, op cit.). Automation allows for reduced staffing levels; the social impact of this factor, and the response of unions and management to that impact, is a fascinating study of great interest to this author, but outside the technical scope of this thesis.

i) A theme which runs through many of the more recent papers in the A.T.C. field is the suggestion that improved railway automation allows the achievement of higher route capacity and/or shorter journey times without new major capital works or environmental damage.
j) When convoys of trains on urban transit systems are considered, there is a basic tendency to instability in that a train which is late is subject to two positive-feedback factors which increase its lateness even more. One is the presence of an increased number of passengers waiting for it at each station, and the second is the fact that without A.T.O. its increased mass causes it to accelerate more slowly for the same tractive effort. With early trains the reverse tends to occur. (This probably accounts for the phenomenon of bunching of surface buses, but there is no published work to support this view.) Automatic control of acceleration, braking and time-keeping tend to prevent bunching.

k) It has been suggested (Forsythe et al., op.cit.), that one of the key requirements for an automated railway system is to be able to implement the results of operations research packages in real time on the railway. One such package in the British Rail J.O.T. (Junction Optimisation Technique). Traction and signalling engineers in Australia have endorsed a suggestion by this author that the most effective operational procedure between control centres and trains is for the control centre to transmit future target times and velocities to the trains, for locations on the track which they have not yet reached. This allows a train-borne system to achieve those targets for each train, which reduces the communications traffic between control centres and trains, and lessens the disruption should any one processor go down. Provision should also be made for the central system to determine the positions and velocities of individual trains on interrogation.

l) Several authorities consider that an automatic speed-holding system is desirable for long-haul trains. (Such systems are common in
France). However, there is a countervailing opinion among some management staff that if the driver's role is reduced to a machine-tending one, his vigilance suffers; this opinion is supported by some psychologists and ergonomists (Andrews, op cit.).

The author's response to this dilemma is to suggest that the driver should remain charged with the prime responsibility for obeying the safety commands of the signalling system, with an automatic back-up such as the British Rail A.W.S. (automatic warning system) should he lapse in this task. This approach is being followed in a semi-automatic project now being worked up in Australia. However, doubts persist as to whether these responsibilities will maintain a driver's vigilance if he spends long periods without doing anything physical.

Some operators, for example Garnier (1979, 28) report reduced wear on rolling stock after A.T.O. is introduced, especially reduced incidence of motor overheating and wheel flats.

In considering the factors above, the Office of the Director-General of Transport in South Australia, in consultation with this author, identified the question of energy conservation as a factor which loomed large in the pre-occupations of Government (especially as all South Australian rail locomotives burn diesel fuel, an increasingly scarce resource). At the same time there was virtually no commitment from railway managements to accord priority to energy conservation if it meant a degradation of service. Research funds were therefore provided to the author to investigate the possibility of using modern digital control technology to promote energy conservation without degradation of service. This thesis reports on appropriate aspects of that work.
We now turn our attention to a review of published work which discusses the specific question of energy consumption on rail journeys.

2.4 Energy Consumption of rail vehicles

In this section we consider relevant material that discusses the energy consumption of rail vehicles, and briefly make comparisons with other modes of transport to set the question in perspective. Although the energy consumption of rail vehicles has been measured and numerically quantified, little work has been done outside academic institutions on the control of vehicles so as to minimise or at least reduce energy consumption, subject to operational constraints. As mentioned previously, most of the academic work has not been applicable to conventional rail vehicles but to novel forms of transport system that are not in service.

Fels (1975, 41) provides comparative energy costs of urban transportation systems of all modes. Her analysis includes direct operating energy consumption, and the energy used in construction amortised over the typical operating life of the vehicle. Table 1 summarises the energy consumption in kWh/(passenger mile) of the various modes for a typical American city of one million people.

<table>
<thead>
<tr>
<th>Transport Mode</th>
<th>Large Auto</th>
<th>Compact Auto</th>
<th>City Bus</th>
<th>Rapid Rail</th>
<th>P.R.T.</th>
<th>Motor Bicycle</th>
<th>Bicycle</th>
<th>Walk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy Consumption</td>
<td>1.9</td>
<td>1.0</td>
<td>0.7</td>
<td>0.7</td>
<td>2.6</td>
<td>0.6</td>
<td>0.1</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Table 1

Energy Consumption (kWh/(Passenger Mile) for various Transport Modes.
Surprisingly, the figures for a city bus system and a rapid rail transit system are similar. This is accounted for by the much heavier mass of the rapid rail system per passenger carried, and the higher average speed (or reduced journey time) associated with the rapid rail system. The amortised energy used during construction is approximately the same in each case, and can be neglected.

Superficially, these results and other published material suggest that the scope for making savings of energy in the case of urban rail transit systems is comparable to that for buses. However, the relative lack of conflicting traffic in a dedicated railway system compared with the near-chaotic situation which occurs when buses share busy city streets with other traffic, make the railway system a more practicable alternative for the application of deterministic optimal control theory. (That is not to say that scope does not also exist for reducing the fuel consumption of urban road vehicles - but such work falls outside the ambit of the author's project).

Turning our attention to the longer-haul freight runs, Baumgerther (1979, 16) established that the energy consumption of rail freight transport, in units of J/(tonne km), is typically one seventh that of road transport, and one twentieth that of air freight.

In practice virtually all rail journey sections are part of a multi-modal journey involving road transport at each end, and the cost and inconvenience of changing mode at each end of the rail sector frequently more than outweigh any notional advantage arising from the better energy-efficiency of the rail mode. It may be that improvements in passenger and freight handling arrangements at road/rail transfer points are the most effective contribution which could be made to
encourage long-haul traffic back to the rail mode, and hence to reduce the total energy consumption in transport.

Bone (1974, 22) contradicts this view in discussing the impact of energy-conservation principles on transportation planning in the U.S.A., and suggests that the main contribution can be made by improving train efficiency on the long haul runs, and by employing operations research techniques to reduce the number of empty freight wagons which are hauled around the country. Packages such as T.O.P.S. (originally developed by the Southern Pacific Railroad Company in California, U.S.A. but subsequently adopted by many organisations including British Rail) can have an impact on the second part of Bone's recipe for improved energy-conservation in freight transport; it is hoped that the work discussed in this thesis may find application in the first part (improving long-haul efficiency).

Mitchell (1977, 93) provides some useful quantitative information on the energy consumption of rail vehicles in the United Kingdom gleaned from various sources. The energy conservation argument for applying modern technology (including railway automation) with a view to persuading freight agents to switch from road to rail is reinforced by his statement that whereas 17% of the world's freight traffic is transported by rail using 4% of the transport energy, 16% is road-hauled by vehicles using 78% of the energy. Of the remaining 67%, less than one tenth is hauled by air freight and the remainder by sea. The advantage of rail over road in the long-haul mode is somewhat exaggerated by these figures; the road-haul energy figure is loaded by the existence of a great deal of short-distance road delivery work in and around cities and rail-heads, which is inherently
inefficient but for which no practicable rail alternative can be envisaged.

Mitchell also analyses the overall energy efficiency of diesel locomotion as compared with electric locomotion, and finds that if one takes into consideration the entire energy chain, including the energy used in refining crude oil and the thermal efficiency of power stations, the two energy efficiencies are similar (22% for electric locomotion, 24% for diesel locomotion).

The significance of the minor difference in efficiency is dwarfed by the fact that electric rail locomotion can use as its prime source of energy the relatively abundant supplies of coal, hydro-electric or nuclear energy, or any of the solar-derived forms of energy which may in future be harnessed to the electric power supply, whereas diesel locomotives require scarce and costly oil.

Considering inter-city passenger traffic, the energy advantage of rail is relatively less great than in the case of freight. This is largely because of the great mass of railway passenger rolling stock. Whereas the passenger payload is 35% of the total mass for a typical road bus it is only 11% in the case of a conventional intercity rail coach. A significant amount of rail's inherent energy advantage is therefore dissipated in hauling around (and frequently having to slow down) large masses of metal. British Rail have recognised this problem, and in the new Advanced Passenger Train the passenger payload is more than 22% of the total mass, and would be close to 35% if the first-class seats, air-conditioning, catering and toilet facilities were removed.

Howard (1979, 63) draws attention to the trade-off between energy
consumption and journey time, which is considered in some detail later in the thesis. On level track with a 2 km spacing between stations, a typical British Rail commuter train can, for a mere 10% extension of journey time, reduce its energy consumption by a staggering 32%; these figures are, however, only achievable if the train is controlled in an optimal way. Without computational assistance (or track-side coasting boards, which are so inflexible that their use has been discontinued in most railway organisations), human drivers tend not to use timetable slack in an effective way; for example, they frequently drive as fast as possible to the next station, then waste slack by waiting until the guard blows the whistle to proceed, thus saving no energy at all.

Garnier and Riodel (1979, 44) show even higher relative energy savings for small reductions in journey time. They also present some surprising figures to justify automatic braking systems in urban rapid transit. A loss of five seconds per braking operation on the Milan subway (and most human drivers do far worse than this through cautiously braking too early) requires up to 20% more energy for driving faster elsewhere in order to maintain the timetable. Unnecessary standing at stations has the same effect - the Japanese are conscious of this and their station-stop times on the Tokyo subway are only half those in London or Paris.

The question of regenerative braking is frequency raised. It is generally agreed that with current (1980) technology the cost and mass of the additional equipment required for regenerative braking do not pay off in terms of saved energy, except in the case of urban rapid-transit systems with high acceleration and retardation rates,
closely spaced stations, and an electricity distribution system which accepts regenerated energy. British Rail estimate that on their most intensive urban service they would achieve a reduction of energy consumption of less than 10% by fitting regenerative braking systems. Vogelsang (1979,131) suggests that for an urban metro system the savings can be as high as 25%, and the current trend is for regenerative chopper-controlled systems to be installed in new metro rolling stock now coming into service. In underground systems, there is an additional energy advantage in that the ventilation fans are not required to remove so much hot air from the tunnels. In this context Hoang et al. (1975, 58) and many other authors have pointed out that, by locating stations on humps, decelerating trains can store some of their kinetic energy as potential energy as they climb towards the station; they then recover it as they accelerate away downhill. Unfortunately, when this concept was tried on the newly completed Montreal subway, it was found that all the hot air in the tunnels tended to accumulate at the raised stations, requiring an increased input of energy to the ventilation fans (especially in summer) which more than offset the gravitational advantage of the tunnel's trajectory. However, the first generation of Montreal subway cars are not regenerative, and it may be that the trade-off will alter once regenerative braking systems are fitted to new generations of cars.

Turning to the simulation of rail vehicles, it is apparent that little attention was paid to the energy implications of various longitudinal control strategies until the impact of oil price rises in 1973. Two analogue-computer train performance calculators (as they were called) which yielded estimates of energy consumption on rail journeys are
described by Hartree (1938, 55) and Murray-Smith (1973, 95). Whereas the former used a very early electro-mechanical differential analyser, the latter exploited a modern hybrid computer with integrated-circuit operational amplifiers.

More recently, digital simulation packages have become available for this work and are used by many railways. A typical digital train performance package is described by Hargreaves (1975, 54).

From the fore-going review of literature concerning energy consumption of rail vehicles, it is clear that if journey times can be tinkered with, considerable scope exists for the application of control theory to achieve substantial savings of energy. In 1979 Lynch (1979, 85), who was then Chairman of the Australian National Energy Advisory Committee, announced that in view of the fact that transportation accounts for 26% of Australia's energy consumption, and that nearly all the fuel so consumed is scarce and costly fuel oil, the goal "Energy-Management in Transport" was to be one of five key objectives of his Committee. It is hoped, therefore, that this thesis is timely.

Having looked at A.T.C. in general, and at aspects of the energy consumption of rail vehicles, we now turn our attention to some of the algorithms that have been implemented by people applying digital computers in the field of automatic train control.

2.5 Algorithms and Software

Software aspects of the application of digital processors to railway A.T.C. tasks are now reviewed.
On the A.T.P. side (often loosely referred to as railway signalling) the use of digital processors in redundant configurations to implement the required logic is now being evaluated in several European countries, and commercial hardware and software packages are offered by such firms as L.M. Erickson (Sweden) and Siemens (W. Germany).

The provision of hardware redundancy to achieve the required reliability is a well-established science, but the problem of software reliability remains a vexed question. If all the processors in a redundant configuration run the same software, then there is no software redundancy. Can one test all A.T.P. software so thoroughly that it is possible to guarantee the absence of any programming error which can lead to a "wrong-side" failure? The author's opinion is that even today (1980) this is a very dubious proposition except in the case of very simple well-structured programmes which have a finite number of different logical paths through them.

The author's prognostication is that this difficulty will shortly be resolved by the provision of multi-processor railway signalling systems in which each redundant set of, say, three processors is dedicated to a very simple logical function, (like the individual dedicated electro-mechanical relays which these systems will replace). It will then be possible to test for both hardware and software integrity of each module. It may be that the individual processor modules will be akin to programmable controllers or programmable logic arrays, and may not incorporate the general-purpose microprocessors with which such organisations as the Research Divisions of British Rail and London Transport currently seem to be pre-occupied.
Another worry with redundant digital signalling logic is that a momentary interference 'spike' (of the type which thyristor-controlled and chopper-controlled locomotives are still inclined to produce) may cause an identical fault or error in all the sections of a redundant system, with the result that the error is not detected.

Forsythe, Marshall and Thomas, in a joint British Rail/Loughborough University of Technology project, are developing an interesting variant of a triple-redundant microprocessor system in which the software (including the redundant 'voting' software) is "time-skewed". A spike hitting all three systems will cause different errors, which cannot fail to be detected. This work will be relevant to A.T.O. and A.T.P. applications, and will be reported elsewhere. Similar work specifically for the A.T.P. application is reported by Cribbens et al. (1978, op cit.).

Other work on the application of digital software to A.T.P. is reported by Nock (1970, 96), Cardani (1977, 32), Mathson (1977, 139), Taylor (1978, 123) and Gelbstein (1980, 48). Of these applications only Taylor's is discussed in the context of A.T.O. as well as A.T.P. On the subject of algorithms and software for A.T.O., it is difficult to find a great deal of relevant work reported in the literature, except in vague terms. This is probably because there is a commercial advantage in keeping good train operation algorithms confidential.

Shirai and Istuhara (1968, 115) describe some of the A.T.O. algorithms developed for the Teito urban rapid transit system. Although implemented in analog form at the time, there are indications that the same control philosophy has carried through into more modern Japanese practice in the
field of digital A.T.O., as described by Yasukawa (1978, 138) and Yagi (1978, 137) for intercity and commuter trains respectively.

Savage (1969, 113) and Pearson (1973, 101) describe what are in effect higher-level "operations-research" techniques for scheduling trains through critical parts of the network. Savage deals with junctions, and Pearson deals also with the scheduling of fast and slow flights of trains on single track (with overtaking loops). It has become clear in recent years that such higher-level control strategies require an effective A.T.O. capability on individual trains in order to be implemented satisfactorily.

Brown (1973, 27), Lang (1975, 81) and Handa (1977, 137) all describe interacting merge control algorithms for novel types of automated guideway transit system under development in the U.S.A., but their work is inapplicable to the author's research because their control algorithms are incompatible with the use of a conventional fixed-block railway signalling system.

Hinman and Pitts (1975, 57) discuss practical headway limitations which are an important feature of control algorithms for P.R.T. systems, and Thomas and Hopkinson (1976, 124) work from the headway limitation imposed by a conventional fixed-block railway signalling system, towards control algorithms which can be used by automatic vehicles operating on such a system. This work was done in the context of the British Rail AUTOWAGON project, a proposal to allow automatic self-propelled freight vehicles to work their way around the British Rail network in between the scheduled traffic. When this work was shelved, to be replaced by the less 'way-out' T.A.C.T. (totally automatic control of trains) and B.R.A.T.O. (British Rail
Automatic Train Operation) research and development projects (Hopkinson and Ashworth, 1977, 78), the relevant ideas from the AUTOWAGON work were carried over into these projects.

Kerr (1977, 73) discusses an interesting approach to the problem of controlling slack action between the couplings in long freight trains. Waves of coupling action travel up and down the train due to changes of tractive effort and the effects of gradient. At certain nodal points very high forces recur, resulting in the danger of broken couplings from fatigue failure. This is a distributed non-linear dynamic problem, handled as a set of difference equations in distance and a set of differential equations in time. This non-linear problem is not amenable to analytic solution, but simulation work by Kerr and his colleagues has clearly shown that if the driver can be given more information about the waves of force travelling up and down his train, he can readily learn to adjust tractive effort accordingly.

This raises the question of whether an adaptive control system could be designed to do the same task. Such a system would not be incompatible with the control strategy discussed in this thesis, for it would result in minor and relatively frequent changes of tractive effort, around mean levels which can be determined from time-keeping and energy-consumption considerations. However, subsequent to Kerr's work, the experience of colleagues of the author on the Mount Newman iron ore railway in Western Australia has indicated that it is uneconomic to provide the required instrumentation transducers at every coupling of a freight train. The author also has doubts as to whether it is possible to devise a suitable criterion of "goodness" for the complex distributed dynamics of this system. Such a criterion
would of course be needed for use in whatever adaptive or self-tuning control strategy is employed.

Nonetheless, the work of Kerr and his colleagues has resulted in a driver-training simulator which has already proved most useful in enabling train drivers to get a better 'feel' for the complex dynamics involved in handling long freight trains. The author has himself used this simulator, and has observed experienced train drivers learning on it. One is left with a strong impression that the well-trained and well-motivated human operator, in a complex ill-defined situation, employs learning stratagems which are far more subtle and effective than the automatic self-tuning algorithms pioneered by such workers as Astrom and Clarke. (Of course there are industrial tasks which self-tuning algorithms will handle excellently, but the control of long freight trains is not yet one of them).

We have touched on a topic (self-tuning control) which is in the front line of the evolutionary thrust of control theory, and which shows signs of breaking through into widespread industrial application (a claim which cannot be sustained for the work of most other modern control theorists since the days of Bode, Nyquist, Nicholls and Walter Evans all of whom completed their main work more than thirty years ago).

We now consider in the train control context other aspects of what is loosely and perhaps erroneously referred to as modern control theory, derived as it is from the pioneering work of historical figures such as Gauss and Lyapunov.

Published applications of optimal control theory to the A.T.O. problem concentrate on the method of dynamic programming and the application of Pontryagin's Maximum Principle. These methods are lucidly explained in text books by Bellman and Dreyfus (1962, 19) and Pontryagin et al
(1965, 105), in which the original papers by the same authors are quoted and interpreted.

Rumsey and Powner (1973, 111) apply the principle of dynamic programming to a cell-following system, in which trains of vehicles are controlled so as to stay within moving cells (each separated by a safe distance from its neighbours) which travel along the route. Unfortunately the existence of non-linear dynamics is neglected, as are practical constraints on the values of state and control variables. The all-important question of selecting appropriate weighting factors for the components of the linear quadratic performance index is not considered worthy of comment (it seldom is). These features mar an otherwise interesting paper in which the application of optimal control theory to an unrealistic transportation system is clearly described.

Further work on the application of dynamic programming to A.T.O., and of other techniques such as linear and quadratic programming, is being carried out by colleagues of the author at the South Australian Institute of Technology, and will be reported elsewhere. In this thesis the optimal control theory which is applied is derived from the work of Pontryagin.

Pontryagin et al. (1956, 104) set out the basis of Pontryagin's Maximum Principle, in which the problem of maximising a functional is dealt with by augmenting the system state equations to include the behaviour of an additional set of variables (the co-state variables) and then maximising a simple functional (the Hamiltonian) which is derived from the original criterion of performance augmented by functions of the state and co-state variables. This procedure is
discussed further in Chapter 5.

Lee (1963, 85) discusses the circumstances under which Pontryagin's principle is applicable in the presence of non-linearities. In the case of the author's approach to the train control problem the journey is broken up into journey sections which are sufficiently short to ensure that the effect of the most troublesome non-linearity (gradient) does not invalidate the principle.

Strobel et al. (1974, 120) describe an attempt to apply Pontryagin's principle to the control of a model railway at the Dresden College of Transport. The problem of determining the initial values of the co-state variables to ensure correct values of the train's end-state (position, velocity and time) is somewhat glossed over in this paper, and realistic constraints on the allowable accelerative tractive effort as a function of velocity are not incorporated. An energy criterion similar to that used by the present author is discussed in this paper.

Janjanin (1979, 67) also applies Pontryagin's Maximum Principle in a specialised train control application, which is the problem of the optimum control of trains so that each of them passes through a junction area in as short a time as possible (thereby increasing the capacity of the junction). He discusses the difficulty of solving for initial values of co-state variables in order to determine the correct moments of switching between modes of control. In the Janjanin scheme the state equations are solved in real time, and the co-state equations in reverse time. The problem of sensitivity of the train's final state to small errors in the initial values of co-state variables is not discussed. The Janjanin approach appears to complement
effectively the Junction Optimisation Technique operation research package described by Savage (1969, 113) for scheduling trains through junctions without trying to minimise their time of passage through the junction area. Burrow and Thomas (1976, 30) have subsequently discussed the performance of this package in practice.

It is one thing to determine what the optimum longitudinal control for a train should be in terms of acceleration or tractive effort. It is another to produce a system which will implement that control, especially during the braking phase when the dynamics of the pneumatic or electro-pneumatic braking system can cause a destabilising effect in what is essentially a deterministic single-input feedback system. Small variations in train resistance and traction or braking system performance inevitably cause velocity and acceleration errors, and these must also be corrected for as the journey proceeds.

For the early analog train control systems a conventional three-term control strategy was adopted successfully, and has been carried through into the digital control philosophy of recent developments. The British Rail three-term control approach is discussed anonymously in (1975, 3) and also by Hopkinson et al. (1977, 61) who arrived at settings for proportional, derivative and integral control coefficients by trial and error. Part of the author's task in connection with the British Rail work has been to provide an analytical basis for the selection of these parameters, and this is done later in the thesis. The nature of braking system dynamics is such that conventional two-term and three-term control approaches are adequate to achieve the desired transient and steady-state performance (see Chapters 6 to 9) and the constraints of industrial life are such that practical railway
organisations will not adopt other approaches until they have to. It is appropriate, however, to be aware of more refined digital control algorithms (which do not all follow in the direct line of succession from the early days of pneumatic analog three-term controllers). A number of interesting single-input single-output algorithms are reported by Auslander et al. (1978, 13) and Bristol (1977, 25) as well as in many recent textbooks. In implementing such algorithms using digital processors the author has adopted with success some recursive digital filtering algorithms first brought to his attention by Gold and Rader (1969, 59). Bristol (1977, 25) also discusses some interesting variations on standard algorithms, that are particularly suitable for microprocessor implementation.

References to other papers in the field of railway automation are made at appropriate points later in the thesis. The reader is reminded that an alphabetical list of references appears at the end of the thesis.

2.6 Conclusion

It is clear that from the time in 1973 when political and economic upheavals forced governments and transport planners to consider the energy-consumption implications of their work, there have been a number of attempts to quantify the energy consumption of rail vehicles, but little attempt by rail operators to apply the principles of control theory in an attempt to minimise it.

In this thesis we therefore extract, from what Cannon (op.cit.) has called the "haze of extraneousness and complexity" which surrounds the field of automation of land transport, two important and related aspects of the energy minimisation problem. One involves the determination
of rail timetables so as to minimise energy consumption, subject to a variety of operational constraints which the timetable planner might wish to take into account, and the other involves the automatic or semi-automatic driving of a train so that it does, in practice, closely follow the optimal timetable previously determined off-line.
CHAPTER 3

TIME-KEEPING CONTROL OF AN AUTOMATICALLY DRIVEN TRAIN

Glossary

$x$  current train position  
$V$  current train velocity  
$t$  current clock-time  
$x_T$  target position  
$V_T$  target velocity  
$t_T$  target clock-time  
$x', V'$  predicted coordinates of any point  
$t_T'$  predicted time of arrival at target  
$t_V'$  predicted time to traverse constant-velocity section of profile  
$t_b'$  predicted time to traverse constant-deceleration section of profile  
$t_j'$  remaining journey-time (predicted)  
$t_e$  time-keeping error  
$|t_a|$  allowable fluctuation of $t_e$  
$x_b$  braking point  
$x_d$  distance-to-go to target  
$V_r$  velocity reference for autodrive

INTRODUCTION

In this part of the study a system was developed for extending the capability of the automatically driven train developed by British Rail (B.R.), so as to include automatic time-keeping.

The AUTODRIVE package evaluated by B.R. in rail-car GEMINI has the capability of driving a train from one target point to another
Figure 3.1. Typical Desired Speed Profile.

Table 1 Data Needed by Train Control System

<table>
<thead>
<tr>
<th>INPUT DATA</th>
<th>TRAIN STATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_p$ = Maximum Permitted Speed (m/sec)</td>
<td>$x(t)$ = train position (m)</td>
</tr>
<tr>
<td>B = Permitted Retardation (m/s$^2$)</td>
<td>$v(t)$ = train velocity (m/sec)</td>
</tr>
<tr>
<td>$X_T$ = Target position (m)</td>
<td>$t$ = clock-time now (hrs, mins, secs)</td>
</tr>
<tr>
<td>$V_T$ = Target velocity (m/sec)</td>
<td></td>
</tr>
<tr>
<td>$T_T$ = Target clock-time (hrs, mins, secs)</td>
<td></td>
</tr>
</tbody>
</table>
while following a typical speed/distance profile shown in Figure 3.1. The train arrives at its target with the correct velocity (zero in the case of a stopping-point), and does so in such a way that constraints (speed restrictions and permitted retardation limits) are satisfied throughout its journey. The existing AUTODRIVE package is discussed in more detail in Section 3.6.

A typical service run (such as the outer-suburban passenger service using electrical multiple units which was proposed for the next phase of British Rail AUTODRIVE trials) consists of a series of targets, some of which are stations. On the approach to each target, data concerning the next target section is communicated to the train, to enable the desired speed profile for the next target section to be constructed. The AUTODRIVE system then follows the new profile, after reaching the current target.

The author's contribution to the British Rail work on automatic train control was to develop algorithms for the on-board digital system, to enable the train to adjust its desired speed profile in such a way that it recovers from delays as quickly as possible when running late, and reduces its average running speed so as to conserve energy when running early. To achieve this, target time \( t_T \) is added to the set of target parameters to be achieved by the train.

### 3.1. Data needed by automatic time-keeping system

The set of target data listed in table 3.1. is the minimum external data set needed for an AUTODRIVE system with automatic time-keeping
capability; it was agreed with B.R. that this set would be used for the author's studies.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Item of Target Data</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_T$</td>
<td>Target position</td>
<td>m</td>
</tr>
<tr>
<td>$V_T$</td>
<td>Target velocity</td>
<td>m/sec</td>
</tr>
<tr>
<td>$t_T$</td>
<td>Target arrival time</td>
<td>sec</td>
</tr>
<tr>
<td>$B$</td>
<td>Deceleration on approach</td>
<td>m/sec²</td>
</tr>
<tr>
<td>$V_p$</td>
<td>Max. permitted velocity</td>
<td>m/sec</td>
</tr>
</tbody>
</table>

Table 3.2. Target data set transmitted to or stored on train.

This target information, together with up-to-date information about the position and velocity of the train, and the current clock-time, complete the information needed by such a system. That this set of data is sufficient for automatic time-keeping of a train with AUTODRIVE capability has been confirmed by discussion with railway engineers in Britain and Australia, and demonstrated by simulation.

3.2. Reasons for automatic time-keeping

Why is an automatic time-keeping capability desirable (perhaps necessary) for an AUTODRIVE train in passenger-carrying service? The answers to this question are listed below.

(a) trains in service are required to adhere (if possible) to the working time-table as modified by the current running plan. Human drivers check their time-keeping 'en route' and take remedial action when necessary.
It will be necessary (or at least desirable) for AUTODRIVE trains to do the same.

(b) If automatic time-keeping is not incorporated, the only scope for making up time will be shortening the station-stop time. If sufficient latitude is built into the working time-table to allow for this, then trains running normally, and on time, will be spending an unnecessary amount of time standing at stations. This is not energy-efficient.

(c) To allow scope for recovering from unscheduled delays, the working time-table must be planned so that average running speeds are less than the permitted maximum speeds.

A train which is automatically driven should run at the preset average running speed only when it is on time. When late, it should be driven "flat out" within the constraints, so as to make up time as quickly as possible.

When early, it should be slowed to a lower average running speed to conserve energy. It should run as slowly as possible consistent with arriving at the next target point "on time".

(d) The incorporation of "target time" in the input data set simplifies the problem of specifying the "desired speed profile" (d.s.p.) for the train-borne system.

Effectively, the task of constructing the d.s.p. in detail is devolved from a central system to the train-borne control system, thus lessening the data-transfer requirement. An adaptive time-keeping algorithm is not possible if the d.s.p. is over-specified by the off-train system.
(e) In the longer term, a system of trains with precise time-keeping control systems allows great scope for those responsible for the smooth flow of traffic in the system as a whole, to regulate traffic in a precise way merely by altering the target-times of certain trains.

For example, gaps can be generated in a convoy of trains approaching a merge point, merely by simultaneously adjusting the target times of the trains behind the required gaps.

Many studies (for example Pearson 1974, Pitt 1975) have shown that the present system of fixed-block signalling (with which B.R. and the Australian railways will be operating in the foreseeable future) is an excellent safety system, but its quantisation makes it a poor instrument for precise regulation of trains in complex traffic situations.

Automatic time-keeping will allow B.R. and other railway organisations to make experiments with more precise schemes for regulating the flow of traffic, but always within the safety envelope of the fixed-block signalling system.

Thus the signalling system can eventually be reserved solely for its prime and overriding function of ensuring system safety, with correction of target-time being used to implement more finely-quantised train-operating tactics within the signalling system's safety envelope.

3.3. The concept of a "supervisory" system for time-keeping control

Various possibilities have been considered by the author for incorporating "target arrival time" in addition to "target velocity" and "target position" as controlled variables.
The most attractive possibility is one which is simple enough to be implemented on-board the train, preferably without the need for additional hardware from that which has already been developed by British Rail and other railway organisations for conventional AUTODRIVE running.

The control problem is considerably simplified once it is recognised that time-keeping corrections to the "set-point" of the AUTODRIVE system need only be made intermittently, rather than continuously. This enables the concept of a "supervisory" control system to be investigated, in which the AUTODRIVE'S velocity control system operates normally most of the time, with corrections to its set-point (desired velocity) being made only when a substantial time-keeping error has accrued.

3.4. A Predictor-Corrector Strategy

Consideration of the time-keeping behaviour of human drivers led to the concept of a predictor-corrector strategy of supervisory control, in which a predictor algorithm regularly computes the estimated time of arrival (E.T.A.) at the next target. If the E.T.A. indicates a time-keeping error greater than a fixed "allowable zone of error" then the corrector algorithm intervenes and give the AUTODRIVE system a command to accelerate or decelerate, until the predictor algorithm again produces a time-keeping error which is at the centre of the allowable zone of error. The AUTODRIVE then receives a new velocity-reference signal (the current velocity) to drive to. The block diagram of Figure 3.3 illustrates this approach.
Figure 3.3. Proposed Automatic Time-keeping System
Before examining the requirements of the predictor and corrector algorithms in more detail, it is necessary to look at the behaviour of the AUTODRIVE system more thoroughly.

### 3.5. The B.R. Autodrive System

AUTODRIVE experiments carried out by British Rail have demonstrated that, given the position of the next target and the velocity to be attained at that target, the train can be driven automatically in order to satisfy those parameters (within an acceptable tolerance); throughout the approach to the target the velocity, deceleration and jerk are constrained so that pre-set levels are not exceeded. These limits depend on track conditions (including gradient, adhesion etc.), train characteristics, and passenger-comfort considerations. (Hollingberry 1979).

The AUTODRIVE system currently installed on the research train GEMINI achieves the target position and velocity parameters, while satisfying the constraints, by following a velocity/position profile such as that shown in Figure 3.1.

Phase 1 of the profile is an open-loop acceleration phase. Progress is determined by the sequencing equipment already installed in the traction system of the train, so that the train accelerates smoothly away from its previous target. (If the previous target velocity is equal to the current speed restriction, which is frequently the case, then the acceleration phase is omitted). Phase 1 is open-loop in the sense that the AUTODRIVE system does not intervene except to round-out the velocity smoothly into Phase 2 as the speed limit is approached.
Phase 2 of the profile is the constant velocity phase. The AUTODRIVE system drives the train according to a driving tactic which has driving and coasting sub-phases (and light brake applications in the event of over-speed due to downhill gradients). During Phase 2 minor fluctuations of velocity are permitted, so as to avoid excessively frequent changes of sub-phase. The permissible fluctuations of velocity can be varied depending on the train's state, according to an optimum driving strategy which is either stored in or computed by the AUTODRIVE system. So far as time-keeping calculations are concerned the train is sensibly proceeding at constant velocity during Phase 2.

