PREFACE

(Including License to Download and Restrictions on Distribution)

Prepared June, 1997, by the Editor, Jack H. Irving

In 1978 Lexington Books (belonging to C. C. Heath and Company) published the book "Fundamentals of Personal Rapid Transit", based on a program of research, 1968-1976, at The Aerospace Corporation, El Segundo, California. I was the Editor and Principal Author of the book and I was assisted by Associate Authors Harry Bernstein, C. L. Olson, and Jon Buyan. Since the writing, Olson is no longer alive and the others have retired from The Aerospace Corporation.

At the time of publication D. C. Heath was the Copyright holder. Several years later when the book went out of print D. C. Heath assigned the Copyright to The Aerospace Corporation, and shortly thereafter I purchased the Copyright from Aerospace. When the book was published the authors and The Aerospace Corporation waived our royalty rights to keep the price of the book at a minimum so that even impecunious students could afford it. I personally bought a large number of copies which I gifted to College and University Libraries across the United States.

The authors and the management of Aerospace felt that PRT (Personal Rapid Transit) is the wave of the future, with its many benefits to the rider (safe, rapid, private, comfortable and low cost transportation) and to the city (low capital and operating cost, pollution free, quiet, improved land use). Therefore, our object was to have the book read as widely as possible, because if enough readers felt as strongly as we did about the virtues of PRT, they might become the constituency which could stimulate the development and widespread installation of PRT systems.

You will imagine my delight when Bob Dunning approached me a while back asking whether he might publish the book on the Internet. It is his plan to publish the book in several installments. I was pleased to give my consent, providing this Preface is attached to and precedes each installment, inasmuch as it grants the right to download, duplicate, and distribute subject to certain restrictions stated in the next paragraph.

PREFACE

(Including License to Download and Restrictions on Distribution)

Prepared June, 1997, by the Editor, Jack H. Irving

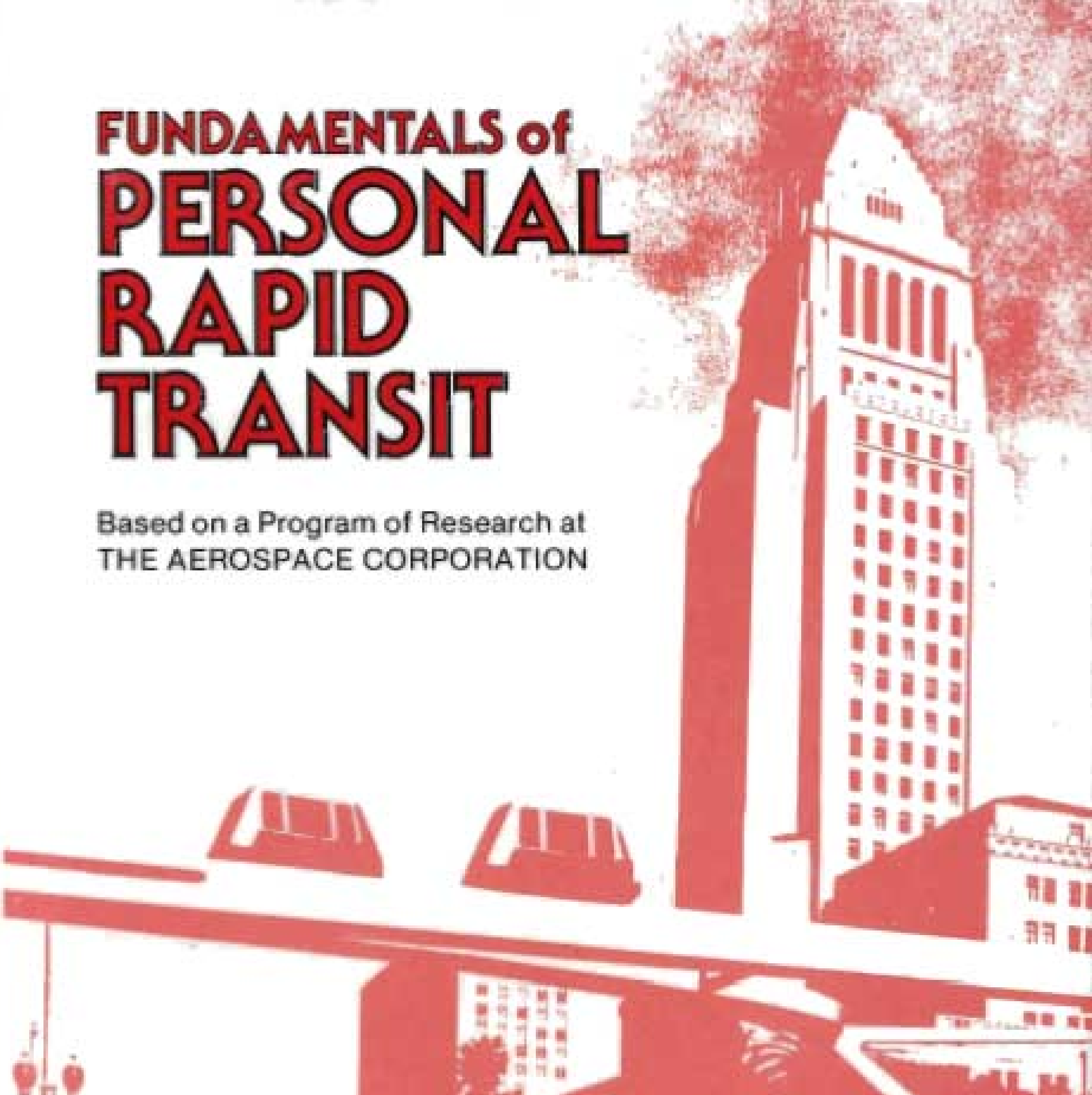
Although the book was written in 1977, I believe that, in general, the analyses made then are still valid today--with the single exception that costs have changed dramatically during that twenty year period. Some costs have come down like those of computer and control systems. Vehicle costs have risen, but possibly less than average costs because of the high degree of automation in their manufacture. The dominant costs, however, were guideway costs, and these might be considerably higher today than in 1977. Operating costs will also be higher, because of the labor costs in operation, maintenance and security and the higher prices for electric power.

All of these changes mean that Chapter 9 needs updating. Chapter 10 on Patronage Estimation will also be using the wrong values for the cost of PRT ridership, the cost of driving a car, and the monetary value that the potential rider places on his own time, but since all of these costs might be increased by roughly the same factor, the conclusions may not change significantly. Also, the cost comparisons in Chapter 11 on PRT Economics and Benefits need updating. However, because the cost on electronics is lower, vehicles up by less than average costs, and because the cost of tunneling and the heavy structures required for subway systems has escalated by a far greater degree, it is likely that the comparison made between heavy rail and PRT would be strengthened in favor of PRT.

Any questions related to the downloading, the fonts to be used, the availability of installments, or questions related to the current status of PRT should be addressed to Bob Dunning at e-mail address: bob.dunning@gmail.com. Requests to use any part of the book in a manner that does not conform to the license granted above should be addressed to me at e-mail address: apprestek@advancedtransit.net.

FUNDAMENTALS of PERSONAL RAPID TRANSIT

Based on a Program of Research at
THE AEROSPACE CORPORATION



Editor and Principal Author
Jack H. Irving
Associate Authors
Harry Bernstein, C. L. Olson, Jon Buyan

FUNDAMENTALS of PERSONAL RAPID TRANSIT

This book is aimed at informing not only the specialist in transportation but the student and interested layman about the concept of Personal Rapid Transit (PRT).

PRT would be a public transit system of small (three- to six-passenger) vehicles traveling automatically on exclusive guideways, separated from street and pedestrian traffic. The traveler and his companions would be assigned a private vehicle, not shared with strangers, to take them on a nonstop no-transfers trip from their origin station to their destination station anywhere in a large urban area. The quality of service would be comparable to that of a chauffered automobile and far superior to that of conventional public transit modes. The capital cost would be much less than that for urban rail and the operating costs far below those for bus.

The concept of PRT goes back several decades, but until this last decade there were many unsolved problems, such as: safely achieving adequate line capacities with small vehicles, vehicle traffic management on large networks, reliability, guideway aesthetics, and system cost. Solutions to these and related problems were worked out in a program of study and research (1968-1976) conducted at and sponsored by The Aerospace Corporation.

In discussing PRT, the book addresses the problem areas one by one and compares alternative solutions. In particular, the approaches worked out by Aerospace, described in some detail, define a PRT concept which is technically feasible, operationally attractive, and economically sound. The book provides a foundation on which development could proceed to make PRT a reality.

Fundamentals of Personal Rapid Transit

Jack H. Irving
Harry Bernstein
C. L. Olson
Jon Buyan

Lexington Books

Photomontage Illustrating PRT Installed in Los Angeles Civic Center

FUNDAMENTALS
OF
PERSONAL
RAPID TRANSIT

Based on a Program of Research, 1968–1976, at
THE AEROSPACE CORPORATION, El Segundo, California

Editor and Principal Author
Jack H. Irving, Ph. D.

Associate Authors
Harry Bernstein
C. L. Olson
Jon Buyan

Lexington Books
D.C. Heath and Company
Lexington, Massachusetts
Toronto

Technical questions relative to the study reported in this book should be directed to the authors. Mssrs. Bernstein, Olson, and Buyan should be contacted at The Aerospace Corporation, 2350 E. El Segundo Boulevard, El Segundo, California (mailing address: Post Office Box 92957, Los Angeles, California 90009). Queries to Dr. Irving should be addressed to Jack H. Irving, 13202 Jonesboro Place, Los Angeles, California 90049.

Library of Congress Cataloging in Publication Data

Irving, Jack H 1920-
Fundamentals of personal rapid transit.

Includes index.
1. Personal rapid transit. I. Aerospace Corporation. II. Title.
TA1207.I782 625.4'4 78-13604
ISBN 0-669-02520-8

Published simultaneously in Canada.

Printed in the United States of America.

International Standard Book Number: 0-669-02520-8

Library of Congress Catalog Card Number: 78-13604

FOREWORD

BY THE EDITOR

This book summarizes the work on Personal Rapid Transit (PRT) carried out at The Aerospace Corporation from 1968 to 1976. It is the intent of the authors that the book be useful to experts in and students of transportation and engineering, but in addition we have tried to make it readable by the interested layman. Our emphasis has been in describing concepts rather than engineering details because we believe that it is the concepts that are important, whereas there may be several alternative engineering implementations of approximately equal merit. The use of mathematics is also minimized and where equations do occur, they can be bypassed without too much loss of meaning.[1]

As a not-for-profit company, The Aerospace Corporation conducts a program of company-sponsored research to apply its talents to solving technical problems of public importance. In 1968 one of my duties as a Vice President of the Company was to recommend such a program of research to the management and Board of Trustees for their approval. It appeared to me that the capabilities of the company could be brought to bear effectively on problems of transportation, and no area of transportation was more critical than urban transportation. There was a tremendous cost in wasted time as people drove to work through congested traffic, and again as they returned home. The automobile was responsible for a disturbing trend in land use, especially in Central Business Districts where up to 70% of the land area was being used for streets, freeways, access lanes, parking lots, and sidewalks. Oil shortages were already apparent and it was clear that alternate sources of energy would soon be needed. In auto-oriented cities, and especially in the Los Angeles area, where The Aerospace Corporation is located, automotive air pollution was becoming a serious problem. Moreover, widespread deployment of

1 We regret that we were not able to use metric units, but the analyses reported here were performed over a number of years in British engineering units (feet, pounds, seconds), often using rounded numbers, and it would seem strained to express those rounded numbers in terms of their metric equivalents.

existing transit modes was not the answer because they did not have the features to attract people out of their automobiles; door-to-door trip times would be substantially longer and comfort and privacy would be sacrificed. Transit modes also were plagued either by high capital costs or high operating costs.

The idea of PRT is not a new one. Basically it is an automated taxicab system — a public transit system of three-to six-passenger vehicles operating automatically on a network of exclusive guideways, separate from street and pedestrian traffic. The traveler and his companions would be assigned a private vehicle, not shared with strangers, to take them on a nonstop no-transfers trip from their origin station to their destination station. In 1966 the U.S. Department of Housing and Urban Development sponsored a study, "Tomorrow's Transportation," by a number of organizations to assess the potential of advanced transportation systems. In particular, the studies of the General Research Corporation and the Stanford Research Institute strongly recommended Personal Rapid Transit as having a potential for a high quality of service that might be competitive with the private automobile. However, there were a number of questions about the technical and economic feasibility of PRT systems. The purpose of The Aerospace Corporation research, then, was to try to find satisfactory answers to these questions.

During the first 18 months of the program, work was confined to paper studies which included system economics, network layout, traffic management, vehicle propulsion and control, and safety. Then, for the next two years, in addition to continuing our paper studies, we embarked on an experimental program to test our ideas on propulsion and control. For this purpose we constructed a one-tenth scale model test track. More recently we have developed a number of digital computer programs for simulating traffic management on large networks and for estimating system patronage; we have also developed models for cost and reliability.

We believe that this work has demonstrated the technical and economic feasibility of PRT. We think that our design approach is a good one, but we make no claim that it is the only way to go. With feasibility established and a good design approach demonstrated, The Aerospace Corporation has successfully completed the original intent of the project and has gone as far as a nonmanufacturing organization of its type should go. It is publishing this book to place its findings before the public with the hope that government agencies and private industry will build upon the foundations laid here. For my part, I have recently left Aerospace and am seeking ways to accelerate the realization of the full potential of PRT.

During the period of the Corporation's PRT project others have not been idle. Most of the work carried on in the United States by manufacturers was on near-cousins to PRT, utilizing larger vehicles in the mode of an automated bus service operating on exclusive guideways. Although there is a good deal of commonality with PRT, there are also important differences. Studies relating to PRT were carried out in the United States at a number of universities, with some of the early important work at the University of Minnesota. In England some very significant studies were carried out on a PRT system called "Cabtrack." But of major significance are three full-scale development and test programs for PRT systems — the "Cabintaxi" system in the Federal Republic of Germany, the "Computer-Controlled Vehicle System" in Japan, and the "Aramis" system in France. Since this book is reporting primarily on the work carried on by The Aerospace Corporation, there will be no comprehensive descriptions of the foreign PRT systems which have recently undergone engineering tests, but there will be occasional references to the technologies utilized in these systems.

It is very encouraging that the Urban Mass Transportation Administration of the U.S. Department of Transportation is now sponsoring a program of Automated Guideway Transit Technology development, which spans the technology of PRT as well as other automated guideway transit modes. The Aerospace Corporation is now conducting additional studies of PRT as a part of that program. Hopefully it will not be too long before the United States will have its own full-scale PRT development and test program.

I am deeply indebted to the management of The Aerospace Corporation, and especially to its President, Dr. Ivan A. Getting, not only for the Company's financial support of the activities reported herein, but also for their continued expression of confidence that we were performing meaningful research in an important area. Although the work was under my general cognizance, specific and devoted leadership came from Mr. Harry Bernstein and his associate, Mr. C. L. Olson. The following is a partial list of the others who made an important contribution to this work:

R. W. Bruce
L. R. Bush
J. R. Buyan
D. J. Cavicchio
W. K. Clarkson
P. Dergarabedian
M. V. Dixon
M. Donabedian
R. B. Fling
G. H. Fuller
F. E. Goroszko
J. F. Grundvig
K. E. Hagen
W. H. Huber
A. L. Johnson, Jr.
J. H. Katz

D. E. Kelley
R. C. LaFrance
V. Larson
R. H. Leatherman
S. E. Levine
G. J. Liopiros
R. A. Mack
S. M. Melzer
S. Miller
A. V. Munson
J. Rossoff
A. Schnitt
B. Siegel
T. H. Silva
A. M. Timmer
B. R. Timmer
T. E. Travis
H. W. Webb
J. D. Wilson

As to this volume, each chapter carries the by-line of its principal author, although there has been a significant interchange of ideas among us.

I would like to thank Mr. Burton Sauer and Mrs. A. R. Pearce for their assistance in editing this book, and Messrs. T. Hamilton and H. Fockler for art and production coordination.

The Editor

July 15, 1977

CONTENTS

FIGURES

TABLES

Chapter 1
SERVICE CONCEPTS

Jack H. Irving

1.1 THE NEED FOR A BETTER SERVICE

Today, in the United States, there is no urban public transit which can be considered a serious competitor with the private automobile — except perhaps in New York City where driving is so difficult and parking costs so prohibitive. According to the 1970 census, the combination of bus[1], streetcar, urban rail, and commuter rail serve only 5.5% of the urban trip miles[2]. During the decade 1960-1970, while urban automobile usage was increasing 74% in passenger-miles traveled, bus dropped 26%, and rail 8½%[2]. Since 1970, with massive infusion of public funds, the downward trend nationally in transit use has been arrested (see Fig. 1-1), but only

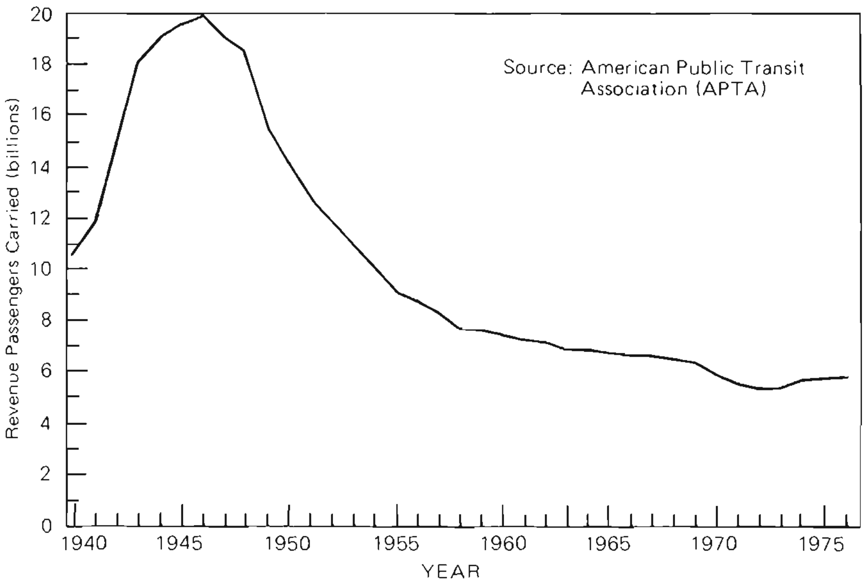


Fig. 1-1. Transit Ridership 1940-1976

1 Includes trolley-bus but excludes school bus.
2 1972 National Transportation Report, U.S. Dept. of Transportation, p. 189.

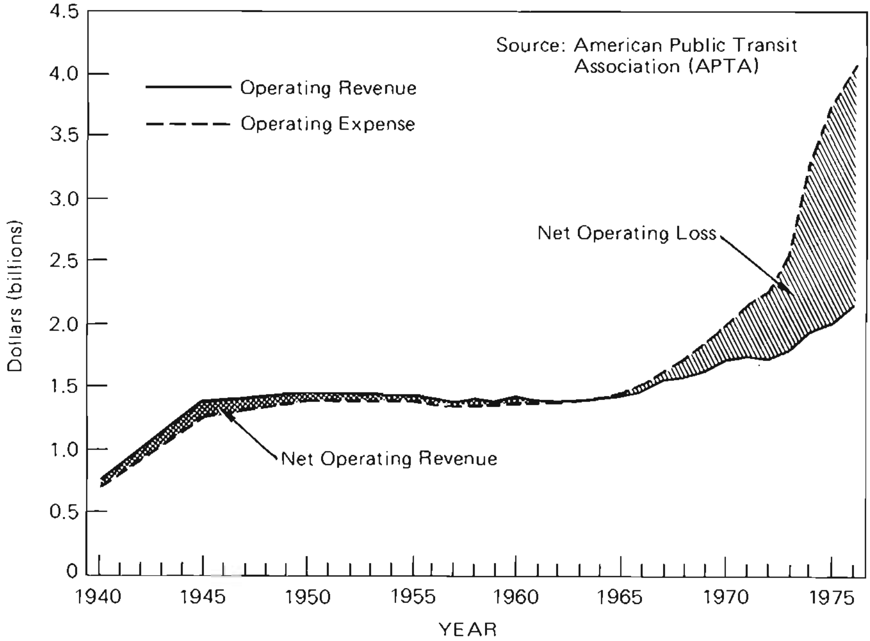


Fig. 1-2. Financial Trends in Transit Operations 1940-1976

with large operating subsidies (see Fig. 1-2). These subsidies are in addition to the very large government subsidies for capital improvement.

The reason for the sad state of public transit is a very basic one — the transit systems just do not offer a service which will attract people away from their automobiles. Consequently, their patronage comes very largely from those who cannot drive, either because they are too young, too old, or because they are too poor to own and operate an automobile. Look at it from the standpoint of a commuter who lives in a suburb and is trying to get to work in the central business district (CBD). If he is going to go by transit, a typical scenario might be the following: he must first walk to the closest bus stop, let us say a five or ten minute walk, and then he may have to wait up to another ten minutes, possibly in inclement weather, for the bus to arrive. When it arrives, he may have to stand unless he is lucky enough to find a seat. The bus will be caught up in street congestion and move slowly, and it will make many stops completely unrelated to his trip objective. The bus may then let him off at a terminal to a suburban train. Again he must wait, and, after boarding the train, again experience a number of stops on the way to the CBD, and possibly again he may have to stand in the aisle. He will get off at the station most convenient to his destination and possibly have to transfer again onto

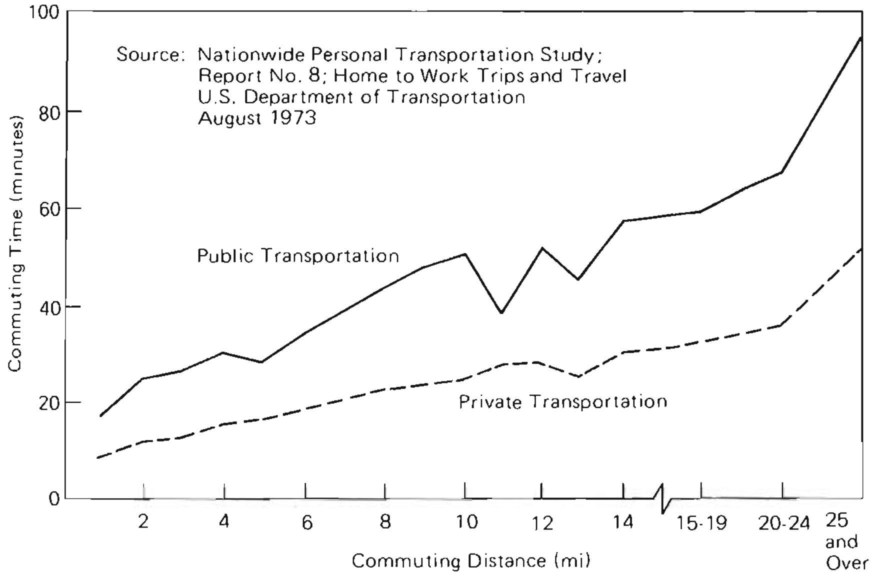


Fig. 1-3. Average Metropolitan Home-to-Work Commuting Time

a distribution system. It is no wonder that in those cities where ample inexpensive parking is available, most of those who can drive do drive.

Quantitatively, the situation is summarized in Fig. 1-3 which compares travel times for the home-to-work trip of those who go by private automobile with those who take public transit. You will note that the transit travel times are almost double those of the automobile. Now, considering that the times for transit given in the figure arose from surveying the relatively few who had chosen transit and not the majority who had rejected transit, we can only imagine what the time ratio would be for the total population.

Of currently available public transit modes, only the taxicab offers a service comparable to that of the private automobile — no transfers are required, one is seated all the way, and the vehicle is not usually shared with strangers. But, one often has to wait for a taxi to arrive and, of course, the taxicab shares with the private automobile the delays of traffic congestion. Unfortunately, from the standpoint of the urban traveler, the taxicab is very expensive because of the high labor component in the operating cost. But, because of its superior service, while transit was falling in the 60's, passenger miles by taxi[3] went up 31% — and with no public subsidy.

What is needed is a transit system which not only will serve the transit captives (the young, old, and poor), and give them a greater

[3] See footnote 2.

mobility, but will have the service features necessary to attract a significant number of those who might otherwise drive an automobile. This is especially important during the peak traffic periods when automobile transportation is plagued with problems of congestion and disproportionate air pollution and energy wastage. Areawide personal rapid transit (PRT) has the promise of providing just such a service. Like the automobile, it is convenient, requires no waiting or transfers, and provides privacy. Yet it has low operating costs, is quiet, reduces air pollution, is energy efficient, and relieves dependence on petroleum.

1.2 CATEGORIES OF AUTOMATED GUIDEWAY TRANSIT (AGT)

Historically the term "Personal Rapid Transit" or PRT referred to a system which might be regarded as an automated taxicab system — a system of small three- to six-passenger automated vehicles for the private use of the traveler and his traveling companions, but not shared with strangers; the traveler is carried nonstop and without transfers from his origin station to his destination station. Later "Personal Rapid Transit" was used to refer to any automated guideway system, regardless of the type of service provided or the size of the vehicles, although typically they were much smaller than conventional rail cars.

To clarify this ambiguity in terminology, "Automated Guideway Transit" or AGT has been proposed as the generic term, and "Personal Rapid Transit" or PRT is again restricted to its original historical meaning. This convention, recently adopted by the U.S. Department of Transportation, will be used here.

Automated Guideway Transit (AGT) is thus defined to be any transit system carrying completely automated vehicles on fixed guideways along an exclusive right of way. The guideways may be underground, at ground level, or elevated, but in any event they are grade-separated from street and pedestrian traffic, so that such traffic will not penetrate the exclusive right of way of the automated vehicles. The vehicles can be operated as single units or in trains. Three major categories of AGT systems have been defined:

Shuttle-Loop Transit (SLT)

Group Rapid Transit (GRT)

Personal Rapid Transit (PRT)

Unfortunately there is still not much standardization in the definition of these subcategories, nor are the definitions such that they unambiguously span the entire spectrum of AGT systems. Nevertheless, the definitions are useful in ordering AGT services from the most conventional mass transit services (SLT) to the most personalized (PRT).