On approaching the pre-computed braking point the AUTODRIVE systems makes a smooth (jerk-limited) transition onto the braking curve of Phase 3. Phase 3 is a constant-deceleration phase (parabolic on the velocity-distance profile but linear on a velocity-time profile), which brings the train to its next target at the correct velocity. Once again, a "rounding-out" mode is employed so that jerk and acceleration are zero (or very small) at the target. In Figure 3.1. the jerk-limited transitions are not shown. The effects of jerk-limiting on time-keeping performance are negligible, and are not included in the analysis.

To summarise, the AUTODRIVE system is, in principle, a closed loop system which keeps track of the train's position and continuously corrects the velocity $V$ so that at any point the velocity is equal to that required by the desired speed profile (with a small tolerance). In conventional control systems block-diagram form it may be represented by the block diagram of Figure 3.4.
The box labelled AUTODRIVE is the heart of the control system. Hardware and software are already well-developed and documented elsewhere (Thomas 1980, Ashford 1977).

The algorithms developed by the author for automatic time-keeping, and (later in the thesis) for energy-minimization, exploit the existence of AUTODRIVE systems such as that of Figure 3.4., which are capable of maintaining the velocity of the train close to that determined by the reference input $V_r$.

### 3.6. Prediction algorithm

The control strategy developed for automatic time-keeping involves the prediction of time-of-arrival at a target (normally a station stop, signal at red or a speed restriction), from a knowledge of the current train state, the system parameters, and the target data.

In general, the time $t_j$ taken to reach a target along any profile from the current train state to the target state may be computed from
the formula
\[ t_j = \int_{x}^{x'} \left( \frac{1}{v'} \right) dx', \]
where \( x \) is current position, \( x' \) is any future position, \( v' \) is any future velocity

While this formula will cope with any predetermined profile (not necessarily an analytic one) to the target, it involves a numerical integration algorithm. For the B.R.A.T.O. plot scheme (Altrincham-Wilmslow outer-surburban journey) the on-board processor will lack the speed to carry out this iterative numerical computation quickly enough; however this approach is one that should be borne in mind for implementation if and when it becomes economically justifiable to instal more powerful computing hardware on the train.

The author's approach to automatic time-keeping takes advantage of the known general form of the velocity profile produced by the British Rail AUTODRIVE system (and by other similar systems in use in Germany, France and Japan). With systems of this type it may be assumed that, once the open-loop acceleration phase is completed, the AUTODRIVE will ensure that all sections of the journey profile from the current state to the target will be either:

(a) constant velocity curves, or
(b) constant deceleration curves

In practice, the autodriver, for sound reasons, permits small fluctuations of velocity above and below these idealised sections of the journey profile, but simulation and train test results show that the final time-keeping errors resulting from these fluctuations are quite negligible in the context of the agreed specification for time-keeping tolerance (+10 seconds to zero seconds).
The reference velocity of the AUTODRIVE system may therefore be used as an estimate of the mean velocity for time-keeping purposes, once the open-loop acceleration phase of a journey has been completed.

The advantage of this approach (in the context of this chapter) is that the prediction algorithm takes the form of a set of simple algebraic formulae which have an analytic solution, and therefore the use of a time-consuming numerical integration algorithm is avoided.

Later in the thesis, when it is shown that for minimum-energy control a coasting phase should in general be inserted before the braking phase, the time-keeping algorithm now under discussion is retained; if for any reason (such as an unforeseen disturbance to running or an inaccuracy in the system model used for determining the optimum control trajectory) the prediction algorithm indicates that the train is going to arrive late at its target, then the optimal control trajectory is abandoned at the appropriate moment and the train is driven under automatic time-keeping control for the remainder of that journey section, so that it arrives at the next target on time.

The algebraic prediction algorithm is now discussed. The predicted clock-time at the target \((t_T')\) is given by

\[
t_T' = t + t'_v + t'_b
\]

where \(t = \) clock-time now
\(t'_v = \) predicted constant-velocity time
\(t'_b = \) predicted constant-deceleration time.

The constant-deceleration time \((t'_b)\) is given by

\[
t'_b = (v_r - v_T)/B
\]
where \( v_r \) = current reference velocity

\( v_T \) = target velocity

\( B \) = constant deceleration value

In order to predict the constant-velocity time we first calculate the braking point \( (x_b) \), which is the distance at which the transition from constant-velocity to constant-deceleration should be made

\[
x_b = x_T - (v_r + v_T) \frac{t'_b}{2}
\]

The predicted constant-velocity time is now given by

\[
t_v' = (x_b - x)/v_r
\]

where \( x \) = current position

The prediction algorithm, which is repeated once every computing cycle, uses the above equations in the order 3.3., 3.4., 3.5., 3.2., and yields the estimated time of arrival at the target \( (t'_T) \).

3.7. Correction Algorithm

We now discuss the software which was developed for deducing the time-keeping error \( (t'_T - t_T) \), and using it to adjust the speed-holding velocity reference so as to reduce the time-keeping error to within the ten-second tolerance at the braking point.

To summarise it, the permissible time-keeping error is allowed to fluctuate between limits \((\pm t_a)\) corresponding to typical autodrive velocity fluctuations without any intervention. But if the time-keeping error \( (t_e) \) exceeds the allowable limit \((-t_a)\) on the 'fast' side, then the corrector issues on over-riding COAST command, so that the train eventually slows to a speed
which will get it to the target on time, at which point the
autodrive is given a new velocity-reference \( (v_r) \) corresponding
to the current velocity, and the over-riding \textit{COAST} command is
cancelled. If, on the other hand, the limit of time-keeping
error is exceeded on the "too slow" or "late" side, then an
overriding "\textit{DRIVE}" command is issued, until train speed again
picks up to a value which will get the train to the target on
time.

Good performance is obtained when the logical strategy makes a
decision for each condition represented by one of three current
train control modes (coasting, driving or speed-holding) and
one of four time-keeping error-zones (too early, early half of
allowable zone, late half of allowable zone, and too late).
There are thus twelve combinations to consider, and these twelve
are shown in the four columns and the first three rows of the \( 4 \times 4 \)
Karnaugh map of Figure 3.5. The four bits (A and B for time-
keeping zone, C and D for control mode) have been carefully chosen
to represent real physical conditions so that the four unused combinations
in the bottom row represent logical impossibilities.

These four combinations can thus be regarded as "don't care" from
the logic point of view, and ones or zeros can be allocated to some
of them in the Karnaugh map, to allow simplification of the resultant
Boolean statements.

There are three possible outcomes of the logical decision (coast,
hold this velocity reference, or drive) and therefore two bits P and Q
are necessary to determine the outcome, which is a command to the
AUTODRIVE system. The P and Q values for each condition are shown at the intersection of the relevant row and column in Figure 3.4.

The bit allocation is as follows:-

**Time-keeping**
- \( A = 1 \) means "early". \( A = 0 \) means "late"
- \( B = 1 \) means "very". \( B = 0 \) means "slightly"
("very" means "outside the tolerable zone")

**Command time-keeping**
- \( C = 1 \) means "coasting"
- \( D = 1 \) means "driving"
- \( C = D = 0 \) means "following AUTODRIVE"

**Outcome commands**
- \( P = 1 \) means "coast"
- \( Q = 1 \) means "drive"
- \( P = Q = 0 \) means "follow AUTODRIVE"

These bit allocations were chosen to ensure that only one bit changes at a time during the progress of the train. If there is a change of command bits P or Q during one computing cycle, the corresponding C or D bit is altered for the next computing cycle.

The presence of hysteresis in the system (the allowable zone of time-keeping error before remedial action is taken) ensure that interruptions to the normal operation of the AUTODRIVE system are made only intermittently. If no disturbances to running occur, the train remains within its allowable zone of time-keeping error throughout a journey section, and only one value of reference velocity is used for the speed-holding phase.

The bit allocations for P and Q in Figure 3.5. are chosen to ensure that no remedial action is taken while the train remains within its allowable zone of time-keeping error. Once it goes outside that
zone the appropriate over-riding command to the AUTODRIVE system
drives it back to the centre of the zone, before reverting to
AUTODRIVE with a new velocity reference.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>A</th>
<th>B</th>
<th>A</th>
<th>B</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early (very)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Early (slightly)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Late (slightly)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Late (very)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>CD</th>
<th>Driving</th>
<th></th>
<th>Speed holding</th>
<th></th>
<th>Coasting</th>
<th></th>
<th>Not poss.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>01</td>
<td>1 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 1</td>
</tr>
<tr>
<td></td>
<td>1 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 1</td>
</tr>
<tr>
<td></td>
<td>1 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 1</td>
</tr>
<tr>
<td></td>
<td>1 1</td>
<td>X X</td>
<td>X X</td>
<td>X X</td>
<td>X X</td>
<td>X X</td>
<td>X X</td>
<td>X X</td>
</tr>
</tbody>
</table>

P is left bit    Q is right bit.

Figure 3.5. Truth Table for P and Q.

We now look for simple Boolean expressions for P and Q, which will give
the required truth table. The sparseness of ones (eight ones compared
with sixteen zeros) suggests that a more economical Boolean expression
will be found by looking for adjacent ones and using AND statements,
rather than by looking for adjacent zeros.

To make the derivation of Boolean statements clearer, separate maps are
presented in Figure 3.6. for bits P and Q. In each case the "don't
care" bits are allocated so as to simplify the Boolean expression.
Using normal KARNAUGH map simplification techniques, the Boolean expressions given in Figure 3.6. were derived, and all possible combinations checked to ensure that no dangerous or absurd results arose from the allocation of arbitrary values to the DON'T CARE bits.

Interpreting the simplified Boolean maps in English, the COAST command is given (P bit set) when the train is either very early, or slightly early and already coasting. The DRIVE command is given (Q bit set), when the train is either very late, or slightly late and already in DRIVE mode. In each case the system switches to the FOLLOW AUTODRIVE mode, with the current speed as the velocity reference, when it reaches the velocity corresponding to the centre of the allowable time-keeping band. The FOLLOW AUTODRIVE mode is held until further disturbances to running (if any) cause the time-keeping error to accrue until it is once again outside the allowable zone of error.
The near-symmetry and simplicity of the two maps, and the fact that the "don't care" bits have sensible allocations, give confidence that this design is as economical as can reasonably be expected.

The corresponding Boolean statements are

\[ P = A(B + C) \] \hspace{1cm} 3.6
\[ Q = \overline{A}(B + D) \] \hspace{1cm} 3.7
\[ C_i = P_{i-1} \] \hspace{1cm} 3.8
\[ D_i = Q_{i-1} \] \hspace{1cm} 3.9

It should be noted that the practical implementation of the logic for \( P \) and \( Q \) with a microprocessor can be either by storing the Karnaugh map in 12 locations of a PROM or ROM (Figure 3.7), or by means of logical decisions in an assembly-language or high-level program (Figure 3.8). While the latter is intellectually more elegant, the former is faster at run-time. The gate circuit used in the simulation is also given in Figure 3.9.

\[
\begin{array}{cccccc}
A & B & C & D & P & Q \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 1 & 0 & 1 \\
0 & 1 & 1 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 1 & 1 & 0 \\
1 & 1 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 0 & 0
\end{array}
\]

Figure 3.7. Look-up table for \( P \) and \( Q \).
3.8. Sensitivity of predicted time-keeping error to velocity disturbances.

In the previous section the velocity reference for the autodrive ($V_r$) has been used as an estimate of mean velocity for prediction purposes. This has been shown to be valid both by simulation and actual running, when the reference velocity is held constant most of the time and
fluctuations of actual velocity around the reference velocity are small.

Even in cases (such as up-hill running of rail-car Gemini with a flat battery) when the true velocity has been persistently less than the velocity reference, so that the estimate of mean velocity has tended to be high, the time-keeping algorithm has been shown to be robust, in the sense that as the train falls behind time the reference is edged upwards, raising the average velocity and so compensating for the error in a manner akin to integral action.

The time-keeping algorithm, with velocity reference \( (V_r) \) as the estimate of mean velocity for time-keeping purposes, is therefore satisfactory in the role for which it was originally designed, namely as a supervisory system for occasionally altering the velocity set-point of an AUTODRIVE system in the speed-holding mode.

Later in the project, when journey profiles incorporating a coasting phase were being investigated, it was considered desirable to retain the time-keeping algorithm even during coasting, so that in the event of a prediction of lateness at the next target, coasting would be abandoned at the appropriate velocity and a short period of constant-speed running inserted in the profile to correct the time-keeping error.

There being no velocity reference for the autodrive during coasting, actual velocity was used for the time-keeping algorithm. Effectively the time-keeping system may be thought of as asking the question "If the current velocity were to be used as a reference for the autodrive from this point until the braking trajectory is encountered,
would the train arrive at its target late?"

During coasting (from an upper velocity previously determined off-line) the answer to the above question will normally be NO. As soon as it becomes YES, speed-holding is initiated so as to ensure an on-time arrival at the target. A typical complete profile over such a journey section would be as shown in Figure 3.9.

![Figure 3.9. Typical profile of train journey]  

Figure 3.10. Typical profile with speed-holding sub-phase.

At point $x_s$ the decision to revert from coasting to speed-holding is made.

It was found that, even during coasting, minor fluctuations of measured velocity occurred; these showed up as variations of predicted journey time ($t'_j$), whose magnitude is dependent on velocity and distance-to-go ($x_d$). Over long target sections, and/or at excessively low speeds, simulation results showed that these fluctuations in predicted journey time could be sufficiently large to result in excessively frequent intervention by the time-keeping algorithm.

It was therefore proposed that the tolerance of allowable time-keeping error ($\pm 5$ seconds at a target) could be relaxed under
conditions which led to excessively frequent intervention of the
time-keeping algorithm, provided that the error could be brought
down to within the final tolerance well before the braking point
was reached.

To explore this proposition a sensitivity analysis was carried
out.

Let \( S \) seconds/(m/sec) be the sensitivity of the predicted journey
time \((t'_{j})\) to small perturbations of velocity \((v)\) at any point \((x)\).
Assuming for the moment that all targets are stopping targets, we
obtain

\[
t'_{j} = \frac{v}{2B} + \frac{x_{d}}{v}
\]

where \( x_{d} \) is distance-to-go \((x_{T} - x)\)
\( B \) is deceleration on approach to target.

\[
S = \frac{\partial t'_{j}}{\partial v} \bigg|_{x_{d}} = \frac{1}{2B} - \frac{x_{d}}{v^{2}} \quad 3.11
\]

Equation 3.11 confirms that the sensitivity of the predicted time-
of-arrival \((t'_{j})\) to small variations of velocity is high at low
speeds and when \( x_{d} \) is large. It is normally negative.

Figure 3.10 shows a graph of predicted journey-time \((t'_{j})\) against
velocity \((v)\) for a range of journey distances. Superimposed on
the graph are isoclines of constant sensitivity \( S \), showing that
the values vary from zero to \(-20\) seconds/(metre/sec) for the range
of journeys under consideration. It should be noted from equation
3.11 that sensitivity \( S \) is always negative, its value reaching zero
at the braking point given by

\[
x_{d} = \frac{v^{2}}{2B} \text{ metres} \quad 3.12
\]
Time remaining (secs)

\[ t_j = \frac{v}{2B} + \frac{x_T}{v} \] seconds

\( (B = 0.6 \text{ m/sec})^2 \)

Sensitivity curves are isoclines of constant

\[ \frac{\partial t_j}{\partial V} = \left[ \frac{1}{2B} \frac{x_T}{V^2} \right] \]
Beyond this point no further corrections are possible. The isocline equations are obtained by eliminating the variable \(x_d\) from equations 3.10 and 3.11, yielding

\[
t_j = (1/B - S)V \text{ seconds}
\]

Normal AUTODRIVE running on GEMINI showed that fluctuations of measured velocity likely to cause undesirably frequent changes of time-keeping error do not exceed \(\pm 1\) m/sec. Satisfactory performance of the time-keeping algorithm will therefore be obtained with the time-keeping tolerance \((t_a)\) set at

\[
|t_a| = 5 \text{ seconds for } S > 5 \\
|t_a| = 5 \text{ seconds for } S < 5
\]

Figure 3.12 shows the general form of the graph of allowable time-keeping error \((t_a)\) plotted against distance \((x)\), assuming that mean velocity does not change during the journey.

From equation 3.11 the slope of the converging boundaries of the allowable zone are \(\pm (1/v^2)\), and the parallel portions commence at distance \(x_p\), given by

\[
(x_T - x_p)/v^2 = (1/2B) + 5
\]

Typically, with \(x_T\) at 3 Km, \(V = 10\) m/sec, \(B = 0.8\) m/sec\(^2\), \(x_p = 2438\) m. Thus, even with this low velocity and long journey-section, the funnel shaped graph of Figure 3.12 allows time-keeping errors to be reduced to within the final tolerance of \(\pm 5\) seconds well before the braking point is reached.

![Figure 3.12 Allowable time-keeping error as a function of distance.](image)
The funnel-shaped graph of Figure 3.12 turns out in practice to have a further operational advantage. Once the parallel-sided portion of the funnel is reached (and this is several hundred metres before the braking point for all the journey sections under consideration) the sensitivity of time-keeping error to small variations of velocity continues to reduce, while the allowable time-keeping error remains constant. The effect of this is that intervention by the time-keeping algorithm (alteration of the velocity reference) becomes less and less likely after point $x_p$ is passed, and in practice such intervention is unlikely to occur. (Changes of velocity reference late in the journey are undesirable from a passenger-comfort point of view, and have little effect on arrival time anyway).

To summarise this section, the sensitivity of time-keeping errors to normal fluctuations of velocity has been analysed, and found to be a function of velocity and distance-to-go. A formula has been developed for varying the allowable time-keeping error as the journey proceeds; as a result the time-keeping algorithm performs satisfactorily for the full range of journey sections under consideration, and corrections to the velocity-reference of the AUTODRIVE system do not occur close to the braking point.

3.9. Simulation of Time-keeping Algorithm

In order to demonstrate the proposed system for automatic time-keeping, and to study energy-consumption implications of two alternative types of journey profiles, a model of the proposed system was set up on the Membrain Digital Differential Analyser in the Department of Electrical and Electronic Engineering at Loughborough University. A modified version of this model was later set up on the EAI 580 analog computer
at the S.A. Institute of Technology, and used for the optimal control study discussed in Chapter 4, as well as for other work in the field of automatic train control. Results from this model will be presented in K. Tyler's thesis.

3.9.1. Description of the Model

The dynamics of the train were simulated, using parameters which had previously been determined from characterisation trials with specific rail vehicles.

At the heart of the model are two integrators set up to represent the non-linear longitudinal dynamics equation for the rail vehicle.

\[ T(v) = M' x + R(v) x + Mgsin\theta(x) \]  \hspace{1cm} (3.17)

where

- \( M' \) = effective mass of vehicle (Kg)
- \( M \) = true mass of vehicle (Kg)
- \( g \) = acceleration due to gravity (m/sec\(^2\))
- \( sin\theta(x) \) = gradient (uphill positive)
- \( T(v) \) = tractive effort (a function of velocity) (N)
- \( R(v) \) = train resistance (a function of velocity) (N)

The non-linear functions for tractive effort and train resistance (functions of velocity) and gradient (a function of distance) were stored in function generators. In the case of resistance the function used was the Davis formula.

\[ R(v) = K_0 + K_1v + K_2v^2 \]  \hspace{1cm} (3.18)

where the three \( K \) coefficients were determined by fitting a quadratic curve to the resistance/velocity graph determined during characterisation tests of the rail vehicle. (Parkin 1977).
The ramping output of an integrator with constant gain was used for the time variable \(t\). Time-keeping equations 3.2., 3.3. and 3.5. were incorporated in the model, as was the equation for the braking point (3.4.). Thus the estimated time of arrival \((t'_T)\) was available as the simulated journey proceeded.

The logic for the corrector model was incorporated using gates (Figure 3.8) and appropriately connected comparators for determining the degree of earliness or lateness, taking into account the variable tolerance on time-keeping error discussed in Section 3.8.

Additional comparators were incorporated for detecting the maximum permitted velocity \((V_p)\), and detecting when the braking point \((x_b)\) is reached for entry into the braking phase.

Scaling coefficients were chosen to enable the model to be exercised over a range of journey profiles. These included typical test runs that might be carried out on the Mickleover test track, as well as on the journey profiles which would occur on the Altrincham - Manchester route of the B.R.A.T.O. pilot scheme.

The model at the S.A. Institute of Technology has also been used for a study of proposed semi-automatic driving systems for a new urban light rail system under development by the South Australian State Transport Authority.

### 3.9.2. Types of journey profile simulated

Logic was incorporated into the simulation so that two different types of journey profile could be investigated (A and B in Figure 3.13).
Both types of journey have three distinct phases. In the type A journey the acceleration phase is followed by an essentially constant-speed phase, which is continued until the train reaches the point where it should brake with constant deceleration to its target.

The type B journey differs only in respect of the central phase. Acceleration is continued until a predetermined point, after which the train is permitted to coast until it either:

(a) encounters the braking point as before.

(b) reaches a speed below which the time-keeping algorithm indicates that it will be late at its target, in which case it reverts to a speed-holding mode.

The ability to incorporate type B journey profiles in the model was incorporated as a modification late in the project because, as will be seen in the chapter on optimal control, Pontryagin's Maximum Principle indicates that a type B profile is in fact the best profile for minimum energy consumption.
In both the type A and type B profiles a comparator was incorporated to limit the velocity to the maximum permitted velocity \( (V_p) \) over any section of track.

In both the type A and type B models it was assumed that an autodriver existed which was capable, during the speed-holding and braking phases, of holding the velocity of the train to within ± 0.5 metre/second at all times. Work on Gemini, reported by P.D. Thomas (1980) had already demonstrated that this was feasible. (Although minor fluctuations of velocity occur during automatic running of a real train, the velocity reference to the autodriver is an excellent estimate of the mean velocity).

Provision was also made for up to three gradient changes during a target section to be incorporated if required.

3.9.3. Summary of results from the model

The time-keeping algorithms have been tested over a range of distances from 500 m to 5 Km, and a range of target times corresponding to average speeds over the journey section of from 10 to 40 m/sec. In all cases, with profiles of type A and type B, the simulated trains arrived within ± 3 seconds of the specified target arrival times, except where target arrival times were so early that they were less than the minimum-time profile (subject to velocity and acceleration constraints) could achieve. In the latter cases the trains automatically followed the minimum time profile subject to the constraints.

When the simulation of the time-keeping algorithms had been carried to the stage where there was sufficient confidence in the time-keeping control system to justify its implementation on the British
Rail test train, the model was modified for use in connection with the minimization of tractive energy. This work is reported in later chapters. The model has also been used to demonstrate energy-saving manual driving practices to railway engineers in South Australia. See the last section of Chapter 5 for a summary of these practices.

3.10. Application of predictor-corrector algorithms

The time-keeping algorithms discussed in this chapter have been successfully applied (in slightly modified form) to the British Rail test train since the author left Derby to return to Australia. They have also been used in a semi-automatic system under development at the S.A. Institute of Technology, in which application the logical outputs P and Q have been employed to put up appropriate driver's displays. These two applications will be reported in the theses of P.D. Thomas and K. Tyler respectively.

3.11. Chapter Summary

In this chapter a description has been given of the work carried out by the author in designing and developing new control algorithms to enable an automatically driven train to arrive at a target position not only at the correct velocity, but also at the correct time.

Satisfactory predictor and corrector algorithms have been developed. These were designed to complement and exploit the already-existing hardware and software developed for the British Rail Automatic Train Operation pilot scheme. They should also be applicable to a wide range of other automatic and semi-automatic trains in service in various parts of the
world. The algorithms operate by correcting the velocity-reference
to the AUTODRIVE system when time-keeping errors accrue beyond a value
which is itself a function of velocity and distance-to-go. A
sensitivity analysis has been carried out to determine what this
function should be.

The validity of the work has been demonstrated by simulation and by
limited test running to be reported by P.D. Thomas (1980).

In subsequent chapters these algorithms will be modified so as to
incorporate the optimal control of velocity and acceleration as
determined by Pontryagin's maximum principle, while maintaining
satisfactory time-keeping performance.
CHAPTER FOUR

PARTITION OF THE TRAIN CONTROL PROBLEM

INTRODUCTION

In Chapter 1 it was seen that the various command and control systems in a railway organisation may be considered in the context of a hierarchical structure (Milroy, 1980).

The basic objective of the research work described in this thesis was to develop a system which would enable trains to be driven so that they satisfy timetable and other operational constraints in an energy-efficient manner. The two main functions of the control hierarchy which can strongly influence energy-consumption are, firstly, the timetable planning function and secondly, the on-board control of traction and braking.

The basic timetable planning function is carried out, not in real time, by staff in railway headquarters. Minor modifications to the working timetable may be made by staff in control centres, in order to recover from traffic disturbances. Many factors need to be taken into account in planning timetables, including such matters as traffic forecasts, rolling stock availability and performance, and staff rosters. These factors are discussed in more detail by Milroy (1980), Williams and Henry (1975), Yoshihisa (1974), and Iida and Koga (1974). The procedures discussed by these workers are more in the realm of operations research than of control engineering. However, the output from a timetable planner is in the form of a table of key points along the journey (hereinafter referred to as target points) and of times at which these targets are to be reached (target times).
There are usually several or many possible timetables which will satisfy the operational constraints, and in Section 4.2, we discuss the broad principles of a computer program which will enable a timetable planner to select the most energy-efficient set of target times for the key points on the journey, from the range of options available. This program does not need to run in real-time.

The on-board control of traction and braking is a real-time task, and in a later chapter we discuss the broad principles of algorithms which will run in real-time on a train-borne digital processor; this system either advises the driver, or operates automatic sub-systems on the train, so that the train conforms with the optimal timetable previously determined off-line, and does so in a way which minimises energy consumption over each journey section, subject to over-riding safety constraints and the requirement to arrive at the next target on-time.

In arriving at this partition between the tasks to be handled by the off-line time-table planning routines (which minimise energy over complete journeys consisting of a number of sections) and the tasks to be handled by the on-train processor (which considers only one section at a time), advice was taken from railway engineers in the South Australian State Transport Authority (Rail Division), the Australian National Railways, and the Melbourne Underground Rail Loop Authority.

The partition of the tasks takes into account the relative computing power of the main-frame computers available for timetable planning, compared with that of the eight-bit microprocessors (Motorola M6800 or Intel 8085) which were candidates for the train-borne role...
when this work was carried out in 1978. This partition of tasks also minimises the amount of data which needs to be transmitted between control centres and trains, when changes to the working timetable are made by control centre staff in order to recover from traffic disturbances.

The two separate functions (time-table planning and on-board control) are now considered in more detail.

4.1. Timetable Planning Program

In this section we discuss a program developed by the author for producing the optimal timetable discussed in the previous section. In addition to producing a table of target positions and the times at which they are to be reached, it also produces target velocities and a list of points on the journey at which coasting is to be initiated if the train is running on-time. (This data is required for the on-train system).

4.1.1. Inputs to the Timetable Planning Program

The following procedure is carried out by the timetable planning person, to generate the data needed by the program. It may be assumed that approximate timings for each section of the journey are known from experience; some of the target times will be fixed by operational requirements, and others are 'free'.

(a) The journey is broken into sections, and at the end of each section a target point is specified. Target points may be:
  Signals possibly at red
  Stopping points (e.g. stations)
  Commencements of new speed restrictions (i.e. changes of permissible maximum speed $V_p$).
Negative changes of gradient (i.e. points where the grade becomes less uphill or more downhill).

The first three types of target arise from the nature of railway operations. The fourth arises from the nature of the optimal control algorithms used in the program. It will be seen in Chapter 5 that the normal drive-coast-brake sequence for optimal control may not occur if negative changes of gradient occur within a journey section.

(b) Target velocities are specified. These are the velocities which the train should have on arrival at each target. Target velocities are zero in the case of stopping points and signals possibly at red. In the case of changes of speed restriction, target velocities are equal to whichever is lower of the old and new permitted maximum speeds ($V_p$).

It has been verified by simulation that in all cases except where the train is able to run without traction throughout the new journey-section, energy is wasted by slowing the train below the permitted maximum speed at a target point - the train has to be run faster in both the preceding and following sections to maintain time. The special case is handled by checking whether a train can omit the acceleration phase through a journey section; if so, the target at the commencement of that section is cancelled and the two sections treated as one. (It will be clear from the discussion of the optimum control algorithm in Chapter 5 that journey sections may be merged provided the
sequence drive - coast - brake for optimal control does not have to be broken at any point).

(c) Target times are allotted for each target. Some target times are firm constraints and are specified to satisfy operational requirements (e.g. the need to arrive on-time at merging junctions, or at passing loops on single-line railways). Other target times are 'free'; for these, initial estimates or guesses may be made on the basis of the planning person's experience. The program will automatically adjust these times to values which minimise the energy consumption over the total journey.

It should be noted that the program allows the human time-table planner to investigate the energy-consumption implications of variations to planned timetables. For example, he may wish to relax the total journey-time by a few percent, to determine the consequent energy-savings. The program will automatically allocate the new timetable 'slack' to those journey sections which produce the largest resultant energy-saving.

(d) The permitted maximum speed ($V_p$) over each journey section is specified. This will not necessarily be the same as the preceding or following target speed (it may be higher than both). This is an upper constraint on velocity during a journey section rather than at a target, and will be discussed again in Chapter 5.
In addition to the above data, which is fed in by the timetable planning person, the program requires the following additional data which will normally be available on file, as it is likely to be fixed for months or years on end.

(e) The gradient at each point on the track. Normally the gradient is constant for a considerable distance along the track, and the information may therefore be stored as a table of distances at which gradients change, and new values of gradient.

(f) The coefficients in the Davis resistance formula for the class of train. The Davis formula for the decelerative force caused by running resistance is

\[ F_r = a + bv + cv^2 \]  \hspace{1cm} (4.1.)

On most routes the coefficients \( a, b \) and \( c \) are approximately constant except at very low speeds. On routes with much curvature the values of these coefficients may vary significantly with distance. If the variations are significant they may be stored in the form of a table, as in the case of gradient. Significant increases of the coefficients also occur in tunnels.

(g) The tractive effort of the traction system when it is accelerating as hard as possible subject to the constraints. The tractive effort is normally a function of velocity, and may also be stored as a look-up table.

(h) The normal service deceleration rate. (This should be as large as practicable subject to the need to avoid wheel-slip or wheel-slide under conditions of low adhesion).
Typical values are 0.8 m/sec\(^2\) for commuter trains, 1.1 m/sec\(^2\) for rapid transit stock with hanging-strap for standing passengers. (Gebherd 1972).

(i) The mass and effective mass (including rotational inertia) of the train. Mean expected values are used.

It should be noted that there will be stochastic variations of the parameters mentioned in sub-sections (f) (g) and (i) above which will be imposed as a result of variations of passenger or freight load, side-winds, and variations of traction system performance due to voltage changes or state of maintenance.

It is a well known principle of stochastic control theory that the best that can be done in such circumstances is to estimate the mean values of the variable parameters, and then to design on the basis of those mean values. Such estimates should be made by the timetable planner for each journey section.

Railway engineers consulted by the author have suggested that the variance in all of these parameters should be quite small, except possibly in the case of train mass for the light rail vehicle under study, which can vary between nineteen and twenty-three tonnes depending on passenger load. Once the vehicles are in service the accuracy of estimation of passenger load should improve markedly when diurnal traffic patterns become established and are measured.

It will be clear from the following sections that the above data will enable the optimal control algorithms to be applied over the complete journey. At the heart of the algorithms is the train performance equation which will allow the state of the train to be updated once per computing cycle

\[ M'\ddot{x} = F - R(\dot{x}) - Mgsin\theta(x) \]  

\((4.2\).

\]
where \( M' \) = effective mass (Kg)
\( M \) = mass (Kg)
\( F \) = tractive force (N)
\( R \) = train resistance (N)
\( g \) = acceleration due to gravity (m/sec\(^2\))
\( \sin \theta \) = gradient (positive uphill)

4.1.2. **Function of the Time-table Planning Program**

The time-table planning program goes through a number of possible journey profiles in sequence searching for the optimal profile. In going through this sequence a variation of Bellman's Principle of Optimality is exploited (Bellman 1958). This may be restated in the context of this problem as follows:

"If a complete journey profile over several sections is to be optimal, then the journey profile over each individual section must also be optimal".

This statement, which appears almost trite when expressed in this way, leads almost directly to the principle of dynamic programming, (see for example Prime 1969 for a particularly clear exposition of the method). Dynamic programming methods are being investigated by colleagues of the author in a rail context, but in this thesis they are passed over in favour of a less general method which is more concise for this particular problem. The method exploits the principle of optimality as expressed above, and also makes use of the following two principles which have been shown by simulation to be true for all the journey sections under consideration.