The following subsections will define and discuss some of the salient features of these three categories of AGT. In addition, we shall define a promising hybrid of PRT and GRT. In Sec. 1.6 we discuss a related service mode, Dual Mode Transit (DMT), which utilizes vehicles that can travel automatically on an AGT guideway and can also be driven manually on streets and highways.

1.2.1 Shuttle-Loop Transit (SLT)

Shuttle and loop transit systems are the simplest types of AGT systems. The simplest type of shuttle system contains only a single vehicle or a single train of vehicles moving back and forth on a single guideway. There will, of course, be stations at each end of the guideway and there may be intermediate stops as well, as illustrated in Fig. 1-4. This type of system is the horizontal equivalent of an

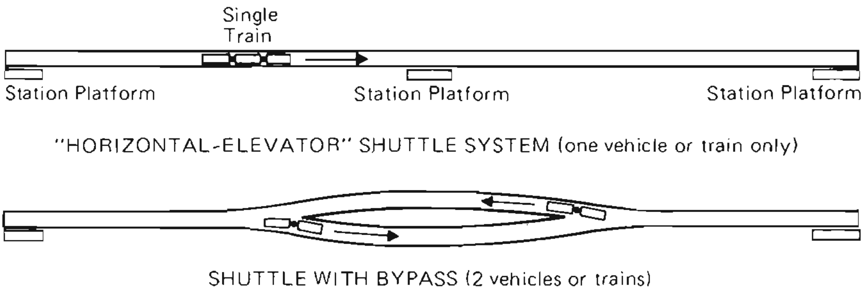


Fig. 1-4. Plan View of Two Variations of Shuttle Transit

automatic elevator. A somewhat more complex version could use two or more vehicles or trains with bypasses so that the oppositely-directed vehicles can pass each other. An example of the latter which involves two stations and a single bypass is the Ford Motor Company's ACT (Automatically Controlled Transportation) system for Dearborn, Michigan, and for the airport at Hartford, Connecticut.

In simple loop systems, vehicles or trains of vehicles stop at each station on a guideway loop, as illustrated in Fig. 1-5. A guideway loop is any closed path. An example of a loop system is the Westinghouse Electric system installed at the Seattle-Tacoma Airport.

Because the stations in a loop transit system are on-line and not off on sidings as often employed in the more complex GRT and PRT systems, the headway or time between passage of one vehicle or train and the next is typically 60 sec or greater. This gives time for the first vehicle or train to unload, load, and clear a station area before the next one arrives.

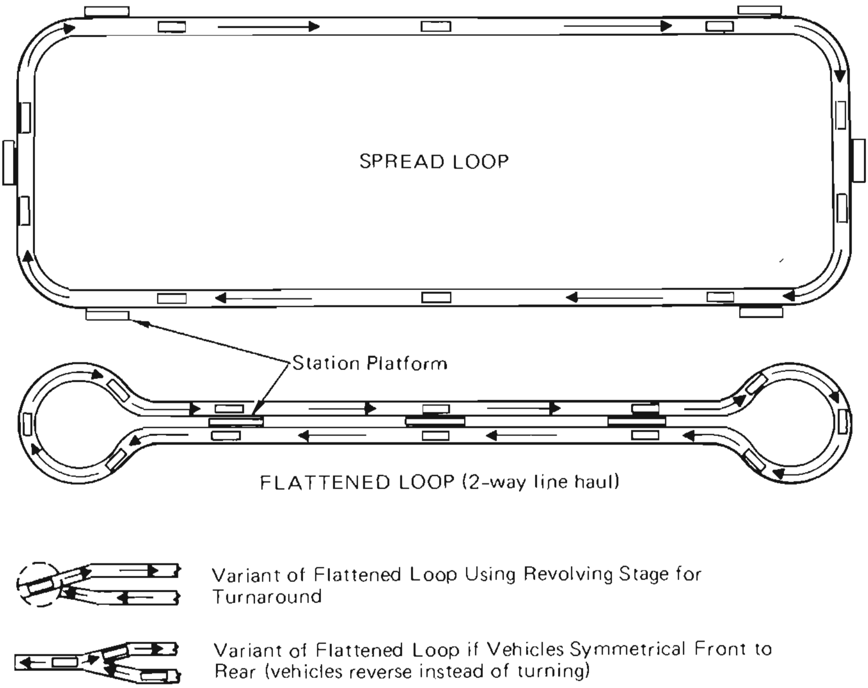


Fig. 1-5. Several Examples of Loop Systems

There is ambiguity in recent government publications as to whether or not an automated two-way line-haul system is to be regarded as SLT. A line-haul configuration may certainly be viewed as a flattened loop, as illustrated in Fig. 1-5. The figure also shows that instead of the vehicle or train turning by traversing an arc at the end of the line-haul run, variants could use a turntable, or, if the vehicles are symmetrical front to rear, a single spur for reversing the vehicle's or train's motion. One government publication classifies the two-way line-haul as a shuttle with two lines. Still another introduces a fourth category of AGT, Automated Rail Transit (ART), although this term has the unfortunate connotation of implying a rail suspension system, where, of course, any kind of suspension from rubber tires to magnetic suspension might be used.

In any event, the line-haul configuration with on-line stations shares with loop systems the necessity of operating at headways of 60 sec or greater.

Very little more will be said of SLT systems in this volume, although, in discussing the costs and benefits of an area-wide PRT system, conventional line-haul rail will be used as a basis for comparison.

1.2.2 Group Rapid Transit (GRT)

Group Rapid Transit (GRT) systems may be considered as an automated bus or jitney service where a passenger must share a vehicle with others. As with SLT, the vehicles can run alone or in trains, but GRT makes greater use of vehicle switching. Usually stations will be off-line on sidings so that a vehicle or train may stop at selected stations but bypass others. Because most stations are off-line, GRT may use shorter minimum headways than SLT, typically ranging from 3 to 30 sec. Many GRT systems involve branching so that more than one route may be available from the departure station. However, the possibility of a person being required to transfer from one GRT vehicle to another is not precluded. Typical vehicles accommodate from 10 to 70 passengers, some of whom may be standing. Examples of GRT are the Boeing system in Morgantown, West Virginia, and LTV's Airtrans system at the Dallas-Fort Worth Airport.

One common characteristic of all GRT systems is that they require that passengers wait at their origin (departure) station. For the nonstop origin-to-destination service, described below, a traveler arriving at the origin station must wait for a car to be dispatched to his destination station. Such dispatchings must not be too frequent if the vehicle is to have a satisfactory seat loading. If the service involves multiple stops, then a traveler must wait for the right vehicle to come along and pick him up.

Following is a partial listing of the great many possible variations in the type of GRT service.

Service Type A. Nonstop Origin-to-Destination Service

As with PRT, a GRT system may provide a nonstop service between pairs of stations. In GRT, however, the vehicle is shared among strangers. This has the advantage over PRT of decreasing the number of vehicles required. However, for a shared vehicle nonstop origin-to-destination service it is necessary for a patron to wait at his departure station until enough other persons have arrived going to his destination station to warrant the assignment of a vehicle to that trip. Clearly the larger the size of the vehicle[4] being used, the longer the necessary waiting times to realize a sufficient passenger load. An analysis presented in Sec. 1.5 shows that with many stations in a regional system, it may be impossible, regardless of waiting time, to achieve adequate passenger loads. This is not a promising alternative as the principal service in a metropolitan region, but may find some limited use between heavily traveled station pairs or as a part of the

[4] As used here the term "vehicle" could mean a single vehicle of a given passenger capacity, or a number of entrained small vehicles to achieve a required passenger capacity.

single transfer service (type E) described below, or in the PRT/GRT hybrid service described in Sec. 1.2.4.

Service Type B. Scheduled Stops Service

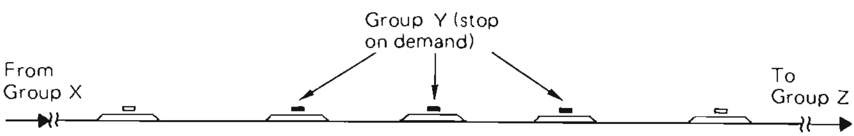
This and the next two types of GRT service obtain a satisfactory seat loading by making multiple stops to pick up and/or discharge passengers. In this type of service each vehicle has a predetermined pattern of station stops, although different vehicles may have different scheduled stops. LTV's Airtrans system in the Dallas/Fort Worth Airport is an example of this type of service. Its disadvantage is the inefficiency of making unnecessary stops. A further disadvantage is the inflexibility of the service to accommodate to varying demand or slipped schedules resulting from malfunction. Its advantage relative to certain other GRT services is that it works on a relatively predictable schedule, allowing the passenger to plan his arrival at the departure station just a few minutes ahead of his vehicle's scheduled departure time.

Service Type C. Preassigned Routes and Possible Station Stops with Actual Stops on Demand

In this service concept, a vehicle is assigned a prescribed set of those stations which it will pass on its route, stopping at any station of the set only if a passenger is to be let off or to be picked up for other stations in the prescribed set. (The prescribed set need not be all stations that it passes.) This is more efficient than service type B because of the elimination of unnecessary stops, but does not permit the traveler to plan his arrival at the departure station just ahead of trip departure because of vehicle schedule uncertainties.

Service Type D. Priority Assigned Group-to-Group (or Cycle-to-Cycle) with Stops on Demand

In this service concept a vehicle travels nonstop from a group of origin stations (say, Group X) to a group of destination stations (Group Y). When the vehicle arrives at Group Y, it is then assigned its next destination group (Group Z) on a priority basis. The assign-



ment could be made in a sequential manner, that is, Group Z could be chosen from a sequence, with each vehicle as it arrives at Group Y being assigned to the next group in the sequence. Alternatively, the stations in Group Y could be polled and the assignment made on the basis of the destination group of the longest waiting passengers, the number of passengers waiting for each group, the cumulative waiting time of passengers destined for each group of stations, or a combination of these factors. The vehicle now would pass every station in Group Y, stopping only to discharge passengers destined for a Group Y station or to pick up passengers going to a Group Z station.

A special type of group, of particular importance in one-way networks, is the cycle of stations. A cycle is a group of stations which can be entered at one of several points, and all stations in the cycle can then be passed without repetition. This is illustrated in Fig. 1-6.

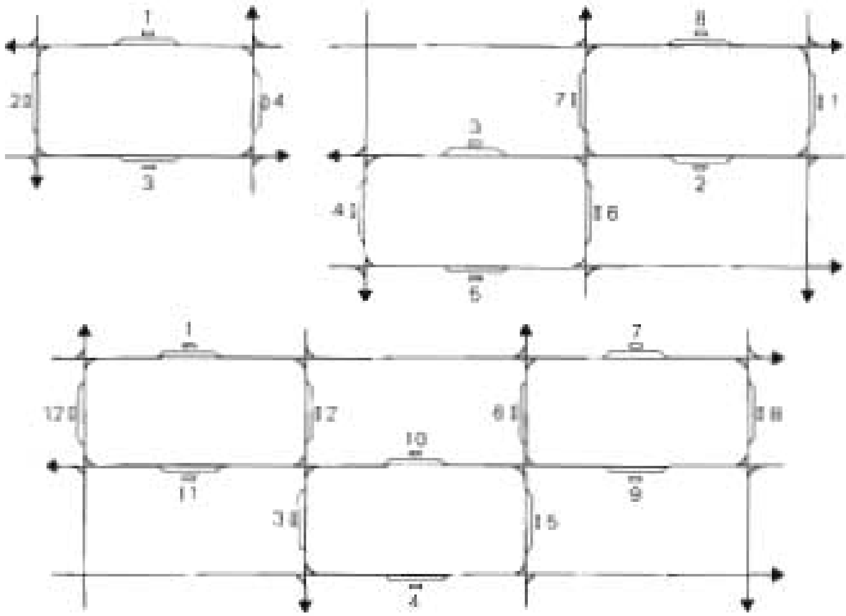


Fig. 1-6. Examples of Cycles

In a group-to-group (or cycle-to-cycle) service, there is no need for all groups (or cycles) to have the same number of stations. A heavily used station, such as might be found at certain work centers, might constitute a group or cycle by itself. A plan was proposed for Gothenburg, Sweden, which, during the evening peak hour, would send shared vehicles from CBD stations — one-station groups — to suburban station groups, each consisting of seven stations. At a CBD station each loading bay was associated with a particular list of suburban station groups, and these groups were sequentially assigned to empty vehicles as they arrived for boarding.

This service scenario and its variants can be used with vehicles of any size. For a given size vehicle there is a trade-off between average service frequency and the number of stops to be made; i.e., the fewer stations in a group or cycle, the longer the service interval required if the operation is to achieve reasonable seat loadings. Of course, the larger the vehicle the more stops or the longer the waiting time will have to be to achieve acceptable seat loadings.

Service Type E. Single Transfer Service — Grid Structure or Intersecting Loops

In its simplest form, this type of service would utilize a grid structure with a number of lines running north-south and others east-west. Vehicles would not transfer but would remain on their assigned lines. In general, a person would have to take one north-south and one east-west vehicle, with a single transfer required. Stations still would be off-line, and generally located at all crossings of east-west and north-south lines.

Since patrons boarding on any line would have only a limited choice of stations (say, 10 or 20), instead of widely scattered stations over the region, shared car service with reasonable seat loadings is easier to arrange. On some lines shared vehicle nonstop service (Service Type A above) from origin to transfer station or from transfer station to destination station might be available. Other lines might employ scheduled stop (Service Type B) or demand stop (Service Type C) service.

A grid system of this type could be configured with crossing two-way lines (if aesthetically acceptable) or with sets of crossing one-way loops. If a two-way line is used, a number of intermediate vehicle-turnaround points should be established to avoid having always to send the vehicles to the end of the line, a procedure which is inefficient because the traffic density requirement is usually highest in the middle.

1.2.3 Personal Rapid Transit (PRT)

PRT vehicles are intended for the private use of a "party" (traveling companions, traveling together by choice, but not including strangers). Typical vehicle designs have maximum capacities of between three and six passengers, all seated. Stations, with few exceptions, are off-line and the network usually includes extensive branching. The vehicles are switched from line to line and carry the passengers from their origin station to their destination station without intermediate stops; no passenger transfers are required.

One variation in PRT service may be worth considering. This variation would allow two or more friends boarding the same vehicle to designate different destination stations, providing they are generally in the same direction; the vehicle would proceed directly to the closest destination and then go on to the others.

In contrast to GRT, where passenger-waiting at the origin station cannot be avoided, there is the possibility in PRT of empty vehicles queued at every station so that passengers can board immediately for departure.

During peak traffic hours the most frequent use of PRT is likely to be the home-to-work and work-to-home trips. Therefore vehicle occupancy should be similar to that experienced in the automobile of today; namely, typical average occupancies ranging from 1.1 to 1.5 persons/vehicle. Somewhat higher occupancies might result if voluntary "PRT pooling" is encouraged by charging a fare by the vehicle rather than by the passenger. Because of the low average occupancy, high capacity per line can only be achieved by operating vehicles at short headways. Therefore, PRT systems are typically designed to have minimum headways[5] of less than 3 sec and possibly as short as ¼ sec.

1.2.4 Hybrid PRT/GRT Service

There may be instances when a hybrid PRT/GRT service is advisable. We are thinking here of a system which is predominantly PRT but where, during the rush hours, vehicle sharing (among strangers) will be required for certain selected regions, such as the CBD. This could come about if, due to capacity limitations, PRT could not handle the traffic into the CBD during the morning rush hours (or the traffic out of the CBD during the evening rush hours) without increasing the number of lines and stations beyond what is aesthetically acceptable. (Such capacity limitations are discussed in Sec. 1.4.) In some historic sections of cities, elevated lines may not be acceptable, requiring that PRT lines be placed underground. Then, for cost reasons, the number of lines and stations will be limited. As a result it may be necessary to limit the number of vehicles entering such underground stations so as to preclude true PRT service and require some mandatory vehicle sharing.

One means of achieving vehicle sharing would be to have vehicles on their way to the CBD in the morning stop to pick up other passengers headed for the same destination station. Unfortunately, this requires that two different procedures be used at each suburban station. There would be immediate boarding for those not going to

[5] When headways get this short, it becomes necessary to define the word "headway" quite precisely. Let us define the headway between two vehicles at a particular point on the guideway as the interval of time between the instant when the first vehicle's nose (forwardmost point) passes that particular point and the instant when the nose of the second passes the same point.

the CBD, while those passengers going to the CBD would have to be queued to wait for the right vehicle to come along.

An alternative means for accomplishing the vehicle sharing is to have one "hub" or "transfer" station in each suburb. If a traveler is going to any destination not in the CBD, he will take a PRT vehicle from his local residential station directly to his destination. If his destination is in the CBD, he will take a PRT vehicle to the closest "transfer" station where he will transfer to a vehicle, shared with others, which will travel nonstop to his CBD destination station. Similarly, in the evening the passenger will take a shared vehicle to the transfer station where he will transfer into a PRT vehicle for private service to his home station.

The shared vehicle might be identical to the PRT vehicle (say, with six passengers seating capacity) or it might be two or more such vehicles linked together, or it might be a single longer vehicle but so configured that it was compatible with the PRT guideways and control system. The transfer station might require a branching to separate the PRT vehicles from the vehicles being used for shared service. All stations except the CBD stations and the transfer stations would operate in the much simpler PRT-only mode at all times.

1.3 PRT FOR AREAWIDE URBAN TRANSPORTATION

Although in ensuing chapters we will be discussing the various facets of PRT in some depth, we here present a cursory description of a typical PRT system which may be envisioned as threading throughout a metropolitan area. Our emphasis will be on the service features.

Although a PRT network will of necessity be built in modules over a period of more than a decade, its full benefits will be realized only when it has been extended over a large part of the metropolitan area. Then the PRT service will be easily accessible for almost all trips.

Most lines would be elevated over city streets, because that is the least expensive means of deploying service, but some might be underground and some at ground level. The latter, for instance, might be located in the median strip of a freeway (expressway). We think of lines in residential neighborhoods as being only on the arterial or shopping streets, typically spaced about one-half mile apart, which would mean a maximum walk of two or three city blocks from any point to the closest line. In central business districts (CBD's) or other activity centers, the lines would be much closer, not only for easier access, but also to provide extra line capacity, and especially to provide extra station capacity. (Stations might typically be spaced two blocks apart in CBD's.)

Where lines cross they would be at different elevations to avoid traffic interference[6] but they would be connected by turn ramps to allow a vehicle to turn from one line onto the crossing line. In most areas the guideways would form a one-way network in which any particular arterial street would carry a line only in one direction, say north, and the next parallel arterial would carry a line running south. In this way one minimizes the investment per street, minimizes the visual impact and shadowing, and only two turn ramps are required per intersection, as contrasted with eight ramps at an intersection of a two-way network. Network characteristics are discussed more fully in Chapter 2. The frontispiece is a photomontage of half of a one-way network intersection near Los Angeles City Hall.

Each line in a PRT network has a characteristic speed; all lines do not necessarily operate at the same speed. When a vehicle is to enter a station, it leaves the through-line at line speed and decelerates on the siding, and when it leaves the station it accelerates on the siding before entering the through-line. The higher the line speed, the longer the siding must be.

Turning speeds should be kept low enough to allow a banked turn with a small radius of curvature, permitting the turn ramp to be constructed within existing street intersections without land acquisition. If a vehicle is turning from a high-speed line to a crossing low-speed line, it will decelerate after leaving the high-speed line but before making the turn. Similarly, if a vehicle is turning from a low-speed line to a high-speed line, it will first turn and then accelerate before merging into the high-speed traffic. Turning from one low-speed line to another may require no change in speed, while turning from one high-speed line to another may require deceleration before the turn and then reacceleration before merging.

To achieve sufficient distributed station capacity in a CBD, the stations must be close together. This is difficult to achieve when line speeds are high and station sidings long. Consequently, in most instances CBD line speeds will be restricted to 20 mi/hr or less. At these speeds it usually will be possible to perform the coordinated banked turns without slowing down. In some older cities where the streets are quite narrow, either a lower line speed will be required or slowing at the turns will be mandatory. In suburban areas or along transportation corridors, typical line speeds might range from 30 to 50 mi/hr. Rarely, still higher speeds might be used.

Now let us examine the PRT operation as it might be experienced by a typical user, say a man going from his home to work. He may

6 An exception is Japan's Computer-Controlled Vehicle System (CVS) which in crowded areas would have the two crossing lines at the same elevation. This requires that vehicles on the two lines time-share the intersection.

walk about two city blocks to the nearest PRT station. He need not worry if the vehicles leaving that station are going in the wrong direction because the vehicle he boards will be routed automatically onto lines that will carry it to his destination. Assuming that the guideway is elevated in the vicinity of his departure station, he ascends by elevator to the station level. He then takes a plastic card from his pocket, which can be either a PRT credit card or, if he has not established credit with the system, a cash card. The reason for having a travel card which identifies the patron is discussed in Sec. 1.7. He inserts the card into some electronic trip-ordering equipment and pushes buttons indicating the number of his destination station. There are display maps to assist him, or, if he knows the address of his destination but not how to find it on the map, he can be assisted by an information operator accessible by a special telephone. If he has a credit card, the trip will be billed to him automatically (once a month as with telephone bills), but if he has a cash card, he is asked to deposit coins for the amount of the fare. Then, in either case, his card is magnetically encoded with the number of his destination station. (If he is a stranger in town, there is a vending machine at the station from which he may obtain a cash card.)

The traveler then takes his magnetically encoded card and walks to the closest gate on the station platform where a vehicle will be waiting for his use. Dipping the card into a slot next to the gate enables the gate and vehicle door to open simultaneously so that he can enter. (If the vehicle is in the process of pulling up to the gate, the gate will not open until the vehicle has come to a stop.) The magnetic reader at the gate then informs the vehicle of its destination station. After the passenger is seated, the door closes and the vehicle automatically merges into traffic on the through-line. Station operations are discussed at greater length in Chapter 3.

After the passenger has boarded his vehicle and the door has closed, the vehicle is under automatic control until it reaches the platform of the destination station. The automatic controls must include means for controlling vehicles in the station area, on straight sections of guideway, and at intersections and merges. Safe separation from neighboring vehicles is maintained at all times. A number of possible control strategies have been considered and these are discussed in Chapter 4. The question of how to choose routes to minimize travel times consistent with not overloading line capacity is discussed in Chapter 5. (Chapter 5 also considers the question of how to dispatch and route the flow of empty vehicles to ensure that an adequate queue of empty vehicles is maintained at each station.)

Once the passenger arrives at his destination station, the door opens and he alights. He then descends in the elevator to the street level and has less than a block to walk to work. If he works in a major office building or plant, the station may even be integrated into his work facility.

In summary, upon walking to his departure station, the passenger finds a queued empty vehicle waiting for him. He travels nonstop to his destination station at speeds averaging perhaps 30 to 40 mi/hr. After arriving he is only a very short walk from his place of work.

How does the trip compare with going by private automobile? The automobile is more accessible, but in transit it is held up by traffic congestion. Altogether the door-to-door travel times are comparable. Traveling by PRT will be somewhat less costly and the passenger is free to enjoy the view, to read, or to otherwise spend his time productively. Altogether, PRT must be considered a viable competitor with the private automobile.

1.4 LINE CAPACITY

A critical question is whether PRT can achieve the line capacities necessary to carry a significant fraction of urban traffic during the rush hours.

To establish a basis for comparison, let us examine the capacity of a freeway–say one with four lanes in each direction. Maximum throughput of a freeway occurs at about 30 to 40 mi/hr; in that speed range each lane carries one automobile every two seconds–the four lanes carry two each second. If, during rush hours, each automobile is occupied by 1.25 people on average, a four-lane freeway would be carrying 2.5 people/sec or 9000 people/hr.

Now, the actual capacity requirement for any PRT line will depend on the number and spacing of lines, the population density, and the fraction of the population which uses PRT during the rush hours. For most lines the requirement will be less than that of the freeway, but for some lines it may be comparable. If a PRT line were to have the capacity of a four-lane freeway, and if occupied vehicles were to have the same average occupancy as the automobile, then this would require that a PRT line carry two occupied vehicles/sec. There will also be some empty vehicles on the line, although most empty vehicles will be moving in the opposite direction from the prevailing direction of occupied vehicles. Moreover, there must be some available space on the line to accommodate merging from crossing lines. When all of this is considered, minimum headways of 0.3 to 0.4 sec are required if a PRT line were to be fully equivalent to a four-lane freeway.

The minimum headway that can be achieved is

$$H = \frac{L + S}{V}, \tag{1.1}$$

where

H = minimum headway in seconds,
V = line speed,
L = length of vehicle,
and S = minimum allowable separation between vehicles traveling at a speed V. (Depending on the type of control, S may be a function of V.)

If L and S are measured in feet, then V must be in ft/sec. (If L and S are in meters, then V must be in m/sec.) The quantity $L + S$ is sometimes called "slot size." It is the length of guideway allocated to each vehicle when vehicles are traveling at minimum separation.

At Aerospace, we have made vehicle layouts which show that L will be about 10 ft. In Chapter 6, covering safety, it is shown that if one vehicle is inadvertently decelerating and the next applies brakes after a 0.2 sec delay and with such force that its deceleration will be 15% above that of the failing vehicle, then, independent of line speed, the vehicles will close on each other by about 4 ft (or less) before they start separating again. With a vehicle separation of about 5 ft, they will not collide. With this separation, the slot length will be about 15 ft. Using this slot length at a line speed of 30 ft/sec (20.455 mi/hr or 9.144 m/sec), the minimum headway will be 0.5 sec. At a line speed of 60 ft/sec (40.909 mi/hr or 18.288 m/sec), the minimum headway will be 0.25 sec.

The total passenger "flow" on a line is given by

$$F = \frac{3600}{H} Dfp, \tag{1.2}$$

where

F = passenger flow in passengers/hr,
H = minimum headway in seconds,
D = line "density," the ratio of the number of vehicles to the number which would be carried if all were at minimum separation,[7]
f = fraction of vehicles that are occupied,
p = average party size; i.e., the average occupancy of occupied vehicles.

[7] D may also be regarded as the fraction of slots which are occupied by vehicles, although this definition is not easily understandable when a "car follower" type of control system is used (see Chapter 4).

If one wishes to determine the flow during a morning rush hour on a line directed toward the CBD, then f might be estimated at 0.9 or higher because almost all empty vehicles will be traveling in the opposite direction. (The same is true during the evening rush on lines directed out of the CBD.) In Chapter 4 it is shown that the problem of merges is managed easily if D is kept below about 0.8. If we assume that the average party size is 1.25 passengers, then, for a 60 ft/sec line having a minimum headway of 0.25 sec,

$$F = \frac{3600}{0.25} \times 0.8 \times 0.9 \times 1.25 = 12{,}960 \text{ passengers/hr.} \qquad (1.3)$$

This is almost equivalent to the throughput of a six lane freeway.

Thus far we have discussed the minimum achievable headways and the maximum achievable flow rates on PRT lines, and we have compared them with automobile freeways. There is still the question of whether such high rates are really needed. The author is convinced by a study reported in Chapter 5 that such flow rates are indeed desirable. There we consider a scenario appropriate to the 1990's in which it is assumed that 300,000 people work in downtown Los Angeles and that 50% of them take PRT to work, arriving over a two-hour interval. If, due to voluntary "PRT pooling," arrivals average 1.5 passengers/vehicle, then 50,000 vehicles/hr would arrive in downtown Los Angeles during the morning rush hours. With some care we were able to design a network (described in Chapter 2) capable of carrying this load. In this design almost every street in the CBD carries a 30 ft/sec line with minimum headways of 0.5 sec; the corridors radiating from the CBD are 60 ft/sec lines with a minimum headway of 0.25 sec. The purpose of the computer programs discussed in Chapter 5 is to route the traffic of both occupied vehicles and returning empty vehicles so as to avoid capacity overloads. It is the opinion of the author that had the headway requirements been substantially relaxed, it would not have been possible to avoid such overloads without adding still more lines.

The experience with respect to the Los Angeles CBD may be generalized by reference to Fig. 1-7. Assume that a stylized square CBD measures 1.5 mi on a side. With a one-way network spacing of ¼ mi between 30 ft/sec lines, there would be 12 such lines entering the CBD[8] and 12 leaving. Based on the assumptions that went into calculating Eq. (1.3), but changing the headway to 0.5 sec to be compatible with the assumed 30-ft/sec line speeds, the flow on each

8 As described in Chapter 2, the Los Angeles CBD network achieves some extra capacity by running along selected streets two 30-ft/sec lines in the same direction and supported on the same columns. Altogether the network is equivalent to bringing in 16 lines.

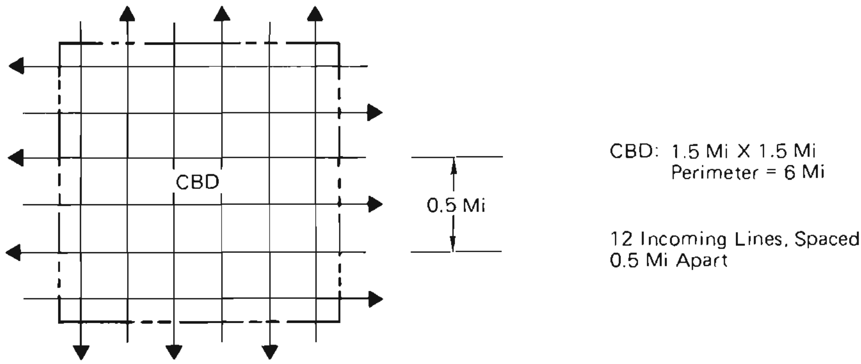


Fig. 1-7. Stylized Central Business District for Estimating PRT Capacity Requirements

line would be 6,480 passengers/hr. The 12 lines would carry a flow of 77,760 passengers/hr. If the average party size were the assumed 1.5 instead of 1.25 as used in Eq. (1.3), the flow would be 93,312 passengers/hr. In practice this could not be achieved, because the flow to most CBD's tends to be quite nonuniform, and some of the lines would not be carrying a full load.

Clearly, if flow rates in excess of 100,000 passengers/hr are required, it may not be possible to accommodate them all with a pure PRT service. One interesting possibility is the hybrid service described in Sec. 1.2.4.

Although we have just established that for application of PRT to a broad metropolitan region there will be a requirement for headways of less than one second, it should be understood that on most lines (for example, in suburban areas) the actual average headways will tend to be much larger. When PRT is applied to a smaller town with, say, a population of less than 500,000, minimum headway requirements might be several seconds instead of a fraction of a second.

At Aerospace, as described in Appendix B, we have operated a one-tenth scale model PRT system at headways of 0.5 sec. In Japan, the CVS (Computer-Controlled Vehicle System) is being operated in the range of one to two seconds minimum headway. In the Federal Republic of Germany, the "Cabintaxi" system has operated at 1.0 sec headway.

Because GRT can achieve high line capacities without using short headways, we shall now digress from our discussion of PRT to evaluate GRT for areawide urban transportation. Before returning to other aspects of PRT we will also briefly consider Dual Mode Transit (DMT).

1.5 PROBLEMS WITH GRT FOR AREAWIDE URBAN TRANSPORTATION

In the previous section we illustrated how PRT applied to areawide urban transportation must use headways of less than one second to achieve sufficient line capacity. In contrast, GRT may use significantly longer headways. For example, a GRT carrying 25 passengers/vehicle and operating at 10 sec headway provides a line capacity of 2.5 persons/sec, or alternatively a GRT system carrying 10 passengers/vehicle and operating at 4 sec headway also provides a line capacity of 2.5 persons/sec. The reader will recall that this is the throughput of the four-lane freeway we previously used for comparison. Moreover, it certainly is true that for the same total passenger trips GRT requires many fewer vehicles than PRT. The problem lies in the quality of service that can be provided by GRT operating on a large network with many stations.

If a transit system is to be a viable competitor to the private automobile, capable of attracting people who might otherwise drive, then like the automobile it should be readily accessible. Initially we shall make the assumption that, as with PRT, walking will be the principal access to GRT. At the end of this section we shall reexamine the assumption of walking access and try to define a more likely role for GRT.

Let us consider a metropolitan region having 100 sq mi within its developed area. If the population of this area is 1 million people, then the average population density is 10,000 people/sq mi, or about 16 people/acre. This implies a mixture of single and multiple family dwellings. For reasonable walking access there would need to be 4 stations/sq mi.[9] This GRT system would thus require about 400 stations in residential areas. Adding stations in the CBD and other activity centers, and at industrial sites, might bring the total to 500 stations. For a large metropolitan region with several million people, there could be a requirement for more than 2,000 stations in a mature system.

With 10,000 people/sq mi and 4 stations/sq mi, there will be 2,500 people, on average, within the influence of a given station. Since the work force is about 40% of the population, there will be roughly 1,000 workers. In studying PRT we have estimated that approximately 25 to 30% of the workers will go to work by PRT. The others will drive. Certainly the modal split for GRT will be lower, in part due to a door-to-door total travel time which is substantially greater than that of PRT, and in part due to the lack of the privacy which is provided by PRT. But, for purposes of this calculation we shall give GRT a modal split of 25%. Using 25%, we find that 250 workers will take the GRT to work, departing over a 2 hr period. Let us assume that

[9] In Sec. 2.1 (Fig. 2-2) it is shown that with an ideal arrangement of 4 stations/sq mi, there is a maximum walking distance of 3/8 mi and an average walking distance of just under 1/4 mi.

50% or 125 of these trips are destined for 50 stations, with the other 125 to the remaining 450 stations.

First we consider the feasibility of a nonstop origin-to-destination service (Service Type A). There clearly is no feasibility of nonstop ride-sharing for the 125 workers going to some of the 450 scattered destination stations. As to the 125 workers going to the 50 more frequently used destination stations, some 2.5, on average, are going to each station, but only rarely could they share rides because their departure times span a 2 hr period. On those rare occasions when two travelers arrive at the departure station at the same time, it is not wise to require them to share a vehicle. There may be safety in numbers, but here we have only two passengers, which may lead to a security problem. Even if there is no security problem, the ride may be an unpleasant experience for one of the two passengers, which may impel that passenger to return to the privacy of his or her automobile. (On the other hand, if only PRT service is offered, and if two persons note that they usually arrive at the departure station at the same time and that they get off at the same destination station, there is nothing to keep them from voluntarily choosing to travel together as a single party. Such voluntary PRT "pooling" could be encouraged by charging a fare by the vehicle rather than by the passenger.)

Thus, to obtain sufficient seat loading it is necessary to make multiple stops to pick up and/or to discharge passengers.

We have analyzed GRT for vehicles carrying 10 passengers on average and operating in the cycle-to-cycle stop-on-demand service mode identified as Service Type D in Sec. 1.2.2. Door-to-door travel times were compared with those for PRT and the personal automobile, which are very nearly equal. The GRT door-to-door time to one of the more frequently used destination stations will vary from the same time as PRT or auto to double that time, with an average multiplier of 1.5. To one of the less frequently used stations, the GRT door-to-door travel time will vary from the time for PRT or auto to triple that time, with an average multiplier of 2.0.

We have also considered scheduled GRT service and found some decrease in the average waiting time but with time lost in unnecessary stops.

The author feels that for a large metropolitan area, if GRT with walking access were to be used, some service improvement would result by having a grid of intersecting two-way lines with a single transfer required (Service Type E). However, not many cities are so regular in their street systems as to be treated effectively by this kind of grid structure. Moreover, there are serious aesthetic questions which must be raised with respect to the two-way lines, requiring up to four lines with sidings, and with respect to two-level stations at intersections.

In summary, it would appear that for a large network with stations spaced for walking access:

a. A nonstop origin-to-destination service (Service Type A) cannot provide significant ride sharing. In those rare instances where a passenger could share, he must not be forced to share because even one unpleasant experience could lose his patronage. (Voluntary sharing on a PRT service should be encouraged, of course.)
b. Cycle-to-cycle, demand-stop, shared vehicles (Service Type D) can achieve seat loadings of around 10 persons per vehicle. However, average door-to-door travel times can be as much as double those for PRT and the automobile, depending on the frequency of use of the destination station.
c. Other types of GRT service may allow modest improvements in these figures, but each such service introduces a new set of problems.

For these reasons, those who have suggested widely deploying GRT over a broad metropolitan area have not thought of it as a serious competitor to the automobile but rather as an improvement over conventional rail systems. When thought of in this way the network may be thought of as a variation of line-haul with some additional branching. Stations are not within walking access of most of the population, which must depend on access by bus or by automobile (kiss-and-ride or park-and-ride). The principal advantage over conventional rail is that, per mile of guideway, GRT is lower cost, which permits greater coverage at the same investment. Thus, GRT service will be improved over conventional rail, both because of more ready access and because fewer intermediate stops are required.

A still more accepted role for GRT (and SLT) is for circulation in certain activity centers such as CBD's, airports, college campuses, large medical centers, and shopping centers. These applications are quite sensible if a city does not have PRT, or as an interim measure until it has PRT. For example, at an airport GRT can take travelers from a parking lot to a terminal or from one terminal to another. Once PRT is operating in a city, a PRT vehicle could perform not only these functions but also it could take a traveler from his airport terminal directly to his hotel in the CBD. Looking forward to the day when areawide PRT service is in effect, the planners of an airport circulation system might do well to consider the idea of building it so that it could be readily retrofitted to become a part of an areawide PRT system when one becomes available.

1.6 DUAL-MODE TRANSIT (DMT)

"Dual-Mode Transit" is the generic name for any system in which vehicles can travel under automatic control on a guideway and also can be operated manually on the city streets. Although DMT is not strictly a category of AGT, no discussion of AGT service concepts would be complete without a comparison with Dual Mode Transit.

We might distinguish broadly between two categories of DMT according to the size of the vehicle and the nature of the service provided. The first category might be called "Group Dual-Mode Transit" (GDMT) and would use vehicles whose size might be anywhere in the range from a minibus to that of a large bus; they would be operated on the city streets by a bus driver. The other category might be called "Personal Dual-Mode Transit" (PDMT) and would use small three- to six-passenger vehicles, automatically controlled on the guideway and privately driven on the city streets.

In the typical use of GDMT, the bus driver would circulate the vehicle around a particular suburb, picking up passengers going to the CBD. He then would drive the vehicle onto an on-ramp of the automated guideway where he would leave it and board another vehicle to make another collection trip throughout the suburb. The vehicle which he had left at the automated guideway would be taken under automatic control into the CBD, possibly making automatic stops at other stations en route. In the CBD the vehicle would circulate to let off passengers and pick up others, all automatically. It would then return to an off-ramp in a suburb where a bus driver would take over manual control.

An alternative approach, which in most cases would be almost the equivalent of GDMT in its quality of service, and certainly would be simpler, would be to use a manual-only bus or minibus to perform the collection in the suburb and then to take the passengers to a transfer station in the suburb where they could board an SLT or GRT going into the CBD. The SLT or GRT would circulate throughout the CBD and no further transfers would be required. By introducing this single transfer service, there is no necessity to build the more complex vehicle which would operate both manually and automatically. Moreover, because the automated vehicles would weigh less as a result of not requiring an engine and other components for operating on the streets, the SLT or GRT guideway could be lighter than that for GDMT.

Personal Dual-Mode Transit (PDMT) from a service viewpoint is an extremely attractive concept since it combines the automatic features of PRT with the off-guideway flexibility of the private automobile. Some of its proponents have particularly emphasized the appealing concept of owning your own dual-mode vehicle. But there are several serious operational problems that must be solved before PDMT can be realized in a practical system.

There are two approaches to PDMT. One of these approaches uses bimodal vehicles which are equipped to operate under automatic control on the guideway and can also be operated manually on the city streets. The bimodal vehicle could be privately owned or it could be a vehicle owned by the system and rented by the user. One of the problems with the bimodal approach is that there is no assurance that the vehicle will not have been seriously abused by the driver under manual use, and if the vehicle is privately owned, there is no assurance that it will have been properly maintained. As a result there is the threat that it might have a high malfunction rate when operating on the guideway. One possibility is to put each vehicle through an automatic checkout before it is allowed to enter the guideway. Such a procedure clearly represents additional system complexity and cost.

The other approach is to use what is known as a "palletized dual-mode." In this concept, there are flat-bed vehicles or pallets which operate on the guideway just as a PRT vehicle would operate; in fact they might be identical to the pallets used for containerized freight discussed in Sec. 1.8. At a dual-mode on-ramp one would drive a small automobile onto a pallet where it would be secured. Then the pallet would join the stream of automated traffic. In this variation the automobile need have no capability for automated operation and the pallet no capability for manual operation. Unless, however, the automobile were very small, the length of the pallet would need to be considerably greater than that of a typical PRT vehicle, and the weight of the loaded pallet would be so great as to require substantially heavier and more costly guideway structures.

Probably the most serious problem with PDMT, whether palletized or bimodal, is that of the "off-ramp" in a congested area such as the CBD. The vehicles coming down the off-ramp could easily be delayed by the congestion of the city streets. As a result, the off-ramp would become jammed and additional vehicles would not be able to get off. The one thing which must not be allowed is to have the congestion back up onto the automated guideway because then the traffic on the guideway would come to a halt. The solution to this problem is to design the system so that there are no off-ramps in the CBD or other congested areas.

What does this mean for a vehicle which is privately owned? In that event it would be necessary to have the vehicle stop at a station for dual mode vehicles so that the owner and his guests could alight from the vehicle. The vehicle would then automatically proceed to an automatic parking garage. In the evening, when the owner wished to return home, he would have to go to the dual-mode station, request his car, and wait for it to arrive. It has been suggested that he could call from his office ahead of time so as to reduce the wait at the station; but, if he were to arrive at the station after the car had arrived, it is likely that the car would have to be returned to the garage because otherwise it would block traffic at the station. Clearly such a system is very complex and the necessity to wait for one's vehicle at the station is less than ideal.

In our opinion a far more attractive concept of PDMT is to have publicly owned vehicles, probably carried on pallets. Then there is no necessity to take home the same vehicle in the evening that one took to work in the morning. Moreover, the same vehicle can be used by a number of patrons during the day, and this cuts down on the total vehicle fleet. In this concept, on leaving the office one walks to the nearest PDMT station where queued empty vehicles are waiting. After the boarding, the vehicle travels automatically to an off-ramp in the suburb, where the driver drives his small rented automobile off its pallet, down the off-ramp and to his home. In the morning the process is reversed, and the vehicle is relinquished in the CBD so that it is available for others. The fare charged might depend on the mileage and the length of time for which one retained the vehicle, similar to the common practice with automobile rentals today.

Let us compare this type of PDMT with PRT service and the privace automobile. Its advantage over PRT is that it can extend service into suburban areas beyond the reach of the guideway system. In addition it might be argued that dual mode transit would permit the network to be designed with a greater spacing between lines because one would be less dependent on walking access. These statements have to be examined rather critically. As to the latter point, the line spacing is determined not only by walking access but also by capacity requirements. In the studies which we have made of the possible use of PRT in Los Angeles, there would have been capacity bottlenecks had substantially larger spacings been used. Moreover, not all patrons can drive; for example, the young, the very old, and those who could not afford the price of overnight rental of the PDMT vehicle. Such patrons may be quite dependent on walking access.

With respect to the point about extending service beyond the region covered by the guideway network, an alternative to having guideway off-ramps at the periphery of the network would be to have parking stations for automobiles. Then one could use one's private automobile for park-and-ride access to PRT. With the PRT stations centered in the middle of the parking lot, there is no reason why more than about two minutes need be lost in the transfer.

Alternatively, one could get to the same peripheral station by being driven (kiss-and-ride), or by scheduled bus or dial-a-ride.[10] As a result the complex on-ramps and off-ramps would be eliminated.

One group of investigators,[11] who initially were strong proponents of this type of dual-mode system, argued convincingly that the patronage of a dual-mode system would substantially exceed that of PRT because, in addition to those who would use walking access, there would be many who otherwise would drive their cars to work but who now would prefer to drive to the automated guideway and use it to save time for a portion of their trip. However, when they tested that hypothesis, using their modal-split model, they found that the customers who gained access by walking, coupled with a few going by dual mode, already unloaded the streets and highways to the point where, with reduced congestion, there was no substantial time saving in going part of the way by dual mode rather than driving all the way by private automobile. Consequently, it was their final conclusion that the total modal split for dual mode would not be substantially increased over that of PRT.

Nonetheless, the whole question of PDMT is so little understood at this time that its service advantages over PRT are very difficult to assess with certainty. It is clear, however, that if PDMT is to become a reality, the problems of safe, reliable, short-headway operation must be solved, and, hence, proceeding with PRT systems is clearly a step on the way to dual mode. Whether the additional benefits will justify going the rest of the way is still open to question.

1.7 SECURITY

We now return to the service features of PRT. From the viewpoint of a traveler, personal security is a very important service feature of any public transportation system. System security, i.e., protection of the system from vandalism, is of less direct concern to the average traveler, except that the steps taken to curb vandalism may impinge on the kind of service offered. This section, then, considers two types of security—passenger security and system security.

[10] Dial-a-ride is sometimes called dial-a-bus. The patron telephones a dispatcher and lets him know both the origin and destination of the desired trip, the number of persons in his party, and when they wish to be picked up. The dispatcher, possibly assisted by a computer, will plan the routes to pick up the party as close as possible to the requested time. When there are widely scattered origins and destinations, the service is very costly and wastes much time for the passengers on board. If, however, all passengers have either a common origin or a common destination, such as a PRT or GRT station, somewhat more efficient service can be provided.

[11] William Hamilton and Ben Alexander of General Research Corporation.

PRT is inherently secure for passengers for two basic reasons: (1) a vehicle is not shared with strangers; (2) there is no reason for loitering on the station platform since empty vehicles are queued and waiting for passengers. SLT and GRT are less secure because the vehicles are shared with strangers and because it is necessary to wait on the platform for a vehicle to come along. But the very factors that make PRT secure for the passenger reduce system security through introducing opportunities for vandalism.

In defining equipment and procedures for coping with personal security and the threat of vandalism, there is a delicate balance which must be reached relative to how much invasion of privacy can be tolerated. We will try to define an approach which minimizes the invasion of privacy consistent with exercising what we believe to be acceptable security measures. It should be borne in mind that the solution for one city is not necessarily the best for another.

1.7.1 Passenger Security

We consider the question of passenger security as beginning when the passenger enters the station premises. Clearly the larger security problem is the walk through the city streets on the way to or from the station, but this is beyond the responsibility of the system designer or system operator.

In many cities it will be advisable to have closed circuit television surveillance of the station platform and, perhaps, of the elevators leading to the station platform. As noted above, so long as there is a queue of empty vehicles available at the station, boarding can take place immediately, and the station platform will be occupied only transiently by travelers. Occasionally someone may wish to wait for a friend with whom he will be sharing a vehicle, but because the platform will be nearly empty, the TV surveillance of suspicious loiterers should not be difficult.

It has been suggested that each vehicle should contain closed circuit TV to ensure passenger security, but we believe this to be an unnecessary invasion of privacy. Because of the private use of PRT vehicles, the only threat after boarding a vehicle might come from the forced entry by a potential assailant. The closed circuit TV surveillance of the station platform is a partial protection against forced entry, but we shall shortly describe an additional precaution that might be taken.

Although the operation of the vehicles is entirely automatic, we imagine each vehicle as containing two buttons for the use of the passenger when necessary. One of these buttons, which might be labeled "Next Station," would bring him into the closest station. He might use this button if he suddenly remembers an appointment which had slipped his mind, or if he has to return home to pick up something he had forgotten. At the station he would request another trip and board another vehicle to his new destination.[12] The other button, possibly marked "Emergency Station," would be used to bring the car to one of several stations placed around the network where first aid and police assistance would be available. Once this button was depressed, the vehicle would go nonstop to the closest such Emergency Station, and the "Next Station" button would become inoperative. If a person became ill during the trip, he would push the Emergency Station button. Another use would be protection against forced entry. If a passenger is boarding a vehicle and finds that another person is attempting to force his way into the same vehicle, then by pushing the button the vehicle will be directed to the closest emergency station. Knowing this to be the case, the mere presence of the button will serve as a deterrent to forced entry.

When a vehicle is approaching its destination station, it might be advisable to play a recorded announcement which alerts the passenger that he should prepare to deboard. When the vehicle comes to a stop at the station platform, the vehicle door will open. The vehicle is weighed, and if, after a specified period of time of about 30 to 40 seconds, the passenger has not deboarded, the door will close and the vehicle will proceed to the closest emergency station. Such a procedure would ensure rapid first aid to those in need. It would also ensure that no potential assailant could linger in a car to harass the next occupant(s).

1.7.2 System Security

Because PRT vehicles are not shared with strangers there is a serious threat that vehicles would be vandalized. To deter such action we feel that in many cities it may be necessary to make at least a temporary record of the identity of passengers traveling in each vehicle. For this reason we recommend that each passenger have a travel card which will identify him when he inserts it into the trip-ordering equipment. Those who use a PRT credit card will be billed once a month. Others will use a "cash card" which, nevertheless,

[12] Alternatively the passenger could specify a new destination en route without stopping at the next station if each car had trip-ordering equipment into which he could insert his plastic travel card (credit card or cash card) and depress buttons to indicate the number of his new destination station. We do not advocate this alternative because, although it would be an improvement from the service standpoint, it clearly adds a significant cost item to each vehicle. Moreover, it also requires procedures for transferring the billing information, if a credit card were used, or for collecting the cash when a cash card is used.

identifies the passenger.[13]

At each station there should be a special vending machine from which a patron can obtain a cash card. This machine would put him in communication with an operator who would take down the patron's name and address. Possibly, the patron would be required to show some identification which could be inspected by the operator through remote video. Possibly, also the patron's picture could be taken. Then the operator would activate the machine to issue a "cash card" carrying the patron's identification. This identification could be anything from a card number which would serve to identify the patron, to possibly his name and address imprinted on the card. The latter certainly is not necessary to identify the patron, but it may convince him that vandalism could be traced to him.

The problem is how to make at least a temporary record of the identity of passengers on each vehicle with a minimum invasion of privacy. It might be useful to make a comparison with the kind of records that are made of telephone calls. Because most of us are billed monthly for our telephone calls we do not consider that our privacy is being invaded when a record is made of a phone call; in fact, we insist on it to validate the charge. By analogy, the people who are billed monthly for their PRT rides will not consider it an invasion of their privacy that a record is kept on what trips they took and when. On the other hand, there may be times when one would like to make a phone call without a record being made of that call, and this is always possible by going to a nearby pay-station phone. Again, by analogy, if one did not wish a permanent record made of a PRT trip, one should be able to pay cash for the trip instead of charging it.

How, then, does one reconcile the protection of anonymity on at least the PRT trips paid for by cash and the need to identify passengers to deter vandalism? We believe that the compromise may be in keeping only temporary records until the vehicle is inspected nightly.[14] If the vehicle passes inspection and has not been vandalized, all detailed records of who used the vehicle during the day would be erased. If the vehicle is vandalized, the list of users

[13] There might be two different ways of using the cash card:

(1) When the card is inserted into the trip-ordering equipment and buttons are depressed to indicate the number of the destination station, the passenger is notified of the fare, and will need to deposit coins, or

(2) A patron may deposit his card into a machine and insert coins or bills for advance payment. The balance is magnetically encoded on the cash card, and when he orders a trip the fare is subtracted from the balance.

[14] We believe that every vehicle should be automatically cleaned periodically, possibly nightly. At the same time there would be an automatic checkout to discover any incipient failures, possibly in redundant components whose failure would not have affected the vehicle's operation. At the same time there would be a visual inspection for vandalism.

would have to be kept. Periodically there would be a correlation study to see whether certain types of vandalism were generally associated with the presence of certain passengers. When the evidence mounted that a particular passenger was probably responsible for the vandalism, he would be sent a letter saying that it might be merely a matter of coincidence but that if the vehicles in which he rode continued to be vandalized, it might be necessary to invalidate his travel card. We believe that the very fact that passengers can be identified will serve as an effective deterrent against most vandalism.

Another system security problem has to do with protection against the improper use of invalid PRT credit cards. When a PRT credit card is inserted into the trip-ordering equipment at a station, the card number should be compared with a list of lost and stolen cards. At the same time there could be protection against expired cards or cards canceled for nonpayment or other reasons.