(a) The journey sections have been chosen so that, within each section, Pontryagin's Maximum Principle may be applied to
determine the optimal profile and the optimum control of demanded acceleration over the journey section. See Chapter 5 for a complete analysis. The routine (referred to as MINENG hereafter) developed in Chapter 5 for solving the optimum control problem over a single journey section is used both in the on-board real-time control system discussed later in the thesis, and (as a sub-routine) in the timetable planning program now under discussion.

(b) The free parameters, which must be varied in the search for the optimal control over the complete journey, are the 'free' target times \( t_T \) previously discussed. (By agreement with railway engineers consulted by the author, the quantisation of target time is ten seconds - greater precision of arrival time at targets is deemed to be an unrealistic requirement).

For all the journey sections under consideration, the relationship between section journey time \( t_j \) and minimum energy consumption \( E_m \) as determined by the MINENG subroutine is a monotonic one, as shown in Figure 4.1.

![Figure 4.1. Relationship between minimum energy \( E_m \) and section journey time \( t_j \).](image-url)
The gradient \( \frac{dE_m}{dt_j} \) is always negative, and moreover the magnitude of the gradient decreases steadily as the section journey time is increased.

It is the latter property which is exploited in this particular timetable planning program, enabling it to converge more rapidly and efficiently to the optimal timetable than would a dynamic-programming or other program which did not exploit this particular property.

The sequence followed by the program in progressively adjusting the free journey-section times is now described.

(a) The complete journey consisting of \( n \) sections as previously defined, is broken into \( m \) supersections, where

\[
m < n
\]

\( (4.3.) \)

A supersection consists of one or more sections, and each of the targets whose target-times are fixed is the end-point of a supersection. The supersections are considered one at a time. The target times for target points within a super-section are, by definition, all free.

(b) For the first supersection, subroutine MINENG is run using the initial estimates of free target times for each constituent section. The resulting minimum-energy values for each section are recorded.

All section journey-times are then increased by ten seconds, and the changes in minimum-energy values are recorded.
The section journey-times are then reduced by twenty seconds (i.e. by ten seconds from the initial values), and the changes in minimum-energy values are again recorded.

The section which records the largest saving of energy for a ten-second relaxation (lengthening) of journey-time is then selected, as is the section which records the least increase of energy for a ten-second tautening (reduction) of journey-time. The free target times for each of these sections are then adjusted accordingly, so that the total journey-time over the supersection is the same as before, but the total energy has been reduced as much as possible. This trade-off between pairs of free journey-section times within the first supersection is repeated until no significant further reductions in energy can be achieved. An important feature of this method is that at each iteration the reduction in total energy over the supersection is less, so that the decision to stop iterating can be clear cut and indeed automatic.

(c) Once the first supersection has had its free target-times 'relaxed' to the optimum values, the process is repeated with each other supersection until the total energy-consumption over the complete journey is minimised. All the target-times are then recorded, to be stored on or transmitted to the train-borne system, together with any other data required by that system. (See Chapter VI).
It will be clear to the reader that initially the timetable planner should 'fix' as few target times as possible, to allow the greatest possible scope for energy-minimisation.

The program developed using these principles has been tested over the journey under consideration using a variety of initial estimates of free target times. In all cases the solution converges steadily to the same optimum situation where the partial derivative \( \partial E_m / \partial t_j \) is close to the same value for each sub-section within a supersection. This is in keeping with the results expected from a knowledge of the variational calculus of Euler and Lagrange. In addition, the algorithm leads to a relaxation of those journey-sections which are close to the minimum-time situation (where the curve of Figure 4.1. is almost vertical) and to a tightening of those journey-sections for which the timings appear to be most 'slack' (to use a railwayman's jargon).

Thus we may be confident, both from the practical point of view of the railwayman and the academic point of view of the optimal control theorist, that we have arrived at a practicable and useful approach to the production of minimum-energy time-tables, despite the non-linear nature of the system dynamics. The algorithms are robust in the presence of uphill and downhill gradients, and therefore represent a significant advance on other work reported in the literature.

4.1.3. Outputs from the Time-Table Planning Program

The following data may be required by the train-borne system. It is output from the time-table planning program in the form
of a table of target and section data, followed by a list of additional general data.

For each target point:

- target position \( (x_T) \)
- target velocity \( (V_T) \)
- *target arrival time \( (t_T) \)
- target deceleration \( (B_T) \)
- estimated effective mass \( (M') \)
- coefficients in Davis formula \( (a, b, c) \)

Within each section:

- point at which permitted max. velocity alters \( (x_p) \)
- new value of permitted max. velocity \( (V_p) \)
- points at which gradient changes \( (x_g) \)
- new values of gradient \( (g \sin \theta) \)
- *point at which coasting is to be initiated \( (x_c) \)

General train data:

Max. accelerative tractive effort as a function of velocity \( (F') \)

It may be noted that only those items marked with an asterisk (target arrival times for free targets and coasting-initiation points for on-time trains) are actually provided by the timetable planning program; these items are in fact crucial so far as saving energy is concerned. However, for logistic reasons it has been arranged that the complete set of data required by the train-borne system, including data which was also required as input for the time-table planning program, is output. In discussions with railway engineers it has been agreed that, should the techniques
developed in this thesis be applied in a railway organisation, it will be desirable for a read-only memory cartridge to be programmed from the output of the time-table planning program, for later use on the train. The arrangement of data in the form described above will facilitate this process.

Other data not required by the train-borne system, but which may be required elsewhere in the railway organisation, may also be output if needed. Examples of such output are:

- Points at which braking is to be initiated ($x_b$)
- Velocity-distance profiles for the optimal journey
- Velocity-time profiles for the optimal journey
- Tractive-effort profiles for the optimal journey
- Mechanical energy used over each section.

With respect to the last item (mechanical energy) it should be noted that this refers to mechanical energy transmitted at the wheel-rail interface, i.e. the output of the traction system. The responsibility for ensuring that the traction system itself is as energy-efficient as possible (especially during the periods of maximum acceleration subject to constraints, for it is during these periods that virtually all of the energy is consumed) lies with the rolling stock engineer. Maintenance practices in the workshops will be the chief determinant of this efficiency, and the time-table planning function cannot influence them. However, if estimates of input mechanical energy are required in the future, then these could be obtained from the timetable planning program provided the efficiency of the traction system was known, as a function of velocity and demanded acceleration.
Detailed analysis of the energy-minimisation routine which is at the heart of the time-table planning program will be found in Chapter 5.

4.2. Chapter summary

In this chapter we have identified the two functions (timetable planning, and control of traction/braking) which most influence energy consumption. The time-table planning program, developed by the author, for minimising energy consumption subject to operational constraints, has been described. This program exploits Bellman's Principle of Optimality to minimise the total energy consumption over a number of journey sections, Pontryagin's Maximum Principle to minimise energy over a single journey section, and the monotonic shape of the graph of energy consumption versus journey time to trade off one journey section against another.

In Chapter 5 we will study the application of a Pontryagin's Maximum Principle over a single journey section, and in Chapters 6 and 7 we will consider the train-borne system which must be used to make the planned optimal journeys 'come true'.
CHAPTER 5

APPLICATION OF PONTRYAGIN'S MAXIMUM PRINCIPLE

INTRODUCTION

In this chapter the theorem developed by Pontryagin (1956) is applied in order to determine the optimal control of traction and braking over a journey section, taking into account constraints on velocity, acceleration and journey time.

The theorem was proved for linear systems with constraints by Pontryagin et al (1959); however, during the early sixties it was applied with success to a number of non-linear systems. Lee (1963) and others have determined criteria that may be applied in order to determine whether the non-linearities in a system are sufficiently smooth to guarantee that the application of Pontryagin's theorem will yield the optimal control, and in this application the journey is broken up into sections which satisfy those requirements, especially in the case of the most troublesome non-linearity, gradient.

In this chapter the principle is applied over one journey section only. As seen in Chapter 4, the work may be extended to ensure that complete journeys are optimal by a new method which makes use of Bellman's Principle of Optimality (1962), so that the procedures developed in this chapter can be embedded in the solution of the larger problem.

During the seventies there has been a trend away from the application of Pontryagin's methods. However, the Pontryagin approach leads
to a concise statement of the general form of the optimal control in the case of both minimum-energy and minimum-time control of a train over a single journey-section. Elsewhere in the thesis the author exploits this conciseness in the design of a simple onboard digital controller which changes its criterion from minimum-energy to minimum-time when the train is late with reference to a timetable previously determined off-line.

Pontryagin's principle has been applied to the vehicle-control problem by Strobel et al (1974). His paper makes linearising approximations in dealing with the non-linear running resistance of the train, in order to arrive at an analytical solution for the initial values of the co-state variables. The paper does not consider the question of whether the variations of gradient over a single journey section are acceptable in terms of the Lee criterion.

The method has also been applied recently by Edwards and Eren (1979) to a specific type of electric drive system. Their paper would apply particularly to a rubber-tyred vehicle running at low speed, but they neglect constraints on traction and braking caused by wheel-rail adhesion limitations, and their state-variable model of vehicle dynamics completely ignores non-linear running resistance, which makes it inapplicable to the general rail problem.

In the following analysis running resistance is assumed to conform to the generally-accepted Davis formula for rail vehicles (a quadratic function of velocity), constraints on acceleration and retardation are considered, and in the case of minimum-energy control the criterion to be minimised is mechanical energy transmitted at the wheel-rail interface, so that the results are applicable to any
type of traction drive.

This work gives a clear indication to operators and manufacturers of rail vehicles as to the modes of driving which should be tuned for maximum efficiency, so that journeys run in accordance with the procedures determined in this chapter will also be optimal in terms of fuel consumed or energy drawn from the electricity supply.

5.1. General procedure for applying Pontryagin's Principle

In this section we review the steps to be followed in applying the Pontryagin principle to control systems in general. In the next section (5.2) these steps will be followed for the train control problem.

5.1.1. Step 1.
Form a dynamic model of the system and express it in state variable form:

\[ x_i = f_i (x,u,t) \]

where \( x_i \) is the \( i \)th state variable
\( x \) is the set of \( n \) state variables
\( u \) is the set of \( m \) control variables
\( t \) is time

5.1.2. Step 2.
Determine the criterion of performance to be minimised and express it in the integral form:

\[ J = \int_0^T F(x,u,t) \, dt \]

where \( T \) is the final time.
5.1.3. Step 3.

Form the Hamiltonian function, which is the derivative with respect to time of the criterion $J$, augmented with $n$ terms of the form $(p_i f_i)$, where $p_i$ is known as a co-state variable.

For the sake of tidiness we may define the criterion $J$ as an $(n+1)$th state variable, in which case the Hamiltonian $H$ is given by:

$$H = \sum_{i=1}^{n+1} p_i f_i$$

where $p_{n+1} = -1$

The essential statements of Pontryagin and his co-workers may be written as follow:

(a) The co-state variable dynamics are given by

$$\dot{p}_i = -\frac{\partial H}{\partial x_i}$$

(b) For a large class of problems, for the criterion $J$ to be a minimum, it is a necessary and sufficient condition that the Hamiltonian $H$ be a maximum.

5.1.4. Step 4.

In the case of unconstrained controls $u$, the Hamiltonian may be partially differentiated with respect to each of the controls $u$, and the results set equal to zero. This will produce dynamic equations for the optimal controls which will be identical to those given by the classic calculus of variations method of Euler and Lagrange.

In the case where there are constraints on the controls $u$, the appropriate sequence of controls $u$ may often be determined by inspection of the Hamiltonian function. In most practical problems the optimum values of $u$ lie either on a constraint or at zero.
5.1.5. Step 5.

In the cases of both constrained and unconstrained controls \( u \), the initial values of the co-state variables \( p \) are not known, and must be chosen in order to satisfy the required end-conditions at time \( T \). Except in simple linear problems there is no general analytic solution for the initial values of the co-state variables. They can in principle be determined by iterative means, by running a dynamic model of order \( 2n \), comprising \( n \) differential equations for the state variables and \( n \) differential equations for the co-state variables. The \( n \) initial values of the co-state variables can, in principle, be adjusted so as to satisfy the required end conditions.

In practice, severe problems of convergence may occur, and in the case of the train control problem the final state of the train is unduly sensitive to small errors in the initial value of the co-state variables.

5.1.6. We have now outlined the steps to be followed in applying Pontryagin's principle to a general problem. In the next section these steps are followed for the train control problem. However, when step 5 is reached, the troublesome problem of determining the initial values of co-states is avoided by making use of other railway-oriented techniques to determine the required moments of switching between optimal values of the control \( u \).

5.2. Pontryagin's Method applied to Minimum-Energy Control of a Rail Vehicle over a Journey Section.

Consider a train to be driven over the typical journey section of
5.2.1. Dynamic Model

Newton's law of motion may be applied thus:

\[ M\ddot{x} = F - R(x) - M \cdot g \cdot \sin(x) \]  

(5.1)

where

- \( M \) = mass (Kg)
- \( M' \) = effective inertial mass (including rotating components) (Kg)
- \( R \) = rolling resistance - a function of velocity (N)
- \( g \) = acceleration due to gravity - a function of distance (m/sec^2)
- \( x \) = distance (m)

Equation 5.1 may be written in state variable form.

where

\[ \dot{x}_1 = \frac{F}{M'} - \frac{R}{M'} - \left(\frac{M}{M'}\right)g \cdot \sin(x) \]  

(5.2)

\[ \dot{x}_2 = x_1 \]

where

- \( x_1 \) = velocity \( v \) (m/sec)
- \( x_2 \) = distance \( x \) (m)
For the sake of simplicity and practicality, and to conform with control systems terminology, we rewrite equation 5.2.

\[ x_1 = u - R' - g' \]
\[ x_2 = x_1 \] \quad (5.3)

where \( u \) = demanded acceleration - the control input (m/sec\(^2\))

\( R' \) = decelerative effect of running resistance (m/sec\(^2\))

\( g' \) = decelerative effect of gradient (m/sec\(^2\))

One of the practical reasons for using the demanded acceleration as the control variable, rather than tractive force, is that on most rail systems the normal upper limits on \( u \) (positive in the case of traction and negative in the case of braking) are independent of mass \( M \). This is a result of the adhesion characteristics of steel wheels on steel rails.

Many of the rail vehicles to which this work might be applied in Australia already have, or can be fitted with, automatic acceleration control or automatic retardation control. In both cases the appropriate control input or set-point is the variable \( u \).

5.2.2. Criterion of performance

As explained previously the criterion to be minimised is mechanical energy transmitted, during traction, at the wheel-rail interface.

\[ J = \int_{x_0}^{x_T} F \, dx \quad \text{for } F > 0 \]
\[ = \int_{t_0}^{T} Fx \, dt \] 
\[ \quad (5.4) \]
The alternative form is used because, referring to Section 5.1.2, we note that for Pontryagin's principle to be applied, the criterion $J$ should be expressed in the form

$$J = \int_0^T F'(x, u, t) \, dt$$  \hspace{1cm} (5.5)$$

Equation 5.5 may be expressed in the form of equation 5.6.

$$J = \frac{1}{2} M \int_0^T (u + |u|) x_1 \, dt$$  \hspace{1cm} (5.6)$$

The term $\frac{1}{2} (u + |u|)$ penalises positive, but not negative values of $u$.

Before moving on we note that we have developed a criterion suitable for vehicles with no recovery of energy during braking. (This is so for the vehicles under study, which have friction brakes only). The criterion can readily be modified for regenerative braking, by adding an additional term to the integrand.

5.2.3. The Hamiltonian

The Hamiltonian function, the maximisation of which will be a necessary and sufficient condition for $J$ to be a minimum, is given by

$$H = -F' + \sum_{i=1}^{n} (p_i f_i)$$  \hspace{1cm} (5.7)$$

where $F'$ is the integrand in equation 5.5.

$f_i$ is the function on the right hand side of the $i$th state variable equation 5.3.

$p_i$ is the $i$th co-state variable.

In this case

$$H = -\frac{1}{2} M x_1 (u + |u|) + p_1 (u - R' g') + p_2 x_1$$  \hspace{1cm} (5.8)$$

5.2.4. The co-state variable dynamics

The co-state variables $p_1$ and $p_2$ are given by

$$p_1 = -\frac{dH}{dx_1} = \frac{1}{2} M (u + |u|) + p_1 \frac{dR}{dx_1} - p_2$$  \hspace{1cm} (5.9)$$

$$p_2 = -\frac{dH}{dx_2} = p_1 \frac{dg_1}{dx_2}$$
Note that in the case of constant gradient over the journey section, \( P_2 \) is zero and \( P_2 \) is therefore a constant. In this case equation 5.9. simplifies to

\[
P_1 = \lambda M(u + |u|) + P_1 (b + 2 cx_1) - K
\]

where \( K \) is the constant unknown value of co-state \( P_2 \)

\( b \) and \( c \) are constants in the Davis formula for rail vehicles, which is

\[
R' = a + bx_1 + cx_1^2
\]

Even in the constant-gradient case we are left with the problem of determining the constant \( K \) and the initial value of \( P_1 \) so as to achieve the end-conditions at time \( T \). If gradients vary during the section, and they may provided that the convexity condition of Lee (1963) is not violated, then we require from the civil engineer a statement of how the first derivative of gradient with respect to distance varies during the journey section. This data is not available in railway organisations; for this and the other reasons alluded to in section 5.2. we now abandon further analysis of the behaviour of the co-state variables \( P_1 \) and \( P_2 \), and turn our attention back to the expression for the Hamiltonian (equation 5.8.), and to the constraints which must be applied to the control \( u \).

5.2.5. The constraints on control \( u \)

If \( u \) were unconstrained, we could apply variational methods as discussed in section 5.1.4., and determine the optimal control \( u_{opt} \) from equation 5.12, which would be true throughout the journey.

\[
\frac{\delta H}{\delta u_{opt}} = 0
\]

where \( u_{opt} \) is the optimal control, and \( H \) is the Hamiltonian of equation 5.8.
We shall see by the end of this section that in fact $u_{\text{opt}}$ lies on one of the constraints for part of the journey at least. (In practice, because railway timetables normally have only a small amount of slack in them, $u_{\text{opt}}$ is on one of the constraints corresponding to minimum-time or 'flat-out' driving for a high proportion of the time). Under these circumstances the effects of small variations from the optimal trajectory cannot be assessed and equation 5.12 cannot be used.

What is the general form of the constraints on $u$? For a large class of electric, diesel-electric and diesel-hydraulic rail vehicles, $u$ is constrained to remain within the shaded area on a diagram similar to that of Figure 5.2.

![Figure 5.2: Velocity-dependent constraints on $u$.](image)

In normal operation the horizontal parts of the boundary correspond to service limits of tractive and braking effort, which ensure that wheel-slip or slide does not occur even under poor condition of adhesion. The approximately hyperbolic section of the boundary results from the fact that eventually tractive effort must be reduced as speed builds up, to avoid exceeding the power rating of the prime mover. (In high-speed trains, a similar hyperbolic section may also apply for negative $u$, to avoid overheating of the brakes). The vertical parts of the boundary relate to the fact that the train
should not be allowed to go astern, nor should it exceed the line-speed limit which is laid down for that section of the track. We may assume that the driver, or automatic sub-systems on the train, will ensure that the control does not pass outside the shaded area in Figure 5.2. The boundaries of this area may change with mass $M$, and with distance along the track, but we may assume that they are fixed during any one journey-section. In particular, if maximum tractive effort is called for when the train is already running at permitted maximum speed $V_p$, then the constraints are such that the acceleration $x_1$ is zero, i.e.

$$u = R' + g'$$ (5.13.)

For the purposes of this chapter, a control command "Accelerate at maximum possible rate subject to constraints" is a single mode of control, even though it may result in three separate sub-modes as the control $u$ traverses the horizontal, hyperbolic and vertical boundaries of the permitted region in Figure 5.2.

In summary, the significance of the preceding paragraphs in the context of Pontryagin's maximum principle, is that constraints on the state variable $x_1$ (velocity $v$) can be handled as constraints on the control variable $u$ provided that sub-systems on the train (or a human driver) can ensure that the constraint boundaries of Figure 5.2. are not violated.

5.2.6. **The general form of the optimal control**

As has been stated in the general discussion of section 5.2., considerable insight into the optimal sequence of values of control $u_{opt}$ may be obtained by inspection of the Hamiltonian, with a view to maximising it as the state and co-state variables alter.
In particular we examine those terms in the Hamiltonian which contain the control $u$ (keeping in mind the constraints on $u$).

In this particular problem, as discussed in section 5.3.4., we lack knowledge of the initial values of the co-state variables, which would be needed if we were to rely on optimal control theory alone to determine the correct moments of switching between optimal modes of control. On the other hand we do have considerable physical and engineering insight into alternative ways of computing the correct switching moments.

The Hamiltonian equation is now repeated

$$H = -\frac{1}{2}m\dot{x}_1 (u + |u|) + p_1 (u - r' - g') + p_2 x_1 \tag{5.8.}$$

where $x_1 = \text{velocity (m/sec)}$

$x_2 = \text{position (m)}$

For a journey such as that depicted in Figure 5.1. we may assume that initially $u>0$, otherwise the train would go backwards or remain stationary. Likewise, we may assume that towards the end of the journey section the train will normally be braked so as to slow down on the approach to the target, i.e. finally $u<0$.

For convenience we therefore consider the Hamiltonian firstly when $u>0$, and secondly when $u<0$. Our consideration now includes all possible values of $u$, including zero (which, of course, corresponds to coasting).

When $u>0$ we consider the sum of those terms in the Hamiltonian which contain $u$. We call this expression $H'$.

$$H' = -M\dot{x}_1 |u| + p_1 |u| = (p_1 - M\dot{x}_1)|u| \tag{5.14}$$
H' and H will be maximised:

(i) When \( \frac{p_1}{x} > Mx_1 \), by making \( u \) as large as possible subject to the constraints, i.e. by accelerating as hard as possible.

(ii) When \( \frac{p_1}{x} < Mx_1 \), by making \( u = 0 \), i.e. by coasting.

This result is of great practical significance, and has been confirmed by simulation. In practical terms, we should always, when accelerating, accelerate as hard as possible subject to the constraints.

When \( u < 0 \)

\[
\begin{align*}
H' = p_1 u &= -p_1 |u| \\
H' \text{ and } H \text{ will be maximised.}
\end{align*}
\]

(i) when \( p_1 > 0 \), by making \( u = 0 \), i.e. by coasting

(ii) when \( p_1 < 0 \), by making \( u \) as large and negative as possible subject to the constraints, i.e. by braking as hard as possible.

Once again we conclude that, when braking, we should brake as hard as possible subject to practical constraints.

For the complete typical journey section of Figure 5.1, we may put the two different sets of results together. Criteria developed for the application of Pontryagin's principle in non-linear cases may be interpreted in practical terms for this type of journey section by reference to equations 5.14 and 5.15. Our assumption that initially \( u > 0 \) corresponds to stating that the initial value of \( p_1 \) is positive and greater than \((Mx_1)\). (We note in passing that
The significance of \( p_1 \) being always negative is that \( p_1 \) must change monotonically with time from its initial positive value towards a negative value. If this is so the journey will pass through up to three phases in sequence, each with its corresponding optimal control \( u_{\text{opt}} \):

\[
\begin{align*}
\text{If } p_1 > Mx_1 & \quad u_{\text{opt}} \text{ positive and as large as possible} & \quad (5.16) \\
0 < p_1 < Mx_1 & \quad u_{\text{opt}} = 0 \text{ (coasting)} & \quad (5.17) \\
\text{If } p_1 < 0 & \quad u_{\text{opt}} \text{ negative and as large as possible.} & \quad (5.18)
\end{align*}
\]

If \( p_1 \) is always negative, then the optimal solution will pass through each phase only once, and the classical 'bang-off-bang' solution is in fact the optimal one.

A point not generally realised, but which is clear from the foregoing analysis, is that even if the variations of gradient are so severe that the above criterion is not satisfied, the three possible settings of \( u_{\text{opt}} \) given by 5.16., 5.17. and 5.18. are still the only valid settings for optimal control, but it is possible that there may be more than one transition to and from the coasting phase as the journey proceeds, as co-state variable \( p_1 \) ranges up and down.
The reader may have spotted two singular solutions. If it happens that $p_1$ remains equal to $mx_1$, or equal to zero, for a substantial period of time during the journey, then the value of $u$ is indeterminate for optimal control. In practical terms these conditions seldom occur and would require a freakish gradient profile to be sustained. If the condition did occur, it would merely mean that the energy penalty involved in not making a smart transition to and from the coasting phase was less severe than usual. The singular solutions may therefore be discounted in designing a practical system to implement the optimal control.

In summary, the general form of the optimal control for a typical journey section such as that of Figure 5.1 is a sequence of three modes:

(a) maximum acceleration subject to constraints
(b) coasting
(c) maximum braking subject to constraints

In the next section we will consider the problem of determining the correct moments at which the transitions between modes should be made, in the absence of sufficient engineering data to enable the initial values and subsequent dynamic behaviour of the co-state variables to be determined.

5.2.7. Optimal switching of control modes

Consideration of the state and co-state variable dynamics (which can be expressed as a fourth-order matrix differential equation comprising the two state-variable equations (5.3.) and the two co-state variable equations (5.9.) has provided valuable insight into
the general form of the optimal control, and into the possible sequence of the three optimal modes.

However, in this non-linear problem the determination of the moments of switching between modes by analytic solution of the equations for the required initial values of co-state variables is impractical, as discussed in section 5.2. Even if an iterative approach is used, simulation has shown that the final state of the train (and in particular its stopping point) is unduly sensitive to small errors in the initial values. In addition, even if the initial values are correctly determined by these means at the start of the journey section, variations in the dynamics of the train can cause the wrong final state to be reached. A typical cause of such variations is side-wind, which increases the coefficients in the Davis formula for running resistance. The longer the journey-section, the greater the errors.

An alternative means of determining the switching moments is therefore required, which will yield the correct final state of the train (target distance, target velocity and target time) within acceptable tolerances.

For the typical journey section under consideration there are two switching decisions to be made. The first makes the transition from driving to coasting and the second from coasting to braking.

5.2.7.1 Second switching decision

A practical approach to the second switching decision has been described by P.D. Thomas (1979) in discussing the British Rail
Automatic Train Operation pilot scheme. This has been validated for non-optimal journeys during several hundred A.T.O. trials on the Mickleover test track near Derby U.K. In the next paragraph it is shown that this approach is also valid in the optimal control case.

Consider a train coasting towards a target, and approaching the point at which a decision to enter the braking mode must be made. Let

\[ x_T \] = target distance
\[ V_T \] = target velocity
\[ x_B \] = unknown braking point
\[ x \] = current position
\[ v \] = current velocity
\[ B \] = required deceleration rate when braking

The algorithm used in the on-board processor computes the required braking point for the train's current velocity, checks to establish whether the current position has reached that point, and initiates braking if required. This algorithm requires the existence of an automatic system (or a human driver) to ensure that once braking has been initiated, the average deceleration during the braking phase is maintained equal to the required rate \( B \), so that the train reaches its target velocity close to the target point \( x_T \). Automatic braking sub-systems which achieve this performance are already in service in Britain, France, Germany, Japan, U.K. and U.S.A. (Milroy 1979).

The equation for the braking point is

\[ x_B = x_T - (v - V_T)^2/(2B) \]  
(5.19.)
Equation 5.19 is well-known to railway engineers. A reader versed in control theory will note that for fixed target parameters \((x_T\) and \(V_T\)) and variable velocity, this equation yields a parabolic switching trajectory in the phase-plane; see Figure 5.3.

![Figure 5.3. Second Switching Trajectory](image)

All the journeys approaching \((v_T, x_T)\) must switch into the braking phase when their trajectories in the phase-plane (shown dotted in Figure 5.2.) reach the parabolic switching trajectory.

An algorithm for determining the correct time at which to switch from the coasting to the braking phase may now be indicated in flow-chart form. The algorithm may be used either in a simulation situation or in real-time on the actual train.
Enter

Measure train state $(V, x)$

Compute $x_B$ (Eqn. 5.19)

Is $x \geq x_B$?

Y: Continue coasting

N: Initiate braking phase

Return

Figure 5.4. Flow-chart for second switching decision.
The algorithm of Figure 5.4. clearly does not involve any consideration of co-state variables. Requiring knowledge only of the current train state and target parameters, it is effective either when the train is on or close to its optimal trajectory, or when disturbances to running have caused a departure from the optimal trajectory. We note in passing that this algorithm may also be used when there is no coasting-phase, which will normally be the case when minimum-time rather than minimum-energy is the criterion in use.

5.2.7.2. First switching decision

Discussion of the first switching decision (from the acceleration phase to the coasting phase) has been delayed until now because the algorithm developed in the preceding sub-section (5.2.7.1.) will be invoked.

The first switching decision should be made at the moment when co-state variable $p_1$ becomes less than or equal to the train momentum $M x_1$. As in the case of the second switching decision, we seek a practical algorithm which will avoid the need to solve for the co-state variable dynamics.

Unlike the second switching decision, there is no analytic solution to the problem of determining the correct moment of switching from acceleration to coasting. Previous researchers [for example Kokotovic and Singh (1969), Sahinkaya and Sridhar (1972), Edwards and Eren (1979)] have either ignored non-linear functions of velocity and distance in tackling this problem, or have made linearising approximations which are not in general valid for a train running over a complete journey section with
realistic gradient variations. It was decided after consultation with railway traction and signalling engineers in the South Australian State Transport Authority, that any practicable system for making the first switching decision would have to involve a tolerably accurate prediction of the train's future journey profile during the coasting and braking phases.

It was also decided that, in the absence of an analytic solution, an iterative approach to the problem of determining $t_c$ would be necessary. In the case of the real-time control system on the train, this would require the repetitive running of a model, faster than real-time, from the train's current state through to the end of the journey section, with $t_c$ being adjusted iteratively until the required target conditions are achieved.

In 1977 a decision was made that this approach was feasible for the off-line time-table planning model, which did not need to converge to its solution in a real-time environment; this model could be run on a powerful main-frame computer such as the Cyber 73 at the S.A. Institute of Technology. The values of $t_c$ so obtained could be stored on the train-borne system and used by trains running on-time.

A post-graduate student (Mr. K. Tyler) working under the author's supervision is currently studying alternative methods of solving the complete optimal control problem (including the determination of coasting times $t_c$) in real-time on the train. His work will be reported elsewhere.

The remainder of the discussion in this section therefore applies
to the determination of $t_c$ in an off-line timetable planning situation.

In order to arrive at a practicable algorithm for determining $t_c$, an investigation was made of the relationships between coast-initiation time $t_c$, target arrival time $t_T$ and energy consumption $E$. This work was carried out using the train performance simulation, initially on the Membrain digital differential analyser at Loughborough University, and latterly (by K. Tyler) under the author's supervision on the EAI 580 analogue computer at the South Australian Institute of Technology.

Results for all the journey-sections considered will be presented by K. Tyler in his thesis, now in preparation. The relationships between coast-initiation time, energy consumption and journey time are, however, all of the general form shown in Figure 5.5, and Howard (1979) has recently confirmed that this is so, provided that the train is driven optimally over the journey section. This means that a period of coasting should be inserted between the 'driving' and 'braking' phase.

It is clear from the results that, if the switch into the coasting phase is made too late, unnecessary energy is consumed and the train arrives at the target unnecessarily early. If, on the other hand, the switch into the coasting phase is made too early, less energy is consumed but the train arrives too late at its target.
For all the journey-sections under consideration there is a monotonic relationship between coasting-initiation time $t_c$ and energy consumption $E$, and the gradient $(dE/dt_c)$ is always positive. Similarly, there is a monotonic relationship between coasting-initiation time and journey-time over the section $(t_j)$, and the gradient $(dt_j/dt_c)$ is always negative. From an engineering point of view it is impossible to conceive of a situation whereby these gradients could change sign. Figure 5.5. shows typical graphs of energy consumption $(E)$ and journey-time $(t_j)$ plotted against coast-initiation time $(t_c)$.

![Figure 5.5. Relationships $t_j$, $E$ and $t_c$.](image)

It is clear from the diagram that, provided the gradients $dE/dt_c$ and $dt_j/dt_c$ are always positive and negative respectively, we can state that for minimum-energy control the switch into the coasting phase should be made as early as possible.
consistent with arriving on-time at the target.

An iterative scheme for determining $t_c$ is therefore used. First the minimum-time journey is run (no coasting-phase). A short-duration coasting phase is then inserted, and the change in journey-time recorded $(\Delta t_j)_i$. The sensitivity of journey-time to the coast-initiation time is computed, and a new estimate of coast-initiation time is made. The process is repeated until the journey-time is within a small tolerance (usually 5 seconds) of the target journey time, at which point the process is terminated and the required coast-initiation time, and the corresponding energy consumption are recorded.