Finally we come to the matter of sabotage. The inaccessibility of elevated guideways (or underground guideways) may be a partial protection against the casual or thoughtless prankster. But, as with any transportation system, there is no feasible way of preventing a well-organized group of terrorists from sabotaging a part of the system. We believe, however, that it is possible to build a system so that there is no "heart of the system" whose loss would paralyze the entire operation.

Quite independent of the question of sabotage is the problem of coping with failures in various system components, including computers and power sources. Here again, to reduce the vulnerability of the operation, it is desirable to use distributed intelligence, redundant components, and backup power. Thus, the very design philosophy which minimizes the impact of natural failures also minimizes the impact of sabotage.

1.8 FREIGHT MOVEMENT

So much of the emphasis on urban transportation is placed upon the movement of people, especially when transit alternatives are being considered, that freight movement often is ignored. Yet freight constitutes a significant portion of street and highway traffic, and is responsible for a disproportionately large share of urban congestion. One need only try to travel across-town in midtown Manhattan to recognize the truth of this statement.

Because conventional transit systems are not used for urban transportation of goods, the problems of freight movement usually are kept quite separate from discussions of transit systems. But, when we are discussing automated guideway transit systems, there is no reason why the guideway cannot also be used for the transport

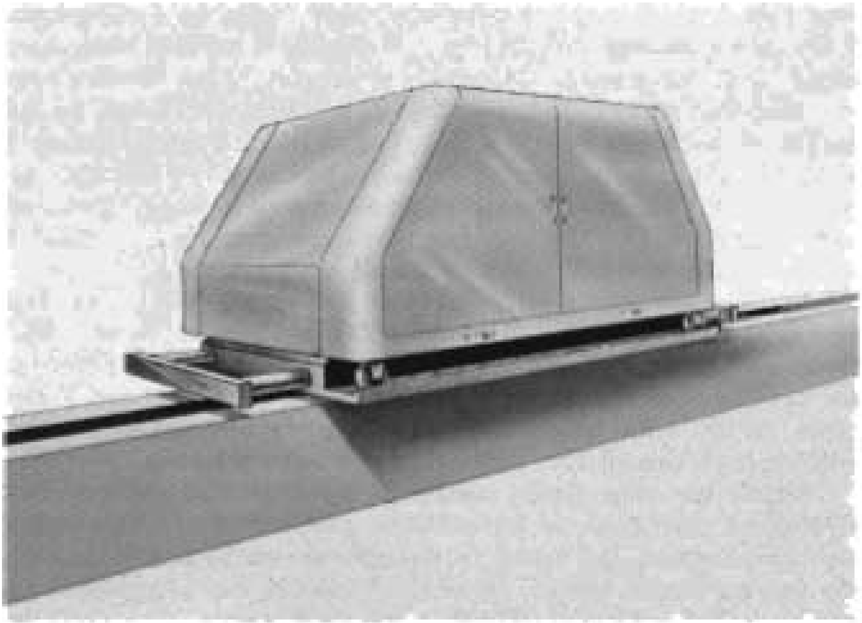


Fig. 1-8. Palletized Movement of Light Freight

of light freight. Clearly there are types of freight movement to which AGT is not well adapted. For example, it could not be used effectively for delivering a load of bricks or sand to a construction site, nor is it well adapted to delivering gasoline to a filling station. However, most of the freight movements within a city are the movements of mail, packages, cartons, crates, and, in general, packaged material with weights of at most a few hundred pounds. This constitutes light freight which could be moved very readily on an automated guideway with an appropriately designed vehicle, e.g., a palletized vehicle such as shown in Fig. 1-8, providing the guideway is readily accessible from both the origin and destination. Freight depots might be on special sidings at such places as branch post offices, warehouses, department stores, and shopping centers. Shipments to neighborhood stores might go to a neighborhood depot where merchandise could be sorted into bins or stalls for each individual store. There could then be periodic deliveries, or else each store could make its own pickups.

AGT also can be used for solid waste movement. The Airtrans system at the Dallas-Ft. Worth Airport is designed to move waste and other material.

Substantial developments have taken place in freight movement as a part of the Computerized Vehicle System (CVS) under development in Japan. On the outskirts of Tokyo they have built a test track consisting of 4.5 km of guideway on which they are testing both passenger and freight vehicles. Their freight vehicles are built to support a removable freight container. They also have built a prototype freight station, different from their passenger station, to automatically handle the containerized freight.

Chapter 2
NETWORK CONFIGURATIONS

Jack H. Irving

2.1 WALKING ACCESS

In Chapter 1 we pointed out that PRT stations should be so placed that they are within a short walking distance of a large fraction of the population. Thus, close station spacing would be used in the more densely populated portions of the city, but, of course, no city could afford to place a large number of closely spaced stations in its sparsely populated areas.

The proximity of stations will depend also upon other factors such as the average income level in a residential neighborhood. For example, it might well be cost effective to have closely spaced stations in a low income area where patrons would depend mainly on walking access; it might not be cost effective to use closely spaced stations in a higher income area. In the higher income area, even if stations were reasonably closely spaced, many patrons might prefer to use their automobiles to gain access to a PRT station because they would save a few minutes of their time, which they value highly. With a limited capital budget, it obviously is a difficult problem to determine where the stations are needed most.

Some cities are laid out in a quite regular gridlike fashion; Salt Lake City is the prime example, but the west side of Los Angeles is also quite gridlike. Other cities have almost randomly oriented streets. Therefore, it is not possible to give a simple formula as to how to place stations to make them most accessible for walking. Nevertheless, it is instructive to consider a segment of regular grid to get a feel for the distribution of walking distances for various patterns of station placement. This is done in Figs. 2-1 and 2-2.

In Fig. 2-1 we consider a PRT system with lines along arterials spaced 1/2 mi apart in both the north-south and east-west directions. It is assumed that the city streets used for walking access within the elementary PRT grid square are also oriented either in the north-south or east-west directions. Two different patterns of station placement are compared. In the upper diagram there are four stations per

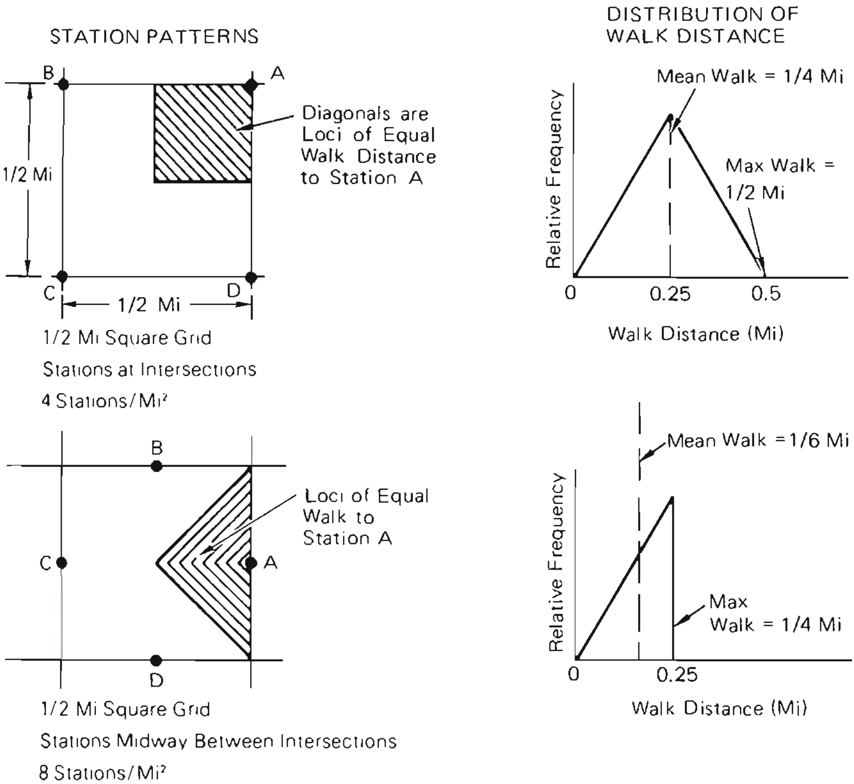


Fig. 2-1. Walking Distance for Various Station Patterns – Square Grid

square mile placed one at each guideway intersection. In the lower diagram there are eight per square mile at the midway points between the intersections. Within the grid square the shaded area is that part which is closer to Station A than to the others. The diagonals within the shaded area represent loci of equal walk distance from Station A and the length of each locus is proportional to the relative area within the elementary grid having the walk distance with which the locus is associated. Under the assumption that there is a uniform population density over the grid square, the relative frequency of any particular walk distance is proportional to the length of the locus associated with that walk distance. For each pattern, the relative frequency is plotted against the walk distance. It is seen that for the upper configuration, with four stations, the distribution function ranges from zero to a maximum walk of 1/2 mi and is symmetrical about the average walk of 1/4 mi. For the lower configuration, with eight stations, the right half of the distribution disappears. We now have 1/4 mi as both the maximum and most probable walking distance, but with a mean walking distance of 1/6 mi.

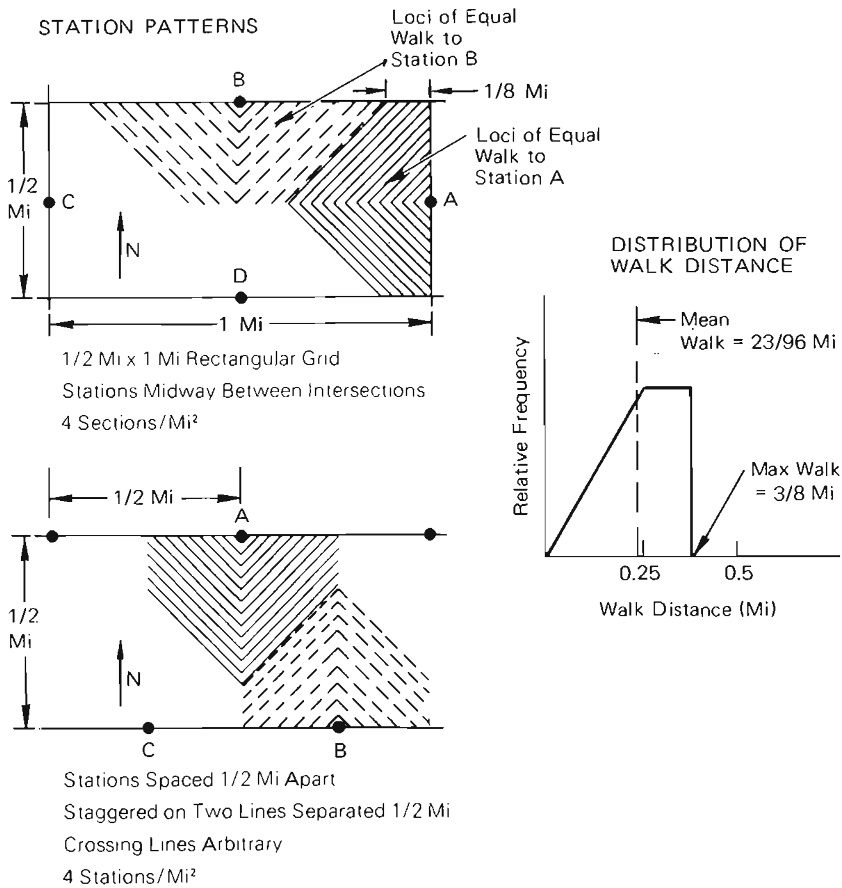


Fig. 2-2. Walking Distance for Various Station Patterns – Nonsquare

Figure 2-2 illustrates that it is not necessary to have lines closely spaced in both directions to achieve short walking distances. We are here assuming that the prevailing direction of traffic flow is in the east-west direction and that lines traveling in these directions are again spaced 1/2 mi apart. The upper diagram illustrates a configuration in which the north-south lines are spaced 1 mi apart and stations are placed midway between the intersections on both east-west and north-south lines. The lower diagram illustrates a configuration in which all stations are on the east-west lines with 1/2 mi spacing between stations, but the stations on the upper line are staggered with the station on the lower line. In this diagram it is immaterial where the north-south lines cross so far as walking distance is concerned. In both diagrams patrons living in the area shaded with solid lines are closest to Station A and those in the area shaded with broken lines are closest to Station B. The relative frequency increases linearly up to a walking distance of 1/4 mi and then is flat out to the maximum distance of 3/8 mi. The mean is 23/96 mi, which is just under 1/4 mi.

In all of the relative frequency plots of Figs. 2-1 and 2-2, it was assumed that the population was uniformly spread over the area covered by a station. In practice, however, multiple family units are more likely to be on the main arterials, where the PRT lines are installed. As a result the average walking distance would be somewhat less than that computed here. To offset this, there will be some irregularity in the patterns of station placement which may have the effect of lengthening the average walk.

2.2 EMPLACEMENT AND ALIGNMENT

Although most PRT guideways will be elevated over city streets, it is also possible to place the guideways underground or at ground level.

2.2.1 Underground Emplacement

Underground emplacement is very costly compared with elevated structures and would tend to be used only where necessary for aesthetic reasons. One example would be that of preserving the architectural integrity of an historic neighborhood. In a study of Automated Guideway Transit for Gothenburg, Sweden, it was felt that the historic walled city should not have elevated guideways, and consequently subways were planned for that sector. The rest of the city has less historic significance and it was felt that elevated guideways would not clash with the prevailing architecture.

We found another example of where underground guideways might be required in our preliminary studies of the Twin Cities of Minneapolis and St. Paul. In the CBD's of those two cities there are a large number of "skyways," which are walking bridges crossing the city streets between the second floors of office buildings and stores. An elevated guideway along one of these streets would have to be installed at the level of the third floor to clear the skyways. Where two such lines intersected, one could dip to the level of the second floor at the intersection, but only if the skyway were not close to the intersection. If skyways cross both streets close to the intersection, then one of the two guideways would have to be elevated to the fourth floor at the point of crossing. To avoid such clumsy configurations it might be necessary in places to use underground guideways. Unfortunately this also means underground stations, and consequently the excavation can be quite costly although considerably cheaper than would be involved with conventional rail.

Another disadvantage with underground emplacement is the relatively long time of construction compared with that required for elevated structures. Construction could involve either tunneling or cut and cover. Although the latter may be cheaper in some instances, it can also be very disruptive during several years of construction. Because of the high cost and likely disruption associated with underground structures, their number should be minimized. Then, to provide sufficient station capacity within the CBD, it probably would be necessary to use shared vehicle service to this area at least during the peak traffic hours. The hybrid system described in Section 1.2.4 would seem particularly attractive under such circumstances.

2.2.2 Ground-Level Emplacement

The problem with ground-level emplacement is the difficulty of separating guideway traffic from automotive and pedestrian traffic. If the guideway is not elevated at least where it crosses streets, then it is necessary to bridge automobile and pedestrian traffic over the guideway. There may be places where such bridging already has occurred and in that event ground-level guideways might be employed. One such example might be at a freeway (or expressway) where the guideway could be built on the median strip or on the shoulder.

Let us consider two cases — the freeway that is in a cut below the city streets, and the freeway that is generally elevated and bridges city streets. For existing freeways installed in cuts below the city streets, the median strip will generally not be satisfactory because it usually will have columns that support the street bridges passing over the freeway. The median strip of an elevated freeway can be quite satisfactory except, of course, any branching from the guideway would require that the branching line become further elevated before turning across the lines of traffic. The elevated freeway poses a difficult problem to elevated guideway lines installed on crossing streets because, when those surface streets go under the freeway, the guideway supporting columns must become much higher to clear the traffic on the freeway.

Another possibility for ground-level guideways would be when two remote communities were connected together by lines requiring very little branching, if any, through the intermediate rural areas. In that event one might afford to elevate the guideway only locally when it passed over crossing highways.

Where ground-level guideways are used, it will be necessary to take special precautions, such as fencing, to minimize access by children at play, by mischief makers, or by domestic animals.

2.2.3 Elevated Guideways—Aesthetics

As indicated earlier, most guideways will be elevated over city streets. For this to be acceptable, a great deal of attention must be placed on aesthetics. The guideway should be as narrow as possible to minimize shadowing and visual intrusion. Because of the importance of this consideration, we at Aerospace have emphasized "monorail-type" support systems rather than flat roadbed types as might be required by a four-wheeled vehicle.

There are two types of monorail, one where the vehicle hangs from the guideway beam in an underhung suspension, and one where the vehicle is supported below in an overriding suspension. In the overriding monorail the bottom of the guideway beam must be about 17 or 18 ft above the street level so that the highest street-driven vehicles will have ample clearance under the guideway. With the underhung configuration, the bottom of the PRT vehicles must be at this height, and the guideway from which they are suspended should be 5 to 6 ft higher. In this configuration the beams are generally supported from above and the supporting column will be at least 8 to 9 ft higher than the columns necessary to support the overriding configuration. Moreover, whereas the columns in the overriding configuration can be centered under the beam, in the case of the underhung configuration, the columns must be to the side, with a cantilevered support of the beam. The overall impact is that the columns must be higher and thicker, and consequently more costly and aesthetically intrusive than columns for the overriding configuration. (Also, as described in Chapter 7, there is greater mechanical complexity when suspending from above, especially at switching sections.) For these reasons we at Aerospace have mainly considered the overriding monorail in our studies of elevated guideway alignment and aesthetics. As a result of design considerations, discussed in Chapter 7, of vehicle suspension and propulsion and of beam stiffness, we have estimated that the beam cross-section would be approximately 3 ft high and 2.5 ft wide.

Another aesthetic consideration is the number of guideways on a given street. With a typical one-way network configuration, there will be only a single through line on any street. Considering also the siding leading into a station, this would mean a maximum of two guideways. In contrast, a two-way system would have traffic in both directions and, including station sidings, there might be as many as four guideways over segments of a street. This is one of several considerations, discussed in Sec. 3.3, for preferring one-way networks for elevated guideways.

Elevated guideways may be aligned either over the curb line or over the center of the street. In either case the guideway may be so designed that, in general, it will require no acquisition of right-of-way. By limiting the speed on turn ramps and using superelevation (banking the turns), it is possible to keep the radius of curvature sufficiently small to accommodate the entire turn ramp without interfering with corner buildings. Station platforms can be narrow structures along station sidings, and again can be installed without requiring land acquisition, in most cases. Alternatively, stations can be integrated into existing building structures and the sidings either brought into or alongside such buildings for easy access.

When aligned over the curb, supporting columns would be imbedded in the sidewalk area. This will cause minimum interference with street traffic, but, if the sidewalk is narrow, the vehicles may be so close to office windows as to create a visual nuisance. We believe it is possible to design the suspension, propulsion, and braking systems to be extremely quiet so that there need be no noise nuisance.

Alternatively the guideway could run down the middle of the street, installed in a center divider about 3 ft wide. Care would have to be taken to protect the columns from being struck by a street-driven vehicle. This type of installation probably is incompatible with a street allocated to one-way traffic because of the difficulties associated with lane changing.

Still a third possible alignment has been suggested by Cerney Associates in a study of Minneapolis.[1] The columns would be installed between the parking lane and the rest of the street. Here the columns do not interfere with through traffic, but they make access to the parking lane more difficult.

One approach to trying to get a better appreciation of aesthetic impact is to prepare photomontages such as that shown in the frontispiece. This shows one of the two turn ramps at a one-way network intersection. The reader will note that of the two intersecting guideways, one is along the curb line and the other is along the center of the street.

Still further appreciation of the visual impact may be obtained by the construction of architectural models. For this reason The Aerospace Corporation sponsored the construction of a 1/160 scale model of several blocks of downtown Los Angeles. We were particularly interested to find out if PRT would fit into this area, with its high buildings, and we were interested to compare the curb-line and street-center alignment options. Photographs of this model are shown in Figs. 2-3 through 2-6. Figures 2-3 and 2-4 show two views of a

[1] Minneapolis People Mover Study, Bechtel, Inc., April 1973.

one-way PRT guideway aligned over the center of Olive Street. Supporting columns are 60 ft apart and are imbedded in the center divider. In the foreground there is a station platform adjacent to the siding. Three empty vehicles are waiting in the input queue just beyond the station platform. The station platform is reached by ascending an elevator in the building shown to the right of Fig. 2-3 and then crossing on a footbridge over to the platform area.

Figure 2-5 shows an alignment over the curb on Grand Avenue. The line to the left is the through line and the one to the right is the station siding. The station platform between the two buildings can be reached either by taking an elevator from the street level or by access from the second floor of the taller office building. In the background one can see the crossing line on 6th Street.

Figure 2-6 shows some details of guideway alignment and support. It is interesting to compare the shadow cast by the PRT guideway with that cast by the buildings, even near noontime, to recognize how little PRT will be responsible for blocking out the sunlight.

Another difficult alignment problem arises when a guideway is to be installed on tree-lined streets. With tree varieties limited in height, the problem would be easily handled, but replanting a street with such varieties might evoke strong reaction regarding preservation of the character of the street and/or the life of existing trees. In some cases, the guideway could be aligned along the centerline of the street, except, of course, where the trees span the street. If located along the centerline, the guideway might be supported by columns embedded in a center divider, or alternatively, supported by arches from the curb-line. In any event, a good deal of ingenuity may be required to find acceptable solutions.

One potential aesthetic benefit of installing a PRT system should be noted — the possibility of modifying and improving the character of selected streets. If, indeed, the transit system is able to attract sufficient patronage, and thus significantly decrease the flow of street traffic, then it may be feasible to eliminate street traffic on certain streets. That would allow the conversion of the street into a walking/shopping mall, or perhaps a linear park with trees, flower beds, and reflecting pools, or perhaps to provide bicycle lanes. If a street cannot be eliminated, it might still be made narrower for street traffic, allowing introduction of bicycle lanes or gardens. Improvements can also be made by incorporating street lighting into the guideway structure, thus eliminating lighting poles.

A significant advantage of elevated PRT is the rapidity and ease with which it can be constructed. Both guideway sections and columns would be prefabricated. After laying the foundation for the columns, the erection of columns and beams can be done very

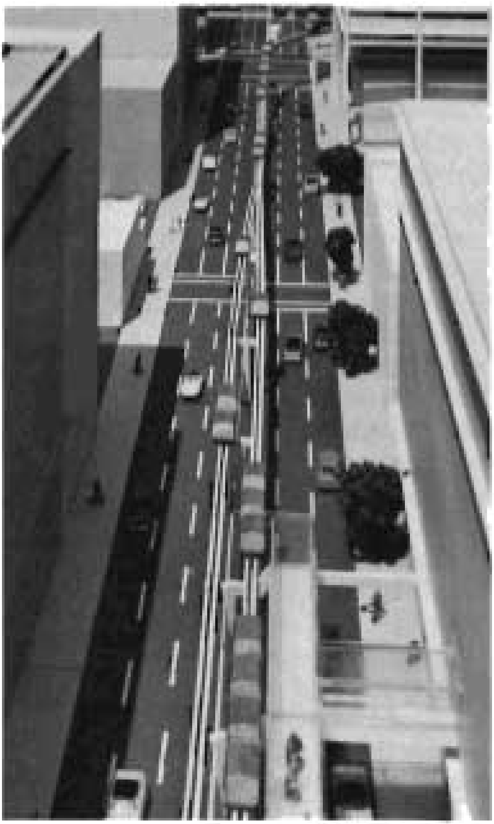


Fig. 2-3.
Typical Street-Center
Alignment for a
One-Way PRT

Fig. 2-5.
Typical Curb-line
Alignment for
One-Way PRT

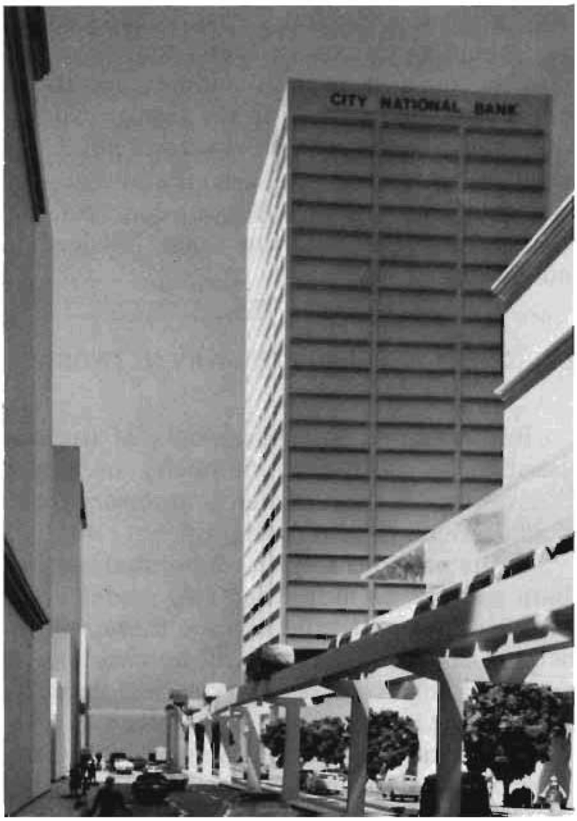
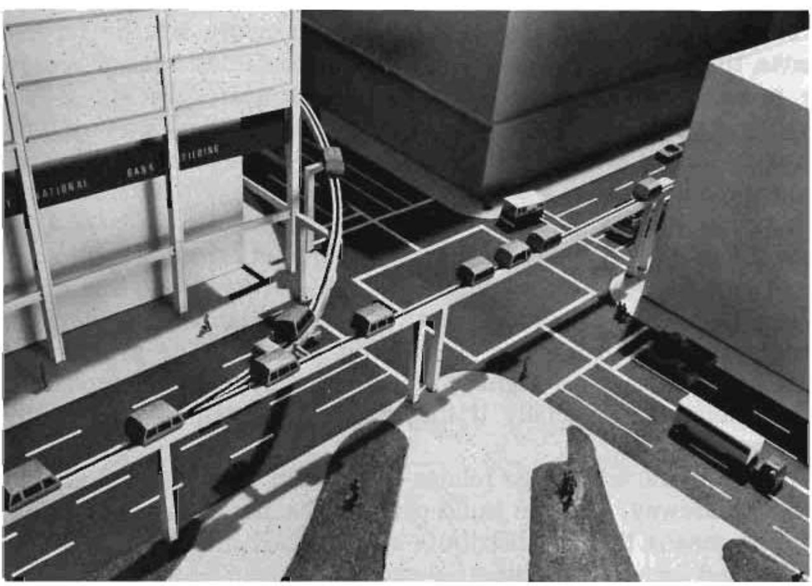


Fig. 2-4. Street-Center Alignment Viewed from Below

Fig. 2-6. Branching Guideways Showing Typical Support Details



rapidly. In the Federal Republic of Germany, DEMAG, in erecting the Cabintaxi test track, found that they could erect columns in 60 minutes and a beam in 90 minutes, and they were using beams much longer and heavier than in the design which we have been studying. This ease of construction contrasts notably with the experience for heavy rail systems where, even if above ground, it usually is necessary to create large forms and pour concrete, and, of course, where the track is underground, the times involved in construction are still much longer.