On the journey-sections under consideration the FORTRAN programme written by K. Tyler to implement the above strategy converges after four iterations at the most. This is an improvement on other schemes reported in the literature, all of which require the use of co-state variable calculations and tend to be unstable in the presence of downhill gradients.

A flow-chart for determining $t_c$ (the time at which the coasting phase is initiated) is given in Figure 5.6. Note that the subroutine JOURNEY is a digital simulation of the state variable equations, with coasting initiated at $t_{ci}$, and the braking point $x_b$ computed using the algorithm discussed in Section 5.2.6.1.
Initialise \((i = 0)\)

Input target data

Run min-time journey

Store journey time \(t_{jo}\) and braking time \(t_b\)

\[ t_{co} = t_b - 0.1 t_b \]

\[ i = i + 1 \]

Run new JOURNEY

Store journey time \(t_{ji}\)

\[ \Delta t_j = t_{ji} - t_{j(i-1)} \]

\[ S = \frac{t_{ji} - t_{j(i-1)}}{t_c(i-1)} \]

\[ \Delta t'_j = t_j - t_{ji} \]

\[ \Delta t'_c = \frac{\Delta t'_j}{S} \]

\[ t_{ci} = t_{c(i-1)} - \Delta t'_c \]

\[ I_S^{\Delta t'_j} < 5 \]

Output journey parameters

END

Figure 5.6. Flow-chart for first switching decision.
5.2.8. **Summary of minimum-energy control**

In Section 5.2. we have developed the general form of the optimal control of traction and braking for minimum mechanical energy at the wheel-rail interface, by applying Pontryagin's Maximum Principle to this constrained problem. To avoid problems in determining the initial values of co-state variables, algorithms have been developed using practical railway engineering considerations, to determine the transitions between acceleration, coasting and braking for typical journeys such as that shown in Figure 5.1.

In Section 5.3. we again apply Pontryagin's principle, this time for the minimum-time journey.

5.3. **Minimum-Time Control**

5.3.1. **Introduction**

Discussions with railway managements in Australia have made it clear that the objective of conserving energy on an individual train must take second place when the train is running late. Indeed, on a busy network or on single-line track where time-of-arrival at passing loops is a critical factor, the waste of energy through other trains having to be stopped after disturbances to the working timetable dwarfs any savings that might be made by running a single train in an energy-efficient manner (but late).

In this system the minimum-energy criterion is therefore abandoned in the event of late running, and replaced by the minimum-time criterion, so that the train concerned gets back on to the working time-table as soon as possible. The decision to switch from one criterion to the other is made on the basis of the predicted time-of-arrival at the next target. A small 'hysteresis' is built into the decision to avoid excessively frequent switches of criterion.
Typical hysteresis is ten seconds for a rapid-transit urban system, and twenty seconds for a heavy-haul freight train.

5.3.2. Analysis

The general form of the optimal control for the minimum-time criterion is now developed.

The train dynamics are given, as before, by equation 5.3.

\[ \begin{align*}
    x_1 &= u - R' - g' \\
    x_2 &= x_1
\end{align*} \]  

(5.3.)

where

- \( u \) is the control input - demanded acceleration
- \( R' \) is the decelerative effect of running resistance
- \( g' \) is the decelerative effect of gradient
- \( x_2 \) is longitudinal position
- \( x_1 \) is train velocity

The criterion of performance for minimum-time is

\[ J = T \]  

(5.20.)

where

- \( T \) = journey-time over the section

We now express \( J \) in the integral form required for the application of Pontryagin's principle.

\[ J = \int_0^T 1 \, dt \]  

(5.21.)

Comparing this with the general form

\[ J = \int_0^T F(x_1 u_1 t) \, dt \]  

(5.22.)

we establish that for minimum-time control

\[ F = 1 \]  

(5.23.)
As in the minimum-energy case, the Hamiltonian is given by

\[ H = F + \sum_{i=1}^{n} (p_i f_i) \]  \hspace{1cm} (5.7.)

where

\[ F = 1 \]
\[ f_1 = (u' - R' - g) \]
\[ f_2 = x_1 \]

\( p_1 \) and \( p_2 \) are the co-state variables.

In this case

\[ H = -1 + p_1 (u - R' - g') + p_2 x_1 \]  \hspace{1cm} (5.23.)

The co-state variables \( p_1 \) and \( p_2 \) are given by:

\[ p_1 = \frac{dH}{dx_1} = -p_2 \]

\[ p_2 = -\frac{dH}{dx_2} = p_1 (\frac{dg}{dx_2}) \]  \hspace{1cm} (5.24.)

Comparing equations 5.24 and 5.9, we observe that in this case the co-state variable dynamics are independent of the control \( u \). This is to be expected, for we are no longer 'penalising' the energy-consumption associated with positive values of \( u \).

The Hamiltonian will be maximised when the term \( H' \) is maximised:

\[ H' = p_1 u \]  \hspace{1cm} (5.25.)

Bearing in mind the constraints on \( u \) discussed in the context of Figure 5.2., we conclude that there are two modes of control for the minimum-time journey. For the journey-section under consideration, the sequence will be as follows:

\[ p_1 > 0 \quad u_{opt} \text{ positive and as large as possible} \]  \hspace{1cm} (5.26)

\[ p_1 < 0 \quad u_{opt} \text{ negative and as large as possible} \]
As in the minimum-energy case, we can be sure that there is only one transition of control mode if the variations in gradient are sufficiently mild that $p_1$ is always negative.

The decision to switch into the braking-phase, as before, may be made using the flow-chart of Figure 5.6., which makes use of the following equation:

$$x_B = x_T - (v-v_T)^2/(2B) \quad (5.19.)$$

5.4. Discussion

It has been known for many years and is, perhaps, intuitively obvious that the way to drive a train for minimum journey-time is, as determined in Equation 5.26, to accelerate as hard as possible subject to constraints, and then to brake as late as possible and as hard as possible on the approach to the target.

The same conclusion may be reached from an optimal control viewpoint by considering the journey profile as a phase-plane plot. If, instead of plotting velocity against distance, we plot the reciprocal of velocity against distance, we obtain a graph similar to that of Figure 5.8.

![Figure 5.8. Reciprocal of velocity, plotted against distance.](image)
It is well known that journey time $T$ may be obtained from such a plot by using the expression:

$$T = \int_0^{X_T} \frac{1}{v} \, dx$$

(5.27)

Journey time is therefore the shaded area under the curve in Figure 5.7., and it is clear from the diagram that the area will indeed be minimised by making the reciprocal of velocity as small as possible throughout the journey, which of course means that velocity should be as great as possible subject to the constraints, and subject to the requirement to brake in order to achieve the required end-state or target conditions.

As with many other applications of optimal control theory in industrial situations, the application of Pontryagin's principle to the minimum-time journey, resulting in the statement summarised in 5.26., merely goes to confirm that procedures evolved by engineers over many years without reference to optimal control theory are in fact optimal or close to it.

However, the fact that the results summarised in 5.26. for the minimum-time case do make practical sense, lend reassurance to the reader (who may be surprised at the conciseness of the analysis) that the results derived for the minimum-energy problem are also valid. The author and his colleagues have confirmed by simulation that, for all the journey-sections under consideration, the minimum-energy and minimum-time optimal control strategies summarised in equations 5.16, 5.17, 5.18 and 5.26 are in fact optimal in the sense that we have been unable to find any variations from the optimal control that result in reduced energy.
Those findings of this Chapter which will be carried over into the design of the time-table planning software and the real-time control system are now summarised.

5.5. Chapter Summary

5.5.1. For minimum-time control, and for minimum-energy control, the optimal control sequence for the journey sections under consideration will commence with a period of maximum acceleration subject to the constraints (this may include a stretch of constant-speed running when the line speed-limit is reached).

5.5.2. For minimum-time control, and for minimum-energy control, the optimal control sequence will end with a period of maximum deceleration subject to the constraints. For a specified service deceleration rate $\beta$, equation 5.19. may be used to calculate the switching-point $x_b$.

5.5.3. For minimum-energy control, a coasting phase will be inserted into the optimal control sequence. The time of switching the coasting phase may be determined by an iterative scheme such as that flowcharted in Figure 5.6. The general principle is to coast as early as possible without arriving late at the target.

This conclusion resolves the question of whether a period of running at a constant average speed is more energy-efficient than accelerating to a higher speed and then coasting. (It is not).

5.5.4. When the control switches from minimum-energy to minimum-time (which will occur when the train for any reason runs late), all that is required is the deletion of the coasting phase.
For operational reasons related to passenger-comfort and the problems of handling long freight-trains, a decision has been taken that if these principles are employed on South Australian railways, lateness of the train after the coasting phase has been initiated will be handled by inserting a period of constant-speed running, rather than by allowing a short period of maximum-acceleration closely followed by heavy braking. This decision is sub-optimal from an energy point of view, but the energy-penalty is only slight. The two profiles are shown in Figure 5.9.

![Figure 5.9. Profiles for lateness after coasting.](image)

5.5.5. In general, the application of the above principles will enable railway vehicles to adhere to their timetables more closely, and to do so in an energy-efficient way. In other chapters these principles are applied to the problem of taking energy-consumption into account in planning a timetable, and to the design of a train-borne system to ensure that the "slack", which is built into all railway timetables to allow for recovery from disturbances to running, is used in an energy-efficient way.
INTRODUCTION

In this chapter we discuss, in general terms, the on-train control system developed in order to drive the train over each journey section in such a way that it satisfies the optimal timetable discussed in Section 4.3., and does so in the most energy-efficient way subject to operational constraints.

It was seen in the analytical work of Chapter 5 that the general form of the optimal control over one journey section of the type under consideration is part or all of the following sequence:

(a) A period of maximum acceleration subject to the constraints. These include the line speed limit, the requirement that the train not proceed astern, limits on acceleration and retardation imposed as a result of wheel-rail adhesion considerations, and a limit on accelerative tractive effort (which varies with velocity) caused by power limitations in the traction system.

(b) A period of coasting, during which the train gradually slows as its kinetic energy is dissipated in overcoming running resistance.

(c) A period of retardation, during which the train decelerates towards its target at a service deceleration rate which should
be as large as possible subject to operational constraints.

From now on, the three modes will be referred to as the DRIVE, COAST and BRAKE modes respectively.

6.1. Inputs to the On-train Control System

As discussed in Chapter 4 the following data must be available to the real-time control system on the train for it to drive the train to its target in an optimal way.

- Target position \( (x_T) \)
- Target velocity \( (V_T) \)
- Target arrival time \( (t_T) \)
- Target deceleration \( (B_T) \)
- Permitted maximum velocity \( (V_p) \)
- Point at which coasting is to be initiated \( (x_c) \)

In addition, the following data pertaining to the train's current state must be available. (The design of instrumentation to provide this data is discussed in Chapter 9).

- Current position \( (x) \)
- Current velocity \( (v) \)
- Current time \( (t) \)

Other data mentioned in Chapter 4 such as gradient information, would be needed if a complete train performance model were to be run (faster than real-time) on the train. In the train-borne system discussed in this thesis variations of gradient are handled as disturbance inputs to a feedback control system, and the system parameters are designed so that these disturbances can be handled without foreknowledge of them.
6.2. **Functions of the on-train control system**

The functions of the control system are now listed and briefly discussed. Detailed discussion of the design of sub-systems and algorithms to achieve the required functions will take place in subsequent chapters. The functions are:

(a) to make the required switching decisions between the three modes (DRIVE, COAST and BRAKE) at the correct moment of time, so that the train arrives on time at its target position, and is travelling at target speed when it does so.

(b) to ensure that constraints on velocity, acceleration and retardation are not violated. The vehicles under consideration are fitted with a satisfactory automatic acceleration control system which controls accelerative tractive effort as a function of velocity.

It is necessary to design a control sub-system for establishing and maintaining the correct average deceleration rate ($\beta_T$) during the braking phase, and a speed-holding sub-system to avoid exceeding the line speed limit.

(c) to ensure that, when running late, the train abandons the minimum-energy criterion, and runs according to a minimum-time criterion until it is back on schedule. It has been seen that the required action is to abandon the coasting phase when lateness is detected. If lateness is detected after the COAST mode has been initiated, the situation is handled by travelling at the appropriate constant speed to arrive at the next target on-time, rather than by
inserting a short DRIVE phase immediately before the BRAKE phase. Although the latter course is marginally more energy-efficient, it is considered to be operationally undesirable because of its effect on passenger comfort, and on the longitudinal dynamics of long heavy-haul freight trains.

6.3. Outputs from the on-train control system

The various control algorithms are run through once per sampling interval. On receipt of a timing pulse from the real-time clock, the processor updates its measurements of train state (distance, velocity and current time), runs through the control algorithms, and outputs either a traction command, a braking command, or a coast command. In the case of the semi-automatic implementation, the system outputs a display to the driver. This will take the form of one of the following words:

- DRIVE \((V_p)\)
- BRAKE \((B_T)\)
- SPEED \((V_h)\)
- COAST

In the case of the first three indications, the five-letter word will be followed by a number which will represent either the line speed limit in m/sec or km/hr., the required retardation rate in m/sec\(^2\) or \(\%g\), or the speed to be held in m/sec or Km/hr. In the event of the driver's display altering, an audible gong is sounded to advise the driver of that event. After completing its computing cycle, the processor waits for the next timing pulse.

In the case of the fully automatic implementation, the visual
indications are still given to provide information to the driver as to what modes are engaged.

Three-term control algorithms are employed to maintain the required velocity during speed-holding, and to maintain the required deceleration on the approach to the target. The design of these algorithms is discussed in Chapters 7 and 8.

Extensive experimental work has been carried out by the Train Control Group at British Rail Technical Centre (Derby) on the implementation of these three-term control algorithms. The author assisted in this work during 1976 and 1977. The results will be published by P.D. Thomas in his doctoral thesis.

6.4. Choice of sampling interval

The choice of sampling interval turns out to be not particularly critical in most digital train control applications. An upper limit of about one second is imposed by the need to arrive at a stopping target with an accuracy of ± 1 metre, with deceleration rates of up to 1 metre/sec\(^2\). (If the last measurement is taken a second before stopping, the train will have a velocity of approximately 1 m/sec at that time). Simulation work and experimental results on GEMINI indicate that these figures are reasonable for electro-pneumatic braking systems on commuter trains.

A lower limit of about 150 msec is imposed by the time taken by the processor (INTEL 8085 in the South Australian application) to carry out all the machine-language instructions involved in one computing cycle. If it had been necessary to reduce the limit a faster but more expensive processor could have been used,
or the task could have been split between two processors, but these steps proved unnecessary.

Variations of sampling interval between those limits had negligible effect on control algorithm performance (provided the actual sampling interval was incorporated in the control parameters). However, it will be seen in Chapter 7 that the effect of the analog data hold at the output of the analog-to-digital converter can be neglected (at least in an approximate analysis) if the sampling interval is less than one-fifth of the longest analog time-constant in the system, and the train velocity is nicely smoothed under these circumstances by the low-pass-filter effect of the longitudinal train dynamics.

For the above reasons, a sampling interval of 0.5 secs was chosen, allowing a margin of about 350 msec for later expansion of the task to be carried out by the digital processor during each computing cycle.

In a review of automatic train control projects in several countries, the author has noted (Milroy 1979) that all project engineers have, for a variety of reasons, settled on sampling intervals of 0.5 seconds or 1.0 seconds for their digital systems, with the lower figure being favoured in recent years.

6.5. Choice of data word length

The word length required for each variable is determined by the resolution required (i.e. the quantity of the variable represented by one bit of the digital number) and the maximum range of the variable. In Table 4.1. below, the first two columns have been
specified as a result of discussions with railway engineers, and the choice of word length has been restricted to multiples of 8 bits.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Resolution</th>
<th>Maximum Value</th>
<th>No. of discrete levels</th>
<th>Chosen Word Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance</td>
<td>± 0.5 m</td>
<td>± 32 Km</td>
<td>64,000</td>
<td>16</td>
</tr>
<tr>
<td>Velocity</td>
<td>± 0.5 m/sec</td>
<td>± 100 m/sec</td>
<td>200</td>
<td>8</td>
</tr>
<tr>
<td>Acceleration</td>
<td>± 0.5 m/sec²</td>
<td>± 2 m/sec²</td>
<td>80</td>
<td>8</td>
</tr>
<tr>
<td>Time</td>
<td>± .5 sec</td>
<td>± 2 hours</td>
<td>14,400</td>
<td>16</td>
</tr>
</tbody>
</table>

Table 4.1. Choice of data word length for system variables.

With two of the four variables requiring double-precision operations on an 8-bit microprocessor, it was a moot point at the time the decision was taken on which microprocessor to use (1978), as to whether an 8-bit or a 16-bit processor should be selected. In fact the 8-bit Intel 8085 processor was selected, and a floating-point package was obtained from the Australian Defence Research Centre (Salisbury) for the algebraic calculations.

The author would recommend at the time of writing (February 1980) that a 16-bit processor be used for future work in this area, and an Intel 8086 has been purchased for a student project to be carried out under the author's supervision.

It is significant that the required word lengths (16 bits for distance and time, 8 bits for velocity and acceleration) turn out to be the same for a heavy load freight run in Western Australia as they are for the urban rapid transit system in Melbourne.
6.6 Intel 8085 Implementation

A project student working under the author's supervision (Mr M Grivell) has recently undertaken a project to implement a prototype system for on-train use, using an Intel 8085 microprocessor.

His report is attached as Appendix 1.
CHAPTER 7

ANALYSIS AND DESIGN OF THREE TERM CONTROL SYSTEM

FOR BRAKING PHASE

INTRODUCTION

The successful implementation of the optimal control strategy (developed in previous chapters) requires an automatic driving system which is capable of:

(a) driving a train at a constant reference velocity (speed-holding).

(b) decelerating a train at a constant rate of retardation on the approach to a target (i.e. following a velocity reference which itself decreases linearly with time).

In practice the braking system phase imposes a more stringent requirement on the control systems, and for that reason the design of the controller for the braking phase is discussed in this chapter.

Three-term controllers (having proportional, integral and derivative action) have been used for the control of the braking phase both in Japan (on the Shinkansen network) and in U.K. (on the British Rail experimental train at Mickleover). Satisfactory values of the controller parameters have been arrived at by a process of trial and error during testing of experimental vehicles. In this chapter design procedures for one-, two- and three-term controller algorithms are developed analytically, using root locus techniques and a linearised model of the train dynamics.

7.1. S-plane model of longitudinal train dynamics

In this chapter a linearised analog model is used, and the required controller transfer functions are developed from the point of view

*On less busy railways in Australia a semi-automatic system is proposed.
of the train, whose longitudinal dynamics are continuous in nature. (In subsequent chapters the controller equation will be implemented in discrete-time form).

The analog model of controller and train is shown in Figure 7.1.

![Figure 7.1. Analog Model of Longitudinal Dynamics](image)

In a practical control system the reference velocity $V_r$ is itself a function of distance; during the braking phase, however, a ramping signal is a good approximation to what occurs in practice and may be used for design purposes. The controller transfer function $G_c$ is given by

$$G_c = K_p (1 + K_d s + K_i / s)$$  \hspace{1cm} 7.1.

or

$$G_c = K_p K_d (s^2 + (1/K_d)s + (K_i/K_d)/s)$$  \hspace{1cm} 7.2.

Equation 7.2 shows that, with full three-term control, the controller transfer function has two zeros, and a pole at the origin of the $s$-plane.

If $K_d$ is set equal to zero, the controller transfer function is written

$$G_c = K_p (s + K_i) / s$$  \hspace{1cm} 7.3.
In the s-plane there is a single zero on the negative real axis, and a pole at the origin. This is two-term (proportional plus integral) control.

If both $K_d$ and $K_i$ are set equal to zero than the controller transfer function becomes simply

$$G_c = K_p$$

This is proportional control.

The transfer function of the controlled plant (the train) is given by

$$G_p = \frac{K_b}{s + c}$$

In Figure 7.1. the braking dynamics are approximated by a first order lag, $c$ is the reciprocal of the time-constant of the approximately exponential braking force which is applied when a step change is made in controller output (b). The resulting deceleration $a$ is integrated with respect to time to yield the velocity $v$. A further integration to yield distance occurs, but this integrator is outside the control loop.

In practice, for a complete dynamic model, the simple linear model of Figure 7.1. should be modified to take account of the following additional factors:

(a) non-linearity in the train's braking system, resulting in gradual variation of the parameters $K_b$ and $c$ over time.

(b) disturbances to the train's acceleration caused by variations in running resistance and gradient.
(c) dead time (sometimes referred to as transport lag) in the braking system. On the modern rolling stock used for the South Australian study dead time is negligible. This may not be so for older rail vehicles with pneumatic braking systems.

The approach taken by the author has been to select the controller parameters using the simple linear model, without taking into account factors (a), (b) and (c) above, and then to use simulation methods to establish whether the non-linearities, disturbances and dead time are sufficiently severe to require variation of the controller parameters.

It has been found that the simple linear model produces good "ballpark" first estimates of controller settings.

7.2. Analysis and design of the braking control system

Utilising the linearised model of Figure 7.1., in which the braking system is characterised by a dominant single time-constant \( \frac{1}{c} \) seconds in the transfer function which relates controller output \( b \) to acceleration \( a \), the closed loop behaviour of the system can readily be predicted (and if necessary compensated) using root locus methods.

7.2.1. Proportional control only

With proportional control only, the locus of the closed loop poles of the system will have the classic second-order form of Figure 7.2. The system will never be unstable, and any desired damping ratio \( \zeta \) can be achieved by setting \( k_p \) to such a value that \( \cos \theta = \zeta \).
Note that for $\zeta < 1$ the time-constant of transient oscillations will always be $(2/c)$ seconds, i.e. twice the dominant time-constant of the braking system.

Figure 7.2. Root locus (no time delay, proportional control only).
For any desired damping ratio $\zeta$, the frequency of transient oscillations is given by $\omega_d$ radians/sec. on Figure 7.2., where

$$\omega_d = \omega_n \sqrt{1 - \zeta^2}$$

The value of $K_p$ to achieve these closed loop pole locations is given by the fact that, on the root locus, the magnitude of the vector GH(s) must be unity.

In this case

$$\frac{K_p}{K_b} \frac{1}{\omega_n^2} = 1$$

which yields

$$K_p = \frac{\omega_n^2}{K_b}$$

The value of the steady velocity error during the braking phase is obtained by assuming that the velocity reference is ramping steadily during the braking phase, that is to say the required acceleration $a_r$ is constant. This is normally the case on the approach to a stop for an automatically driven train.

Considered as a velocity control system, there is a ramp input and the system is of Type number 1. Thus we expect a steady state velocity error $V_e$ of magnitude.
where \( K_1 \) is the steady state "gain" for this type 1 system, given by

\[ K_1 = \frac{K_p K_b}{c} \]

A suitable design procedure using this model is therefore as follows:

1. Consult the specification. Establish the required damping ratio \( \zeta \) for transient disturbances. Confirm that a dominant time constant \( T \) equal to twice the braking system time constant will be satisfactory. If so, proceed to step 2. If not, derivative action will be required (section 5.2.3.).

2. Locate the required closed-loop poles in the s-plane (\( s' \) and its conjugate in Figure 7.2.).

In polar form

\[ s' = \omega_n \sqrt{1 - \arccos \zeta} \]

where

\[ \omega_n = \sqrt{\frac{K_p}{K_b}} = \frac{c}{2\zeta} \]

Alternatively, in cartesian form

\[ s' = -\zeta \omega_n + j\omega_n \sqrt{1-\zeta^2} \]

Again, considering the real part in Figure 7.2.

\[ -\zeta \omega_n = -\frac{c}{2} \]

This also yields

\[ \omega_n = \frac{c}{2\zeta} \]

3. Evaluate \( K_p \), using equation 7.2.

\[ K_p = \frac{\omega_n^2}{K_b} \]
(4) Check that the steady velocity error $V_e$ is within the specification,

$$V_e = \frac{a_r c}{(\omega_n)^2}$$  \hspace{1cm} 7.16

If so, the design of the control system is complete. If not, use integral action compensation as outlined in section 7.2.2.

7.2.2. Proportional + integral control

This design procedure is invoked when the specified settling time (or dominant time constant) and damping ratio of transient oscillations can be achieved with proportional control alone, but the resultant steady state velocity error in the braking phase is too great.

It should be noted that, while the introduction of integral action theoretically reduces the steady state velocity error to zero, it should be avoided if possible as the resultant velocity control system is of Type 2. Such systems are notoriously difficult to stabilise, and with any significant braking system dead-time in the loop the total phase lag around the loop is likely to exceed $360^\circ$ under a wide range of conditions. This point is discussed further in Chapter 6.

Referring to Figure 7.1, the open loop transfer function with integral control is

$$GH(s) = \frac{(1 + K_f/s)(K_p K_b)}{s(s + c)}$$  \hspace{1cm} 7.17

$$= \frac{K_p K_b}{s^2(s + c)}$$  \hspace{1cm} 7.18

The resulting root locus is of the form shown in Figures 7.3(a) and 7.3(h).
7.3.(a) Root locus (proportional + integral action)

\[ K_i < c. \]

7.3.(b) Root locus (proportional + integral action)

\[ K_i > c. \]
We note that the new open loop zero must be to the right of the pole at \( s = -c \) if the system is to be stable, i.e. \( K_i < c \). Also, the locus is to the right of that in Figure 7.2, for all \( K_i \).

A further point of practical interest is that in addition to the two dominant closed loop poles for which the design is carried out, there is a third pole on the real axis between the pole at \( s = -c \) and the zero at \( -K_i \). This determines the time-constant of the integral action in the closed loop system.

The following design procedure invokes the requirement that

\[
\frac{1}{|GH(s)|} = \Pi
\]

at all points on the root locus to determine a value of \( K_i \).

\[
|GH(s)| = 1
\]

is then invoked to determine the corresponding value of \( K_p \).

Finally, the location of the real-axis closed loop pole is checked, to ensure that the integral action does not occur too slowly.

If a satisfactory design is not achieved, then the procedure of Section 7.2.3. should be followed, (p + i + d control).

(1) Establish the required damping ratio \( \zeta \) and time constant \( T \) from the specification.

(2) Locate the required dominant (complex) poles in the s plane \( (s' \text{ in Figure 7.3}) \).

Note that the relationship between \( K_p, \omega_n, c \text{ and } \zeta \) which existed in the proportional case no longer applies, owing to the presence of an open loop zero in the system.

Evaluate \( \omega_n, \omega_d, \theta \text{ and } \zeta \) for the closed loop poles.
(3) Compute the angular contribution $\psi$ of the open loop poles at $s'$, where
\[ \psi = 2(\pi - \theta) + \phi \]  
\[ \phi = \arctan \left( \frac{\omega_d}{c - \zeta \omega_n} \right) \]  
\[ (4) \] Hence compute the leading phase angle ($\eta$) to be contributed by the open loop zero at $s'$. This is given by the requirement that the total angular contribution at $s'$ be $-\pi$ radians
\[ \eta = \pi - \psi \]  
\[ (5) \] Locate the zero, if possible on the real axis in the left half-plane, but to the right of $-c$, so that the correct value of $\eta$ is achieved. The location of the zero gives $-K_i$.

The procedure is visualised graphically, but the use of a calculator or computer program to carry out the required trigonometry will speed the solution.

For example, $\eta$ is given by
\[ \tan \eta = \frac{\omega_d}{K_i - \zeta \omega_n} \]  
where \[ \omega_d = \omega_n \sqrt{1 - \zeta^2} \]

or \[ K_i = \left( \frac{\omega_d}{\tan \eta} \right) + \zeta \omega_n \]  
\[ (7.22) \] A stable solution is not possible if $K_i > c$
i.e. if $\omega_d/\tan \eta > (c - \zeta \omega_n)$

If a stable solution is not possible, procedure 7.2.3. (with derivative action) must be followed.
The procedure of step 5 has been spelt out step by step for clarity, but the process of determining $K_i$ may be consolidated. From equations 7.15 and 7.16.

\[ n = \pi + \arctan \left[ \frac{\omega_d}{(c-\zeta \omega_n)} \right] - 2\theta \]  

\[ K_i = \frac{\omega_d}{\tan n} + \zeta \omega_n \]

(6) $K_p$ is now determined by invoking the magnitude requirement that $|GH|_s = 1$ at $s'$

\[ |GH|(s) = \frac{K_p K_b L_1}{\omega_n^2 L_2} \]

Therefore $K_p = \frac{\omega_n^2 L_2}{(K_b L_1)}$

where

\[ L_1 = \sqrt{(K_i - \zeta \omega_n)^2 + \omega_d^2} \]

and

\[ L_2 = \sqrt{(c - \zeta \omega)^2 + \omega_d^2} \]

Note that for stability $L_1 < L_2$, therefore $K_p$ must always be greater than in the proportional control case for the same closed loop poles.

(7) Finally, the location of the third closed loop pole must be checked out. Let it be at $s = -p$.

The characteristic equation of the closed loop system is

\[ 1 + GH = 0 \]

i.e. \[ s^2(s+c) + K_p K_b (s + K_i) = 0 \]

This is a cubic of the form

\[ s^3 + ds^2 + es + f = 0 \]

where \[ f = K_p K_b K_i \]

But the product of the closed loop poles must also be equal to $-f$, for the characteristic equation
s also given by
\[(s + p) (s^2 + 2\zeta \omega_n s + \omega_n^2) = 0\]

Therefore \[p = K_p K_b K_i/\omega_n^2\]

The value of this third pole should be checked to ensure that the time constant of the closed loop integral action \((1/p)\) is not excessive. If it is, then the procedure in 7.2.3. should be used.

7.2.3. Proportional, Derivative and Integral Controller

The use of derivative action is necessary if the dominant time constant \((2/T)\) for the closed loop system is required to be less than \((2/c)\) seconds. In other words, the root locus of Figure 7.3. may require to be pulled to the left.

The three term controller (or control algorithm) has transfer function
\[G_c = K_p K_d (s^2 + (1/K_d)s + K_i/K_d)/s\]

In the s-plane, therefore, two open loop zeros are contributed, and in general these zeros may be positioned anywhere in the s-plane by suitable choice of \(K_i\) and \(K_d\).

The effect of the zeros is clear from the root-locus diagram of Figure 7.4.

Figure 7.4. Root Locus for p+i+d control
The root locus is pulled sharply around to the left, and its two branches terminate on the zeros as $K_p \rightarrow \infty$. Note that for small $K_p$ the system may be unstable.

As before, suitable locations for the closed loop poles at $s'$ are determined from the specification, and the zero positions are chosen so that the total angle contributed at $s'$ by the open loop singularities is $\pi$ radians. $K_i$ and $K_d$ are thus determined, and $K_p$ follows from the magnitude criterion.

Unlike the previous design procedures there are an infinite number of possible locations for the two conjugate zeros. Each results in different design values for $K_p$, $K_i$, and $K_d$, but all combinations locate the closed loop poles at the correct point $s'$. The locus of possible zero positions is a circle centred at a point on the real axis.

The choice of zero location may be a combination of:

(a) engineering judgement based on experience.

(b) the resulting DC open loop gain, which will determine the steady state error with which this type 2 system will follow a parabolic input command of the form $V_r = Kt^2$.

(c) the location of the third closed loop pole.

However, in practical terms, neither (b) nor (c) are particularly critical factors; this system follows ramp inputs with zero error, and the effect of the third pole is a short-duration transient which is always of shorter time-constant than the main braking system time-constant ($1/c$). It should be noted that without derivative action this third pole is to the right of the open-loop
pole at \(-c\), whereas with derivative action it is to the left, (and therefore results in a faster transient).

The author has experimented with a root locus plotting package which also produces closed loop step responses on command, and has found that in terms of simulated train performance the zero location is non-critical provided the dominant closed loop poles are correctly located.

A satisfactory rule-of-thumb is to position the zeros on a vertical line in the s-plane at \(-\frac{2}{\zeta \omega_n}\), i.e. twice as far to the left of the \(j\omega\) axis as the required closed loop poles, and to compute the vertical \((j)\) component which satisfies the angle criterion at \(s'\).

The design procedure may therefore be summarised as follows:

1. From the specification, determine the best location \((s')\) for the dominant pair of closed loop poles. Evaluate \(\zeta, \omega_n, \omega_d, \theta\). (See Figure 7.4.)

2. Evaluate the total angular contribution \((\psi)\) of the open-loop poles at \(s'\).
   \[
   \psi = \left[\phi + 2(\pi - \theta)\right]
   \]

3. Evaluate the angle \(n\) to be contributed by the controller zeros.
   \[
   n = \Pi + \psi = (\Pi + \phi - \theta)
   \]

4. Arbitrarily choose the real part of the zero coordinates.
   As discussed previously, locate the zeros on a vertical line at
   \[
   \sigma = -2\zeta \omega_n
   \]
The zero locations are therefore at

\[-2\xi \omega_n \pm jd\]

It remains to evaluate d.

5) Fix the zeros in conjugate positions on the vertical line so that their combined angular contribution at \(s'\) is \(n\) radians.

\[n = \arctan [(\omega_d + d)/\xi \omega_n] + \arctan [(\omega_d - d)/\xi \omega_n]\]

This is a transcendental equation (it has no analytic solution) but may readily be solved for \(d\) by iterative means using graphical considerations and a calculator or computer.

6) The zeros are now located at \(s = -2\xi \omega_n \pm jd\)

\[s = -2\xi \omega_n \pm jd\]

i.e. the numerator quadratic is

\[s^2 + 4\xi \omega_n s + (4\xi^2 \omega_n^2 + d^2)\]

Solve for \(K_i\) and \(K_d\) by comparison with equation 7.2

Hence \(K_d = 1/(4 \xi \omega_n)\)

and \(K_i = (4\xi^2 \omega_n^2 + d^2)/4\xi \omega_n\)

7) Solve for \(K_p\) using the magnitude criterion as before

(See Figure 7.4.)

\[K_p = (\omega_n^2 L_3)/K_d K_b L_1 L_2\]

8) Check the location of the third closed loop pole at \(-p\) (as before) by considering the last term in the closed loop equation.

\[(s^2 + 4\xi \omega_n s + 4\xi^2 \omega_n^2 + d^2) + s^2 (s+c) = 0\]

This must factorise into

\[(s^2 + 2\xi \omega_n s + \omega_n^2) (s - p) = 0\]
Thus \( -\omega_n^2 p = 4\xi^2 \omega_n^2 + d^2 \)

\[ p = -4\xi^2 - (d/\omega_n)^2 \]

If \( p \) is too close to the \( j\omega \) axis then shift the zeros further to the left and repeat the procedure from step 4.

(9) If required, check the steady state error \( V_e \) for a unit parabolic reference input.

\[ V_r = \frac{1}{2} t^2 \]

The steady state error is

\[ V_e = \frac{1}{(s^2 K_{DC})S} \rightarrow 0 = c/(K_pK_b) \]

where \( K_{DC} \) is the steady state "gain".

If the steady state error is too large then shift the zeros further to the right and repeat the procedure from step 4.

7.3. Application of design procedures

These design procedures have been applied to typical light rail vehicles, and the resulting control systems have performed satisfactorily when modelled.

7.4. Chapter summary

In this chapter formal design procedures based on the root locus have been developed, enabling the parameters of continuous one-, two- and three-term controllers to be determined for braking systems, assuming linearised models without dead-time.

In the next chapter the general three-term continuous controller algorithm will be re-interpreted as a discrete digital control algorithm suitable for implementation on a general-purpose microprocessor.
CHAPTER 8

Z-PLANE ANALYSIS OF TRAIN CONTROL SYSTEM

INTRODUCTION

We have seen in Chapter 7 that a root locus design in the s-plane may be carried out for the linearised train control system if the 'transport lag' or pure time delay associated with the braking system is negligible compared with other system time-constants.

In practical train control systems, especially when the actuating sub-system for the brakes is predominantly pneumatic in nature, the transport lag may be substantial compared with other system time-constants. Even if a compensatory controller such as the Smith predictor (1956) is used, there is a possibility of mismatch between the time-delay used in the controller and the actual value of transport lag in the plant, which can drift under conditions of environmental variation (e.g. temperature change). See Koepke (1976,76)

It is therefore necessary, in many applications, to take transport lag into account in the design of controllers for maintaining constant deceleration during the braking phase.

The use of s-plane methods is unsatisfactory; the s-plane transfer function $G_T$ for a time delay or transport lag of $T$ seconds is

$$G_T = e^{-sT}$$  \hspace{1cm} \text{(8.1)}

which results in non-rational functions of $s$ in the system transfer functions.
It is possible to incorporate transport lag fairly readily into frequency-response design techniques using

\[ G_T = e^{-j\omega T} \]

In the train control application, however, frequency response techniques are not particularly useful. An important part of the specification during the braking phase relates to the transient response of the system following the step change in commanded acceleration \( a_r \), which occurs when the braking point is reached. Attempts were made by the author to reinterpret transient response specifications in frequency response terms, using concepts such as gain margin and phase margin. However, the clear relationship which exists between transient behaviour and frequency response in the case of simple second order systems (or systems whose s-plane transfer functions are dominated by second-order poles) lacks validity when appreciable transport lag is present within the control loop. Designs which yielded respectable values of gain margin and phase margin in frequency response terms, turned out in simulation to yield intolerable step responses in most cases. The frequency response approach was therefore abandoned.

In this Chapter we apply Z-transform methods, developed by Tou (1959) for the control of sampled-data systems and subsequently extended by many authors. As well as coping satisfactorily with values of transport lag which are integer multiples of the sampling interval, the design procedures developed in this chapter yield, directly, discrete-time digital control algorithms which are suitable for implementation on any typical general-purpose digital processor.
In the next chapter a new method is proposed for dealing with transport lags which occur in a control loop and which are not integer multiples of the sampling interval, making use of a concept of partial poles in the z-domain. As well as being useful for the train control problem, this concept should find application in many other fields.

8.1. Derivation of z-plane transfer function

8.1.1. Block diagram of autodrive system

Figure 8.1. shows a conceptual block diagram for the train control system during the braking phase. (Work carried out at British Rail had indicated that this phase is the critical one from a control systems point of view).

![Block diagram of autodrive system as a sampled-data velocity control system.](Image)

Control calculations are carried out on each of a sequence of discrete values of velocity error $V_e$. 
The output of the controller is converted to an analog value at each sampling instant, and the analog value held constant until the next sampling instant.

The brake system is actuated by the controller output voltage.

The output of the brake system is a controlled decelerative force, which, together with external disturbance forces due to gradient or train resistance, cause the train's velocity to decrease. The last block in the diagram is therefore an integrator, to show the relationship between acceleration and velocity.

8.1.2. Transfer functions of analog elements

The linearised s-plane transfer function of the braking system \((G_B)\) is given by

\[
G_B = \frac{K_b a}{s+a} \ (m/sec^2)/V
\]

Eqn. 8.4. is valid if the transport lag \(T\) is negligible; \(a\) is the reciprocal of the braking system time constant, and \(K_b\) is the steady state deceleration achieved per volt of brake actuating signal. If the transport lag \(T\) is significant, then

\[
G_B = K_b a e^{-ST}/(s+a)
\]

8.5.

The transfer function \(G_H\) of the zero-order data hold is given by

\[
G_H = \frac{1 - e^{-ST}}{s}
\]

8.6.

and the transfer function of the integrator is

\[
G_I = (1/s)
\]

8.7.
When designing a controller to meet the transient response specification we neglect the presence of the disturbance inputs $D$ due to gradients and train resistance. The s-plane transfer function for all the analog elements is obtained by combining equations 8.5., 8.6. and 8.7., to yield:

$$ G_A(s) = \frac{1 - e^{-sT}}{s} \cdot \frac{K_b a}{s(s+a)} \cdot e^{-sT} \, (\text{m/sec})/V \quad 8.8. $$

Assuming for the moment that transport lag $T$ is negligible, we have

$$ G_A(s) = \frac{1 - e^{-sT}}{s} \cdot \frac{K_b a}{s(s+a)} \quad 8.9. $$

Following the procedure given in any standard treatment of Z-transform methods, for example Tou (1959), we now transform the s-plane transfer function $G_A(s)$ into the Z-domain, to yield the transfer function $G_A(z)$ which relates sampled values of velocity to sampled values of the brake actuating voltage.

It has been shown by a number of authors that in the case of a zero-order hold followed by a general s-plane transfer function $G(s)$, the corresponding z-plane transfer function is given by

$$ G(z) = (1 - z^{-1}) \underline{Z}(G(s)/s) \quad 8.10. $$

where $\underline{Z}$ means "the z transform of".

In this case, therefore, we obtain

$$ G_A(z) = (1 - z^{-1}) \underline{Z} \frac{K_b a}{s^2(s+a)} \quad 8.11. $$

Evaluating the z-transform using standard tables. (Tou 1959), we obtain

$$ G_A(z) = (1-z^{-1}) \cdot K_b \frac{T z/(z-1)^2 - (1-e^{-aT})z/a(z-1)(z-e^{-aT})}{(z-1)^2} \quad 8.12. $$
This simplifies to
\[ G_A(z) = K_b (z - 1) - K_b \frac{(1 - e^{-aT})/a(z - e^{-aT})}{a - e^{-aT}} \]  
8.13.

Equation 8.13 is tidied into pole-zero format by placing its two terms over a common denominator, yielding
\[ G_A(z) = K_3 (z - z_1)/(z-1)(z-p_1) \]  

where \( z_1 = \frac{(1 - e^{-aT})}{aT - 1 + e^{-aT}} \)
\[ p_1 = e^{-aT} \]
\[ K_3 = \frac{K_b (aT - 1 + e^{-aT})}{a} \]

Equation 8.14 is the crucial design equation as it enables the open loop poles and zeros associated with the zero-order-hold and braking system dynamics to be located in the z-plane for any values of braking system time-constant (a), sampling interval (T), and braking system gain coefficient (Kb).

We now note that if there is a transport lag \( T \) seconds, its s-domain transfer function is
\[ G_T(s) = e^{-sT} \]  
8.15

Equation 8.15 may be written
\[ G_T(s) = (e^{-sT})^{T/T} \]  
8.16

The corresponding z-domain transfer function is
\[ G_T(z) = z^{-T/T} \]  
8.17

If the ratio \((T/T)\) is an integer (K) then the transport lag introduces K poles at the origin of the z-plane. If the ratio is not an integer then in general, the method of advanced z-transforms may be required. However, for the particular problem of designing a digital controller to meet a specified closed-loop transient response the author has proposed a
new method involving the concept of partial poles. This is discussed in Chapter 9.

8.1.3. Evaluation of $G_A(z)$ for a specific class of train

As part of a study undertaken using data provided by the South Australian Department of Transport, the author has designed digital control algorithms for a light rail vehicle (the modern version of a tram) offered by A.S.E.A. Pty. Ltd. (Australia).

The relevant data for the application of equation 8.14 are

- Braking System Coefficient $K_b' = 1.1 \times 10^4$ Newton/V
- Effective Mass $M = 19$ Tonne
- Braking System Time Constant $(1/a) = 1.47$ second
- Transport Lag (Type A system) $T = 0.2$ second
- Transport Lag (Type B system) $T \approx 0.0$ second (negligible)
- Sampling Interval $T = 0.5$ second

One of the purposes of the study was to establish whether a satisfactory controller could be designed with the cheaper Type A brake system, or whether a fast-acting Type B system with negligible transport lag should be specified.

Before applying equation 8.14 we note that the constant $K_b'$ in the data is expressed in terms of Force per unit of actuating voltage, whereas the constant $K_b$ is expressed in terms of acceleration per unit of actuating voltage. We therefore have

$$K_b = \frac{K_b'}{M}$$

$$K_b = .58 \text{ (m/sec}^2\text{/V}$$

Numerical values for the other constants used in equation 8.14 are obtained from the data, and are:
8.2. Interpretation of transient response specification in the z domain

As has been stated earlier, one of the important aspects of the performance of an automatically controlled train is the transient response after entry to the braking phase. Tests on Gemini showed that there tends to be a response term closely similar to that of an underdamped second-order system. Variations of velocity error occur, of the form

\[ V_e = A_e e^{-\sigma t} \cos(\omega_d t + \phi) \]  

In the s-domain this response corresponds to specific locations of a pair of complex conjugate poles, and in Chapter 5 the required locations were specified in terms of any two of the following parameters

(a) A vertical line of constant \( \sigma \), where \( \sigma \) is the reciprocal of the timeconstant of the response's exponential decay

(b) A horizontal line of constant \( \omega_d \), where \( \omega_d \) is the damped natural frequency of transient oscillations

(c) A radial line (from the origin of the s-plane) of constant \( \zeta \), where \( \zeta \) is the damping ratio. The angle \( \theta \) of the radial line to the horizontal reference is given by

\[ \theta = \pi - \arccos \zeta. \]
Figure 8.2. shows that specification of any two of the three closed loop parameters $\sigma$, $\omega_d$ or $\zeta$ specifies the required closed loop pole locations.

![Diagram showing lines of constant $\zeta$, $\omega_d$, and $\sigma$.](image)

Figure 8.2. $s$-plane, showing lines of constant $\zeta$, $\omega_d$, and $\sigma$

Are there corresponding lines in the $z$-plane which can be specified, in order to ensure that the sampled response corresponds to specific values of $\zeta$, $\omega_d$, and $\sigma$?

Consider

$$s = \sigma + j\omega_d$$  \hspace{1cm} (8.21)

$$z = e^{sT} = e^{\sigma T} e^{j\omega_d T}$$  \hspace{1cm} (8.22)

A line of constant $\sigma$ is a line of constant distance $e^{\sigma t}$ from the origin, i.e. a circle of radius $e^{\sigma t}$. The special case of marginal stability ($\sigma = 0$) is of course the unit circle; normally a negative value of $\sigma$ will be specified, and this corresponds to the requirement that the dominant pair of closed loop poles lie on a circle in the $z$-plane whose radius is less than unity.

Lines of constant $\omega_d$ are, from equation 8.22, lines of constant angle in the $z$-domain, the relevant angles being $(\pm \omega_d T)$ radians from the horizontal. The required closed loop pole locations are simply the intersection of the specified $\sigma$ circle and the specified pair of $\omega_d$ radial lines.
Lines of constant damping ratio $\zeta$ are more complex in the $z$-domain.

From $s$-plane theory we know that

$$\sigma = \frac{-\zeta \omega_d}{\sqrt{1-\zeta^2}}$$  \hspace{1cm} 8.23.

Equation 8.22 may therefore be written

$$z = e^{-q \omega_d} e^{j \omega_d T}$$  \hspace{1cm} 8.24.

where

$$q = \frac{-\zeta T}{\sqrt{1-\zeta^2}}$$  \hspace{1cm} 8.25.

For constant $\zeta$ and $T$, $q$ is a constant, and equation 8.24 shows that, as $\omega_d$ is increased, a logarithmic spiral is traced out.

To avoid having to consider logarithmic spirals in the $z$-plane the author has adopted the practice of interpreting a specification containing damping ratio $\zeta$ in terms of $\omega_d$ and $\sigma$. This can always be done using equation 8.23.

Figure 8.3. shows a typical circle of constant $\sigma$ and lines of constant $\omega_d$, locating the specified closed loop poles in the $z$-domain.

![Figure 8.3. z-plane, showing lines of constant $\omega_d$ and $\sigma$](image)
If the undamped natural frequency $\omega_n$ is specified instead of $\omega_d$, then a conversion may be made using

$$\omega_d = \sqrt{1 - \zeta^2} \omega_n \quad \text{(8.26)}$$

The relationships between the locations of closed loop poles in the s-plane and the z-plane may be confirmed by considering the two relevant transforms for time functions of the form (refer to standard tables).

$$x = e^{at} \cos(bt + \phi) \quad \text{(8.27)}$$

In each case the transform's denominator factorises into a pair of complex poles, but the geometrical relationships between the two sets of poles in the s-plane and Z-plane respectively are not clear without the foregoing discussion.

8.3. Root loci in the z domain

In section 8.1, we have developed a technique for locating the two open loop poles and one open loop zero for the train control system in the z-plane.

In section 8.2, we have seen that, in order to achieve a specified sampled transient response it is necessary to place the two closed-loop poles at specified points in the z-plane, and we have considered simple methods of determining the required locations.

In this section we apply the root locus method of Evans (1948) to the problem of determining appropriate controller parameters. While the literature shows that root locus techniques were widely applied in the nineteen-sixties and early seventies, using the s-plane for analog systems, there has been little application of the method to the design of single-loop digital controllers, which are now becoming
common in this era of distributed microprocessor control.

In fact all the methods which were developed by Evans and his successors for the s-plane may be applied in the z-plane. In particular, the author has used these methods to analyse the digital algorithms which were developed by Thomas (1980) et al. for real-time control of rail vehicles.

Whereas the early users of root locus techniques made use of sophisticated graphical construction techniques, the methods used by the author (as in Chapter 5) exploit the calculator and digital computer rather than the compass and spirule to solve the design problem. Versatile computer-aided design packages for the application of root locus techniques to complex problems have been developed at the University of Manchester Institute of Science and Technology, at Cambridge and Reading Universities, and at Imperial College; unfortunately the author did not have access to such a package, and has therefore confined his attention to a limited number of control algorithms.

8.3.1. Proportional control (braking phase)

Figure 8.4. Root locus - proportional control
Figure 8.4 shows the root locus for the system whose z-domain transfer function was developed in the previous section, assuming that the controller is a simple variable gain $K$.

The closed loop transfer function $p(z)$ is given by

$$p_z = \frac{KG_A}{1 + KG_A} \quad 8.28$$

where

$$G_A = \frac{K_3 (z - z_1)}{(z-1)(z-p_1)} \quad \text{refer } 8.14$$

The root locus is the locus of roots of the characteristic equation

$$1 + KG_A = 0 \quad 8.29$$

or

$$(z-1)(z-p_1) + K K_3(z-z_1) = 0 \quad 8.30$$

which is

$$z^2 - z (1 + p_1 - KK_3) + (p_1 - KK_3 z_1) \quad 8.31$$

When $K$ is increased this system becomes unstable, at a value of $K$ given by

$$K K_3 (-1-2-z_1) \quad \frac{(-1-p_1)\cdot-2}{(-1-p_1)\cdot-2} = -1 \quad 8.32$$

i.e.

$$K = \frac{2 (1 + p_1)}{(1 + z_1)} \quad 8.33$$

8.3.2 Effect of increased transport lag

Figure 8.5 shows the root locus (again with $K$ as the variable parameter) for the same system with a transport lag corresponding to one sampling interval.

Figure 8.5 Root locus for system with transport lag $T$ seconds
Note the effect of the new open-loop pole at the origin, which causes the locus to move outside the unit circle (and therefore into an unstable regime) for a much lower value of $K$ than is the case without transport lag.

Similarly, Figure 8.6. shows the root locus for the same system with a transport lag of two sampling intervals.

![Root locus for system with transport lag 2T seconds.](image)

Figure 8.6. Root locus for system with transport lag 2T seconds.

The value of $K$ at which the loci of the closed loop poles move into the unstable region (outside the unit circle) is lower again.

In practice, not only does the maximum stable gain of the basic uncompensated system reduce rapidly with increased transport lag, but the design of a controller to achieve a desired closed loop response becomes more difficult. In sinusoidal frequency response terms, the phase lag $\psi$ caused by transport lag $T$ is, at any frequency, given by

$$\psi = -\omega T \text{ radians}$$

The controller algorithm has, in effect, to provide phase lead to compensate for this tendency.
8.3.3. Effect of increased sampling interval

It is of interest to comment in passing on the effect of increasing the sampling interval $T$, or in other words, of reducing the sampling frequency.

In a system without transport lag the effect is well known - the longer the sampling interval, the less stable the response. But with transport lag present the effect of changing the sampling interval is more complex. On the one hand there is the usual destabilising tendency; reference to equation 8.14 shows that this is accounted for mathematically by the changed positions of the open loop zero and pole ($z_1$ and $p_1$).

But, on the other hand, there is a countervailing reduction in the effect of the poles at the origin (which are directly attributable to transport lag), when the sampling interval $T$ is increased. This occurs because the magnitude of the ratio ($T/T$) reduces with an increase of sampling interval $T$. (See Equation 8.17).

The author has found that no general rules apparently exist to indicate which of these counteracting effects will predominate in a particular case - the answer depends on the precise location of the open-loop poles, as well as on the values of gain and transport lag being considered, and on the position of the desired closed-loop poles.

It is possible to draw root locus diagram in the z-domain with sampling interval $T$ as the variable parameter and all other parameters held constant. Such a diagram is, however, of academic interest only in this context because, in the practical design situation, changes of sampling interval result in
consequential changes of controller parameters in order to maintain the desired closed loop responses.

The author has resolved the question of selecting the appropriate sampling interval in closed loop system with transport lag, by using a rule of thumb in the first instance (sampling interval roughly five times shorter than the shortest time-constant being considered), and then repeating the analysis and design with variations in sampling interval, to see whether the resulting design can be improved.

This lack of precise formulae for determining the effect of changes in sampling interval when transport lag is present, is in marked contrast to the clear formulae which exist in the open loop case without transport lag. See for example Tou (1959), Kuo (1978).

8.4. Design procedure for two-term controller

It became apparent during trial running on Gemini that integral action in the controller is a desirable feature, provided that the destabilising effect on the transient response is not excessive.

The reasons for incorporating integral action in the controller are:

(a) so that sustained disturbances entering the closed loop system (for example a sustained increase in running resistance due to flanging on a curve) are ironed out and do not result in persistent velocity error.
so that, considering the response of the system during the braking phase from the point of view of the continuous system dynamics of the train, the velocity control system will be of type 2. If this is so then the train will follow a ramping velocity reference input (which is approximated during the braking phase) with zero steady state error.

Experience on Gemini, and with simulation of the control system dynamics, indicates that with appropriate integral action present in the controller, the complete effects of gradient changes and of running resistance are dealt with as they occur, without the autodrive having to have 'foreknowledge' of their magnitude.

In this section design procedures are discussed, for a digital proportional-plus-integral control algorithm, for the case of a braking system with negligible transport lag. In a later section the effects of transport lag will be taken into account.

8.4.1. Z-plane transfer function

As before, the transfer function of the train and its braking system (including the zero order hold after the analog-to-digital converter), is given by

\[ G_A = \frac{K_3(z-z_1)}{(z-1)(z-p_1)} \]

The controller transfer function is derived using the z-transform corresponding to rectangular integration. (The more complex integration transform corresponding to trapezoidal integration was also tested but gave no significant improvement, probably because any changes were swamped by the relatively coarse quantization of the eight-bit digital processor).
For the controller:

\[ G_c = K_p + \frac{K_i T z}{z-1} \]  \hspace{1cm} (8.36)

or

\[ G_c = \frac{(K_p + K_i T) z - K_p}{z - 1} \]  \hspace{1cm} (8.37)

Expressing the controller transfer function in pole-zero format

\[ G_c = \frac{K'(z-z_2)}{z-1} \]  \hspace{1cm} (8.38)

where

\[ K' = K_p + K_i T \]
\[ z_2 = \frac{K_p}{K_p + K_i T} \]

Assuming that the dynamics of the feedback path are sufficiently fast to be approximated by a unity feedback system, the open loop transfer function is given by

\[ G = KK'(z-z_1)(z-z_2) \]
\[ (z-1)^2(z-p_1) \]  \hspace{1cm} (8.39)

The dominant closed loop poles of the transfer function may be positioned at the appropriate points in the z-domain (if possible), by selecting the controller zero \( z_2 \) so that the root locus angle criterion is satisfied at the desired closed loop pole locations, and then applying the magnitude criterion. These steps yield two equations for the values of controller parameters \( K_p \) and \( K_i \).

In general, the form of the resulting locus is as shown in Figure 8.7.

![Figure 8.7. Root locus for typical braking system with proportional and integral control](image-url)
The detailed application of the design principle is not discussed here because the methods involved are closely similar to those used in the s-plane design procedure discussed, for the continuous controller, in Chapter 7.

Note that, in addition to the two dominant closed loop poles which must lie within the unit circle, there is a third pole situated on the real axis. The location of this pole will determine the time constant with which sustained disturbances or changes of gradient are ironed out by the integral action of the controller.

If the design procedure for proportional and integral control yields a negative value of $K_i$, then $z_2$ will lie to the right of and outside the unit circle, which means that the third closed loop pole will always be outside the unit circle, and a satisfactory design cannot be achieved.

Under these circumstances the incorporation of derivative action may achieve the desired response. The design procedure for an appropriate three-term digital controller is closely analogous to that used in the s-plane case, and is not discussed in detail here.

The $Z$-plane transfer function used by the author in the three-term case is:

$$
\frac{K_i T_z}{K_p + \frac{K_i T_z}{z-1} + \frac{K_d}{T_z}}
$$

which simplifies to:
where:

\[ \frac{pz^2 + qz + r}{z (z-1)} \]  

\[ p = (K_p + K_i T + K_d / T) \]  

\[ q = -(K_p + 2K_d / T) \]  

\[ r = K_d / T \]

Note that, as in the case of the s-plane three-term controller design, p, q and r may be selected so as to locate the controller zeros appropriately (equation 8.41) after which the coefficients \( K_p \), \( K_i \) and \( K_d \) may be calculated using equations 8.42, 8.43 and 8.44.

8.5 Chapter Summary

In this chapter we have examined the effects of transport lag in the braking sub-system of the train on the controller design. We have noted that the effect of a change of sampling interval is not clear-cut when transport lag is present. We have described the design of one-term and two-term controllers in the Z-domain and shown that the design of a three-term controller is also possible.

The procedures of this chapter apply to transport lags corresponding to an integral number of sampling intervals. In the next chapter, we propose a technique for the design of Z-plane controllers in cases where the transport lag is a non-integer number of sampling intervals.
CHAPTER 9
THE CONCEPT OF PARTIAL POLES IN THE Z-DOMAIN

INTRODUCTION

The analysis and design of closed loop deterministic systems which incorporate analog to digital converters (A/D), digital to analog converters (D/A), digital signal processors and multiplexors is made difficult when time delays occur which are not integer multiples of the sampling interval (T). A typical configuration for a control system is shown in Figure 9.1.

This chapter describes a concept of partial poles in the Z-domain which enables many of the effects of such delays to be calculated without resorting to complex analytical procedures.

To illustrate the concept a design example is provided for a train braking system similar to that discussed in Chapter 8, but with significant pneumatic transport lag. The system is treated as an acceleration control system.

9.1. Causes of delay

Time delays frequently arise in systems such as that of Figure 1, or in simpler configurations where digital filters are employed in analog
systems, from one or more of the following causes:

(a) There is a delay of less than one sampling interval between the time when sampled data is read in to the digital processor and the time when output data is written out.

(b) There is a transport lag ($T_t$) in the analog plant.

(c) Multiple analog signals are sampled simultaneously so as to avoid 'time skew', and then held for a period of time while a multiplexor scans the signals sequentially ($T_m$).

(d) A/D and/or D/A conversion times are significant ($T_c$).

If the system is linear or can be represented adequately at any time by a linear model, then for the purposes of analysis such delays may be commonly lumped together into a total delay ($T$).

$$T = T_t + T_m + T_c$$

This will be so if by a valid process of block diagram or signal-flow chart manipulation the delay elements can be made consecutive. This is normally true if all the delay elements are within a single loop.

9.2. Previous methods of analysis

If the resulting total delay $T$ is not an integer multiple of the sampling interval ($T$), then it has in the past not been considered possible to tackle the analysis or design of such systems by simple $z$-plane methods because non-rational functions of $z$ arise.

One stratagem which has been developed to cope with the analysis of such systems when $T < T$ is the use of a fictitious sampling frequency higher than the actual sampling frequency. But this stratagem fails if the ratio $T/T$ is not an integer.
Another strategem is the use of the advanced Z-transform, which involves the calculation of system variables in between samples and is unwieldy. Such an approach should not be necessary, if as is usually the case, the input and output samplers are effectively synchronized even though time delays exist in the process. Faced with these difficulties, most engineers with the problem of designing or predicting the behaviour of such systems (even in the simplest cases of single-input single-output systems) have resorted to digital simulation, or to numerical solution of the system state equations, which in this context amounts to the same thing.

9.3. The concept of partial poles in the z-domain

If a block diagram of the total system is visualized or drawn it may be broken into sections of four different types.

(a) Elements of the continuous system between samplers whose transfer functions are rational functions in the complex variable s. These may be lumped together into single transfer functions and then each such s-plane transfer function may be Z-transformed to produce a corresponding Z-domain transfer function which relates sampled values of input to sampled values of output for that section.

(b) Discrete sections of the system; these may include sample and hold elements and a digital processor. It is fortunate that a large class of algorithms used for digital filtering and control are readily represented by Z-domain transfer functions. This is clear when one considers that most digital algorithms employed for the purpose, including those which approximate differentiation and integration with respect to time, involve only the elemental steps of adding, multiplying by a constant and shifting or delaying by one
sampling interval, which is represented in the z-domain by the operator $z^{-1}$.

(c) Time delays which are integer multiples of the sampling interval. These may be represented in the z-domain by a corresponding number of poles at the origin, each pole representing a delay of one sampling interval.

(d) Time delays of less than one complete sampling interval. It is postulated that these may be represented by partial poles in the z-domain. If the time delay is equal to $q$ sampling intervals where $0 < q < 1$, then for convenience of interpretation the partial pole may be represented graphically by a sector of a circle whose angle corresponds to the value of $q$. For example in Figure 9.2a sector angle of $90^\circ$ would correspond to $q = .25$.

Figure 9.2 Z-plane pole-zero constellation showing partial pole

Thus a complete representation of all the elements comprising the loop transfer function $GH(z)$ is obtained, and is represented graphically in the $z$-plane by a number of complete zeros, a number of complete poles and a partial pole at the origin. All the conventional design procedures (including the use of residue methods and of root locus concepts to calculate closed loop responses) which involve computing the influence
of poles and zeros at specific points $z$ elsewhere in the plane may now be applied, except that it is now postulated that the magnitude contribution of the partial pole at any point $z$ is $z^{-q}$, rather than $z^{-1}$ in the case of a complete pole at the origin.

Similarly the angular contribution of the partial pole is postulated to be $q\theta$ (where $\theta$ is the angle shown in Figure 92) rather than $\theta$ in the case of a complete pole.

9.4 Theoretical considerations

A time delay of one sampling interval $T$ is represented by the $s$-plane transfer function $e^{-sT}$, which transforms to $z^{-1}$ in the $z$-plane. A time delay $qT$ is represented by the $s$-plane transfer function $e^{-sqT}$ or $(e^{-sT})^q$. The corresponding $z$-transform is $z^{-q}$. When carried over into polar operations in the $z$-plane the directed line segment from the origin to the point in question should have its magnitude raised to the power $q$ and its angle multiplied by $q$. Note that when $q$ is made equal to an integer, e.g. 0, 1 or 2, the procedure described is identical to conventional pole-zero methods, and produces the correct results.

One physical reason why a similar concept of partial poles does not apply in the $s$-domain is that a pole at the origin ($s^{-1}$) corresponds to the process of integration with respect to time. The concept of partial integration is in this context physically meaningless whereas the concept of a partial time delay is related to physical phenomena which do in fact occur in digital or hybrid systems.

It is the author's opinion that control engineers have tended to carry over their techniques involving integer numbers of poles and zeros from the $s$-plane to the $z$-plane without realising that the concept of a non-integer number of poles also has meaning in the latter case.
9.5. **Experimental validation**

The design exercise described in the next section has been carried out for a number of different values of time delay. The resulting system has been modelled using an EA1580 analog computer for the analog plant and an EAI Pacer minicomputer to implement the time delay and digital controller algorithms with 12 bit resolution. The parameters of the resulting output responses from the hybrid simulation have, as expected, been indistinguishable from those predicted using partial poles.

9.6. **Design example**

A design example now follows in which the form of the step response of a practical closed loop system incorporating a transport lag is specified. The specification is translated into required locations of the dominant poles of the complete system (including the digital controller). The loop transfer function is represented in the Z-plane by a pole zero constellation which includes a partial pole, and the controller parameters are calculated using the postulated method so that the root locus for the complete system does pass through the required points in the Z-domain.

9.6.1. **Introduction**

A digital controller is to be designed to control the deceleration of a railway vehicle during the braking phase.

The configuration of controller, electro-pneumatic braking system and longitudinal train dynamics may be represented by the block diagram of Figure 9.3. Any time-delay in the controller is negligible compared with the pneumatic transport lag.
Figure 9.3 Block diagram for braking control system

The time delay $T$ is caused by the delay in propagation of a pneumatic brake-pressure signal through the system. The time constant $a^{-1}$ seconds is associated with changes in brake-cylinder pressure during application or release of the brakes. The force tending to decelerate the train may be affected also by disturbances arising from gradients $M \sin \theta$ and train resistance $R$. These two effects vary non-linearly with velocity and distance.

It has been shown that a control system employing integral action can correct satisfactorily for changes in train resistance and gradient, which are effectively disturbances to the system. A digital proportional-plus-integral controller is therefore to be designed.

9.6.2. System parameters

- $K = 1.1 \times 10^4$ (braking system gain factor)
- $M = 19$ tonnes (Effective mass)
- $a = 0.68$ seconds$^{-1}$ (reciprocal of braking system time-constant)
- $T = 0.2$ seconds (pneumatic transport lag)
- $T = 0.5$ seconds (sampling interval)
9.6.3. Specification

It is known from experimental tests that the critical aspect of braking system performance is the transient response following either a step change in reference input (desired acceleration), or a step change in gradient. Resistance changes are gradual.

The response in each case has the form of a damped oscillation (See Figure 9.4) and so it may be assumed that the response is dominated by two complex conjugate poles in the s-domain.