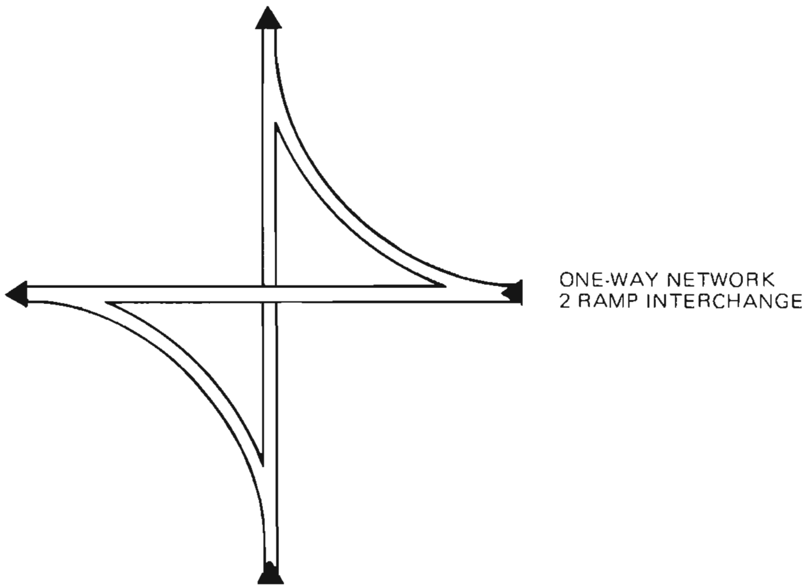
2.3 ONE-WAY VERSUS TWO-WAY NETWORKS

Briefly stated, the advantages of the one-way network may be classified as aesthetics, simplicity of control, and economics. Its disadvantage is that it often is necessary for the vehicle to "circle the block," which lengthens the trip.

In the previous section we pointed out that in a one-way network there will not be more than two guideways on any one street. Away from stations and intersections there will be only one line. In the vicinity of a station there will be two lines, the through line and the station siding. Near an intersection there may be two lines — the through line and portions of the turn ramp.

In contrast, a two-way network will have two through lines, one in each direction, and could have up to two sidings. In addition there could be two station platforms, one on each siding. Consequently, it is clear that on most streets a one-way network will be far less intrusive than a two-way network. The aesthetic difference is even more pronounced at an intersection. As illustrated in Fig. 2-7, a one-way network has only two turn ramps. Where two two-way lines cross, eight ramps are required, if all possible turns are to be allowed. Not only would this be a terribly costly and intrusive structure, but it introduces some very difficult merge problems. In Chapter 4 we discuss the work which Aerospace has done on controlling vehicles at intersections to permit the merging of vehicles from one line on to another. We have restricted our activities to the control problem for one-way intersections, which is difficult enough. The problem of controlling merges with eight connecting ramps would appear to be overwhelming, especially if the system is to degrade gracefully on malfunction.

The economic advantage relates to the fact that for each mile of two-way guideway, one can build over 1-1/2 miles of one-way guideway. This means that to distribute a prescribed number of stations at designated locations, substantial capital cost can be saved by inter-



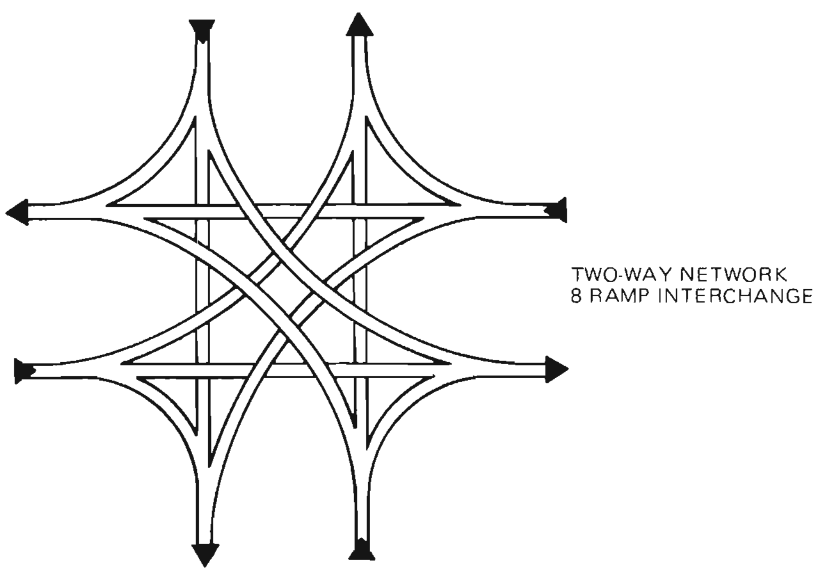


Fig. 2-7. Comparisons of Intersections of One-Way and Two-Way Networks

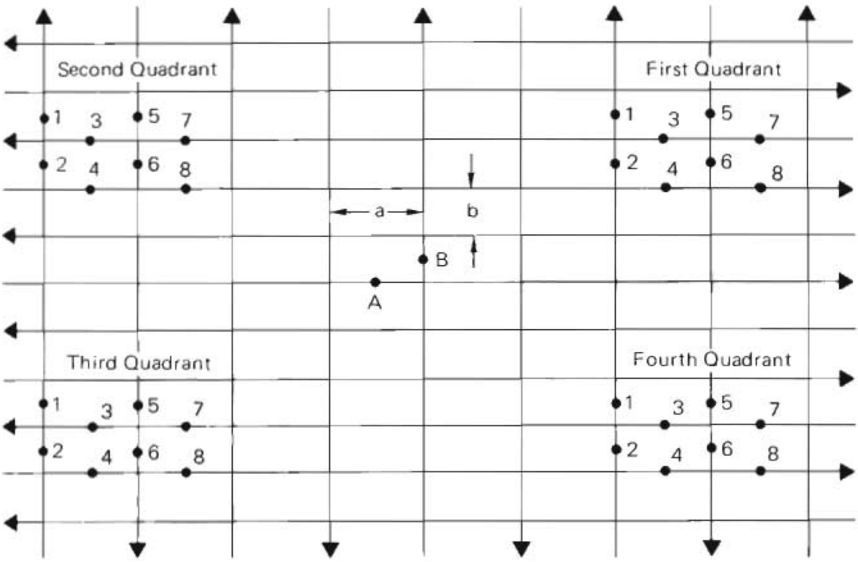
connecting the stations in a one-way network. Alternatively, with a fixed capital budget, a greater portion of a city can be covered.

To estimate the penalty in travel time when using a one-way network in contrast to a two-way network, we shall again idealize to a regular rectangular grid pattern, as indicated in Fig. 2-8. The dots in the figures represent the locations of stations, which are understood to be on sidings. The larger line spacing is "a" and the shorter spacing "b". We consider trips leaving from two different stations — Station A in the middle of one of the longer segments and Station B in the middle of one of the shorter segments. Trip destinations are to 32 different stations, 8 in each quadrant. This number is required to take into account both the direction of the line on which the destination station is located and the direction of the crossing line just upstream of the station. It will be noted that in the analysis we ignore destinations which are near continuations of the lines going through Stations A and B. Although the results can be somewhat anomalous for such stations, such results would not affect the averages significantly because, in a very large network, the anomalous stations would represent only a small portion of the total.

We also consider return trips from the 32 stations to Stations A and B. For each of these 128 trips (2 origins × 32 destinations + 32 origins × 2 destinations), the table in Fig. 2-8 shows the penalty in path length relative to what it would have been had all lines been two-way. It is seen from the grand averages tabulated that the average penalty to or from Station A is 3/4 a + 5/4 b, and the penalty to or from Station B is 1/4 a + 7/4 b, giving a general average of 1/2 a + 3/2 b. Let us consider a numerical example. Assuming that "a" is 1 mi and "b" is 1/2 mi, the general average penalty would be 1.25 mi. At a speed of 30 mi/hr, this would introduce a time penalty of 2.5 minutes.

Actually the time penalty is not as great as this analysis would imply because it assumes that a patron necessarily returns to the same station from which he departed in the morning. Let us imagine that a commuter lived an equal distance from Station A and Station B, and was going to Station 6 in the fourth quadrant. Using our numerical example, if he departed from and returned to Station A, there would be no penalty in the morning, but a penalty of 3.5 mi in the evening. If he departed from and returned to Station B, the penalty would be 1.5 mi in the morning and 1.5 mi in the evening. It would be better to depart from A in the morning and return to B in the evening, with no penalty in the morning and only 1.5 mi in the evening. Similar arguments can be used for any of the 32 stations if one is approximately an equal walk distance from two or more of these.

Since the PRT lines are presumed to be on arterials, the corner



SECOND QUADRANT					FIRST QUADRANT				
Station	From A	To A	From B	To B	Station	From A	To A	From B	To B
1	a	b	0	2b	1	0	a + 3b	0	6b
2	a	3b	0	4b	2	0	a + b	0	4b
3	a	a + 2b	0	a + 3b	3	a + 2b	a	a + 2b	3b
4	2a	0	a	b	4	0	2a	0	a + 3b
5	a + 3b	0	3b	b	5	b	a	b	3b
6	a + b	0	b	b	6	3b	a	3b	3b
7	a	a	0	a + b	7	a	a	a	3b
8	2a + 2b	0	a + 2b	b	8	0	2a + 2b	0	a + 5b
Ave	$\frac{5}{4}a + \frac{3}{4}b$	$\frac{1}{4}a + \frac{3}{4}b$	$\frac{1}{4}a + \frac{3}{4}b$	$\frac{1}{4}a + \frac{7}{4}b$	Ave	$\frac{1}{4}a + \frac{3}{4}b$	$\frac{5}{4}a + \frac{3}{4}b$	$\frac{1}{4}a + \frac{3}{4}b$	$\frac{1}{4}a + \frac{15}{4}b$
THIRD QUADRANT					FOURTH QUADRANT				
1	a + 3b	0	2b	0	1	3b	a + 2b	6b	0
2	a + 5b	0	4b	0	2	b	a + 2b	4b	0
3	a + 2b	a	b	a	3	a	a + 2b	a + 3b	0
4	2a + 4b	0	a + 3b	0	4	0	2a + 4b	3b	a + 2b
5	a + 2b	3b	b	3b	5	0	a + 3b	3b	b
6	a + 2b	b	b	b	6	0	a + 5b	3b	3b
7	a + 2b	a + 2b	b	a + 2b	7	a + 2b	a + 2b	a + 5b	0
8	2a + 2b	0	a + b	0	8	0	2a + 2b	3b	a
Ave	$\frac{5}{4}a + \frac{11}{4}b$	$\frac{1}{4}a + \frac{3}{4}b$	$\frac{1}{4}a + \frac{7}{4}b$	$\frac{1}{4}a + \frac{3}{4}b$	Ave	$\frac{1}{4}a + \frac{3}{4}b$	$\frac{5}{4}a + \frac{11}{4}b$	$\frac{1}{4}a + \frac{15}{4}b$	$\frac{1}{4}a + \frac{3}{4}b$

GRAND AVERAGES			
From A	$\frac{3}{4}a + \frac{5}{4}b$	From B	$\frac{1}{4}a + \frac{7}{4}b$
To A	$\frac{3}{4}a + \frac{5}{4}b$	To B	$\frac{1}{4}a + \frac{7}{4}b$
General Ave	$\frac{1}{2}a + \frac{3}{2}b$		

Fig. 2-8. One-Way Network Distance Penalty – Stations Midway Between Intersections

where two such lines cross may be a nucleus of commercial activity. In that event there may be a desire to place a PRT station near the intersection. In a one-way system there is then some merit in splitting the station into two parts, as shown in Fig. 2-9; i.e., a departure platform and an arrival platform. It will be noted that they have been placed so that one can depart along either of the intersecting lines and one can arrive from either also.

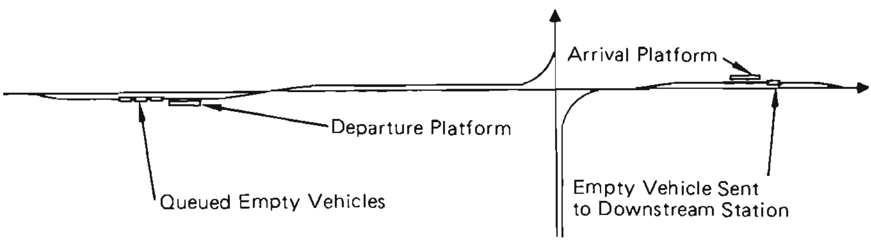
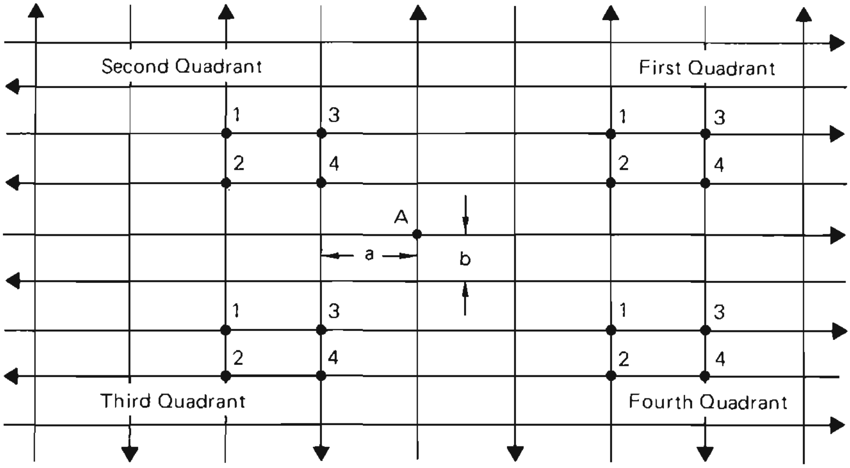


Fig. 2-9. Schematic of Split Station at a One-Way Intersection (not to scale)

Now imagine that we have an entire network of such stations (Fig. 2-10). Again "a" is the larger spacing and "b" the shorter in the rectangular grid. As before we can derive the distance penalty relative to a two-way network. Now we need only consider trips to or from station location A (having two separated platforms) and in each quadrant only four station locations need be considered. The distance penalties are given in the table of Fig. 2-10, where it is seen that the average penalty to or from station location A is "b". If the shorter spacing is 1/2 mi as in our earlier example, then the penalty either way is this amount, and at 30 mi/hr the time penalty is just 1 minute.

This configuration with split stations at intersections (Fig. 2-10) has lower one-way penalties than does the earlier configuration of stations midway between intersections (Fig. 2-8). There are only two station locations per square mile in Fig. 2-10 instead of the four stations per square mile in Fig. 2-8, but each of the station locations will involve almost twice the investment of the nonsplit station (by doubling switches, sidings, platforms, elevators), so that overall costs are comparable. The clear disadvantage is in lengthening the walk distance, as shown in Fig. 2-11. Here, the maximum walk is 3/4 mi and the average walk is 3/8 mi, assuming uniform population density. (These should be compared with the upper diagram in Fig. 2-2.) But, if the population density or, more precisely, the trip activity, is focused very much at the intersection, then the average walk may not have been lengthened and this type of configuration is worth consideration.



SECOND QUADRANT			FIRST QUADRANT		
Station	From A	To A	Station	From A	To A
1	0	0	1	0	4b
2	0	2b	2	0	2b
3	2b	0	3	0	2b
4	0	0	4	2b	2b
Ave	$\frac{1}{2}b$	$\frac{1}{2}b$	Ave	$\frac{1}{2}b$	$\frac{5}{2}b$
THIRD QUADRANT			FOURTH QUADRANT		
1	4b	0	1	0	0
2	2b	0	2	2b	0
3	2b	0	3	0	2b
4	2b	2b	4	0	0
Ave	$\frac{5}{2}b$	$\frac{1}{2}b$	Ave	$\frac{1}{2}b$	$\frac{1}{2}b$
Grand Average From A = b Grand Average To A = b					

Fig. 2-10. One-Way Network Distance Penalty – Split-Platform Stations at Intersections

It is our opinion that the strong advantages in aesthetics, control simplicity, and economics of one-way networks far outweigh their disadvantage of adding 1 to 3 minutes to the average trip time. There will, of course, be some circumstances where two-way lines will be required. For example, if it is necessary to get from one side of a range of foothills to the other, there may be only a single pass for the guideway, and a two-way guideway will need to be constructed. Under such circumstances, however, branching can generally be minimized in that region.

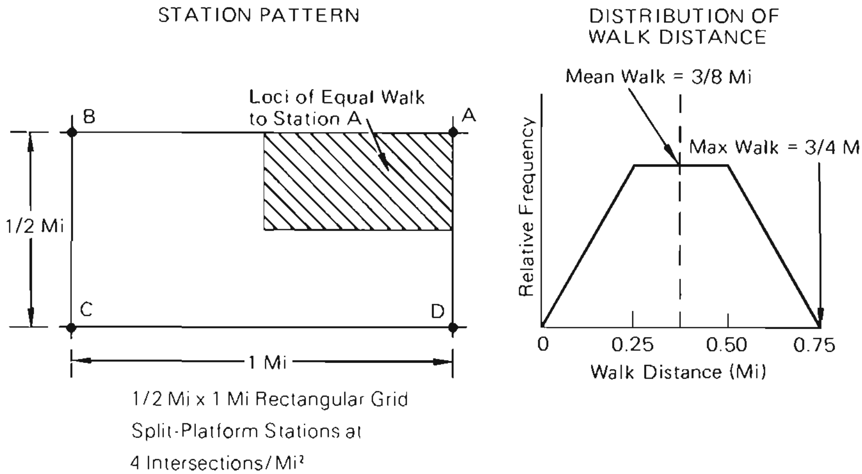


Fig. 2-11. Walking Distances to Split-Platform Stations at Intersections of a Rectangular Grid

2.4 RELATIONSHIP BETWEEN NETWORK CONFIGURATION AND LAND USE

There are two ways in which PRT network configurations interact with land use and transportation needs. One is the requirement placed on the network configuration to meet current and anticipated trip patterns that arise out of existing or planned land use. The other, far more subtle relationship, is the influence that the network configuration and transportation services will have on future growth patterns.

Current trip patterns can, of course, be measured. Anticipated trip patterns can be estimated on the basis of planned land use, as, for example, the location of industry and the corresponding distribution of employee residences. One then hypothesizes a PRT network to carry these trips and performs a modal split analysis which estimates how many of the total trips will be taken on PRT versus those which would go by private automobile or other transit mode. In Chapter 10 we describe The Aerospace Corporation's computing programs by which the modal choices of a large number of travelers are analyzed. Each trip is considered on a door-to-door basis. The choices are based on accessibility of PRT stations, PRT travel times, fare structure, automobile travel times, automobile costs, and the income level of the traveler. These programs provide not only the overall modal split but they estimate as well the activity on each line and at each station. As a result they provide a great deal of insight into which lines and stations are not cost effective, and where additional capacity or accessibility may be required. This allows the planner to redesign his network configuration and station placement to meet the needs of the existing and anticipated trips more effectively.

As for the influence that the network configuration and transportation service will have on future growth patterns, it is important to lay down networks in such a way that they will not foster haphazard and undesired developments nor lead to a dull homogeneity of the metropolitan area. This can be done by differentiating the type of network configurations that are planned for different areas, while still connecting them into one compatible system. For example, in Central Business Districts or other activity centers, good circulation should be provided with short walking distance so as to have good access throughout these areas. Such circulation centers might be interconnected by less-fine-grained networks in transit corridors or by a more skeletal line-haul configuration, depending upon objectives and needs.

In residential areas, PRT lines should only rarely, if at all, be put on single-family residential streets, but rather along the arterials passing through residential areas. Less circulation should be provided in residential areas to discourage their conversion to commercial use (although zoning should continue to be the main control over land use). Lines can generally be somewhat farther apart in residential areas than in activity centers. An example is the treatment of the western section of Los Angeles shown in Fig. 2-12 where better circulation is provided in such activity centers as the Wilshire region, Hollywood, the Miracle Mile, Beverly Hills, Century City, and Westwood, with less circulation in areas which are primarily single family residential.

PRT should bring about the happy medium between the very intense land use brought about by urban rail, and the urban sprawl brought on by complete reliance on the automobile. Urban rail usually brings with it severe congestion. It is well known that if a corridor has already been heavily congested by automobile traffic, building a subway line under that corridor will not relieve the congestion but will intensify it. This occurs because the effect of bringing in the rail encourages additional construction of high-rise buildings for easy access to those who use the rail system. But most people continue to drive their automobiles, and because of the added activity, additional automobiles are brought into the corridor. Because of the advantages to an individual or an industrial or commercial establishment of being within the PRT network area, PRT might be expected to have an inhibiting effect on the city spreading out into rural areas; yet it should not encourage the kind of inefficient, extremely high-density land use that is brought on by installation of urban rail.

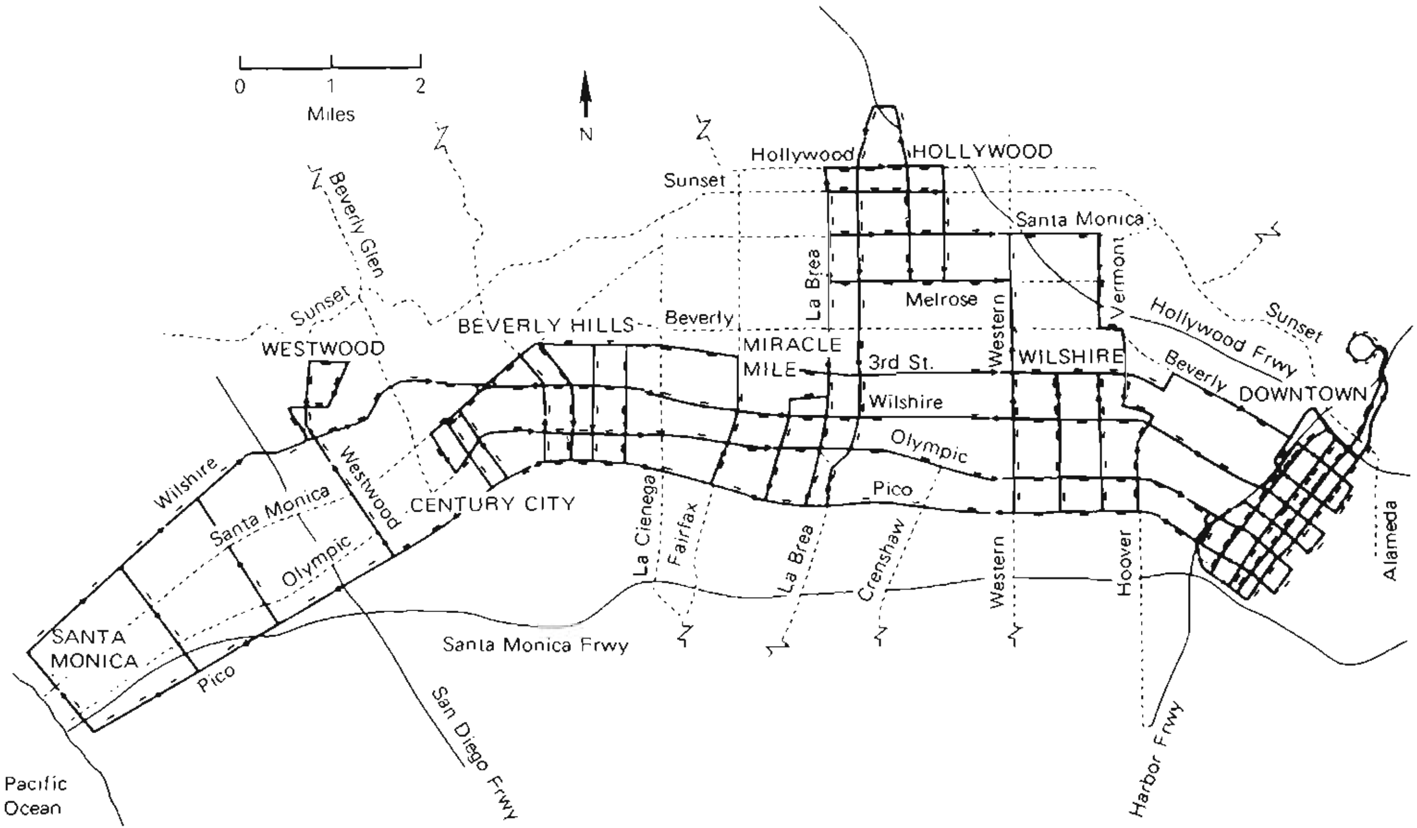


Fig. 2-12. PRT Network for West Los Angeles

2.5 INFLUENCE OF CAPACITY REQUIREMENTS ON NETWORK CONFIGURATION

Requirements in both aggregate line capacity and aggregate station capacity can influence network configuration. For example, we might start the design process by laying out PRT lines and stations along a heavily traveled corridor. Then, on performing a modal-split analysis, we might find that the estimated demand for service along those lines was greater than their capacity. In that event it would be necessary either to increase the capacity of the hypothesized lines by changing speed or vehicle spacing, or to run more lines down the corridor, probably along parallel arterials.

In Central Business Districts line capacity may also be important, especially with line speeds limited. But aggregate station capacity is of even greater importance. The capacity of a simple station with a single platform on a siding depends on the length of the siding, the length of the input queue, the method of operation, and the performance goals. These factors are all discussed in Chapter 3. A simple station of this kind can almost always be installed without land acquisition. One way to increase capacity is to use a number of parallel sidings, each with a platform, but this will require land acquisition and possibly the construction of a large terminal building. Alternatively, one can increase aggregate capacity by having a larger number of the simpler stations. The problem is how to fit enough of them, along with necessary intersections, into a small dense CBD.

We will illustrate the design processes by describing an exercise carried out at The Aerospace Corporation to examine a dense network in downtown Los Angeles. The scenario projected forward to a time in the 1990's when the working force downtown will be approximately 300,000 people. For purposes of the illustration it is assumed that by that time there will be a substantial PRT network throughout the metropolitan Los Angeles area, and that 50% of the downtown workers will arrive by PRT from nine corridors and three outlying parking lots. It is further assumed that they will arrive over a 2 hr period, loaded 1.5 per vehicle. Consequently, during the morning peak period, 50,000 vehicles per hour will arrive downtown. The question is whether it is possible to have sufficient stations and line capacity to handle this flow.

At the outset we obtained data on the number of trips destined to each city block, based on existing and planned office space. This assisted us in distributing the stations to meet the capacity requirements in each subregion of downtown. The assumed speed for the downtown region was 30 ft/sec or about 20 mi/hr. At this speed it is possible to make a coordinated banked turn within the confines of available street widths. With the main line speed and station capacity requirements known in each station, platform and siding lengths can be obtained by the methods described in Chapter 3. The

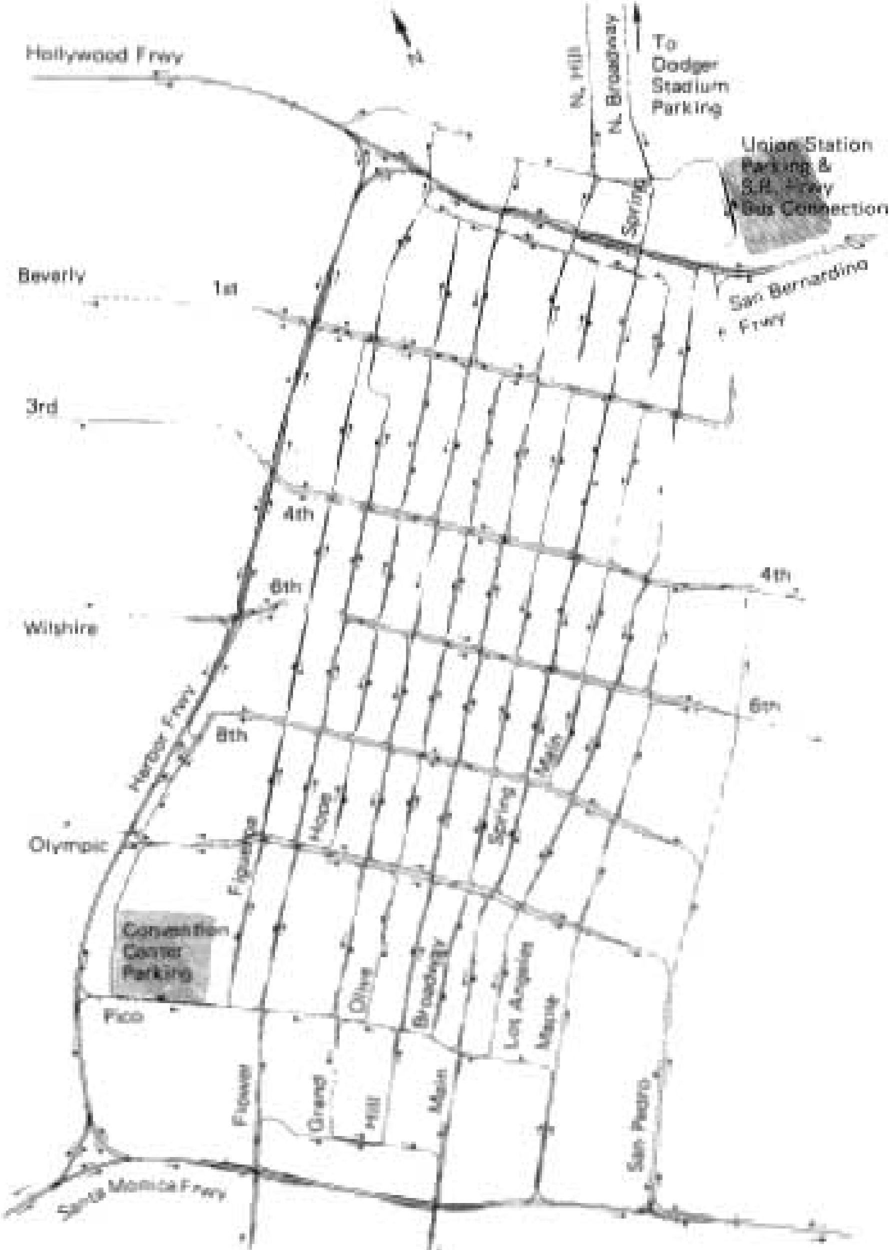


Fig. 2-13. Downtown Los Angeles Network

siding lengths then limit the proximity of neighboring stations. To obtain sufficient distributed capacity, a majority of the downtown streets would need PRT lines.

We next noted that if full grade separation were provided at every intersection of two streets containing PRT lines, the turn ramps (and intersection maneuvering regions on the main lines) would occupy a significant fraction of the linear space and would tend to restrict the space available for station sidings. To avoid this problem we proposed the configuration shown in Fig. 2-13 which is mostly at one level.

PRT stations were placed at three parking lots in the downtown area — at Dodger Stadium, the Convention Center, and Union Station. In addition, there are 58 other downtown stations which, with one exception, are all located along the north-south streets, since these streets have longer blocks. Lines on east-west streets, with short blocks, merely serve as feeder lines to the stations.

On the incoming corridors the line speed is 60 ft/sec (about 40 mi/hr), but each line splits up into two 30 ft/sec lines as it approaches downtown. Each of the utilized east-west streets carries two lines in the same direction, which are presumed to be supported on the same columns, and which have frequent transitions between the two. This is more clearly illustrated in the expanded drawing of Fig. 2-14 where it can be seen that, in principle, the whole network could be constructed at one level without grade-separated crossings. However,

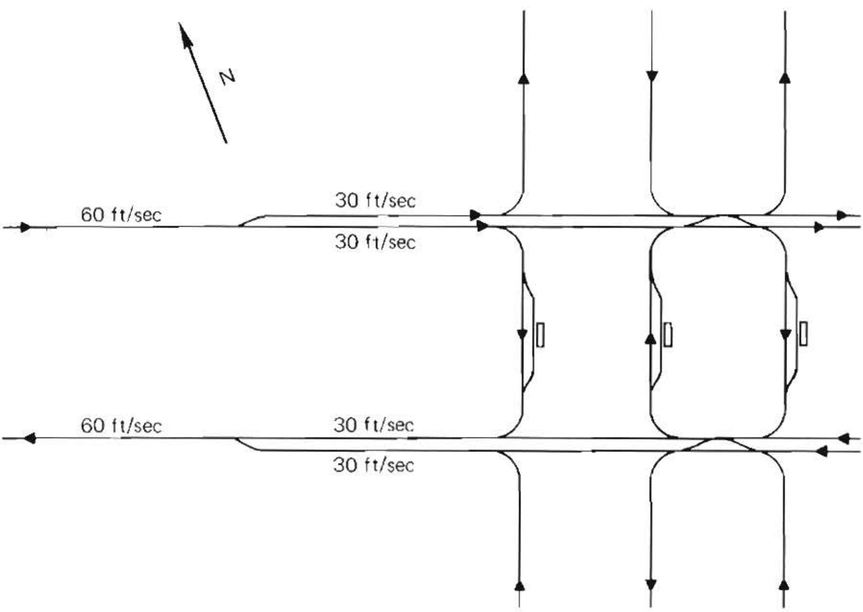


Fig. 2-14. Typical Detail of Downtown Los Angeles Single-Level Network

a north-south through trip would then be quite tortuous. Therefore, we compromised by putting in a few grade-separated intersections to facilitate through north-south traffic. It also will be noted from Fig. 2-13 that the direction of flow on most north-south lines has been chosen to facilitate the flow of occupied vehicles from the west and the return of empty vehicles to the west, because the heaviest traffic corridors lie in that direction.

It might be noted that once a few grade-separated north-south lines have been included, it may not be necessary for all stations to be on sidings. This is illustrated in Fig. 2-15 where Station *A* can be on-line. If Station A were replaced by two sequential stations, then they should both be on sidings.

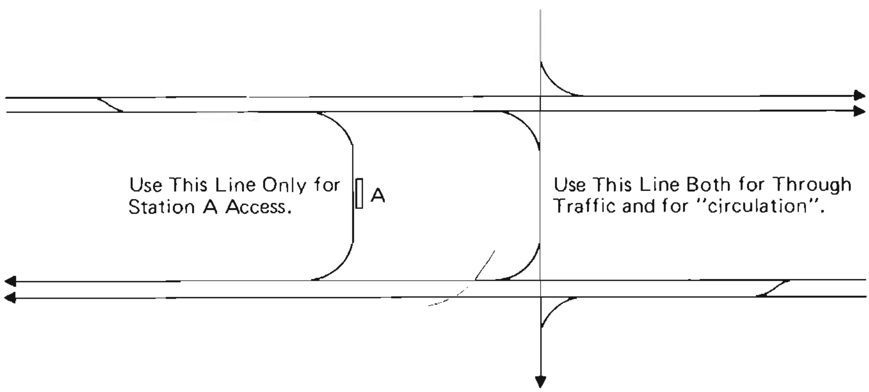


Fig. 2-15. Use of On-Line Station

It is of interest that the network of Fig. 2-13, being largely at one level, was designed to integrate into some planned pedestrian bridges similar to "skyways" of the Twin Cities.

Another approach aimed at working a large number of stations into a small CBD is that taken in Japan's Computer-Controlled Vehicle System (CVS). The designers envision a coarse grid of express lines extending over the metropolitan area. Typical line speeds would be 60 km/hr (about 37 mi/hr). Where these lines cross each other they would be at different elevations, in the usual manner. Within certain coarse grid squares there would be a fine grid, with all lines at one elevation. By using only one elevation it is, of course, necessary for vehicles on crossing lines to time-share an intersection, but the intersections can be made smaller because there is no need to change altitude. It is claimed that the fine grid spacing can be as small as 100 meters (328 ft) and that along any line, stations may be spaced 100 meters apart.

The problem with this approach is that it requires time-sharing the intersections and this introduces the possibility of a broadside collision. When a collision is imminent, it can be averted by warning the oncoming vehicle to apply emergency brakes. In CVS the emergency braking is accomplished by clamping onto a part of the guideway structure; the brake is explosively deployed. The emergency braking deceleration is 2 g. To avoid being thrown out of their seats when this sudden braking occurs, passengers are seated backwards.

Although we think that the CVS design has many clever features (see, for example, Sec. 4.6.7), we cannot endorse line crossings without grade separation because of the risk of broadside collisions. If crossing lines are at different elevations, then emergency braking decelerations can be much lower. Moreover, in contrast with the single-level emergency deceleration used in CVS, we feel that it is important to use a variable emergency deceleration so that no greater deceleration is used in any circumstance than is necessary to cope with the safety threat. This matter is explored further in Chapter 6.

2.6 RELATIONSHIP BETWEEN NETWORK CONFIGURATION AND SERVICE DEPENDABILITY

No matter how carefully the hardware components are chosen, no matter how much redundancy of critical components, and no matter how perfect the maintenance program, there will still be vehicle and other system failures. It is important, however, that when these failures occur they have a minimum effect on service dependability.

One way of protecting service dependability is to have a procedure wherein a failed vehicle is pushed by the vehicle behind it to an emergency siding where the people in the failed vehicle can transfer to a spare vehicle, after which maintenance personnel will pick up the failed vehicle. This procedure would ensure that service is not interrupted; vehicles behind the pushing vehicle lose at most a few seconds, and the parties in the failed and pushing vehicle are delayed by at most a few minutes.

This procedure, however, is not always feasible if the failed vehicle cannot be pushed or if the failure is some sort of blockage of the guideway. In that event, to avoid an accident, succeeding vehicles on the line would have to be brought to a stop and then detoured onto alternate paths to their destination. As soon as possible, the failed vehicle or blockage would have to be removed from the guideway, possibly by using street-driven equipment.

In a PRT network, rich in lines, there are usually many routes for getting from one point in the system to another, and if the operational strategy quickly reroutes the affected vehicles following a line blockage, then a high degree of service dependability can be maintained. In contrast, however, for those areas of a system which are line-haul in nature, a single line blockage could interrupt the entire flow down the corridor in question. There are, however, ways around this problem. Let us consider the example of an alignment along a freeway or other right-of-way on which it is practical to install a two-way guideway system. Now consider a third line being added between the other two. This line could serve as a bypass in the event of failure on either of the other lines. It would, of course, require frequent interchanges between the middle line and the other two. The method of line clearing is illustrated in Fig. 2-16 where four time "snapshots" are presented.

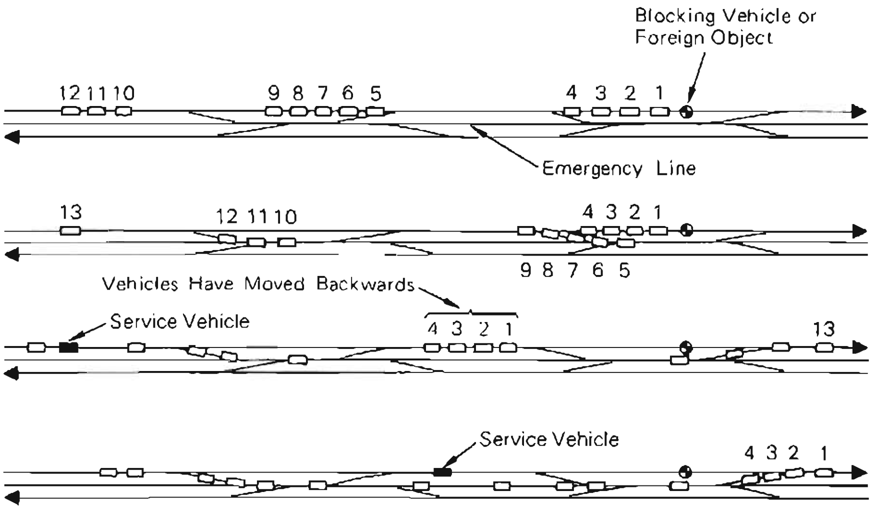


Fig. 2-16. Elevated, Two-Way Line-Haul Guideway – Sequential Events during Emergency Operations

The problem is even more severe where there are blockages on subway lines, for here we have not only the problem of arranging for traffic to bypass the blocked area, but there is also the problem of getting access to the blocked area for rescue and repair. For a one-way subway line, it may be advisable to connect the station sidings together, as illustrated in Fig. 2-17, into a continuous through line so that in the event of a failure on the main line the siding can be used as a bypass of the blocked area. At most, one station would have to be shut down to avoid interference with through traffic.

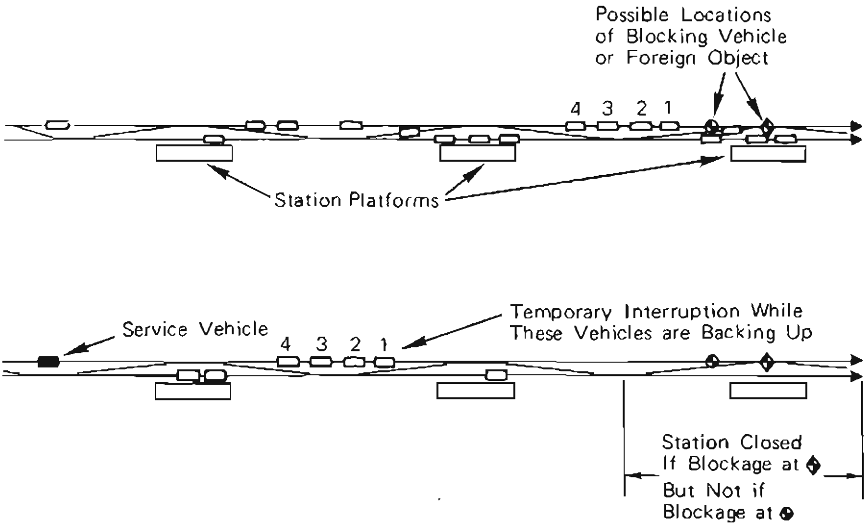


Fig. 2-17. One-Way Subway Line with Continuous Station Off-Line – Sequential Events during Emergency Operations

Chapter 3
STATIONS

Jack H. Irving

3.1 STATION TYPES

In this section we shall examine some of the general station types which are possible with PRT. They will be classified in accordance with the number of platforms, their arrangement on sidings, and whether the vehicles are stationary or moving when they are being deboarded or boarded. The variant known as a "docking" station also will be discussed.

3.1.1 The Single-Platform Station on a Simple Siding

The simplest station type, and one of the most useful, consists of a single platform adjacent to a simple siding. The same platform is used for deboarding and boarding, although in some modes of operation the vehicle, after being deboarded, may move forward before being boarded. The length of the platform depends on the maximum number of vehicles which must be simultaneously accommodated for deboarding and boarding. We refer to each vehicle location as a "berth." The number of platform berths, N_P, may typically vary from 1 to 3 in a residential station, and from 3 to 24 in a CBD station. Each berth should be about 1 ft longer than the vehicles to allow a 1 ft separation. Thus, for a 10 ft long vehicle, the berths would be about 11 ft.

Figure 3-1 is a sketch of the siding for this type of station. From left to right the sections of the siding are:

1. the entrance section, which accommodates branching from the through line onto the siding and accommodates deceleration
2. the input queue section, which provides a number of "input queue slots," the same length as the platform berths, for vehicles waiting to approach the platform
3. the platform section, with its N_P berths
4. the output queue section, where vehicles can be queued awaiting available space on the through line
5. the exit section, a mirror image of the entrance section.

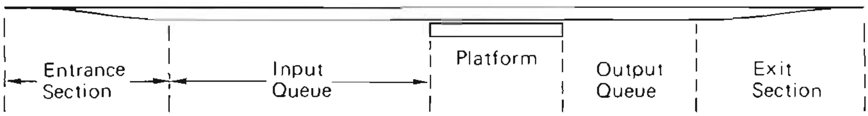
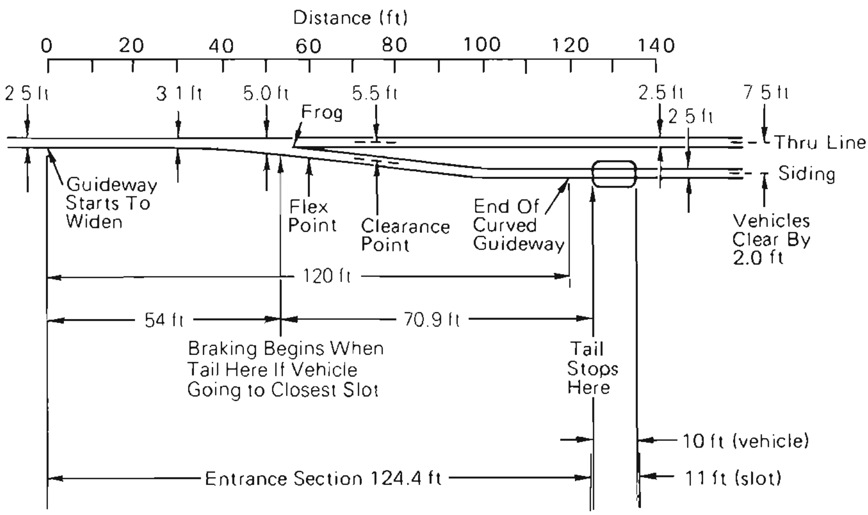


Fig. 3-1. Simple Siding for Single-Platform Station

Figure 3-2 is a plan view of an entrance section for a line speed of 30 ft/sec. It is based on an analysis carried out in Appendix A, Sec. A.3, and on the numerical assumptions listed in Fig. 3-2, which are quite conservative for all passengers seated.



Line Speed	30 ft/sec
Maximum Lateral Acceleration	3.75 ft/sec² (0.116 g)
Lateral Jerk	3.75 ft/sec³ (0.116 g/sec)
Braking Deceleration	8.05 ft/sec² (0.25 g)
Braking Jerk	8.05 ft/sec³ (0.25 g/sec)
Guideway Width	2.5 ft
Vehicle Width	5.5 ft
Lateral Clearance Between Vehicles	2.0 ft
Vehicle Length	10.0 ft
Input Queue Slot Length	11.0 ft (vehicle centered)

Fig. 3-2. Plan View of a Typical Entrance Section for a Line Speed of 30 ft/sec

In addition, it is assumed that a vehicle which is to stop in the closest input queue slot will start braking when its tail is 1.8 sec into the entrance section. As a result of this assumption, when the tail passes the clearance point,[1] the vehicle will be only 0.5 ft behind

[1] It will be noted from the figure that although the guideways diverge for 120 ft, vehicles on the two lines would clear once the tail of the one on the siding has passed the "clearance point" which is only 75 ft downstream of the onset of divergence. (The 120 ft is merely the line speed multiplied by 4 sec, the time to carry out a lateral displacement of 7.5 ft. This time depends only on lateral jerk and acceleration, and is independent of line speed.)

where it would have been with no braking. Clearly, this poses no hazard.

By allowing the divergence and braking regions to overlap, the total entrance section is only 124.4 ft long. As the figure shows, had we kept divergence and deceleration as sequential operations, the entrance section would have been 66 ft longer.

As noted, the exit section is just the mirror image of the entrance section, based on the assumption that the longitudinal acceleration and jerk (rate of change of acceleration) used in the exit section are numerically equal to the braking deceleration and jerk.

Figure 3-3 is a plot of the length of the entrance (or exit) section as a function of line speed, with all other assumptions as indicated above (see Eq. (A.14) of Appendix A). For the higher speeds, the quadratic term in the braking (or acceleration) distance dominates and the length grows rapidly with line speed. For line speeds below 27.37 ft/sec, the braking (or acceleration) can all take place within the divergence (or convergence) region, and the curve becomes a straight line, representing the length of this region.

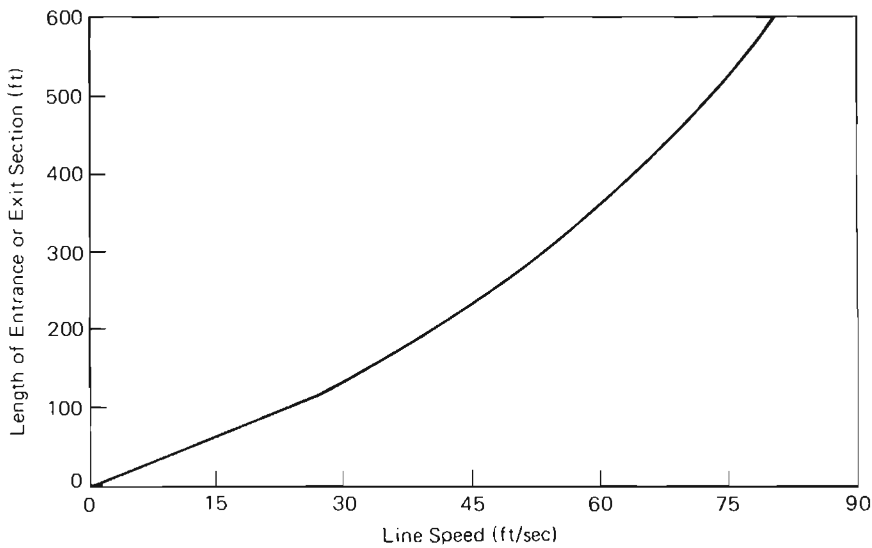


Fig. 3-3. Length of Entrance or Exit Section versus Line Speed

The input queue section of Fig. 3-1 not only provides a temporary storage site for occupied vehicles waiting to approach the platform but it also provides a place to queue empty vehicles. In Sec. 3.2 it is shown that, to obtain good performance of an activity-center station during morning and evening peak traffic periods, the input queue should be approximately twice as long as the platform and the output queue equal in length to the platform.

During peak traffic hours at a busy station there is a continuing flow of vehicles and the station works very well. But when the station is not busy, and especially in a residential area, the configuration of Fig. 3-1 has a substantial weakness. Imagine that it is just before 7:00 a.m. and a number of empty vehicles are queued in the input-queue section awaiting residents of the area on their way to work. Now imagine that one person who works in the residential area or one of its neighborhood stores arrives by PRT. To bring his vehicle to the station platform (assumed small), it would be necessary to advance almost all of the empty vehicles past the platform, and thus "waste" them. One way around this problem will now be discussed.

3.1.2 Single-Platform Station on a Siding with Two Entrances

To avoid wasting empty vehicles when bringing in an occupied vehicle one merely needs to have two entrances to the siding, one for empty vehicles and one for occupied vehicles. This is illustrated in Fig. 3-4. When this type of configuration is used at a residential station, it is possible to keep empty vehicles queued during the morning peak, even though some occupied vehicles might be arriving.

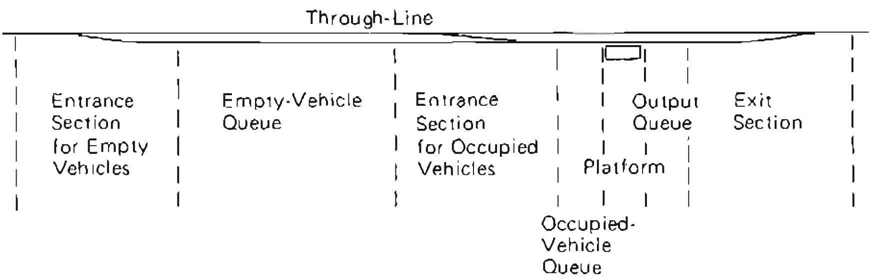


Fig. 3-4. Siding with Two Entrances for a Single-Platform Station

Because of the low arrival rate of occupied vehicles during the morning rush hours, and because they need not be queued for long, the occupied-vehicle input queueing space can be very short, perhaps two or three slots in length. In contrast, the empty-vehicle queueing space must store a significant number of empty vehicles if the queue is not to be depleted with surges in demand. The empty-vehicle queueing space must also be substantially larger than the average length of the empty-vehicle queue if there is to be room for empty vehicles that arrive during periods of a lull in the demand. This is especially true if empty vehicles are not readily accessible from nearby vehicle storage facilities, but, on the other hand, if a storage facility is very near, then the capacity for storing empty vehicles at the station can be reduced.