![Figure 9.4 Typical response following step change in Commanded Deceleration](image)

The desired output response is therefore specified in terms of the classical s-domain second order characteristic equation.

$$s^2 + 2\zeta\omega_ns + \omega_n^2 = 0$$

Appropriate criteria to be satisfied are:

- Decremental time constant \((1/\zeta\omega_n)\) 1.5 seconds
- Damping ratio \((\zeta)\) 0.7
- Hence \(\omega_n = 1/(.7 \times 1.5) = 1.05\) rad/sec.
9.6.4. Design procedure for proportional and integral controller

1) Redraw block diagram with single transfer function between samplers (s-domain). See Figure 9.5.

\[ \frac{K_p + K_i}{sT} \frac{1 - e^{-sT}}{s} \frac{0.58}{s+6.8} e^{-0.25} \]

Figure 9.5. Simplified block diagram. (s-domain)

2) Transform block diagram into the z-domain (Figure 9.6.)

\[ \frac{K_p + K_i T}{z - T} \frac{(1-z^{-1})}{(z-1)(z-0.712)} \]

Figure 9.6. Control system block diagram (z-domain)

3) Find GH(z) in factored form.

\[ \frac{(K_p + K_i T)(z - K_p/(K_p + K_i T))}{(z - 1)(z - 0.712)} 0.4 \]

\[ = \frac{0.288 C (z - K_p/C)}{(z-1)(z-0.712) z^{0.4}} \text{ where } C = K_p + K_i T \]

4) Sketch a pole zero plot in the z-plane (Figure 9.7).

Figure 9.7 Pole-zero plots in z-plane (note 0.4 poles at origin).
5) Identify required closed-loop pole position in z-plane.
   From the specification the s-plane characteristic polynomial is:
   \[(s + 0.67)^2 + 1.05^2\]
   The corresponding z-plane characteristic polynomial is:
   \[z^2 - 2ze^{-0.395}\cos 0.5025 + e^{-0.67}\]
   which is:
   \[(z - 0.627)^2 + (0.345)^2\]
   Thus the required closed loop pole locations are:
   \[z = +0.627 \pm j0.345\]

6) Sketch a new open-loop pole-zero plot augmented with the required closed-loop pole locations (Figure 9.8.)
   Evaluate the phase-angle contributions of the open loop poles and zero at one of the desired pole positions.

\[
\phi(GH) = \arctan(0.345/(0.627-Kp/c)) - 0.4 \times \arctan(0.345/0.627) + 90 + \arctan(0.345/0.035) + 90 + \arctan(0.345/373)
\]

Figure 9.8. Pole-zero plot showing upper closed loop pole position
7) In order for the closed loop poles to be correctly located

\[ \angle(GH) = -180^\circ \]

i.e. \[ \arctan \left( \frac{0.345}{0.627 - Kp/c} \right) = 0.4 \times 28.82^\circ + 76.16^\circ + 42.76^\circ \]

\[ = 130.5^\circ \]

\[ \frac{0.345}{0.627 - Kp/c} = -1.173 \]

\[ Kp/c = 0.921 \]

Thus the use of the angle criteria has determined

\[ Kp/c = Kp/(Kp + KiT) = 0.921 \] \hspace{1cm} (1)

8) In order to determine \( Kp \) and \( Ki \) a second equation is required.

Use the magnitude criterion at the closed loop pole location.

\[ |GH| = 1 \]

\[ |GH| = \frac{0.288(Kp + KiT) \times (0.345^2 + 0.913^2)^{1/2}}{(0.345^2 + 0.627^2)^{1/2} \times (0.345^2 + 0.085^2)^{1/2} \times (0.345^2 + 0.373^2)} \]

\[ = 0.288(Kp + KiT) \times 0.976 \]

\[ = 0.874 \times 0.355 \times 0.508 \]

Equating \( |GH| \) to unity we have

\[ Kp + KiT = 0.561 \] \hspace{1cm} (2)

9) Determine \( Kp \) and \( Ki \) from (1) and (2)

\[ \frac{Kp}{Kp + KiT} = 0.921 \]

\[ Kp = 0.921 \times 0.561 = 0.517 \]

\[ KiT = 0.561 - 0.517 = 0.044 \]

\[ Ki = 0.088 \]

Thus the required controller settings are

\[ Kp = 0.517 \]

\[ Ki = 0.088 \]
9.7 CONCLUSION

The concept of partial poles in the z-plane has been described, to handle aspects of the analysis and design of sampled-data systems in which there is a time-delay which is not an integral number of sampling intervals.

The validity of the concept has been discussed and the principle has been used in a typical design exampled with satisfactory results.

In summary, the z-transform of a time delay of duration q sampling intervals is \( z^{-q} \), where q is any positive number. In the z-plane this may be represented by an integral number of poles at the origin and one additional partial pole at the origin.

When considering the polar contribution of a partial pole at another point in the z-plane, the magnitude contribution \( M \) is given by

\[
M = \ell^q
\]

and the angular contribution \( \phi \) is given by

\[
\phi = q \theta
\]

where \( \ell \) is the length and \( \theta \) is the angle with respect to a horizontal reference of the directed line segment from the partial pole to the z-plane point under consideration.

9.8 Footnote

The author has recently found that an alternative graphical visualisation is sometimes helpful. A partial pole at the origin of the z-domain may be considered as a complete pole and a partial zero. The significance of this variation in outlook is that, whereas
closed loop poles migrate in the z-plane when a variable parameter (usually gain) is altered, closed loop zeros (and partial zeros) frequently remain fixed.
CHAPTER 10

INSTRUMENTATION FOR AUTOMATIC TRAIN OPERATION

INTRODUCTION

This chapter discusses the factors which should be taken into account in designing the instrumentation required for measuring the state of a train under automatic driving. Specifically, the measurements of longitudinal position, velocity and acceleration are discussed. Other variables, for example brake cylinder pressure, traction current or torque, and gradient, may also need to be measured for a specific control requirement, but in this chapter attention is confined to the fundamental problem of determining the longitudinal position and its derivatives.

The instrumentation is discussed in the context of the B.R.A.T.O. pilot scheme, involving the automation of the driving function of an outer-suburban passenger service employing AM4 electrical-multiple-unit stock, but has more general application. Where choices have had to be made regarding the required precision with which variables should be measured, the B.R.A.T.O. requirements have been taken as a guide. Experience with the T.A.C.T. program (which involved the automation of the driving function of battery rail-car Gemini) has been a major influence.

The recommendations of this chapter should therefore be relevant in a wide range of future applications of train control, but may need to be modified somewhat in the case of high-speed intercity trains, or of low-speed freight-handling operations involving precision stopping.
10.1 GENERAL REVIEW OF TACHOMETER INSTRUMENTATION

10.1.1 Analog or Digital?

Tachometry (the measurement of angular velocity of functions of that variable) is a well-established and well-documented field of instrumentation. We now review the various types of analog and digital tachometer which are available. Broadly, analog tachometers rely on the principle of electro-magnetic induction. Rotation of the instrument's rotor produces a voltage (a.c. or d.c. depending on the design) which is proportional to the rate at which lines of magnetic flux are cut by the armature conductors. For A.T.O. systems in which a digital processor is employed, an analog tachometer's output must be passed through a sampler and analog-to-digital converter. Other disadvantages of analog tachometers in a train-borne application are as follows:

- the wide range over which accurate speed measurement is required (0.2 to 40 m/sec for the B.R.A.T.O. pilot scheme) exceeds the range obtainable accurately from commercially available analog tachometers
- either the signal is swamped by noise at the low-speed end or it exceeds the required voltage (normally 10 V) at high speeds
- the effects of electromagnetic interference (e.g. that caused by rapidly-changing traction currents) cause severe problems in analog tachometer systems
- if an analog velocity signal is used, it is not possible to derive from it sufficiently accurate acceleration and position signals, due to noise problems in the differentiation case and
inadequate voltage range and/or inaccuracy of integration in
the position case

the effects of temperature, ageing and accidental demagnetisa-
tion (e.g. due to shorting of the armature terminals) can alter
the effective flux density of the magnetic field, and thus the
accuracy of the measurement.

On the other hand, with digital tachometry, recent work on the
test railcar (Gemini) has demonstrated that it is quite possible
to derive position, velocity and acceleration data using the
sampled data derived from a single transducer. Various algorithms
for carrying out the required processing of the sampled data stream
are discussed by Forsythe (1979, 125) and have been applied success-
fully by British Rail.

With a digital design the required velocity range is readily achieved,
and the measurements are immune to noise and electromagnetic inter-
ference appearing in the measurement channel; the effects due to
ageing of components and materials are negligible in the case of
optical digital tachometers, and can be tolerated without loss of
accuracy in the case of electromagnetic digital instruments.

Temperature changes do alter the effective diameter of the steel
wheels whose angular velocity is being measured, regardless of
the tachometric system being used, but this effect is tolerable
for velocity and acceleration measurement, and can be reduced to
an acceptable level in the case of position measurement by the use
of fixed position reference markers in the track sufficiently often
to avoid unacceptable accumulations of the position error.
The acceleration signal derived from the tachometer by digital processing is very much less contaminated with stochastic fluctuations (noise) than that from an analog accelerometer on Gemini, to the extent that little or no filtering appears to be required.

If, however, filtering of the resultant velocity and acceleration data is required, then by the use of software alone such filtering is readily carried out. Gold and Rader (1969, 50) give clear procedures for developing digital software by which conventional filtering (low-pass, high-pass, band-pass and notch) can be carried out. Lewis (1975, 81) has applied this work in the specific context of railway vehicle instrumentation.

In addition to conventional filtering (which could in some cases be carried out by hardware processing of raw analog signals) Kalman and later researchers have developed exclusively digital techniques for the specific case where stochastic disturbances influence the behaviour of systems whose process dynamics are known. This work has been applied in the chemical process control field by McGreavy and Gill (1975, 49) to a process whose transfer function contains a combination of time delay (due to transport lag) and first-order exponential lag. These process dynamics are very akin to those of an automatically controlled train during the (most critical) braking phase. Some work on Kalman filtering in a train control context has been carried out by London Transport, but is not reported in the literature.
Whilst work with Gemini does not indicate that Kalman filtering is needed to cope with the stochastic disturbances imposed on the measurement channel, there is no doubt that such techniques could be useful in future applications, and their availability is a further powerful incentive to the use of digital techniques.

In summary, the existence of a variety of digital filtering options presents a strong argument for digital rather than analog instrumentation. It is significant that all automatic train control systems developed since 1960 (of which the Tokaido Line in Japan, London Transport's Victoria Line and San Francisco's Bay Area Rapid Transit are significant examples) have used digital tachometry, and the use of analog tachometers for any future system is unlikely.

10.1.2 Types of Digital Tachometer

Accepting that digital tachometers are preferable to analog instruments in the context of train state measurements for automatic train control, it is appropriate to consider the various types of digital tachometer which are available for this purpose. They may be grouped into five main classes:

a) capacitive
b) active electromagnetic
c) passive electromagnetic
e) absolute optical
f) incremental optical
10.1.2.1 Capacitive tachometers employ the variations in capacitance which occur when the air-gap between two electrical conductors is allowed to alter. These variations may be converted to variations of natural frequency of a tuned circuit, or variations of impedance of a circuit fed by constant-frequency a.c.

Capacitive transducers are ineffective in environments subject to harsh shock and vibration - even flexing of coaxial cable connecting them to the rest of the instrumentation system has been observed to invalidate the measurements - and no capacitive tachometers have been accepted for rail vehicle applications.

10.1.2.2 The active electromagnetic transducer is essentially a variable-reluctance device, in which a sensing coil is wound around a magnetic circuit whose mmf is provided by a permanent magnet.

The effective air-gap in the magnetic circuit varies as a toothed steel wheel rotates (typical variation 0.5 mm to 3 mm). The resulting flux variations induce an e.m.f. in the coil.

These devices are approved by both British Rail and London Transport; they are robust and give a reliable pulse stream at speeds varying from 0.5 to 100 m.p.h.

Their disadvantages are:

a) although the frequency of the output pulse stream is linear with speed, the amplitude is speed-dependent and tends to be obscured in noise at speeds below 0.5 m.p.h.; they are therefore unsuitable for position measurement, as the train may move
slowly without the change of position being registered.

b) the waveform of the pulse stream is characterised by sluggish leading and trailing edges, in contrast to the sharp square-wave of an optical device. This becomes significant when acceleration is to be derived from the pulse stream, as the exact moment of transition cannot be precisely registered.

c) the maximum number of poles (teeth) per revolution is about 120. This may not be adequate for the A.T.O. requirement, especially with regard to acceleration measurement.

d) heat and vibration tend to cause deterioration of performance due to reduction of the permanent magnet’s m.m.f.

If velocity only were to be measured, then the active electromagnetic tachometer (as used on Class 47 locomotives) would be a good solution, despite its drawbacks.

10.1.2.3 The passive electromagnetic system employs a similar magnetic circuit but detects the presence or absence of a metal body (e.g. a steel or aluminium cooling-fan blade) in the air-gap. The m.m.f. is provided by an external constant-frequency power supply and, in the absence of a metal body in the air-gap, a capacitor in series with the inductance of the device is adjusted for series resonance (low impedance). When the metal body is present in the air-gap, the impedance rises because of the increased inductance and/or effective resistance of the device.

The resultant variations of impedance are detected, processed and converted to a pulse stream whose frequency is a measure of the
speed of rotation of the device to which the metal bodies (e.g. fan-blades) are attached. This type of transducer works well down to zero speed, does not suffer from loss of m.m.f. due to temperature and vibration, but has the other disadvantages of the active variable-resistance tachometer mentioned previously. In addition, it is unlikely that a resolution of more than 50 cycles per revolution can be attained, which is too low for acceleration measurements to be derived. It is, however, suitable for both position and velocity measurement in an A.T.O. system.

This type of system has been used in Class 56 locomotives (in which application the passage of traction motor cooling fan blades was detected). It is, presumably, immune to the electromagnetic noise in such an environment.

10.1.2.4 Incremental optical tachometers (or encoders) produce a stream of pulses at a frequency proportional to velocity, by means of a rotating disc whose alternate opaque and transparent segments regularly interrupt a light path.

From work undertaken by British Rail using a proprietary incremental optical encoder with an output of two pulse streams in quadrature, it appears likely that longitudinal position, velocity and acceleration can all be obtained to the required accuracy for automatic train control (assuming the existence of periodic position reference markers installed in the track). In addition, direction of motion can also be readily obtained.

Simple optical encoders with slotted discs generating a pulse stream by interrupting the light path between a light-emitting diode and a
photo-transistor have been constructed and used for experimental purposes by the Instrumentation, Signalling and Automation sections of the Research Department of British Rail. Of these, two have failed due to contamination of the opto-electronics by grease and/or oil.

Proprietary models, suitably enclosed to protect against such contamination, and having up to 1000 poles per revolution, are available from several sources of supply in the U.K. Those manufactured by Ferranti Ltd. are the only ones made in the United Kingdom which satisfy stringent British and NATO military specifications for use in such applications as gun-laying on military tanks. Other suitable instruments are supplied by Rayleigh Instruments (made in W. Germany) and Vactric Baldwin) (made in U.S.A.) and by other manufacturers.

10.1.2.5 Absolute optical encoders are made by the same manufacturers as the incremental units discussed in 2.2.4, but produce a uniquely coded digital output representing angular position for each of up to 1024 angular segments.

For many applications this is an advantage over the incremental encoder, because the counting of pulses from a position reference is not required in order to determine angular position (as it is in the case of the incremental encoder). However, the absolute encoder is of no advantage in train control applications, because counting of the number of wheel revolutions from a fixed position reference on the track would still be required in order to determine longitudinal position.
As incremental encoders are cheaper than absolute encoders, they are to be preferred for train control purposes.

The main drawback of optical encoders is the need for external power supplies, and the limited life of the light source. Some redundancy may be required in order to detect failures safely.

10.1.2.6 Recommended tachometers

From the discussion above, and after study of relevant manufacturer's literature, the recommended tachometers to give complete information on position, velocity, acceleration and direction of motion to the required standard for automatic control are incremental optical encoders.
10.2 ERRORS IN TACHOMETER-DERIVED VELOCITY AND ACCELERATION SIGNALS

10.2.1 Introduction

This aspect of the instrumentation study examines the accuracy with which velocity and acceleration can be derived from a stream of tachometer pulses (typically produced by an axle-mounted magnetic or optical digital tachometer).

Apart from mechanical defects, relevant parameters which influence the accuracy are:

a) the resolution of the tachometer, i.e. the distance travelled for each pulse (q metres).

b) the time period over which tachometer pulses are counted in order to produce samples of distance (h seconds).

It is shown in the following sections that the selection of these parameters is quite critical, and depends largely on the expected values of acceleration, jerk (rate of change of acceleration) and higher derivatives which are likely to be encountered during the measurement. The error curves exhibit sharply defined minima, and it is the author's opinion that previous attempts within British Rail to measure acceleration from tacho pulse streams have failed because of injudicious selection of these parameters rather than as a result of a basic defect in the measuring technique.

In this section of the report a general discussion of the relevant factors is followed by an analytical section, from which a design procedure is established.
10.2.2 Mechanical Effects

Effective 'noise' is introduced into the tacho pulse stream by such factors as eccentricity of the wheel and/or tachometer, mechanical defects (backlash and/or springiness) in the coupling between axle and tachometer, and apparent variations in effective wheel diameter due to conicity and track misalignment. (Creep effects are quite negligible for velocity and acceleration, and it is assumed that wheel slip or slide is avoided by the use of unmotored and unbraked axles for tachometry).

This mechanical 'noise' has been considered in separate studies, and from them it has been deduced that it results in a lower useful limit of tacho resolution of approximately .004 m for a typical train with wheel circumference of 2 m.

If, in a particular A.T.O. project, it should turn out that mechanical effects are preventing the effective use of digital tachometers, then the following steps can readily be taken to alleviate the mechanical problems:

- **Excess Conicity**
  - a) Use of an axle with non-conical wheels (or wheels with less than normal conicity).
  - b) Averaging of tachometer data from wheels at both ends of an axle (this will reduce conicity effects but not track misalignment effects).

- **Wheel Eccentricity**
  - c) Selection of low eccentricity wheels for the tachometered axles.
  - d) More frequent replacement of tachometered wheels, if uneven wear causes eccentricity to become unacceptable.
e) Modification of instrumentation software so that acceleration is averaged over an integral number of wheel revolutions. (This would involve small-diameter wheels to avoid an excess sampling-time (h) at low speeds).

f) Mechanical redesign of the tachometer so that its rotor is rigidly attached to the axle-end.

If very high precision of acceleration measurement is required, then a special independently-suspended lightly-loaded non-conical wheel with integral tachometer could be used. This will not be necessary for any envisaged A.T.O. schemes, and could be incorporated at the rolling-stock design stage of any future proposal.

In summary, for the proposed B.R.A.T.O. system it seems likely that mechanical engineering factors will not prevent the measurement of acceleration to the required accuracy for control purposes. Gradual variations of wheel diameter may be compensated for by an automatic calibration routine.

10.2.3 Errors Arising due to Digital Differentiation Algorithms

The simplest algorithm for determining velocity is a two-sample formula:

\[ v = (x_i - x_{i-1}) / h \]
where: \( v \) = velocity
\( x_i \) = current position count
\( x_{i-1} \) = previous position count
\( h \) = sampling interval

This formula effectively estimates the velocity at the mid-point between the two most recent samples, and is in error by an amount dominated by \( ha/2 \), where \( a \) is the acceleration. By using three samples instead of two, a quadratic curve is effectively fitted between the three points, and the dominant error depends on the jerk (rate of change of acceleration).

Similarly, a four-sample algorithm (which fits a cubic to the distance curve) has no errors due to acceleration or jerk, but has an error term due to the rate of change of jerk (\( d^3x/dt^3 \)).

In general, an \( n \)-sample algorithm eliminates errors up to and including that arising from the \((n-1)^{th}\) time derivative of distance. Similar considerations apply to the algorithms for determining acceleration.

A clear derivation of the algorithms, and a discussion of the dominant errors of this type for algorithms up to the five-sample case, are presented by Forsythe et al (op cit). The error is quantitatively analysed by considering the extent to which the Taylor series approximation for the function \( x = f(t) \) is truncated. Forsythe has neatly generalised the analysis for the \( n \)-sample case using matrix methods. In the following discussion this type of error is referred to as the "Taylor series truncation error", \((Et)\).
A second type of error arises from the fact that the distance samples are quantised, due to the finite resolution of the tachometer (q metres/pulse). The measured distance sample is $x_i$, corresponding to an integer number of tacho-pulse counts, whereas the actual distance traversed from an initial starting point is $(x_i + q)$ metres. The quantisation errors in an ensemble of distance samples are uniformly distributed between 0 and q metres. The worst case arises when all terms of positive sign in the algorithm have error q and all terms of negative sign have zero error (or vice versa). From now on this second type of error is referred to as the worst-case quantisation error ($E_q$).

For all the algorithms the total worst-case error ($E$) in estimating velocity is obtained from the sum ($E_q + E_t$). It will be clear that whereas $E_t$ increases with increasing sampling-interval $h$, $E_q$ reduces with increasing $h$. In the next section we discuss clear-cut design formulae for minimising the total error.

### 10.2.4 A Real-Time Example (A.T.O)

From the three-sample real-time acceleration formula:

$$A_i = h^{-2} (x_i - 2x_{i-1} + x_{i-2})$$

the truncation error is given by

$$E_t = j_m h$$
where $j_m$ is the maximum jerk.

The quantisation error is given by:

$$E_q = 2qh^{-2}$$

where $q = \text{tacho resolution}$. 

Therefore $E = j_m h + 2qh^{-2}$

There is thus a trade-off, and for a particular resolution $q$ and maximum expected value of the relevant derivative (jerk in the example above) there is a sharply-defined minimum worst-case error, which is obtained by appropriately selecting the sampling interval ($h$).

For given values of $j_m$ and $q$, the value of sampling interval $h$ for minimum total error $E$ is obtained simply by differentiating $E$ with respect to $h$ and setting the result equal to zero.

Similarly, for a specified value of maximum allowable $E (E')$ and a given value of $j_m$, the appropriate values of resolution $q$ and sampling interval $h$ may be obtained as follows.

In the case of the three-sample acceleration algorithm considered above, the required sampling interval ($h$) is given by:

$$h = 2E'/(3j') \text{ seconds}$$

and the required tacho resolution is given by

$$q = h^3j'/4 \text{ metres}$$

where $E'$ = maximum acceleration error

$j'$ = maximum value of jerk.
Application of these design principles by British Rail has shown that, if errors due to mechanical effects may be neglected, measurement of velocity and acceleration sufficiently accurate for A.T.O. purposes may be obtained by digital differentiation of the distance samples.

10.2.5 A Non-Real-Time Example (Characterisation Tests)

A second example illustrates two important points which are usually neglected by those designing for digital tachometry:

a) if real-time measurements of velocity and acceleration are not required for on-line control, then non-causal algorithms (using 'future' as well as 'past' sample of distance) may be used on the recorded data. These "non-causal" algorithms result in reduced errors, enabling lower-resolution tachometers to be effectively used.

b) For measuring velocity and acceleration when the higher derivatives are very small (e.g. during coasting or when running at constant speed), the appropriate sampling intervals are quite long (10-30 seconds for typical railway applications). The author has noted that previous experimenters have tended to select an arbitrary sampling interval which is too short, resulting in very large errors due mainly to quantisation effects swamping the small differences in velocity which are being looked for.
A requirement existed to measure the deceleration during coasting of an AM4 electric multiple unit. (The results were to be used for assessing train resistance as part of the characterisation tests for the B.R.A.T.O. pilot scheme). 'Noisy' measurements from a seismic accelerometer were available, from which a rough estimate of the coefficients \( b, c \) and \( d \) in the following formula were obtained:

\[-a = b + cv + dv^2\]

where: \( a \) = acceleration (m/sec\(^2\))
\( v \) = velocity (m/sec)

It was known that the 'seismic' data contained errors due to a constant non-zero gradient as well as noise.

In order to select parameters for a three-sample tachometric acceleration formula, an estimate of likely maximum rate-of-change of jerk \( (dj/dt) \) was required. This was obtained by double differentiation of the acceleration formula, yielding

\[\left(\frac{dj}{dt}\right)_{\text{max}} = 2d.\]

It should be noted that the error due to jerk is zero, compared with the corresponding causal algorithm in which it is dominant. Following the same procedure as before, the worst-case error is:

\[E = 4q/h^2\]

and \( h' = 4\sqrt{2q/(dj/dt)_{\text{max}}} \)
Only one tacho was available, which when mounted on the AM4 axle-end gave a resolution of 0.098 metre/pulse. The above expressions yielded a sampling interval of 25 seconds, and a worst-case error of just under 0.004 metre/sec², which was acceptable. The tachometry system was installed and satisfactory results obtained and reported by Parkin (1976, 100).

To summarise, in this example unsatisfactory 'seismic' accelerometer data was used to form a rough estimate of train resistance, from which an estimate of likely rate-of-change of jerk was obtained. Although of doubtful accuracy, use of this quantity in these design equations yielded satisfactory acceleration results from a tachometer of relatively coarse resolution together with a 'non-causal' three-sample acceleration algorithm. (In the final analysis, a reasonable measure of track gradient was also obtained by subtracting the 'seismic' acceleration from the tacho acceleration - this was unforeseen spin-off from the test results). These results are reported in more detail by Thomas, Milroy and Forsythe (1979, 125).

10.3 Chapter Summary

To summarise this section, design procedures have been developed to minimise the errors which arise from the use of digital differentiation algorithms for determining train velocity and acceleration, and two practical examples (one involving the specifications for the B.R.A.T.O. pilot scheme and the other the measurement of very low accelerations during coasting) have shown that the output of a digital tachometer may be processed to yield satisfactory acceleration data from multiple-unit rolling stock.
Further practical results from these algorithms will be reported by P D Thomas in his doctoral thesis, which is now in preparation.
CHAPTER 11
SUMMARY, COMMENTS AND SUGGESTIONS FOR FURTHER WORK

11.1 Summary

This thesis has described the following aspects of research and development work carried out by the author in the field of railway automation.

* A review of published work in the field of Automatic Train Control (which includes Automatic Train Protection and Automatic Train Operation). The review concentrates on applications of and algorithms for digital processors in this field. See Chapter 2 and the list of over 130 references which follows this chapter.

* The design of supervisory control algorithms which adjust the set-point (desired velocity) of an automatic driving system so as to ensure on-time arrival at target points. See Chapter 3.

* The design of a time-tabling planning program which will minimise energy consumption over a journey consisting of several sections. This program exploits a variant of Bellman's Principle of Optimality, and assumes that the optimal control for each section can be determined. See Chapter 4.

* The study of the optimal control strategy which will minimise energy consumption at the wheel-rail interface, over a single journey section.
A technique for determining the optimal control is developed, which may be used either in the time-table planning program, or in the on-line control of a train. See Chapter 5.

* The design of controller algorithms to ensure that the velocity of the train follows a pre-determined profile (especially in the case of the braking phase). S-plane methods are used. Transport lag is not considered. See Chapter 7.

* The design of digital controller algorithms which can be used in the presence of transport lag, where the value of the transport lag is an integer number of sampling intervals. Z-plane methods are used. See Chapter 8.

* The proposition that non-integer values of transport lag may be handled as partial poles (or as complete poles and partial zeros) in the Z-domain. See Chapter 9.

* A brief treatment of the method developed by the author for deriving sufficiently accurate measurements of position, velocity and acceleration for A.T.O. purposes. See Chapter 10.
11.2 Comments

The work carried out by the author (an academic staff member who worked for a while in a railway organisation) is complemented by the work of P D Thomas (a practising railway control engineer) and K Tyler (a post-graduate research student working under the author's supervision). The thesis of P D Thomas will present many of the results obtained when algorithms of the type developed in this thesis are applied to a practical railway train. The thesis of K Tyler will present many of the simulation results associated with this work, and will evaluate the performance of the optimal control algorithms developed in this thesis against other algorithms based on the principles of dynamic programming and quadratic programming.

Taken together, the three theses will provide a balanced treatment of the application of modern control and digital simulation techniques to the automatic operation of railway trains.

This partition of the work between three researchers has been occasioned by the fact that these studies have been carried out as part of two different A.T.O. projects sponsored by British Rail and the South Australian Department of Transport respectively. All three authors have had to bend their researches to suit the exigencies of these externally funded projects and of their other professional commitments.

The recent award of equipment grants and research scholarships to the amount of $A61,000 to the author's group by the South Australian Minister of Transport, to support the next and more
practical phase of the Australian project is evidence that the research work described in this thesis has prospects of being applied in at least one "real railway".

11.3 Suggestions for Further Work

As mentioned above, it is pleasing that the project work described in this thesis has recently attracted substantial financial support from the South Australian Department of Transport.

The author's institution (the South Australian Institute of Technology) is by its charter particularly concerned with the transfer of technology from the pages of research journals into practical applications in industry. One of a number of stumbling-blocks in this process is the task of convincing a busy engineer in an industrial environment (a railway in this case) that some research findings couched in what is (to him) obscure mathematical jargon contain the seeds of a thoroughly practical down-to-earth project, with which he can become enthusiastically involved. Such an engineer needs to see control algorithms operating on real rotating machines.

A substantial portion of the funds granted will therefore be devoted to a laboratory dynamometer rig, on which it will be possible to mount actual traction systems of realistic scale (up to 30 KW). Arrangements are being provided to load the traction system in such a way that it can mechanically emulate the non-linear dynamics of railway vehicles, including the effects of mass, running resistance and gradients.
The dynamometer will be controlled by a mini-computer based system, which will also manage the data-logging and data-reduction tasks. The mini-computer is to be interfaced to the dynamometer via the standard international data bus known in the U.S.A. as the I.E.E.E. 488 or the G.P.I.B. bus. The mini-computer will also be interfaced to the Institute of Technology's central Cyber 73 computer, initially by a serial link. The dynamometer is now being built by the Australian subsidiary of the Swedish firm A.S.E.A.

In the context of the work described in this thesis, the rig will be used to demonstrate and evaluate potential train control hardware and software. Its expense could not of course be justified for this application alone, and it will also be used for a range of research and development activities in the fields of electrical machines and power electronics, and will be made available to South Australian industry for type testing of rotating electrical machines.

The South Australian State Transport Authority have recently acquired new diesel-electric trains for their Adelaide commuter services, and may in the longer term be acquiring modern electric light rail vehicles for a new dedicated route to the northeastern suburbs.

Using the research described in this thesis as a basis, it will be proposed that a project team be set up, comprising engineers from the State Transport Authority, the author, and an engineer from the State's Directorate of Transport Planning and Research,
to initiate the application of this work on a pilot scheme basis on an operating rail vehicle.

It may be seen, then, that the author's work has reached the stage where it should be taken from the groves of academe out into the fields of railway operations. It is hoped that this thesis will provide the vehicle for that transition.

The research work described here has also opened up a variety of future lines of study which may be appropriate for the next generation of research students in academic institutions. Energy having been minimised at the wheel-rail interface, there is scope for the application of control theory further back in the energy chain, to reduce losses in the electric power distribution system or the diesel prime mover. Scope also exists for developing refined tachometry software, so that the sampling interval and/or the effective number of poles in the tachometer system (both of which can readily be altered under software control) can be modified adaptively as the train proceeds on its journey.

It is appropriate to end a section entitled "Suggestions for Further Work" with words used by Rosenbrock (110, 1975) in addressing the sixth world congress of the International Federation of Automatic Control. His topic was "The Future of Control".

"To attempt to write on the future of control shows, perhaps, more rashness than good sense. Our chief difficulty in envisaging where the current changes in control technology may take us, is in being sufficiently audacious in our conjectures".

THE END
REFERENCES AND BIBLIOGRAPHY


139. Yoshihisa, "Railway Transit Planning Systems", Proc. IFAC, 6th World Congress.
ACKNOWLEDGEMENTS

The work described in this thesis would not have been possible without the support and assistance of a great many people in the following organisations, all of which provided a substantial amount of equipment, facilities, operating data, financial support or technical advice.

British Railways Board (Research Division)
Loughborough University of Technology
South Australian Department of Transport
South Australian Institute of Technology
South Australian State Transport Authority

Technical information was also provided by a number of railway authorities and companies operating in the railway signalling field.

Several people took a particularly important role in my work. They are my colleagues and friends in the Library and in the Schools of Electrical and Electronic Engineering at the South Australian Institute of Technology, my Loughborough University supervisor Bill Forsythe, my Australian supervisor Reg Underdown, British Rail research engineer Peter Thomas, research student Kim Tyler and Professor David Lee. To all these people, and especially to my wife Ann (sine qua non), I offer my special thanks.