In Sec. 5.7 we discuss the problem of maintaining an adequate supply of empty vehicles and present the performance of a station of the type shown in Fig. 3-4, when some 3 to 5 parties/minute are departing from the station. This is a particularly heavily used residential station in contrast to the average. The station simulated had a platform length of 3 slots, an occupied-vehicle queuing space of 3 slots, and an empty-vehicle queuing space of 15 slots, although the average number of empty vehicles queued was only half this number.

The same station, operating during the evening, would not require bringing any empty vehicles to the station, but might require a somewhat longer space for the occupied-vehicle input queue. Consequently, the mode of operation during the evening would be to bring the occupied vehicles onto the siding through the first entrance rather than the second. The last few vehicles unloaded at the platform would remain there for any parties wishing to depart or until replaced by new occupied vehicles to be deboarded.

It also may be advisable to use two entrances to the siding of an activity-center station, but with the second entrance being used only during the nonpeak hours for occupied vehicles. This is illustrated in Fig. 3-5. We are assuming that the station platform has been sized to handle peak-hour traffic in accordance with the procedures to be discussed in Sec. 3.2.

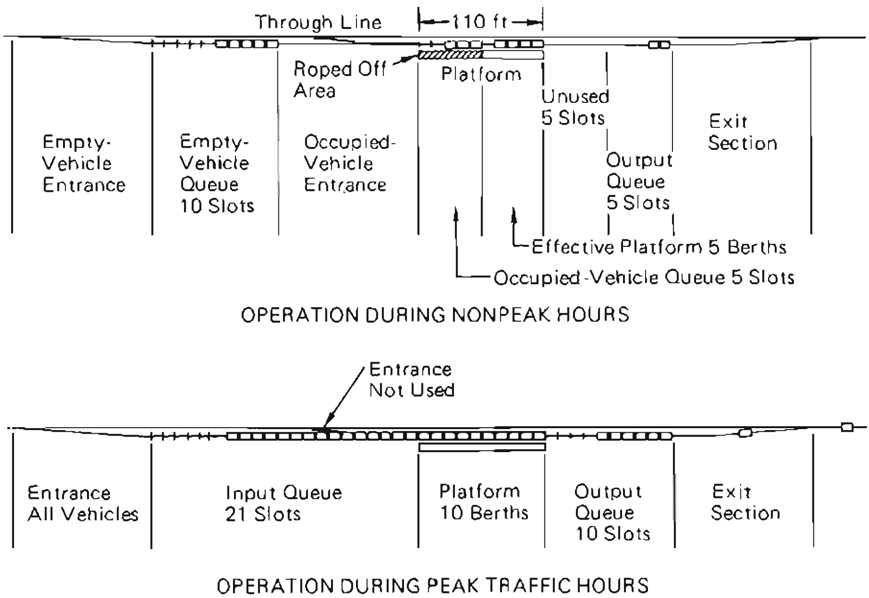


Fig. 3-5. Activity-Center Station with Two Entrance Sections

Let us assume, for purposes of illustration, that the platform has 10 berths. During the nonpeak hours, 5 berths clearly are adequate. Consequently, as indicated in the upper sketch of Fig. 3-5, the last 5 platform gates could be roped off and the berths for these gates could be used to queue occupied vehicles. Just upstream of the platform would be the second entrance section, used for occupied vehicle entrance during the nonpeak hours. With a line speed of 30 ft/sec, the second entrance section would be about 125 ft long. Just upstream of that there would be a queue section with perhaps 10 slots for empty vehicles. During peak hours the second entrance would not be used to admit vehicles and consequently could be used as a part of the input queue, holding about 11 vehicles. When combined with the queuing space that was used for empty vehicles during nonpeak hours, this gives a total input queuing space of 21 slots upstream of the platform. This is ideal for peak-hour operation.

3.1.3 Two-Platform Stations

There are two types of two-platform stations which are worth considering.

One is the kind illustrated in Fig. 2-9 for installation near an intersection of a one-way network. As indicated in that figure, the departure platform is placed so as to allow departure along either outgoing line, and the arrival platform is placed to allow arrival from either incoming line. The departure platform would require an input queue for storing empty vehicles only, and the arrival platform would require a temporary input queue for occupied vehicles waiting to get to the platform. Both platforms would require an output queue for vehicles which have left the platform and are waiting to find available space on the through line. Of course, both platforms would require entrance and exit sections.

The second type of two-platform station is illustrated in Fig. 3-6. It is appropriate for placement midway between two intersections. Both platforms are on a single siding but separated by an empty-vehicle queueing space. Both empty and occupied vehicles come in through the single entrance section, wait only as long as necessary in the input queue, and proceed to the arrival platform. Since no boarding takes place at this platform, all vehicles leaving the plat-

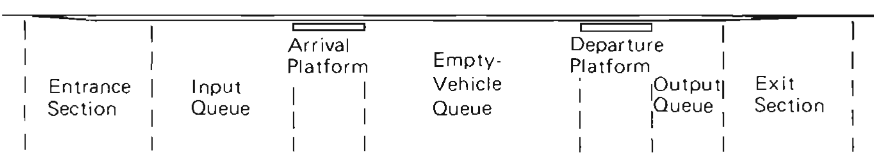


Fig. 3-6. Two-Platform Station on a Single Siding

form are empty and are stored in the empty-vehicle queue. They are called forward to the departure platform, as required, to serve departing parties. As before, the output queue is for temporary storage while awaiting space on the through line. A station of this type requires a somewhat longer siding than the others we have discussed.

Because both of the two-platform stations discussed here require an extra platform and more siding than the stations described in Secs. 3.1.1 and 3.1.2, they clearly will be more costly and therefore would be used only when specially indicated. One of their advantages is that they segregate arriving passengers from departing passengers, which may simplify passenger flow within the station itself. Also, because neither of the platforms has both deboarding and boarding, the average dwell time of vehicles at the platform is somewhat reduced. Consequently, to achieve capacities similar to the one-platform stations, the platforms can be somewhat shorter, i.e., fewer vehicles need to be brought in for a loading or unloading cycle.

A prior Aerospace Corporation paper[2] examined the two-platform station of the type illustrated in Fig. 3-6 as an activity-center station during peak traffic, and concluded that it does not have performance advantages over stations of the type illustrated in Fig. 3-1. However, it is the opinion of the author that the work cannot be considered definitive and there still are some open questions.

3.1.4 The Moving-Belt Station

All of the stations discussed up to now have involved deboarding or boarding a stationary vehicle. Therefore, they have sometimes been called "taxi stations" because the operation is similar to loading a queue of standing taxicabs. In contrast, it is possible to design a station so that people deboard and board slowly moving vehicles from a moving belt whose speed is matched to that of the vehicles.

To illustrate, let us assume as before that the vehicles are 10 ft long and they are spaced 1 ft apart in the station area. If the vehicles and the belt move at 2.2 ft/sec, one vehicle will enter the deboarding/boarding area each 5 sec. The moving belt should be long enough to provide ample time for vehicle deboarding and boarding in all but exceptional cases. Thus it probably is advisable to provide 60 sec for deboarding and subsequent boarding. This would require a belt 132 ft long. In those rare instances where a party had not completed boarding within that period of time, there would need to be sensors

[2] K.J. Liopiros, "PRT Station Operational Strategies and Capacities," *Personal Rapid Transit II*, U. of Minnesota (Feb 1974).

that automatically stop the belt and the stream of vehicles until the party had boarded.

The advantage of this type of station is that it provides continuous operation and that each vehicle automatically gets out of the way of the next. However, it has two significant defects. First, it is substantially more expensive because of the cost of the moving belt and the length of platform required. Second, there is a serious safety problem, that of protecting the passenger from falling into the guideway. For the "taxi stations" the passenger is protected by a continuous wall or fence which separates the station platform from the guideway; the vehicles all stop at fixed berths and the passengers enter through "gates" in the wall or fence. But since the vehicles are moving in the moving-belt station, the only way to protect the passenger would be to have a moving wall with openings through which he could enter a vehicle, and this would further complicate the design.

The throughput of such a station is very easy to compute. In our example the throughput would be one vehicle each 5 sec. However, by using platooning techniques, it is possible to exceed such throughputs at fixed platform stations. Consequently, we see no need for further consideration of the moving-belt station.

3.1.5 Docking Stations

A "docking station" is a fixed platform station which uses a guideway design capable not only of moving vehicles forward (and backward) but sideways. In such a design the vehicle can pass by other vehicles at the platform and then move sideways into a chosen platform berth. When the vehicle is loaded and ready to depart it can move sideways and then bypass the vehicles ahead of it without delay. The ability to dock has been cited as one of the advantages of air suspension.

The real domain of docking stations is their use in GRT systems. The reader will recall that in GRT operation there are a number of people waiting on a platform for different vehicles. The order of arrival of these vehicles will depend on what stops they have had to make and how long each stop took. Therefore, without docking it would be difficult to predict far in advance at what berth each vehicle will stop. Thus, as a vehicle approached a station it would be necessary for the people waiting for it to scurry around to the right berth. But with the ability of sideways motion a vehicle can always stop at a prescribed berth, moving forward from the input queue when that berth is available and bypassing other vehicles at the platform.

There is no such requirement for a PRT system since all vehicles are, from the standpoint of the user, identical. It might be argued that docking PRT vehicles is useful because it would allow departing vehicles to bypass a vehicle delayed by a slow boarder. But on the average, this will not help because the dynamics of moving the vehicles sideways and then forward adds enough time to the typical advancement cycle to more than compensate the occasional time savings. Moreover, the docking station, because it requires a wider guideway, and a particular type of guideway, will present greater aesthetic intrusion and will be more costly.

3.2 PERFORMANCE OF AN ACTIVITY-CENTER SINGLE-PLATFORM STATION

This section will address the question of how long the input queue, platform, and output queue need be at a single-platform activity-center station, and how that station should be operated, to provide a specified "throughput" without unduly sacrificing the quality of service. By "throughput" we mean the number of vehicles deboarded each hour during the morning rush hours or boarded each hour during the evening rush hours.

In 1973 The Aerospace Corporation developed two simulation programs[3] — one to study the performance of a single-platform station of the type illustrated in Fig. 3-1, and the other to study a two-platform station of the type illustrated in Fig. 3-6. During the preparation of this book, the former program was reexamined and found to have several significant errors. In addition, it was felt that a more realistic strategy of operation could be employed. As a result, an entirely new program for a single-platform activity-center station (Program "STATION") was written.[4] This program has provided the performance data that will be the topic of this section.

In Sec. 3.2.1 we treat some preliminaries related to boarding and deboarding and the advancement of vehicles in the station area. In Sec. 3.2.2 we discuss operational strategies. Then in Secs. 3.2.3 and 3.2.4 we examine station performance during the morning rush hours and the evening rush hours, respectively. The criteria of satisfactory performance are quite different for those two periods. The results are summarized in Sec. 3.2.5.

3.2.1 Some Preliminaries

To find the number of vehicles that can be processed each hour by a station it is necessary to know how long it takes for parties to deboard and board and how long to move vehicles from the input

[3] See footnote 2.
[4] Frank Goroszko and J.H. Irving.

queue into the platform area.

Later we shall be emphasizing platoon operation of a station where a number of vehicles are advanced simultaneously from the input queue to the station platform. Then deboarding and boarding begins. When all vehicles at the platform have completed deboarding and boarding (and certain other events have occurred), they will be advanced into the output queue while a new platoon is brought up to the platform. Clearly, the time to get all vehicles at the platform ready to move is paced by that vehicle which takes the longest to deboard and/or board. During the morning rush hours all vehicles going to an activity-center station will be deboarded but only a few vehicles will be boarded. For some platoons there may be no vehicles to be boarded; in that case the pace will be set by the slowest deboarding party. But, when one or two vehicles of the platoon are boarded, one of these vehicles may set the pace even though the deboarding party is quite average in its time to deboard.

Because deboarding and boarding times vary from party to party, both because of the variation in party size and because of the difference in agility in the individual passengers, we cannot use specified boarding and deboarding times but shall assume them to be distributed. Because it is the slower-to-average parties that determine when the platoon is ready to move, the exact nature of the distributions for the short deboarding/boarding times is not important, but the distribution should have a "tail" for the longer times to properly account for the slower parties, even though they are in the minority.

One mathematical form which would seem to have the right shape is the lognormal distribution which appears in Fig. 3-7. (The logarithm of the deboarding or boarding time is normally distributed.) It will be noted that the distributions have the desired tails for the longer times and even behave reasonably for shorter times. To specify a lognormal distribution, two numbers are required — the mean and the root-mean-square (RMS) deviation, the latter being a measure of the average spread. Altogether the simulation accepts four numbers, two for the deboarding distribution and two for the boarding. The numbers which we used are:

	Deboarding	Boarding
Mean Time	8.0 sec	10.0 sec
RMS Deviation	3.0 sec	4.0 sec

We know of no definitive data that can be used to establish these numbers, although we are aware of an experiment carried out by Messerschmitt-Bölkow-Blohm (MBB) several years ago. A number of automobiles (4-door sedans) were driven up to a marked

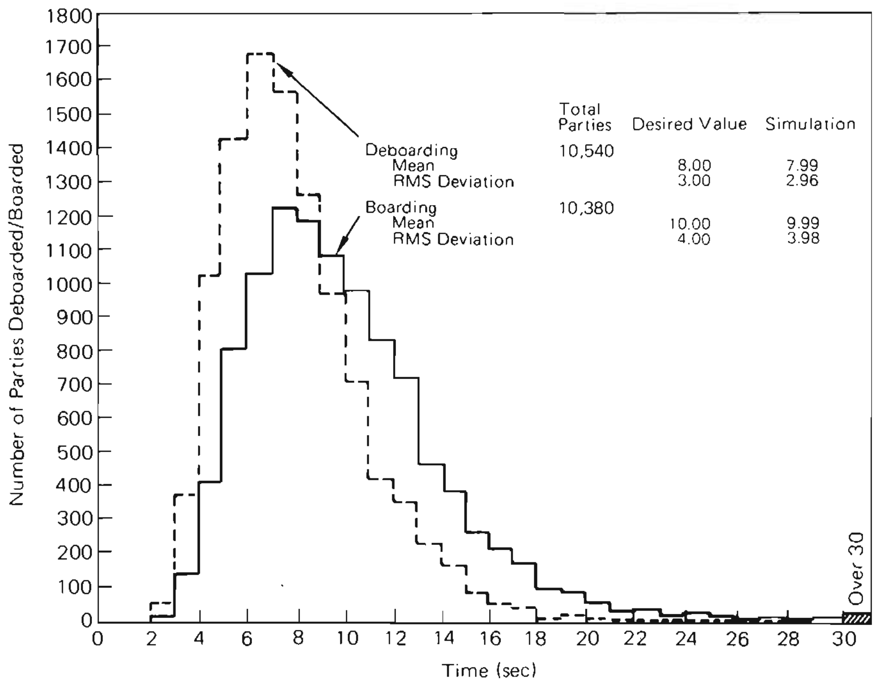


Fig. 3-7. Lognormal Distribution for Vehicle Deboarding/Boarding

area to load into or discharge passengers from the back seat. With an average party size of 1.5, the mean unloading time was 8 sec and the mean loading time was 10 sec. If MBB employees were the subjects of the experiments, it might be argued that they probably were more agile than the representative passenger; but we are dealing here with simulating rush hour traffic at an activity-center station where most of the passengers will be workers on the way to work or on their way home. Of greater importance, it should be far easier to deboard or board a PRT vehicle than the back seat of an automobile. The PRT door would be wider, the vehicle higher, and possibly a portion of the roof will slide away too; crouching should be minimal. Nevertheless, to be on the conservative side we took the MBB means for purposes of our simulation.

Now, we consider the time required to index the platoon forward. Based on limiting acceleration and deceleration to 0.25 g, jerk to 0.25 g/sec, and speed to 22 ft/sec, the time for indexing is plotted in Fig. 3-8 as a function of the number of slots moved.

It will be recalled that we have assumed that input queue slots, station berths, and output queue slots are all 11 ft long. We limited the station speed to 22 ft/sec to keep headway down to 0.5 sec, although this may not be necessary. In addition, we assumed a some-

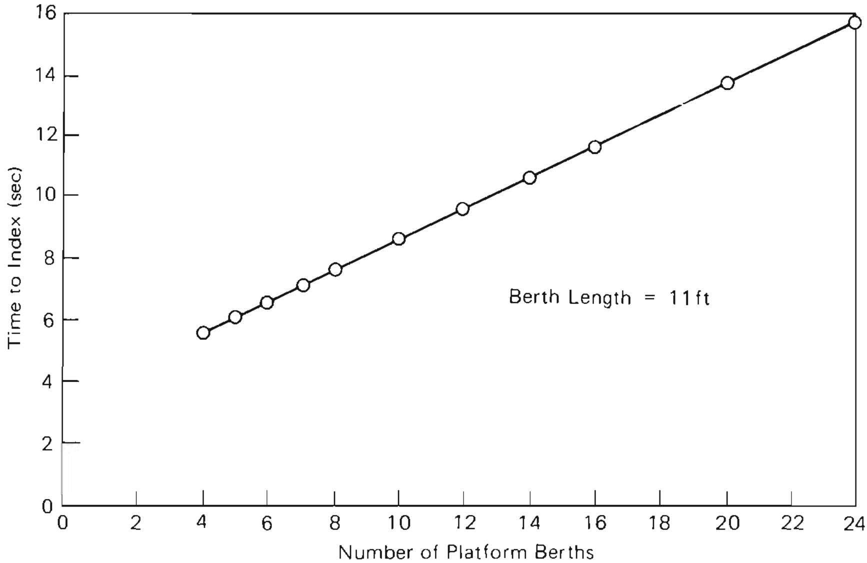


Fig. 3-8. Time to Index

what longer entrance section and exit section than now appears necessary in the light of the discussion of Sec. 3.1.1 and Sec. A.3 of Appendix A. The effect of these conservative assumptions is to slightly underestimate station performance.

3.2.2 Operational Strategies

Most of the early work of Aerospace was based on the strategy of allowing a vehicle to move forward into a station berth whenever one was available. During periods of moderate station usage this means that each vehicle as it arrives would stop in the berth behind the one which was then deboarding or boarding. When the wave of stopping vehicles got to the position of the last berth, the front berths would have been cleared out, but they would not be accessible from the input queue. Consequently, during the deboarding/boarding of the vehicle in the last berth, a short input queue would develop and when the vehicle in that berth moved forward, the input queue would be emptied in a platoon-like movement. This would be followed by another backward-moving wave through the berths. As the traffic rate increases, this kind of operation, which has been referred to as the "trickle operation," approaches that of platooning, and consequently should have a similar performance.

We feel, however, that this type of operation should be ruled out because it may present a safety hazard. Recalling that the vehicles are only 1 ft apart at the station platform, if a vehicle were to overshoot its mark, it could bump another vehicle with a passenger alighting or entering. To avoid this possibility we have restricted our attention to platoon strategies.

In a platoon strategy, entering vehicles also stop in the input queue only 1 ft behind the vehicle ahead, but if they were to overshoot they would strike a vehicle with seated passengers rather than one being boarded or deboarded. At an appropriate time a platoon of vehicles is brought forward from the input queue to the platform as the platoon alongside the platform is advanced to the output queue.

The advancement or "indexing" of the platoons cannot take place, of course, until there is adequate room in the output queue. The process of merging vehicles from the output queue onto the main line is relatively rapid because, in addition to the vacant spaces which were originally on the line, each vehicle which enters the siding creates another vacant space. As a consequence, the time to merge vehicles from the output queue onto the main line is less than the time necessary to bring in an adequate number of vehicles for the next platoon. This being so, the output queue need be no longer than the platform. Only rarely will there be a few seconds wait for the output queue to "clear." (We use "clear" to mean that the last vehicle in the output queue has started its acceleration to merge into the main-line traffic.)

We have considered two different platoon strategies:

Strategy A — index the platoons forward as soon as deboarding and boarding have been completed and the output queue is clear; and

Strategy B — index the platoons forward as soon as deboarding and boarding have been completed, the output queue is clear, and there are sufficient vehicles in the input queue to fill all platform berths.

It will be noted that in Strategy A the average elapsed time ("cycle time") between successive indexings is somewhat less than for Strategy B because it is never necessary to wait for the arrival of vehicles into the input queue. But the platoon brought forward in Strategy B will fill every berth while that of Strategy A is of variable length, depending on the number of vehicles that were available in the input queue at the time of indexing. We reasoned that these should be nearly offsetting factors affecting throughput and therefore the two strategies should lead to about equal throughputs. However, for operation during the evening rush hours, there is a very pragmatic reason for choosing Strategy B.

To understand this reason, imagine that we have a station having 10 berths at its platform and operating under Strategy A. Because of the variable length of the platoon brought in from the input queue there would be no assurance that vehicles would stop at Berths 9 or 10. Especially during periods when there was very little reverse flow, i.e., no arrival of occupied vehicles, the loading process might be completed rapidly and the platoons would be advanced, in accordance with Strategy A, before 10 vehicles were available in the input queue. As a result, there would be a learning process and people would not go to Berths 9 and 10. Then, even when there were adequate vehicles in the input queue to send vehicles to these berths, the vehicles would not be used. (The system could not be designed to wait for people to walk, or even run, from other berths and then insert their travel cards into the cardslots next to Gates 9 and 10 because this would add so much time to the boarding process as to substantially delay the next indexing.) As a consequence, Berths 9 and 10 would fall into disuse and would, in effect, be wasted.

In contrast, when Strategy B is used, all gates can be used equally and there are no preferred gates. One method of operation under Strategy B might be to direct the passenger to a specified gate at the time that he requests his trip. In this way gate usage may be kept evenly distributed.

Although we feel that Strategy B will work better than Strategy A for the evening rush hours, we have no reason to prefer it for the morning rush hours. Indeed, Strategy A may be somewhat better for the morning rush hours, especially during slack periods, because it would not keep people waiting in the input queue as long. But the simulations we have carried out are not directed to slack periods but rather to finding the maximum throughput, and we believe that, when the station is operating near capacity, the two strategies, as remarked earlier, should be nearly equal in their performance for the morning rush. Consequently, because of the pressure of time, we chose Strategy B for our simulation program — both for the evening and morning rush hours.

One other operational aspect is that which was alluded to earlier. People must not be allowed to start boarding after the others at the platform are well along in that process because that would materially delay the next indexing. This means that a person departing from a specified berth must have inserted his travel card into the slot beside the gate and withdrawn it before the vehicle comes to a stop or within a specified short time thereafter; otherwise, the gate and vehicle door will not open.

There are two other elements of strategy which affect only the evening operations. One of these has to do with bringing in an excess supply of empty vehicles to ensure that, even with fluctuations in passenger demand and in empty vehicle arrival, there still will be a supply of empty vehicles adequate to ensure that a long queue of passengers waiting to depart will not develop. As a result of this excess of empty vehicles, there will be many empty vehicles denied access to the input queue because it will be full. This has no serious consequences because the same empty vehicles could then be made available to neighboring stations in the CBD or other activity center to ensure that they, too, have an adequate supply.[5] An unfortunate consequence might be that there would be insufficient room in the input queueing space to accommodate an occupied vehicle that might be arriving during the evening peak traffic. This brings us to the last strategic measure, which is to reserve the last space or two in the input queue for occupied vehicles only.

3.2.3 Operation During the Morning Rush Hours

Operations at an activity-center station during the morning rush hours are much simpler than during the evening rush hours. During the morning all arriving vehicles will be occupied and the only criterion of acceptable service from the traveler's point of view is whether or not his vehicle is allowed to enter the station siding when it comes to the siding entrance. If his vehicle cannot enter the siding because the input queue is full (called a "miss"), the vehicle is required to circle the block and try again.[6] We define "miss rate" as the number of misses divided by the number of vehicles attempting to enter the station siding. The problem, then, of designing a station for morning operations is to find the lowest cost combination of input queue, station platform, and output queue which provides the necessary throughput at an acceptable miss rate.

Although there may be a few parties departing from the activity-center station during the morning rush hours, there are so many vehicles available for them that they will have immediate service. Their need to board, however, will sometimes delay the time when vehicles are ready to be indexed forward, and as a result the station

[5] The excess vehicles serving the CBD or an activity center could be thought of as a circulating reservoir of vehicles for that area during the evening rush hours. Any station requiring one of these vehicles can pull it in off the main line. Because this reservoir of empty vehicles is shared, the number of excess empty vehicles can be a smaller percentage of the total vehicle throughput for the area than would be required without sharing. The number of excess empty vehicles required is proportional to the square root of the demand. Hence, if 16 similar stations shared a common reservoir of empty vehicles, the number of excess empty vehicles routed to that reservoir would only need to be 4 times the number that would have been sent to each individual station without sharing.

[6] Once a vehicle has circled the block, it would be given priority over any of its neighbors that had not circled.

throughput will be diminished.

Before discussing the real problem of operating within acceptable miss rates, we shall first discuss the station throughput under saturation conditions. During the morning rush hours, saturation conditions are those where a very large number of occupied vehicles are trying to get into a station, so many that most of them cannot be accommodated. "Saturation throughput" is the number of vehicles/hr (or parties/hr) deboarded under these conditions. Although the miss rate under saturation conditions would be entirely unacceptable, the saturation throughput is still a useful concept because it serves as a guide in sizing the station to meet specified throughput requirements.

The saturation throughput is plotted in Fig. 3-9 for three values of reverse flow, as a function of N_P, the number of platform berths. As noted above, when there is reverse flow, i.e., when there are also parties to be boarded, the throughput is diminished. Under saturation conditions the input queue is full or nearly full at the time the deboarding and boarding are completed; consequently, even with Strategy B there is only rarely a need to delay indexing while waiting for vehicles to enter the input queue. Because of this, as long as the input queue is a little longer than the platform, the saturation throughput is almost independent of the length of the input queue.