Ian Milroy
Loughborough
July 1980
SOUTH AUSTRALIAN INSTITUTE OF TECHNOLOGY

AUTOMATIC CONTROL OF A TRACTION SYSTEM

USING AN

INTEL 8085 MICROPROCESSOR

M. GRIVELL

A project for the Graduate Diploma in Control and Measurement Systems.

MARCH, 1980.

SUPERVISOR: I.P. MILROY Senior Lecturer, School of Electrical Engineering.
List of Contents

1. Introduction 1
2. Aim of Project 1
3. Specification of Project 2
4. Equipment Available 2
   1. SDK-85 MCS-85 System Design Kit 2
   2. Analog Devices RT1200 - Real Time Interface 3
   3. 3 Phase 415V Siemens Motor Drive Controller 3
   4. Squirrel cage induction motor driving a D.C. generator 3
5. Design Concept 3
   1. Braking point 4
   2. Estimated time of arrival of train at station 4
   3. Train dynamics 4
   4. Block diagram of system 5
   5. Controller Algorithm 5
6. Hardware Design 6
   1. Input to microprocessor 6
      1. Selection of n and sampling interval 6
      2. Input circuit 7
   2. 8080 bus from 8085 bus 8
   3. Buffer area on SDK85 kit 8
   4. Generation of sampling time 9
7. Software Design 10
   1. Overall flow chart 10
      1. Initialisation module 10
      2. Supervisory module 10
      3. Controller module 11
   2. Floating point package 11
   3. Flow charts of program modules 12
(a) Main program

(b) Subroutines

1. CDIST
2. UPDATE
3. XBCAL
4. TECAL
5. CVREF
6. CVE
7. CDIFF
8. CINT
9. CPROP
10. LDAC

8. Results

9. References

<table>
<thead>
<tr>
<th>List of Appendices</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Glossary of Symbols used</td>
</tr>
<tr>
<td>B Mathematical derivation of distance at which braking is to be applied</td>
</tr>
<tr>
<td>C Derivation of transfer function of induction motor</td>
</tr>
<tr>
<td>D Determining Kp, Ki and Kd</td>
</tr>
<tr>
<td>E Differentiation on a digital computer</td>
</tr>
<tr>
<td>F Integration on a digital computer</td>
</tr>
<tr>
<td>G Assembly language listing</td>
</tr>
</tbody>
</table>
1. Introduction

As Australia enters into the period of the 1980's, large increases in the price of fuel have caused the Railway industry to look more critically at increased fuel economy. By reducing the quantity of fuel that the industry uses, considerable cost savings would result and this will allow the industry to either decrease their freight charges or alternatively keep their charges steady thereby becoming more competitive with the road and air transport industries. Ultimately, this would lead to business being attracted away from the road and air industries who are both large consumers of vast quantities of fuel.

One way that the railway industry could economise on the quantity of fuel used is to automatically control the driving function of trains and consequently allow the trains to operate at their optimum performance with regard to fuel economy.

This paper describes a system for automatically controlling the driving function of a train, the automation being digitally processed by an Intel 8085 microprocessor.

The work was carried out for the Electrical Engineering School, the South Australian Institute of Technology located at the Levels campus as part requirement for the Graduate Diploma in Control and Measurement Systems.

2. Aim of Project

The aim of the project was to design and develop the interface hardware to a microprocessor and to program the microprocessor to receive information such as instantaneous velocity, distance, time or any other information so as to automatically speed control the journey of a simulated train.

The simulated train was required to leave a station, accelerate to a set velocity and after coasting or maintaining the set velocity as determined by the program decelerate to stop at a following station at a predetermined time.

The control algorithm used was required to cope with typical disturbances such as adverse gradients. The braking system was to be emulated by a laboratory rig consisting of a variable frequency supply and an induction motor; the train dynamics, resistance and gradients encountered were to be emulated by appropriate loads to the d.c. generator.
3. Specification of Project

The specification of the system was as follows:

(a) The target distance \( X_T \) to be no greater than 5 km with an acceptable error of ± 0.5 m (see Figure 1).

(b) The target time \( t_T \) to have a resolution of 5 s (see Figure 2).

(c) The maximum velocity that the train can travel is to be approximately 20 m/s.

(d) The deceleration \( B \) to be no greater than 0.8 m/s.

With regard to transient response during the braking phase:

(e) The damping coefficient \( \xi \) to be 0.7.

(f) The natural frequency \( W_n \) to be 4 rad/s.

4. Equipment Available

The following equipment was made available by the Electrical Engineering School of the South Australian Institute of Technology located at the Levels Campus, Pooraka.

(1) **SDK-85 MCS-85 System Design Kit**

The SDK-85 MCS-85 System Design Kit is a complete single board 8 bit microcomputer system capable of addressing up to 64 K bytes of memory and having a 6-digit LED display and a 24-key keyboard for direct insertion, examination and execution of a user's program. In addition, the SDK-85 can be directly interfaced with a teletype terminal and the system monitor is included in a pre-programmed ROM for general software utilities and system diagnostics. The Central Processing Unit is an Intel 8085A Single chip 8 bit N channel microprocessor and the Instruction cycle time is 1.3 us with a clock cycle time of 330 ns. Data is transferred on an 8 bit bidirectional tri-state bus which is time multiplexed so as to also transmit 8 lower order address bits. An additional 8 lines expand the system memory addressing capability to 16 bits, thereby allowing 64 K bytes of memory to be accessed directly by the CPU. For a more complete description of the SDK-85 MCS-85 System Design Kit, refer to references 1 and 2.
(2) **Analog Devices RTI 1200 Real Time Interface**

This is a real-time interface that facilitates the task of interfacing analog input and output signals to microcomputers used in applications such as process and machine control, laboratory experimentation test equipment and medical instruments. For a complete description of this device, the reader is referred to reference 3.

(3) **3 Phase 415V Siemens Motor Drive Controller**

This unit is a variable-voltage variable-frequency supply.

(4) **Squirrel Cage Induction Motor** driving a d.c. generator with appropriate instrumentation.

Unfortunately, the Siemens motor drive controller developed a fault during the year and had not been repaired. It was therefore recommended by the supervisors that items 4.3 and 4.4 would be simulated using an analog computer model of the laboratory vehicle speed drive.

5. **Design Concept**

The typical velocity versus distance and velocity versus time profiles of a train departing from a station at time $t=0$ and arriving at the following station at time $t=T$ is as shown in Figs 1 and 2. For simplicity the profiles are drawn assuming that both constant acceleration and deceleration occur.

Referring to both figs 1 and 2, the train begins to increase its speed with a constant acceleration $a$ from $v=0$ at $t=0$ to $v=V$ at $t=t_1$. At this instant in time, the train will have travelled a distance $x_A$ from the originating station. The train then coast or maintains the set speed until $t=t_2$ i.e. $x=x_B$. At this instant in time constant deceleration $B_d$ or braking is applied and the velocity of the train decreases from $v=V$ at $t=t_2$ until the train arrives at the following station at $t=T$ where $v=0$. 
(1) **Braking Point**

In Appendix B it is shown that the distance from the station at which braking is to be applied can be readily computed from the relationship

\[ x_B = x_T - \frac{V^2}{2B_d} \]  

where:

- \( x_T \) = target distance (m)
- \( x_B \) = distance at which braking is to be applied (m)
- \( V \) = current velocity of train (m/s)
- \( B_d \) = deceleration (m/s^2)

(2) **Estimated Time of Arrival of Train at Station**

The estimated time of arrival of the train at a following station or for the train to reach a target velocity \( V_T \) can be computed from the relationship

\[ t_e = t^1 + \frac{x_B - x^1}{V} + \frac{V_T - V}{B_d} \]  

where:

- \( t_e \) = estimated time of arrival of train at station (s)
- \( t^1 \) = current time (s)
- \( V_T \) = target velocity (m/s)
- \( x^1 \) = current position (m)

Note: If the train stops at the following station, then \( V_T = 0 \) and equation (2) simplifies to

\[ t_e = t^1 + \frac{x_B - x^1}{V} - \frac{V}{B_d} \]  

(3) **Train Dynamics**

The acceleration of a train of mass \( M \) is expressed by the relationship

\[ F - Mg\sin \theta - R(v) = M\ddot{x} \]  

where:

- \( F \) = force (N)
- \( M \) = mass (Kg)
- \( \theta \) = angle of gradient (rad)
- \( R(v) \) = resistance of wheels of train on track which is a function of the velocity \( v \). Fig 3 shows the general shape of this function.
R(v) can be expressed mathematically by

\[ R(v) = K + K_1v + K_2v^2 + K_3v^3 + \ldots \]

where \( K, K_1, K_2 \ldots K_n \) are constants

(4) Block Diagram of System

Using equation (4), the block diagram of the system can be readily drawn (see fig 4).

The laboratory rig which is to be automatically controlled is simulated by a 3 phase 415 Volt induction motor driven by a 3 phase motor controller (Siemens unit) which produces a frequency and output voltage that are directly proportional to the input voltage derived from a digital 3 term controller i.e. a digital controller containing proportional, integral and differential control.

The velocity \( V \) is determined by the program and the difference between this velocity and the actual velocity \( v \) of the train is obtained to determine the error velocity input to the 3 term controller. The velocity \( v \) of the induction motor is determined by a shaft encoder that produces \( 2^n \) pulses/revolution where \( n \) lies between 1 and 10 and the appropriate value of \( n \) may be selected. For reasons explained in Section 6, 64 pulses/revolutions were chosen to measure the velocity \( v \) of the induction motor.

(5) Controller Algorithm

The controller algorithm used in the system is as given in equation (6).

\[ \text{Control algorithm} = K_p + \frac{K_i}{s} + K_d s \]

where \( K_p \) = proportionality constant

\( K_i \) = integral constant

\( K_d \) = differential constant

\( s \) = Laplace operator

The control algorithm was digitally computed using the bi-linear transformation

\[ S = \frac{2(Z - 1)}{T(Z + 1)} \]

for converting from the S domain to the Z domain

where \( Z \) = Z transform operator

\( T \) = sampling interval (s)

The method used to determine the values \( K_p, K_i \) and \( K_d \) is given in Appendix D and for the laboratory rig the following values were obtained.
Kp = 1.544
Ki = 13.55
Kd = 0.486

6. Hardware Design

(1) Input to Microprocessor

(i) Selection of n and sampling interval

The shaft encoder produces $2^n$ pulses/revolution where $n$ is any number between 0 and 10 and may be selected. Assuming no slip, the maximum velocity of the induction motor is 1500 rpm and, consequently, in $t$ seconds, the motor will have rotated $25t$ revolutions ($= 9000t^0$). The Intel 8085 microprocessor has an 8 bit data bus and for maximum resolution, 256 bits of information are possible. Hence, the number of bits per revolution $Q$ is given by the relationship

$$Q = \frac{256}{25t} \text{ bits/revolution}$$

$$= \frac{10.24}{t} \text{ bits/revolution}$$  \hspace{1cm} (8)

where $t = \text{time (s)}$

Using equation (8), Table 1 may be produced.

<table>
<thead>
<tr>
<th>Sampling time (secs)</th>
<th>$Q$ bits/revolution</th>
<th>n (for maximum resolution)</th>
<th>no. of degrees between encoder samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.24</td>
<td>8</td>
<td>45°</td>
</tr>
<tr>
<td>0.5</td>
<td>20.48</td>
<td>16</td>
<td>22.5°</td>
</tr>
<tr>
<td>0.25</td>
<td>40.96</td>
<td>32</td>
<td>11.25</td>
</tr>
<tr>
<td>.1</td>
<td>102.4</td>
<td>64</td>
<td>5.625</td>
</tr>
<tr>
<td>.05</td>
<td>204.8</td>
<td>128</td>
<td>2.813</td>
</tr>
</tbody>
</table>

Table 1

For good response of the system, it is essential to choose the sampling time $\Delta t$ such that

$$\Delta t \ll \tau_{\text{rig}}$$  \hspace{1cm} (9)

where $\tau_{\text{rig}} = \text{time constant of laboratory rig}$
A sampling time of 0.1s was chosen for the following reasons:

(1) The value of $\tau_{rig}$ had been deduced from previous experiments and was found to be in the order of 0.25 s.

(2) The train (or laboratory rig) controller was to be effected using a microprocessor and a floating point package was to be used to carry out mathematical operations such as multiplication and division. It was therefore necessary to allow sufficient time for the microprocessor to complete the program run before the next sampling interval. From Table 1, for the sampling interval of 0.1s it is necessary to choose $n = 6$ (i.e. 64 bits/revolution) if maximum resolution is required.

(ii) Input circuit

Fig 5 shows the circuit diagram of the input to the microprocessor. The operation of the circuit is as follows:

On arrival of each pulse from the encoder, the monostable multivibrator (Type 74121) generates a pulse having a 100 nS pulse width. The pulses are fed to the input of two 4 bit counters (74LS193) that are connected in cascade. The two counters sequentially count the number of pulses produced by the monostable multivibrator within the sampling interval. This count is read into the microprocessor at the end of the sampling interval and within 10 uS the CLEAR line goes high (logical 1) for several microseconds to reset the counters to zero. When the clear line has returned to logical 0, the counters are ready to receive information for the next sampling interval.

As the encoder generates 64 pulses/revolutions the minimum time between encoder pulses is approximately .6 ms. So as to ensure that no pulses are lost in the reading process, it is essential that the counters be reset within a time which is very much less than 0.6 ms. This requirement is well satisfied.
(2) 8080 Bus from 8085 Bus

The Analog Devices RTI1200 Real Time Interface used in the project is directly compatible with the Intel 8080 microprocessor bus but not directly compatible with the Intel 8085 microprocessor bus. As an Intel 8085 microprocessor was used, it was necessary to design appropriate interface circuitry so as to make the RTI1200 operate with the 8085 bus. Fig 6 shows the pin connections from the microprocessor to the Bus Interface connector of the RTI1200.

The state of control signals for both memory read and memory write operations in the 8085 microprocessor is given in Table 2.

<table>
<thead>
<tr>
<th>IO/m</th>
<th>RD</th>
<th>WR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEMR</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MEMW</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2

State of Control signals for MEMR and MEMW operations

Using the control signals as given in Table 2, the design of logic circuitry that will generate MEMR and MEMW signals is easily accomplished. The circuit diagram is given in fig 7.

(3) Buffer Area on SDK85 kit

In order to read and write to the RTI1200, it is essential that its addresses should occupy 1K of the 64K memory address range accessed by the microprocessor.

Fig 8 shows the memory map of the microprocessor and it is seen that the RTI1200 must be placed in the memory space where the expansion buffers are enabled i.e. from memory location 8000H to FFFFH. For convenience, it was decided to use 1K by tes of the address range starting from memory location 8000H. As a consequence, the integrated circuits required for the buffer area on the SDK85 kit had to be mounted and the appropriate jumpers either removed or connected (see reference 2).
(4) Generation of Sampling Time

As explained previously, a sampling time of 0.1 s was chosen. To generate this sampling time it was decided to use the timer located in the Programable Input-Output RAM device (Type 8155) mounted in the extension area of the SDK85 board. The timer is a 14 bit down counter that counts the TIMER IN pulses and provides either a square wave or pulse when terminal count (TC) is reached. To program the timer, its COUNT LENGTH register is loaded from the microprocessor first, one byte at a time by selecting the timer addresses. Bits 0 to 13 of the high order count register specify the length of the next count and bits 14 and 15 specify the timer output mode. There are four timer output modes to choose from.

These modes are:

<table>
<thead>
<tr>
<th>Mode</th>
<th>B15</th>
<th>B14</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>generates single square wave</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>generates continuous square waves</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>generates single pulse on TC</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>automatically reloads the counter on TC and continues operation</td>
</tr>
</tbody>
</table>

To generate the sampling interval time, the appropriate count was loaded into the count length register and mode 3 selected. The microprocessor has a 330ns clock cycle time and if each of the 14 bits of the counter was loaded with a 'one' (1) then the maximum sampling time available would be 6.39 ms.

Hence to obtain 0.1 s, it was necessary to extend this time by using two Up/down decade counters. The counters are connected so that one feeds into the other resulting in a "divide by 100" circuit. The circuit diagram is given in fig 9.

When TC is reached, the TIMER OUT signal goes low for one clock cycle (33 us) and then returns to logical 1. This pulse is fed to the RST6.5 interrupt pin on the buffered area of the SDK85 kit and thence to the 8085. The count length register is then immediately reloaded automatically with the appropriate count length and begins the next interval.
7. **Software Design**

(i) **Overall Flow Chart**

The overall flow chart of the system is given in fig 10.

There are 3 main modules:

(i) **Initialisation module**

(ii) **Supervisory module**

(iii) **Controller module**.

(i) **Initialisation Module**

In this module, all parameters used in the program together with the system parameters and the two programmable I/O devices mounted on the SDK85 board are initialised. It should be noted that initialisation is only accomplished at the beginning of the main program and is not repeated during each sampling interval.

(ii) **Supervisory Module**

At each sampling interval, on receiving an interrupt request this module carries out the following functions:

(a) The distance travelled by the train in one sampling interval is read into the microprocessor.

(b) The total distance travelled is computed together with a constant check kept on the remainder.

(c) The total travel time from t=0 is maintained.

(d) The estimated time of arrival $t_{e}$ of the train at the following station is computed from equation 2 and the velocity of the train is either increased, decreased or maintained at the set speed depending on whether the train is late, early or on time.

(e) If the train is late and its speed is equal to the permitted maximum speed then the speed remains at the permitted maximum speed.
(iii) **Controller Module**

This module is entered each time the supervisory module is completed and is also part of the interrupt servicing. In it, the following functions are performed:

(a) The difference between the actual speed of the train and the speed that the train should be travelling for it to reach the following station on time is computed. This velocity difference is known as the error velocity.

(b) The error velocity is input to a 3 term controller algorithm having proportional, differential and integral control, (see section 5.5).

(c) The output of the 3 term controller is loaded into the D/A converter located on the RT1200 to generate the input voltage to the 3 phase 415 volt Siemens motor drive controller.

(d) Return to wait loop in main program ready for next sampling pulse.

(2) **Floating Point Package**

The value of parameters and variables used in the program were of either fixed point or floating point format. Considerable time was spent in attempting to scale the values of each so that mathematical operations such as multiplication and division could be accomplished in fixed point format to yield integer results. This was not found to be satisfactory because of the wide range of parameter values likely to be encountered and consequently a floating point package was used. The floating point package used was a program available as a standard maths package supplied by the Intel Corporation and written by Mr Keith Caserta of the Proctor and Gamble Company Ohio U.S.A. (see reference 9). Prior to using the package it was necessary to test it to ensure that no program errors existed. This was accomplished using the IBM 370/3033 computer located at DRCS. Test programs were written for all operations available on the INTERP 80 package and their execution simulated using the resident Intel 8080 Emulator (see Ref 8).
(3) Flow Charts of Program Modules

The main program was written in 8085 assembly language in a modular form so that it would be more easy to detect programming faults and so that it would be easier to understand. A complete listing of the program is attached as Appendix G.

(a) Main program

The flow chart for the main program is given in fig 11.

Following the initialisation module, the program runs in a WAIT loop until timing interrupt is received (see section 6, Hardware Design). When the interrupt is detected, an unconditional jump is made to an instruction to read in the distance travelled in the sampling interval. The counters are then cleared and are then ready to receive information for the next sampling interval. The program runs through each subroutine and returns to the main code to await the next interrupt.

(b) Subroutines

The following is a brief description of each subroutine used in the program. The variable and parameter names referred to are the same as the symbols for the addresses storing those variables in the program.

(1) CDIST

A flow chart of CDIST is given in fig 12.

This subroutine determines the actual distance that the traction system has travelled and at the same time, maintains a constant check on the remainder from division by 10 (hexadecimal); if, when the remainder is added to the previous total remainder, the result is greater than 10 (hexadecimal), then the distance travelled is increased by 1 and bit 5 of the total remainder is changed to 0.

The distance travelled in a sampling interval (DIST) is fed into the microprocessor. The 4 LSB are removed and stored and the 4 MSB moved into the 4 LSB of DIST and then added to the total distance travelled (TDIST). The 4 LSB of DIST are then added to the total remainder (RDIST). A test is then made to check whether bit 5 of RDIST is 1. If it is, then 1 is added to TDIST and bit 5 of RDIST changed to 0.
(2) **UPDATE**

A flow chart of subroutine **UPDATE** is given in fig 13. The function of this subroutine is to maintain a check on the total time taken for the journey. A resolution of 5 s was specified for the system and as a sampling time of 0.1 s is used then 50 (32H) is loaded into the memory location TINIT in the initialisation mode. In the supervisory module, TINIT is loaded into the accumulator and then tested to see whether it is equal to zero. If it is, 1 is added to the total time TTIME, TINIT is reset to 32H and the program returns for the next instruction in the main program. If TINIT is not equal to zero, then it is decremented by 1 and the program returns for the next instruction in the main program.

(3) **XBCAL**

A flow chart of subroutine **XBCAL** is given in fig 14. This subroutine calculates the distance from the following station when braking is to be applied. The parameters $v$, $V_T$, $x_T$ and $B_d$ are loaded and converted into floating point form. $X_B$ is then calculated using the formula:

$$X_B = x_T - \frac{(V-V_T)^2}{2B_d}$$  \hspace{1cm} (10)

(4) **TECAL**

The flow chart of subroutine **TECAL** is given in fig 15. This subroutine calculates the estimated time of arrival of the train at the following station using the formula:

$$t_e = \frac{X_B - X^1}{V} + t^1 + \frac{V_T - V}{B_d}$$  \hspace{1cm} (11)

where $t^1 = $ current time (s)

$X^1 = $ current distance (m)
(5) **CVREF**

The flow chart of subroutine CVREF is given in fig 16. The subroutine tests whether

(a) \( X^1 > X_B \).

(b) \( t_e > t_T \).

(c) \( t_e < t_T - t_{TOL} \)

where \( t_{TOL} \) = permitted tolerance on time (s)

(a) If \( X^1 > X_B \) then the train has gone past the point at which braking should be applied and consequently the train would stop beyond the station. Hence the speed of the train is reduced.

(b) If \( t_e > t_T \) then the train is late and the speed of the train must be increased.

(c) If \( t_e < t_T - t_{TOL} \) then the train is early and the speed of the train must be decreased.

(6) **CVE**

The flow chart for this subroutine is given in fig 17. The subroutine tests whether the speed that the train should travel to arrive at the station on time VREF is greater than the permitted maximum speed. If this condition is satisfied, then the speed is set equal to the permitted maximum speed. In addition, the subroutine computes the error velocity by subtracting the current velocity from VREF.

(7) **CDIFF**

The flow chart for this subroutine is given in fig 18. This subroutine calculates the rate of change of the error velocity and multiplies the result by the differential constant \( K_d \). Refer to Appendix D for the method used to determine the rate of change of a function on a digital computer.
(8) **CINT**

The flow chart for this subroutine is given in fig 19. The subroutine calculates the integral of the error velocity and multiplies the result by the integral constant Ki. Refer to Appendix E for the method used to determine the integral of a function on a digital computer.

(9) **CPROP**

The flow chart for this subroutine is given in fig 20. The subroutine calculates the product of the error velocity and the proportionality constant Kp.

(10) **LDAC**

The flow chart of this subroutine is given in fig 21. The subroutine sums the results of subroutines CDIFF, CINT and CPROP and loads the result into the D/A converter located on the Analog Devices RTI1200. A voltage is then generated by the D/A converter between 0V and 10V. The output of this module is an analog signal, and is fed to the voltage-to-frequency converter (see Fig 4).

---

**RESULTS**

The interface hardware has been manufactured, and, when tested to check its operation was found to function correctly. At the time of writing a temporary fault existed in the RTI1200 Real Time Interface and as a consequence, an analog voltage was not generated at the output of the D/A converter. It was agreed that this would be corrected by S.A.I.T. technical staff. Static testing of the MEMR and NEWW logic circuitry and all other hardware, however, functioned correctly.

The time spent writing the program was far greater than expected. As mentioned previously, the program was written in modular form and each subroutine was tested using the Intel 8080 Simulator resident on the IBM 370/3033 computer
located at DRCS. All subroutines functioned correctly on this computer. A punched
paper tape of the whole program was then produced and the program transferred to the
Future-data Emulator plus prototype hardware system at the Levels. When run on this
Emulator, all modules functioned correctly and the main program called the modules
in the correct sequence without any obvious error.

There was insufficient time to completely test the system with the laboratory
rig and, by agreement with the supervisors this task has been handed over to a fresh
project student.
## REFERENCES

1. Intel Corporation  
   MCS-85 users manual 1977
2. Intel Corporation  
   SDK85 users manual 1977
3. Analog Devices  
   User's guide for the RTI-1200 Real Time Interface 1977
4. Forsythe, Milroy & Thomas  
   Towards Automatic Train Control S.A.I.T. School of Electrical Engineering Research Paper TRI/79
5. A.G. Bolton  
   Sampled Data Systems - The Z Transform S.A.I.T. School of Electrical Engineering 1978
6. G.E. Biggins  
   Operation of Intel 8080 Macro Assembled S.A.I.T. School of Electronic Engineering
7. R.F. Dancer  
   Development of Software for the Intel 8080 Microprocessor on the IBM/370 under TSO
8. Intel Corporation  
   Interp/80 User's manual
9. Intel Corporation  
   Insite Intel Users Library Vol 5/1 "8080 Floating Point Package with BCD conversion routine" Ref BC5
## Glossary of Symbols Used

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>acceleration</td>
<td>m/s²</td>
</tr>
<tr>
<td>δd</td>
<td>deceleration</td>
<td>m/s²</td>
</tr>
<tr>
<td>f</td>
<td>function of time</td>
<td>-</td>
</tr>
<tr>
<td>fₛ</td>
<td>synchronous frequency</td>
<td>Hz</td>
</tr>
<tr>
<td>F</td>
<td>force</td>
<td>N</td>
</tr>
<tr>
<td>g</td>
<td>gravitational constant</td>
<td>m/s²</td>
</tr>
<tr>
<td>J</td>
<td>moment of inertia</td>
<td>-</td>
</tr>
<tr>
<td>K</td>
<td>conversion constant</td>
<td>-</td>
</tr>
<tr>
<td>Kc</td>
<td>differential constant</td>
<td>-</td>
</tr>
<tr>
<td>Kd</td>
<td>integral constant</td>
<td>-</td>
</tr>
<tr>
<td>Kp</td>
<td>proportionality constant</td>
<td>-</td>
</tr>
<tr>
<td>Kn</td>
<td>conversion constant</td>
<td>-</td>
</tr>
<tr>
<td>n</td>
<td>slope</td>
<td>v/bit</td>
</tr>
<tr>
<td>m</td>
<td>mass</td>
<td>Kg</td>
</tr>
<tr>
<td>nₚ</td>
<td>maximum speed</td>
<td>rev/s</td>
</tr>
<tr>
<td>nₛ</td>
<td>synchronous speed</td>
<td>rev/s</td>
</tr>
<tr>
<td>n⁺</td>
<td>resistance as a function of velocity</td>
<td>-</td>
</tr>
<tr>
<td>S</td>
<td>Laplace operator</td>
<td>-</td>
</tr>
<tr>
<td>tₑ</td>
<td>estimated time of arrival</td>
<td>s</td>
</tr>
<tr>
<td>tₜ</td>
<td>target time</td>
<td>s</td>
</tr>
<tr>
<td>t¹</td>
<td>current time</td>
<td>s</td>
</tr>
<tr>
<td>t₂</td>
<td>time at which braking is to be applied</td>
<td>s</td>
</tr>
<tr>
<td>T¹</td>
<td>torque</td>
<td>Nm</td>
</tr>
<tr>
<td>v</td>
<td>velocity</td>
<td>m/s</td>
</tr>
<tr>
<td>Vin</td>
<td>input voltage to Siemens Motor Drive controller</td>
<td>V</td>
</tr>
<tr>
<td>v</td>
<td>set velocity</td>
<td>m/s</td>
</tr>
<tr>
<td>vₜ</td>
<td>target velocity</td>
<td>m/s</td>
</tr>
<tr>
<td>X¹</td>
<td>current distance</td>
<td>m</td>
</tr>
<tr>
<td>Xᵢ</td>
<td>instantaneous distance from following station at which braking is to be applied</td>
<td>m</td>
</tr>
</tbody>
</table>
## Glossary of Symbols Used (continued)

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>target distance</td>
<td>m</td>
</tr>
<tr>
<td>$e \ dt$</td>
<td>$Z$ transform operator</td>
<td>-</td>
</tr>
<tr>
<td>$z$</td>
<td>Damping coefficient</td>
<td>-</td>
</tr>
<tr>
<td>$\theta$</td>
<td>angle of gradient</td>
<td>rad</td>
</tr>
<tr>
<td>$\ddot{z}$</td>
<td>$a^2x/\ dt^2$</td>
<td>-</td>
</tr>
<tr>
<td>$n$</td>
<td>natural frequency</td>
<td>Hz</td>
</tr>
</tbody>
</table>
APPENDIX B

Mathematical Derivation of distance at which braking is to be applied

In order for the train that is being automatically controlled to stop at the following station at the correct position, it is essential that the distance from the station when braking is to be applied $X_B$ be computed every sampling interval. As the speed of the train varies so also will the distance $X_B$ vary.

Refer to figs 1 and 2

\[
\frac{dv}{dt} = -Bd \quad \text{(B.1)}
\]

\[
\frac{dx}{dt} = v \quad \text{(B.2)}
\]

\[
\frac{dv}{dx} = -\frac{Bd}{v} \quad \text{(B.3)}
\]

The differential equation (B.3) can be readily solved.

Let $v = V$, $x = x_B$

\[
v = \frac{V}{T} \quad x = x_T
\]

\[
\int_{V_T}^{V} v \, dv = \int_{x_T}^{x_B} -Bd \, dx \quad \text{(B.4)}
\]

\[
\frac{V^2}{2} \bigg|_{V_T}^{V} = -Bd \int_{x_T}^{x_B} \quad \text{(B.5)}
\]

\[
\frac{V^2}{2} - \frac{V_T^2}{2} = -Bd (x_B - x_T) \quad \text{(B.6)}
\]

If $V_T = 0$, i.e. the train has stopped, then

\[
\frac{V^2}{2} = Bd (x_T - x_B) \quad \text{(B.7)}
\]

And from equation (B.7), we get

\[
x_B = x_T - \frac{V^2}{2Bd} \quad \text{(B.8)}
\]
Appendix C

Derivation of Transfer Function of Induction Motor

A typical graph of torque vs speed of an induction motor is as shown in fig 22. The induction motor in the laboratory rig was a 3 phase 50 Hz 415V motor with a synchronous speed $n_s = 1500$ rpm and at maximum torque $T_m$, the speed $n_m$ is approximately equal to 1450 rpm. If operation of the motor is assumed to occur over the linear region of the characteristic i.e. region A-B in fig 22, then the slope of A-B is given by the relationship:

$$m = -\frac{T_m}{50} \text{Nm/rpm} \quad (C.1)$$

where $m =$ slope of region A-B

$T_m =$ maximum torque (Nm)

Hence the torque at any speed $N$ can be expressed by:

$$T_1 = T_m \left(\frac{n_s - n}{n_s}ight) \quad (C.2)$$

where $T_1 =$ Torque (Nm)

$n =$ speed at torque $T$ (rpm)

$n_s =$ synchronous speed (rpm)

Noting that the induction motor operated at 50 Hz equation (C.2) becomes:

$$T_1 = T_m \left(30.f_s - n\right) \quad (C.3)$$

where $f_s =$ synchronous frequency (Hz)

The Siemens motor drive controller that drives the induction motor has the following characteristics:

For $V_{in} = 0V$, $f = 0$ Hz

$V_{in} = 10V$ $f = 200$ Hz

Therefore the transfer function of the Siemens motor drive controller is given by:

$$f = 20 V_{in} \text{ Hz} \quad (C.4)$$

Substituting (C.4) into (C.3) gives:

$$T_1 = T_m \left(600 \frac{V_{in}}{f_s} - n\right) \quad (C.5)$$

$$T_1 = 12 \frac{T_m V_{in} - 2\pi T_m W}{3000} \quad (C.6)$$

where $W =$ angular velocity (rad/s)

$V_{in} =$ input voltage to motor speed controller
Given \( a = 12 \ \text{Tm} \)
\( b = \frac{2 \pi \ \text{Tm}}{5000} \)

Then equation (C.6) can be written as:

\[
T' = a \ \text{Vin} - b \ \omega
\]  
(C.7)

The block diagram representation of equation (C.7) is given in fig 23.

The transfer function of fig 23 is easily obtained:

\[
\frac{\bar{W}}{V} = \frac{\frac{a \cdot J_s}{b}}{1 + \frac{b}{J_s}} = \frac{a}{b} \cdot \frac{1}{1 + \left(\frac{J_s}{b}\right)}
\]  
(C.8)

\[
= \frac{K}{1 + \tau_s}
\]  
(C.9)

where \( K = \text{constant} = \frac{a}{b} \)

\( \tau = \text{time constant (s)} \)

\[
= \frac{J}{b}
\]
(C.10)

Tests were previously carried out on the induction motor and values for both \( K \) and \( \tau \) determined. Values for \( K \) and \( \tau \) are

\( K = 12 \)

\( \tau = 0.25 \ \text{s} \)
APPENDIX D

Determining \( K_p, K_i \) and \( K_d \)

Fig 24 shows the block diagram of the system.

To obtain the overall transfer function of the system, it is essential to obtain values for the constants \( K_c \) and \( K_t \).

(a) \( K_c \)

The D/A converter located on the RTI1200 Real Time Interface unit generates 10V when FFH or 255 in decimal is loaded.

\[ K_c = \frac{10}{255} \text{ Volts/bit} \quad (D.1) \]

(b) \( K_t \)

255 bits = 255 pulses/1 s

\[ = 2550 \text{ pulses/sec} \]

As the encoder generated 64 bits/revolution then

\[ 255 \text{ bits} = \frac{2550}{64} \text{ revs/s} \]

\[ = \frac{2 \pi \times 2550}{64} \text{ rad/s} \]

\[ \therefore 1 \text{ rad/s} = \frac{255 \times 64}{2 \pi \times 2550} \text{ bits} \]

\[ = \frac{3.2}{\pi} \]

\[ \therefore K_t = \frac{3.2}{\pi} \text{ bits/rad/s} \]

\[ = 1.019 \text{ bits/rad/s} \quad (D.2) \]

Fig 24 can be redrawn inserting values for \( K_c \) and \( K_t \) (see fig 25).

The open loop transfer function of fig 25 is given by:

\[
\frac{W}{W_{\text{ref}}} = 1.019 \times \frac{0.469}{1+258 s} (K_p + \frac{K_i}{s} + K_d s) \quad (D.3)
\]

\[
= 1.912 K_d \frac{(s^2 + K_p s + K_i)}{s(s+4)} \quad (D.4)
\]

Examination of equation (D.4) reveals that it has two poles occurring at \( s = 0 \) and \( s = -4 \).

The closed loop response of the system is required to have a damping coefficient \( \zeta = 0.7 \) and \( \omega_n = 4 \text{ rad/s} \).
Using these values, the position of the zeroes on the root locus (fig 26) can be readily obtained. These occur at \((-5, 4.8)\) and \((-5, -4.8)\).

The constants \(K_p\), \(K_i\) and \(K_d\) were then calculated using the method detailed in Coughanowr and Koppel "Process Systems Analysis and Control" Chapt 15 and yielded the following results:

\[
\begin{align*}
K_p &= 10.3 \\
K_i &= 23.73 \\
K_d &= 1.03
\end{align*}
\]
Fig 27 shows any function $e$ plotted against time $t$. If $e$ is sampled with sampling time $\Delta t$ where $\Delta t$ is small compared with the rate of change of $e$ and $e_i$ and $e_{i-1}$ represent the $i$th and $(i-1)$th sample, then the rate of change of $e$ with respect to time, $p$, can be represented by:

$$p_i = \frac{e_i - e_{i-1}}{\Delta t} \quad (E.1)$$

Now $e_{i-1}$ is the previous sample of $e_i$ and can be written using the Z transform as:

$$e_{i-1} = 2^{-1} e \quad (E.2)$$

Substituting (E.2) into (E.1) and neglecting the subscripts $i$ we have:

$$p = \frac{(1-2^{-1})e}{\Delta t} \quad (E.3)$$

Equation (E.3) can be implemented in the simple recursive structure of fig 28.
Refer again to fig 27.

The integral of the function $e$ can be readily obtained using trapezoidal approximation. Integration is a summation process and if the area under the curve in fig 27 is found, then this gives the integral of the function.

In the following derivation, the subscripts $i$ and $i-1$ represent the $i$th and $(i-1)$th sample.

Assume that the sampling interval $\Delta t$ is small and that no rapid changes of the function occur during the sampling interval. Then the slope of $A-B$ in fig 27 can be assumed to be linear and:

$$y_i = y_{i-1} + \frac{e_i + e_{i-1}}{2} \Delta t$$

where $y = \int dt$

$t =$ sampling interval (s)

Rewriting (F.1) using the $Z$ transform operator and noting that $e_{i-1} = z^{-1}e_i$ we have:

$$y_i = y_{i-1} z^{-1} + \frac{\Delta t}{2}(e_i + e_{i-1} z^{-1})$$

i.e. $y = \frac{\Delta t}{2}(1 + z^{-1})$ (F.3)

Equation (F.3) can be implemented in the simple recursive structure of fig 30.
**FIG. 1. VELOCITY VS DISTANCE PROFILE**

**FIG. 2. VELOCITY VS TIME PROFILE**

**FIG. 3. RESISTANCE OF WHEELS ON TRACK VS VELOCITY**
Fig. 1: Block Diagram of System
Resistance in Ohms
Capacitance in Farads

Monostable - gives pulse
\(~100\ \text{ns} \) wide
+5V

\[ \begin{align*}
14 & \quad 11 \quad 10 \\
74121 & \\
A1 & \quad A2 & B & GND \\
1 & \quad 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16
\end{align*} \]

1500p

\[ \begin{align*}
5 & \\
CPU & 74LS193 & \\
\end{align*} \]

47K

\[ \begin{align*}
14 & \quad GND & \quad Vcc & \quad CL & \quad 16 & \quad B
\end{align*} \]

\[ \begin{align*}
14 & \quad GND & \quad Vcc & \quad CL & \quad 16 & \quad B
\end{align*} \]

25 Pin plug

To Porta of 2155 on UP Board A16 (Port 21H)

Fig. 5 - Input to UP
### Component Side

<table>
<thead>
<tr>
<th>Signal Name</th>
<th>Mnemonic</th>
<th>Pin</th>
<th>Track Side</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logic Ground</td>
<td>GND</td>
<td>1 2</td>
<td>GND</td>
</tr>
<tr>
<td>+5V</td>
<td>VCC</td>
<td>3 4</td>
<td>VCC</td>
</tr>
<tr>
<td>Logic Power</td>
<td>VCC</td>
<td>5 6</td>
<td>VCC</td>
</tr>
<tr>
<td>+12V (*)</td>
<td>VDD</td>
<td>7 8</td>
<td>VDD</td>
</tr>
<tr>
<td>-5V (*)</td>
<td>VXI</td>
<td>9 10</td>
<td>VXI</td>
</tr>
<tr>
<td>Logic Ground</td>
<td>GND</td>
<td>11 12</td>
<td>GND</td>
</tr>
<tr>
<td>N/C</td>
<td></td>
<td>13 14</td>
<td>INIT/</td>
</tr>
<tr>
<td>N/C</td>
<td></td>
<td>15 16</td>
<td></td>
</tr>
<tr>
<td>Mem. Read</td>
<td>MROM/</td>
<td>19 20</td>
<td>MWTC/</td>
</tr>
<tr>
<td>N/C</td>
<td></td>
<td>21 22</td>
<td>N/C</td>
</tr>
<tr>
<td>Xfer Acknowl.</td>
<td>XACK/</td>
<td>23 24</td>
<td>INH 1/</td>
</tr>
<tr>
<td>Adv. Acknowl.</td>
<td>AACK/</td>
<td>25 26</td>
<td>INH 2/</td>
</tr>
<tr>
<td></td>
<td></td>
<td>27 28</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>29 30</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>31 32</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>33 34</td>
<td></td>
</tr>
<tr>
<td>Interrupt 6</td>
<td>INT 6/</td>
<td>35 36</td>
<td>INT 7/</td>
</tr>
<tr>
<td>Interrupt 4</td>
<td>INT 4/</td>
<td>37 38</td>
<td>INT 5/</td>
</tr>
<tr>
<td>Interrupt 2</td>
<td>INT 2/</td>
<td>39 40</td>
<td>INT 3/</td>
</tr>
<tr>
<td>Interrupt 0</td>
<td>INT 0/</td>
<td>41 42</td>
<td>INT 1/</td>
</tr>
<tr>
<td>Address Bus</td>
<td>ADR6/</td>
<td>43 44</td>
<td>ADR7/</td>
</tr>
<tr>
<td></td>
<td>ADR5/</td>
<td>45 46</td>
<td>ADR6/</td>
</tr>
<tr>
<td></td>
<td>ADR4/</td>
<td>47 48</td>
<td>ADR5/</td>
</tr>
<tr>
<td></td>
<td>ADR3/</td>
<td>49 50</td>
<td>ADR4/</td>
</tr>
<tr>
<td></td>
<td>ADR2/</td>
<td>51 52</td>
<td>ADR3/</td>
</tr>
<tr>
<td></td>
<td>ADR1/</td>
<td>53 54</td>
<td>ADR2/</td>
</tr>
<tr>
<td></td>
<td></td>
<td>55 56</td>
<td>ADR1/</td>
</tr>
<tr>
<td></td>
<td></td>
<td>57 58</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>59 60</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>61 62</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>63 64</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>65 66</td>
<td></td>
</tr>
<tr>
<td>Data Bus</td>
<td>DAT6/</td>
<td>67 68</td>
<td>DAT7/</td>
</tr>
<tr>
<td></td>
<td>DAT5/</td>
<td>69 70</td>
<td>DAT6/</td>
</tr>
<tr>
<td></td>
<td>DAT4/</td>
<td>71 72</td>
<td>DAT5/</td>
</tr>
<tr>
<td></td>
<td>DAT3/</td>
<td>73 74</td>
<td>DAT4/</td>
</tr>
<tr>
<td></td>
<td>DAT2/</td>
<td>75 76</td>
<td>DAT3/</td>
</tr>
<tr>
<td></td>
<td>DAT1/</td>
<td>77 78</td>
<td>DAT2/</td>
</tr>
<tr>
<td></td>
<td>DAT0/</td>
<td>79 80</td>
<td>DAT1/</td>
</tr>
<tr>
<td>Logic Ground</td>
<td>GND</td>
<td>81 82</td>
<td>GND</td>
</tr>
<tr>
<td>N/C</td>
<td></td>
<td>83 84</td>
<td>N/C</td>
</tr>
<tr>
<td>+5V</td>
<td>VCC</td>
<td>85 86</td>
<td>VCC</td>
</tr>
</tbody>
</table>

*Used only if optional ROM is installed.

Note: / following mnemonic indicate active low signal.
N/C indicates no connections made to that pin.
RESISTANCE IN OHMS
CAPACITANCE IN FARADS

+5V

74LS20

14 13 12 11 10 9 8
Vcc

1 2 3 4 5 6 7

MEMW

+5V

MEMR

WR

RD

10/1M

74LS04

Vcc

14 13 12 11 10 9 8

MEMR

TRUTH TABLE

<table>
<thead>
<tr>
<th>10/1M</th>
<th>RD</th>
<th>WR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEMR</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MEMW</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

3

DECODING CIRCUIT FOR MEMR & MEMW

8080 BUS FROM 8085 BUS

RAIL VOLTAGES

PIN 14 +5V

PIN 7 0V

0.12N
Figure 8: Memory Map of SDK85
RESISTANCE IN OHMS
CAPACITANCE IN FARADS

CLOCK 5
CPU

74LS192

4.7K

+5V

0V

Cpu

74LS192

4.7K

+5V

0V

PIN 16

+5V

0.12V

PIN B

0V

FREQUENCY DIVIDER (÷100)

FIG. 9

RAIL VOLTAGES
START

ENTER INITIAL VALUES

ENTER SAMPLING TIME & OUTPUT TO BISS

INITIALISE \( c, x, u \).

ENABLE INTERRUPTS

WAIT

ENTER ISR

READ IN ENCODER COUNT

OBTAIN VELOCITY & DISTANCE FROM ENCODER COUNT

UPDATE TIME \( t = c + T \)

DETERMINE \( V_{REF} \)

SET \( V_{REF} = V_{\text{MAX}} \)

\( V_{REF} \geq V_{\text{MAX}} \) YES \( \rightarrow \)

\( V_{\text{REF}} = V_{\text{MAX}} \) NO

CALC. \( V_{O} = V_{\text{REF}} - V \)

CALC. \( V_{O} \) FROM 3 TERM CONTROL ALGORITHM \( V_{o} = (K_P + \frac{K_I}{S} + K_dS)V_{e} \)

OUTPUT \( V_{O} \) TO FREQUENCY CONTROLLER

RETURN

INITIALISATION

MODULE I

SUPERVISORY

MODULE II

CONTROLLER

MODULE III

FIG 10 OVERALL FLOWCHART
START

LOAD STACK POINTER

INITIALISE PARAMETERS & I/P & O/P PORTS

ENABLE INTERRUPTS

WAIT

ENTER

READ DISTANCE TRAVELLED IN SAMPLING TIME & CLEAR COUNTERS

CALL COIST

CALL UPDATE

CALL XBGAL

CALL TBGAL

CALL CVREF

CALL CVE

CALL CDIFF

CALL CINT

CALL CPROP

CALL LDAC

RETURN

INTERUPT SERVICE ROUTINE WHICH IS ENTERED AT EACH SAMPLING TIME

FIG. 11. FLOWCHART OF MAIN PROGRAM
FIG. 12 FLOWCHART OF CDIST

1. Enter
2. Load distance travelled in sampling time
3. Remove 4 LSB & store (remainder)
4. Divide new result by 16 (10H)
5. Add to total distance travelled
6. Add remainder to total remainder
7. Is bit 5 of total remainder = 1?
   - No: return
   - Yes: add 1 to total distance travelled, change bit 5 of total remainder to 0

Total distance travelled = REMAINDER

Total remainder = TOTAL REMAINDER
LOAD SOH INTO TINIT

ENTER

LOAD TINIT

TINIT = 0H ?

NO

TINIT - 1

RETURN

YES

TTIME = TTIME + 1

SET TINIT = 32H

FIG 13 FLOWCHART OF UPDATE
Fig. 14 Flow chart of XBCAL

ENTER

LOAD \( v, V_T, x_T, B_d \)

\[ x_B = x_T - \frac{(v - v_T)^2}{2Bd} \]

STORE \( x_B \)

RETURN
FIG. 15 FLOWCHART FOR TECAL
**Fig 17. Flowchart for CVE**

1. **ENTER**
2. **LOAD** $V_{REF}$, $V_{MAX}$
3. **$V_{REF} > V_{MAX}$?**
   - **YES** $V_{REF} = V_{MAX}$
   - **NO**
4. **$V_e = V_{REF} - V$**
5. **RETURN**
FIG. 18 FLOWCHART FOR CDIFF
**Fig. 19** Flowchart for CINT
FIG. 20 FLOWCHART for CPROP

LOAD \( K_p, V_e \)

\[ PVE = K_p \times V_e \]

RETURN
Fig. 21 Flowchart of Subroutine LDAC
6.22 TYPICAL TORQUE VS SPEED CHARACTERISTIC OF MOTOR
Fig. 23 Block Diagram of Equation (6.7)
**Fig. 24 Block Diagram of System**

![Block Diagram of System](image)

**Fig. 25 Block Diagram of System**

![Block Diagram of System](image)
FIG 26  ROOT LOCUS OF EQUATION (0.4)

- $l_1 = 1.65 \text{ cm}$
- $l_2 = 11.15 \text{ cm}$
- $l_3 = 5.08 \text{ cm}$
- $l_4 = 7.18 \text{ cm}$
**FIG. 27**  A. GRAPH OF ANY FUNCTION $e$ VS. TIME ($s$)
FIG. 28  BLOCK DIAGRAM OF METHOD USED TO DETERMINE
RATE OF CHANGE OF A FUNCTION ON A DIGITAL
COMPUTER

FIG. 29  BLOCK DIAGRAM OF METHOD USED TO DETERMINE
INTEGRAL OF A FUNCTION ON A DIGITAL COMPUTER
Towards Automatic Train Control

W. FORSYTHE
Lecturer, Department of Electrical and Electronic Engineering, Loughborough University of Technology, U.K.

I.P. MILROY
Senior Lecturer, School of Electrical Engineering, South Australian Institute of Technology

and

P.D. THOMAS
Research Engineer, British Rail Technical Centre, Derby, U.K.

ABSTRACT

The purpose of this paper is to provide a review of recent developments in automatic train control around the world, and to discuss some aspects of railway automation in an Australian context; the discussions indicate that, while full automation may be justified on urban mass-transit systems, automatic control of individual trains (by means of microprocessor-based driving aids) is an appropriate intermediate development for Australian non-urban railway networks. Such a development would lead to improved regulation of traffic and substantial savings of energy.

The paper concludes with a brief theoretical section which should provide a useful starting point for action and signalling engineers who wish to investigate the possibilities of more precise control of traffic than the conventional fixed-block signalling system will allow.

The authors are involved with research and development projects in the field of automatic train control systems at the British Rail Technical Centre (U.K.), Loughborough University of Technology (U.K.) and South Australian Institute of Technology.

INTRODUCTION

Railways in Australia

Ways are the safest forms of land transport from walking (1), and by far the most energy-efficient (2). A railway has much less environmental impact than a road system designed to carry the same passenger or freight capacity.

The 25 KV electrification program proceeds in Australia, railways will increasingly depend on relatively abundant supplies of coal as a primary fuel source, rather than scarce and costly oil by transport planners than in the past and we can expect to see massive investment in existing Australian rail facilities over the next decades at least.

Railway Automation

There is to be a large investment in railways in Australia, and on historical trends it is likely that between five and eight per cent of this investment will be in the area of automation and signalling. During this century, major developments in this area have taken as a result of train accidents, and failures of signalling and automatic braking systems resulted in safety of a high order. New improvements in safety systems can be expected, but the relevant design and development will probably take place outside Australia, in the laboratories of the half-dozen or so international companies which dominate the international railway signalling field.

However, the evolution of railway systems in railways has, in the authors' opinion, reached a stage where there is scope for more precise control of train movements than can be achieved by means of the conventional fixed-block signalling system (which is an excellent safety system but is too coarsely quantised in distance for effective regulation of traffic). The need for better regulation of traffic is particularly apparent in the large metropolitan centres (e.g. in the Sydney circle where 90-second headways are quite common during the commuter peaks, and delays due to conflicting train movements at junctions are common) and on the heavy-haul freight runs where the costs both in energy and in time through trains encountering red signals (e.g. at passing loops on single-line track) are unacceptable. Before discussing technical aspects of automatic and semi-automatic train-control systems which should alleviate or even cure these problems, we present a brief review of recent developments in automatic train control around the world.

1.3. Automatic Train Operation - a brief historical perspective

1.3.1. The Sixties

In 1963, at an international symposium on railway automation, independent papers by Shinohara, Schmitz and Lagershausen (4) indicated that railway authorities in Japan and West Germany were conducting major development projects leading to automation of the driving function. The factors leading to this development are clearly explained in these papers, and were as follows:

(1) the need to control groups of trains to avoid conflicting train movements (e.g. at merging junctions), and the inadequacy of the fixed block signalling system for such fine control.

(2) the need to correct disturbances to the working timetable in such a way as to minimise the propagation of such disturbances through the network.

(3) the fact that, in order to control groups of trains, it is necessary to be able to control
the speed and position of individual trains on the network.

3) the ergonomic problem that, at speeds then envisaged for inter-city passenger trains (180 Km/hr), the reliable detection of wayside signal aspects places undue stress on drivers, especially in poor weather conditions.

At that time the need to conserve energy did not appear to be considered as an important factor; it is probably the major factor today. Since then, it has also been recognised that a solution to the ergonomic problem is to repeat the wayside signal aspects in the driver's cab, and this solution (rather than full automation) has been adopted for many of the high-speed trains now in service in Europe.

In the second half of the nineteen-sixties completely automatic trains were introduced into service, initially on the Tokyo-Osaka run (the bullet train) and on the London Transport metro system (the Victoria line). These systems reflected the state of the art in industrial control systems generally, and had the following features which would not apply to a system being designed today:

1) the instrumentation, communications and control systems were predominately analog in nature.

2) although the hierarchical nature of train control was recognised, little of the control function was devolved out to individual trains. All velocity set-points were computed at a central control room, or stored at fixed locations on the track and transmitted to individual trains as they passed those locations.

3) control of traction and braking (especially traction) was intermittent and "noisy", with only a limited number of discrete levels of tractive effort being available to the control system.

4) the systems worked to a fixed timetable, with no provision for on-line adjustment of velocity set-points to correct for early or late running.

5) the systems did not cope with mixed traffic, being pre-programmed on the basis that all trains were of the same type, with similar traction and braking characteristics.

6) above points, made with hindsight in the text of today's power electronic systems and on-board digital processors, may appear critical. In fact the analog systems introduced in the nineteen-sixties were outstandingly successful and some are still in service.

1.2. The Seventies

During the seventies automatic train operation has been applied throughout the extensive and rapidly-expanding Shinkansen inter-city network in Japan, on a number of newly-engineered routes in Europe. The main change since the sixties has been in the introduction of digital instrumentation, communication and control. Train position and velocity are now sensed by incremental encoders, which deliver a pulse stream at a frequency proportional to velocity, and acceleration may also be derived by digital differentiation and filtering of the data from these encoders (5) rather than from the "seismic" accelerometers fitted, for example, to trains on the Victoria line. Data communications between control centres and trains are now mostly digital.

A special-purpose digital computer was designed by Toshiba (6) for on-board use, and Kuto and Kamada (7) describe in some detail algorithms developed to exploit the availability of a train-borne digital processor for such purposes as:

1) downward adjustment of velocity set-point to correct for early running.

2) logging of train operational data.

3) selective display of information to the driver via C.R.T.

Fast trains with on-board digital processors are now running throughout the Shinkansen network in Japan.

Meanwhile, in San Francisco, U.S.A., the Bay Area Rapid Transit System (8) came into operation in 1972; although plagued by a succession of maintenance problems with the cars, and difficulties in convincing the State authorities that the train protection and control systems were demonstrably "fail-safe", the BART system represents a watershed point in U.S. railroad development, and the digital monitoring and control algorithms now in use are comparable with those in Japan and Europe. BART, however, has not devolved the control task to individual train-borne processors, with the result that a train's operation is heavily dependent on the integrity of communications between trains and control centres. However, the BART administration has been very open in discussing the early failings of the system (9) with the result that other rapid-transit authorities can install their systems having learnt from the teaching of BART. The current state of the automatic rapid-transit business is very healthy, with a large number of cities of over three-million people now placing contracts to demonstrate their conviction that automatically-controlled electric trains are a major part of the answer to their people-moving problem. (Systems are currently being installed in Toronto, Hong Kong and Melbourne). Even in smaller cities such as Adelaide, there is a distinct possibility that partially automated light-rail vehicles (3), similar to trains, but running on dedicated track, will be the preferred solution for linking the downtown centre with suburban transport nodes, at each of which a bus system will be centred.

Meanwhile, the older "metro" systems such as the Sydney network, the London Underground system, the Paris metro and the Chicago Loop are gradually modernising their signalling, control, and passenger-information systems, although constrained by the massive investment which they already have in equipment which was not designed with automation in mind.

In France and Germany automatic control of trains on long distance runs as well as local journeys (10, 11) also developed rapidly during the early seventies. The French have used automatic speed-control on their inter-city trains for many years, and it is a fairly simple evolutionary step
have the set-point (desired velocity) of a chopper-controlled system determined by a supervisory system rather than by the driver.

2.1. Urban Railways

The problems facing transport authorities in Sydney and Melbourne, are in general identical to those facing up to a hundred cities of similar size throughout the world. There is no doubt that extensive investment in modern signalling and automatic-control equipment is justified, to improve traffic capacity and regulation in the inner-city zones without excessive investment in new civil engineering works. Takaoka (14) gives a good summary of the benefits to be gained from modern automatic systems controlling groups of trains, in the context of the Yokohama municipal subway. His article would apply equally to the Melbourne Underground Rail Loop, or the Sydney Circle.

2.2. Non-urban Railways

Australian engineers have always been extraordinarily inventive in dealing with the technical problems arising from the supply of services to a sparsely-distributed population outside the main capital cities.

It is therefore reasonable to suggest that similar inventiveness should be applied to the development of low cost automatic or semi-automatic systems to improve operations on the many thousands of kilometres of single-line railway track which link the States. One of us (IPM) has been fortunate in having the opportunity to discuss this matter informally with traction and signal engineers from railway authorities in Australia and from overseas, who gathered in Australia to participate in the Heavy Haul Railways Conference (Perth, September 1978). We now summarise the factors which seem to us important in applying modern electronic control systems to remote railways.

(1) There is an enormous investment in existing fixed-block signalling equipment on the Railways of Australia network, and on the private railways which haul bulk materials from mine to port in Western Australia. This equipment is extremely effective in its main function of ensuring safe operations. Any new systems installed for the purpose of conserving energy, or for improved regulation of traffic, should be compatible with the existing fixed-block signalling systems and should operate within the safety envelope provided by such systems.

(2) Full automation of these railways is completely out of the question. Because of the prohibitive cost and impracticality of secure fencing of the track it will always be necessary for there to be a person in the cab to provide a lookout. Such a person should carry out other functions on the train, to justify the associated labor cost and to provide job satisfaction.

A system for conserving energy and improving...
traffic regulation is required; this should be designed taking ergonomic factors into account, i.e. it should act as an aid to the person in the cab and his or her skills should be exploited. The safe handling of long freight trains is an example of a skilled activity for which no practicable automatic system can be envisaged.

Accurate control of time-of-arrival at critical points (e.g. passing loops on single-line railways, and merging junctions) is seen as a key control function which is difficult to achieve manually.

A train approaching a key junction under automatic time-keeping control (or driver-aided timekeeping control) should encounter green signals all the way unless there is a safety hazard. Provided all trains in a group have been given appropriate target times (times-of-arrival at various points) and provided that the time-keeping system is operating satisfactorily, then it should not be necessary for the fixed-block signalling system to be used for speed-regulation. Several studies (15,16) have shown quite clearly that the use of the safety signalling system for fine regulation of traffic is unsatisfactory because of the limited number of single-aspects available, and the coarse quantisation in distance.

A concept which has considerable merit for remote railways is that of a supervisory system in a train control centre, transmitting time-keeping targets to the trains in accordance with the working timetable as modified by the current state of the system. An on-board system would compute for each train, appropriate speed commands so that the train arrives at its target points at the correct times (within a small tolerance), and without excessive consumption of energy. This hierarchical concept is in keeping with modern trends in control in other industries (17), with much of the control processing being devolved out to small units in the field (train-borne processors in this application), so as to reduce the need for data-flow between a large central controller and the trains.

In Australia, most railway organisations either have installed or are in the process of installing voice communication systems between control centres and trains.

Such systems can be used for the transmission of target data to enable the on-board control system to compute the appropriate speed commands during the journey. The set of target data is shown in Figure 1.

There are differences of opinion as to whether automatic speed-holding is appropriate or whether the driver should be responsible for maintaining the indicated speed. The authors are of the opinion that, on the long journeys common in Australia, automatic speed-holding should be available to drivers should they wish to use it. Experience in France indicates that drivers do in fact make use of speed-holding systems on long tedious sections of inter-city journeys.

The driver should be able to override the speed-holding system smoothly whenever he or she wishes, and will do so on encountering restrictive signal aspects. It is unlikely that effective timekeeping control can be achieved without an automatic speed-holding system.

![Figure 1 Proposed automatic time-keeping system](image-url)

3 TECHNICAL ASPECTS OF TIME-KEEPING AND SPEED-HOLDING CONTROL SYSTEMS

The authors have for some years been involved with the design, development and simulation of fully automatic train control systems meeting the operational requirements briefly discussed in section 2 above. This work has been carried out by the Train Control Group at the Research Division of British Rail at Derby, U.K.

Theoretical and simulation aspects of this work have been carried out in the Department of Electrical and Electronic Engineering at Loughborough University of Technology, using a Membrain digital differential analyser and ICL1903A digital computer.

Work on a semi-automatic system for Australian conditions is proceeding in the South Australian Institute of Technology's School of Electrical Engineering, using data provided by the S.A. Dept. of Transport. A laboratory rig including an inverter-fed induction motor for traction is being used for the study, with control algorithms being implemented on an enhanced Intel 8085 microcomputer system interfaced to the analog elements by Analog Devices A/D and D/A equipment.

We now proceed to discuss typical control algorithms.
because two of us are involved in control systems design; these are presented in a form suitable for use as case study material in control systems courses.

Figure 1 shows a conceptual block diagram to indicate a possible train-borne control system structure. The predictor section computes the /estimated time of arrival of the train at the next location, taking into account the relevant data for that section of the journey, and current train state (position, velocity and ...corrector section determines whether the train will be late or early at its target, and by how much. Running early is regarded as the major

On the strength of this, the velocity reference for the speed-holding system is computed until the predictor determines that the train will in fact arrive at its target on time (+ tolerances band of ± 10 seconds). The currently used corrector algorithm results in a new journey-time profile if the train is late, but regard to energy-consumption. The train is prevented from running ahead of time because a braking phase is initiated (to a new lower velocity set-point) as soon as earliness is detected. Running early is regarded as the major wastage on railways.

The system assumes that the reference speed is restored by a speed-holding control system; it is not to indicate that providing integral action is used in the speed-holding controller or the driver, the actual speed error is zero and approximately Gaussian in form, and the velocity reference is a valid input to the predicted corrector algorithms. Residual errors are usually "cleaned-up" by new predictions as train proceeds on its journey.

where s is the traction or braking time-constant. and c may be determined by conventional methods so that the dominant quadratic loop poles of the controller and train (including its traction or braking system) are correctly located. A damping ratio of 0.3 time-constant of 3 seconds for the simulated train's closed loop poles was used in design at S.A.I.T., and gave values of Kp, Kd and Kt as 1.67, 0.3 and 1.4 respectively. These values were actually found to be effective on the laboratory train in U.K.

If the key parameters affecting K and h, train resistance and braking coefficient of friction, are not constant but vary deterministically with time. This problem appeared at first sight to invalidate the linear design, but a piece-wise approach (with Kp, Kd and Kt being adjusted as speed varies so as to compensate for the variation of these parameters) appears to be useful. It is significant that this approach would not have been possible with a conventional analog controller.

The controller algorithm

\[ \frac{a s^2 + bs + c}{s} \] (4)

has been implemented digitally. In the S.A.I.T. design a sampling interval of 0.5 seconds has been chosen. This is approximately six times faster than the shortest time-constant in the system, and a decision was taken to implement the algorithm by direct use of the bilinear transformation from the s-plane to the z-plane, using the relationship

\[ s = \frac{2}{z + 1} \] (5)

without compensating for the distortion of the desired frequency response which this transformation produces at higher frequencies. (It is intended to investigate the use of other s to z transformations but this work has not yet been done. (18).

The resulting z-plane controller transfer function is

\[ \frac{dz^2 + ez + f}{z^2 - 1} \] (6)

where \( d = a + b + c \), \( e = 2(c - a) \) and \( f = a - b + c \).

This may be written

\[ \frac{d + ez - 1 + fz - 2}{1 - z^{-2}} \] (7)

and may be implemented in the simple recursive structure of Figure 2.

\[ \text{Velocity Error} \ (Ve) \]

\[ \text{Braking Force} \ (F_b) \]

Figure 2 Z-domain block diagram of controller

The implementation of the controller using a general purpose digital processor is simple when one bears in mind that E is the result of the present computation of \( E, z^{-1}E \) is the result of the previous computation, and \( z^{-2}E \) is the result of the computation before that.

A new value of \( Ve \) is read from the A/D converter each sampling interval, and a value of E computed, using the algorithm

\[ E = Ve - z^{-2}E \]
\[ F = dE + e z^{-1}E + f z^{-2}E \]

This digital controller, which may be implemented on any general purpose microprocessor, provides an effective replacement for a conventional analog three-term controller.

4 CONCLUSIONS

Effective time-keeping control of a train on a typical journey section may be achieved on Australian railways by:
a) the communication of target data for each key point in the journey from central control to the driver.

b) the computation of appropriate velocity-reference commands using, for example, a predictor/corrector algorithm so as to ensure on-time arrival at the target points.

c) the maintenance of the desired speed by the driver, or preferably by an automatic speed-holding system which can be over-ridden by the driver when required.

Algorithms for the train-borne controller have been discussed. Instrumentation to give train position and velocity to the required accuracy is discussed in (5).

It is considered that installation of simple train-borne processors and associated instrumentation on Australia's non-urban railway networks will provide for more effective regulation of traffic, and result in considerable savings of energy.

ACKNOWLEDGMENTS

All three of us are grateful to our employing organisations for the facilities provided. One of us (IPM) acknowledges the award of a grant for overseas staff development leave from the South Australian Institute of Technology. During this period he worked as a visiting engineer at Loughborough University and at the British Rail technical Centre.

REFERENCES


SMITH Put your control where your action is, Inst. and Contr. Syst., April, (1976).