Each of the circled points in Fig. 3-9 was obtained by a computer simulation run. It is seen that by going to sufficiently long platforms,

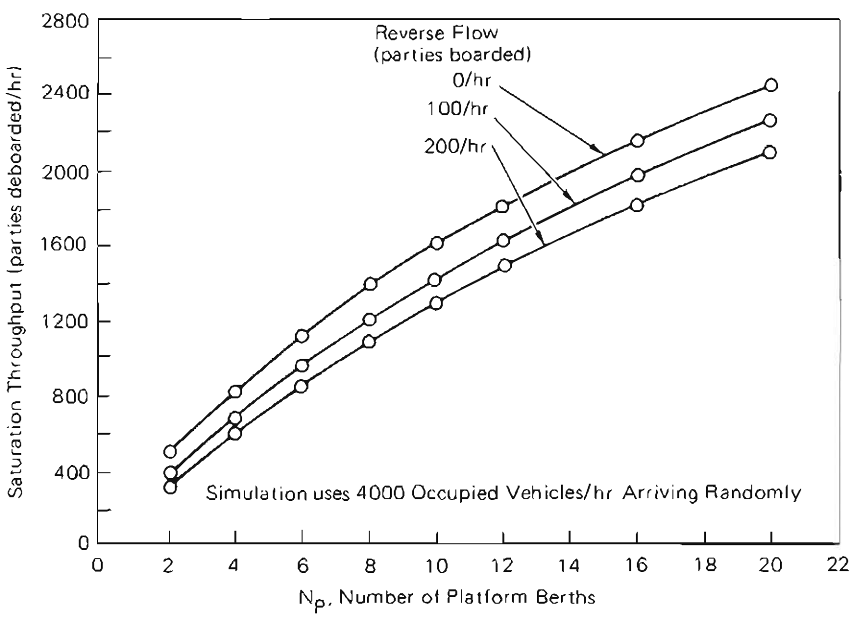


Fig. 3-9. Saturation Throughput for Morning Rush Hours

saturation throughputs can exceed 2000 parties/hr.

Now we turn to the real problem of sizing a station to meet a specified throughput requirement with an acceptable miss rate. Here the length of the input queuing space takes on an important role because the queuing space, if long enough, can store a temporary surge of arriving vehicles and thus lower the miss rate.

We illustrate the design process by considering a throughput requirement of 1000 parties/hr to be deboarded and a reverse flow requirement of 200 parties/hr to be boarded. Referring to Fig. 3-9 we see that there must be 8 platform berths to achieve a saturation throughput of over 1000. (A platform with 7 berths saturates at about 950 parties/hr.) To bring the miss rate down to an acceptable level, it might be expected that a platform with more than 8 berths would be required. Consequently, we investigated platforms with 8, 9, 10, and 12 berths. In addition, to find the impact of using an underdesigned station, we also investigated a station platform of 7 berths. For each of these, we varied the input queue length over a considerable range. The results are plotted in Fig. 3-10. There we plot miss rate versus total station slots.

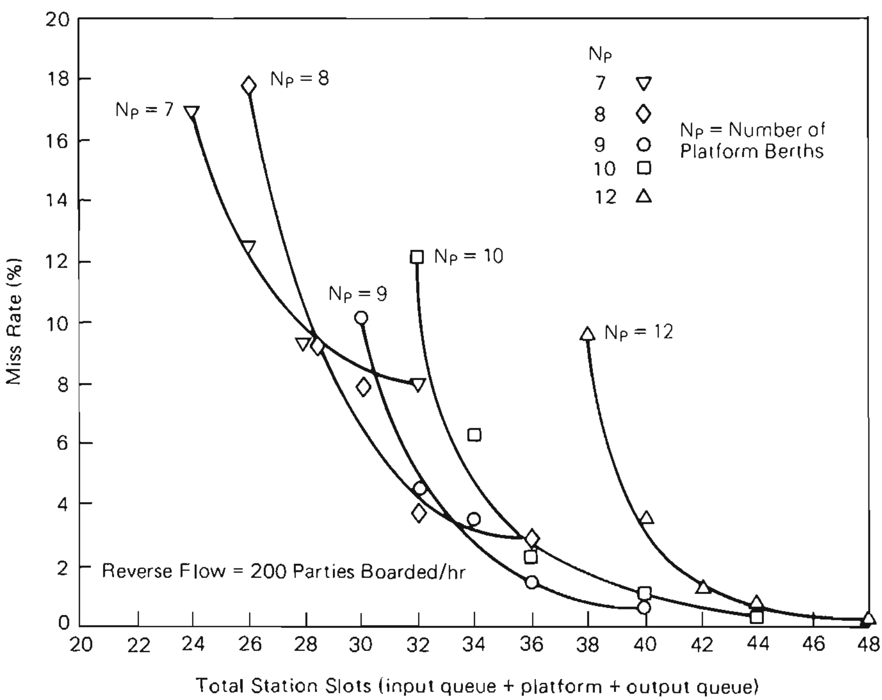


Fig. 3-10. Miss Rate During Morning Rush Hours for a Throughput of 1000 Parties Deboarded/hr

Each of the plotted points represents the average of at least four simulations, each representing two hours of activity. For the higher miss rates there is a considerable spread in the data to be averaged, but at the lower miss rates the spread is quite small.

Figure 3-10 shows that if the miss rate is to be brought below 1%, the station platform must be at least 9 berths long. It also shows that for any specified miss rate the station with 9 berths is superior to that with 10 or 12 berths because it is less costly. Not only is the platform shorter, but the siding is reduced because it requires fewer total station slots in the input queue, platform area, and output queue combined. This is not difficult to understand when it is recalled that we are here working with a fixed arrival rate of 1000 vehicles/hr. If a platform of 9 berths with an adequately long input queue is adequate to handle the throughput at a low miss rate, then platforms with 10 or 12 berths are overdesigned, adding unnecessary slots both at the platform and in the output queue.

It might be noted that if the miss rate is to be less than 1% for the 9-berth platform, the input queue should be at least 20 slots long, corresponding to 38 total station slots.

Figure 3-11 presents similar data for three other throughput requirements — 500, 1500, and 2000 vehicles/hr to be deboarded. Corresponding optimum platforms appear to have 5, 16, and 24 berths, respectively. If the miss rate is to be 1%, input queues should have 10, 28, and 39 slots, respectively.

3.2.4 Operation During the Evening Rush Hours

Operation during the evening rush hours is more complex than during the morning rush hours.

There are two criteria of acceptable service. The principal one will be how long parties must wait before they are able to board. A secondary criterion is that there should be an acceptable miss rate for the relatively few people arriving at the activity-center station during the evening peak traffic. As indicated earlier, this miss rate can be kept under control by reserving the last slot or two in the input queue for occupied vehicles only. We carried out most of our simulations with only one slot reserved, and found that this led to a miss rate of about 1 or 2% for occupied vehicles. With two slots reserved, there are virtually no misses.

As previously discussed, we found that it was necessary to bring in an excess of empty vehicles to ensure against the development of a long queue of waiting passengers. We simulated several cases where the planned arrival rate of vehicles was equal to the planned departure rate of parties. In some of these simulations, depending on how the

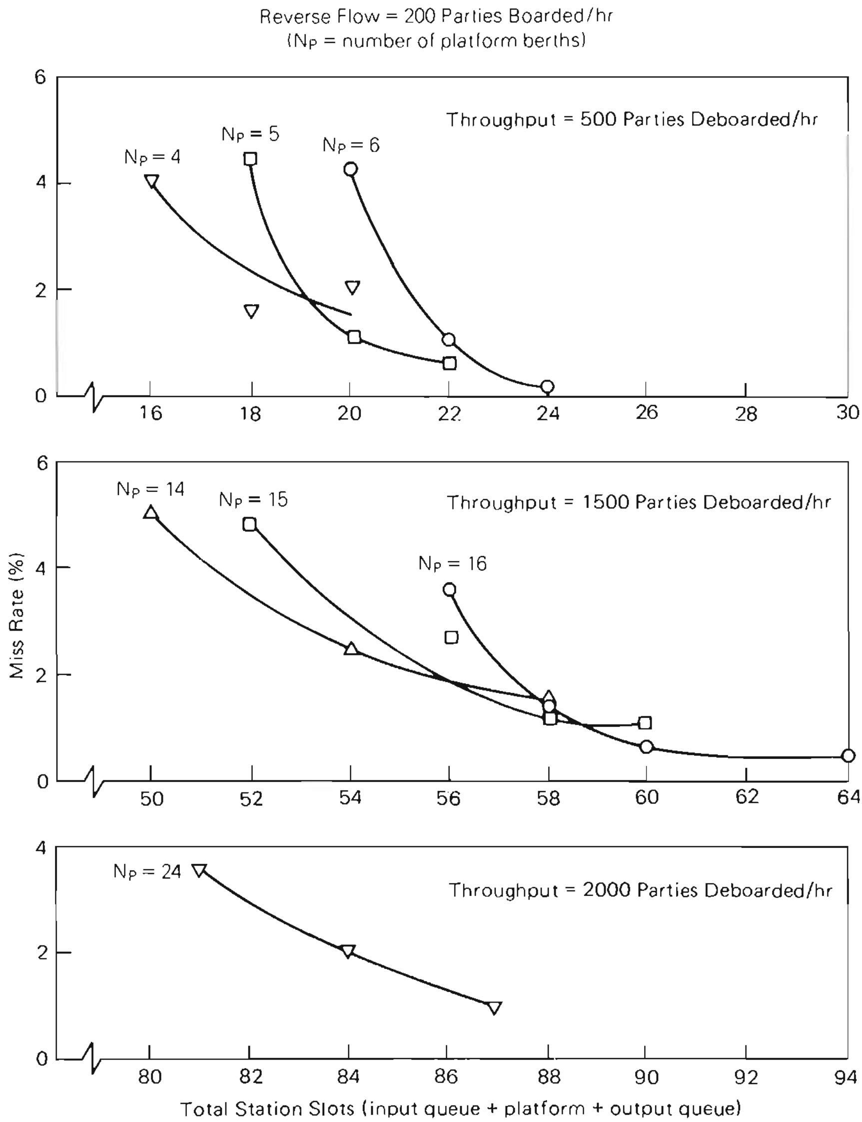


Fig. 3-11. Miss Rate During Morning Rush Hours for Various Throughputs

random numbers fell, vehicle arrivals would stay ahead of demand and no significant queues of waiting parties developed. In other cases, where the random numbers were less favorable, vehicle arrivals fell behind, and long queues of waiting passengers resulted. Once a queue has developed as a result of fluctuations in demand and/or vehicle supply, it is difficult to eliminate the queue and it persists for a long time. The only solution, as indicated, is to order an excess of vehicles.

As might be expected, we also confirmed that for the smaller stations where average vehicle arrival rates are lower, fluctuations become relatively more important and a larger percentage of excess vehicles must be ordered.

As during the morning rush hours, it is useful to find saturation throughput for the evening rush hours as a guide to station sizing. "Saturation throughput" here is defined as the number of parties boarded each hour under saturation conditions. For the evening, saturation conditions imply that vehicle indexing is being paced entirely by the time necessary to deboard and board vehicles, and not by the availability of parties wishing to depart or vehicles to serve them. Fig. 3-12 is a plot of saturation throughput versus N_P, the number of berths at the platform. Again, the throughput is given for three levels of reverse flow, i.e., the number of deboarding parties/hr. Comparing Fig. 3-12 to Fig. 3-9, the throughputs are seen to be somewhat smaller because of our assumption that the mean boarding time is 10 sec while the mean deboarding time is only 8 sec. However, as the reverse flow becomes significant, the pacing vehicles are those which are both deboarded and boarded, and the two saturation throughputs should be approximately equal.

To illustrate the design process for the evening rush hours, we consider a throughput requirement of boarding 1000 parties/hr with a reverse flow requirement of deboarding 200 parties/hr. From Fig. 3-12 it is seen that the platform must have at least 8 berths to exceed

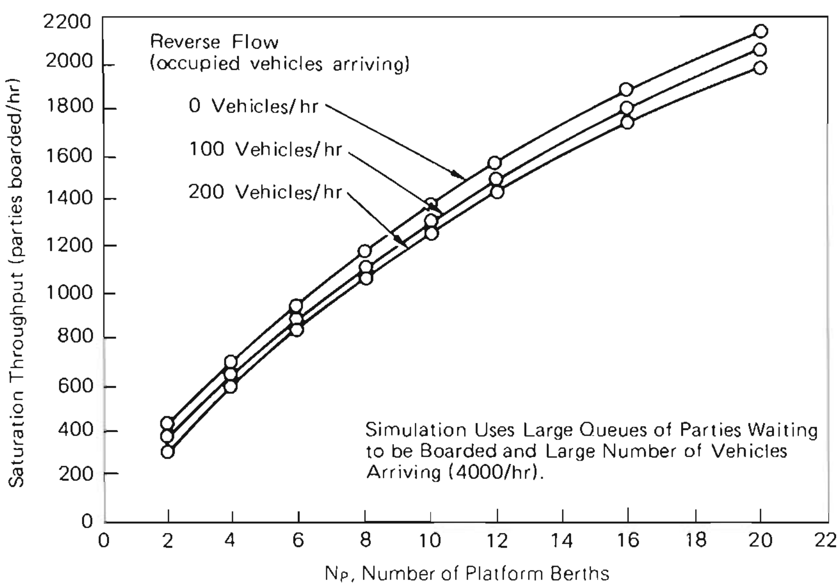


Fig. 3-12. Saturation Throughput for Evening Rush Hours

the throughput and reverse-flow requirement. Consequently we considered stations with platforms having 8, 9, 10, and 12 berths and, as in the morning case, also considered one with 7 berths to find the effect of an underdesigned station. Again, several input queue sizes were investigated for each platform. The results are presented in Fig. 3-13 where the average waiting time is plotted against the total station slots. The waiting time here is defined as the elapsed time from the instant when a person is at the station gate ready to board to the time when his vehicle arrives at the gate. The dashed curves represent simulations in which 5% extra vehicles were ordered, and the solid curves are for 10% extra vehicles.

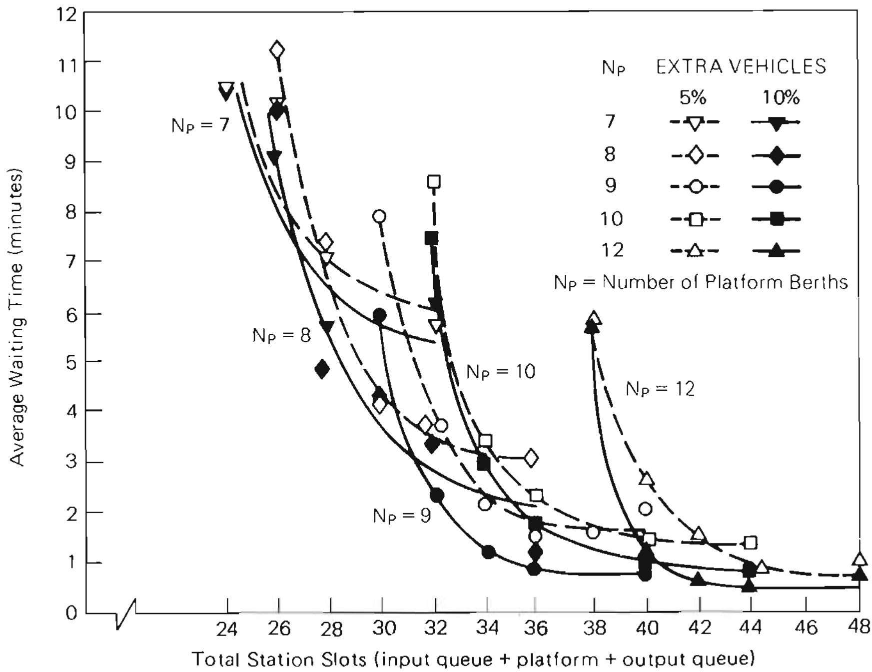


Fig. 3-13. Average Waiting Time During Evening Rush Hours for a Throughput of 1000 Parties to be Boarded/hr

As before, each data point represents the average of at least four 2-hr simulations. There is a very large spread in simulation results when the average waiting time is high. The spread is particularly significant for an 8-berth station, because, depending upon the random numbers, it could either develop a long queue or practically none. With a platform of 9 or more berths, and a reasonable input queue, the data become quite consistent.

Looking at the curves it would appear that 10% extra vehicles are required if waiting times are to be less than 1 minute. One again sees that a station of 9 berths is superior to one of 10 or 12 berths, both in requiring a shorter platform and fewer total slots. The station with 9 platform berths need have only 36 total slots, i.e., 18 slots in the input queue, although there is some improvement for longer queues. (It may be recalled that for the morning rush hours the recommended input queue was 20 slots.)

Figure 3-14 presents similar data for throughput requirements of 500, 1500, and 2000 boarding parties/hr. In all cases the reverse

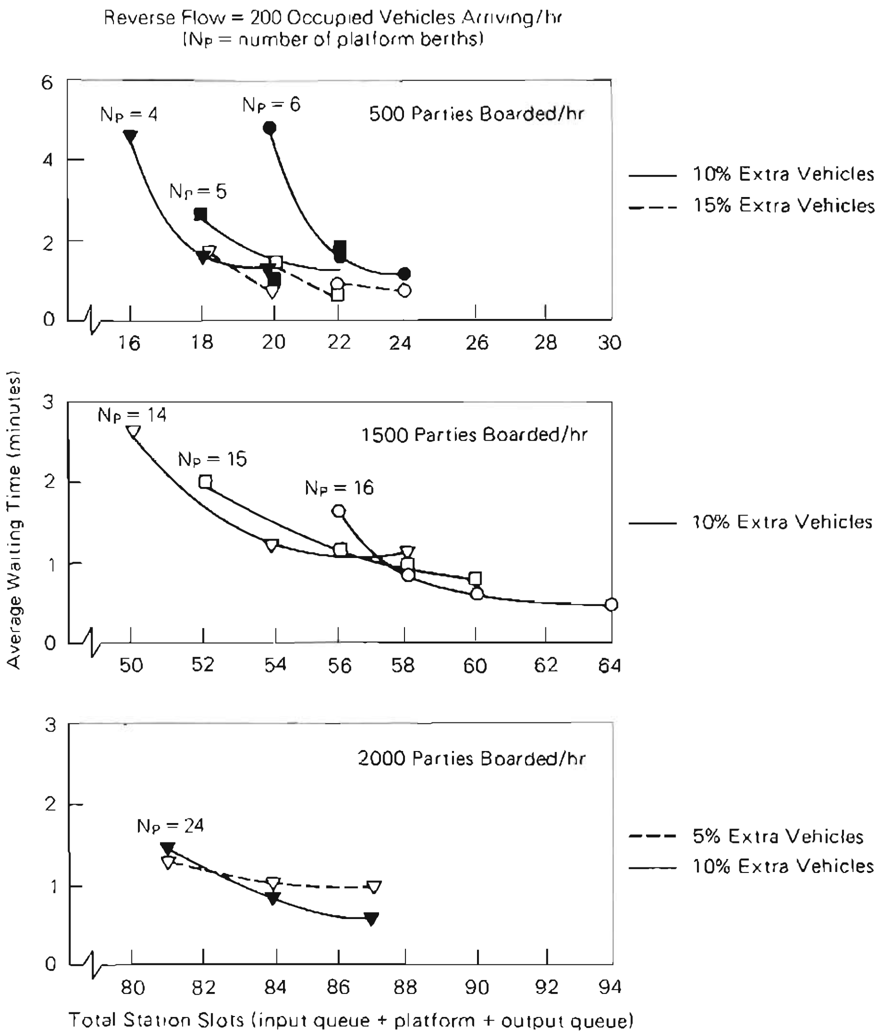


Fig. 3-14. Average Waiting Time During Evening Rush Hours for Various Throughputs

flow was taken as 200 arriving occupied vehicles/hr. It would appear that for a station requiring a throughput of 500 parties/hr, 15% extra vehicles should be brought to the station. A good choice would appear to be a platform having 5 berths (the same as the morning) and an input queue of 12 slots. (The morning required only 10.) At 1500 parties/hr, platforms with 14 or 15 berths appear to be somewhat underdesigned. A good design would appear to be one of 16 berths, with an input queue of 28 slots, which is identical with the morning requirement. For 2000 parties/hr, 24 berths and an input queue of 39 slots appears to be very satisfactory for both morning and evening operations.

3.2.5 Performance Summary

We have presented the results of our simulation of single platform stations in an activity center, both for the morning and evening rush hours. In each case we found the number of platform berths and input queue slots required to achieve specified throughputs without sacrificing the quality of service. For the morning operation we considered a miss rate of less than 1% to represent quality service, and for the evening operation we considered a waiting time of less than 1 minute to be quality service. Remarkably, in spite of the difference of these definitions, we found that the same number of platform slots were required and almost the same size input queues. Taking the larger input queue as being the dominant requirement, the results are summarized in Fig. 3-15.

As has been noted for both the morning and evening rush hours, an underdesigned station will not meet the requirements for quality service but an overdesigned station, though wasteful, will meet the requirements. Consequently, if there is some uncertainty in the throughput requirements at the time a station is being planned for an activity center, it may be wise to slightly overdesign it, to be on the safe side or to allow for growth in demand. If it turns out that the platform is longer than it needs to be, then several of the rear gates could be decommissioned or roped off, and the siding opposite them could be considered as a part of the input queue. Alternatively, since the station platform may be built in modules, with one module per berth, one might not overdesign, but instead add modules as required. This can be done most simply if the siding is sufficiently long from the outset.

3.3 STATION DESIGN CONSIDERATIONS

In Chapter 2 we pointed out that PRT guideways and stations can be underground, at ground level, or elevated. Certainly this will

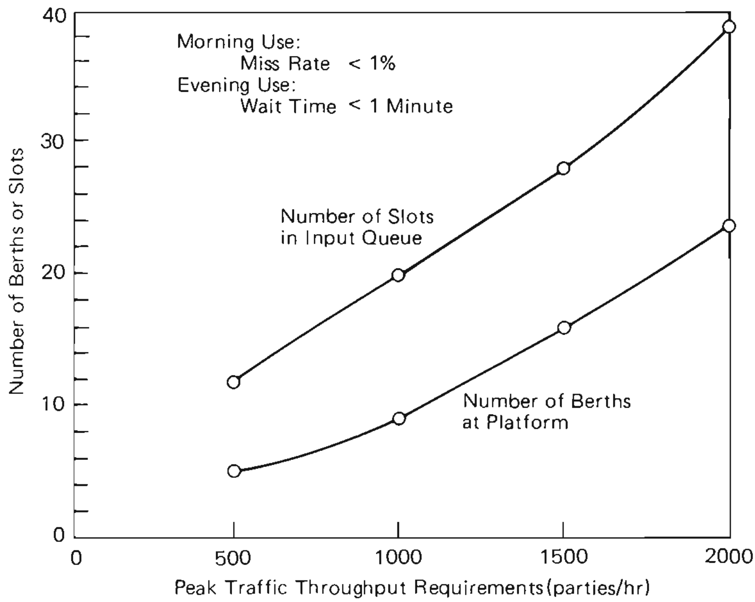


Fig. 3-15. Number of Platform Berths and Input Queue Slots to Achieve a Specified Throughput Without Sacrificing the Quality of Service

be one important factor affecting station design. For the reasons pointed out in that chapter, however, most of the stations will be elevated. Many will be integrated into or adjacent to major facilities such as office buildings, hotels, department stores, schools, sports arenas, or airline terminals. These will be of special design to blend architecturally into the overall facility, but with adequate attention paid to the functional features as well.

The remainder of the stations, and by far the greater number of them, will be elevated stations that stand free from existing structures. There would be significant economy if these were based on modular design and prefabricated parts. The majority of the station structural elements — floor and roof sections, walls, supporting beams, and columns — can be precast or prefabricated and delivered to the construction site in a variety of assembled arrangements to suit the station size and/or available working space conditions.

We have investigated both escalators and elevators as a means to get to or from the elevated structure. Not only are escalators more expensive, but the aged and incapacitated may find them difficult to manage. We much prefer elevators, which can also accommodate wheel chairs, baby carriages, and shopping carts. A single elevator will be adequate for the typical station in the residential areas and the smaller stations in activity centers, but the larger stations in activity centers may require two or more elevators. For the very largest stations an escalator might be used, but one elevator would still be required for those who cannot manage the escalator. In addition, there must be stairs for emergency use.

Figure 3-16 is a sketch of a moderately-sized (6 berths) activity-center station installed over the center of a street. A footbridge carries the passengers from the elevator or stairs to the loading platform. Trip-selection equipment is located on the footbridge. The station

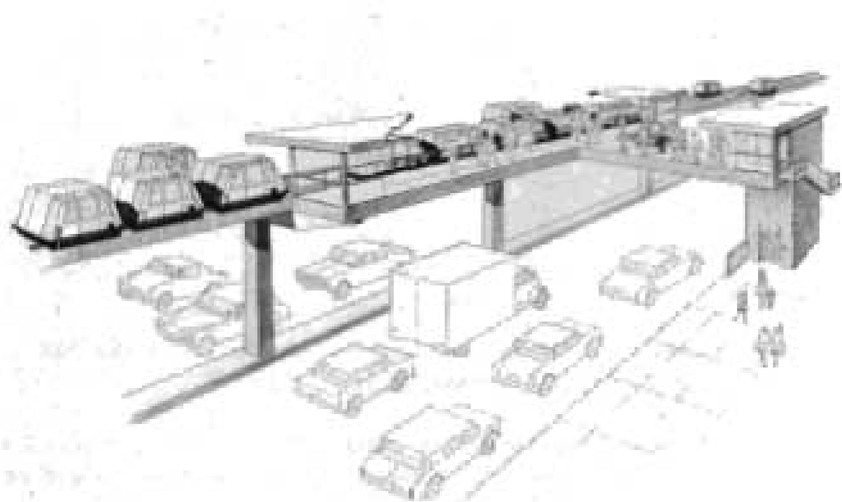


Fig. 3-16. Activity-Center Station Installed Over Center of Street

platform need only be about 8 ft in width. In this design the protective wall comes only to waist height and the sliding gates are indicated in the wall. Figure 3-17 is a design more appropriate to a curb-line installation.

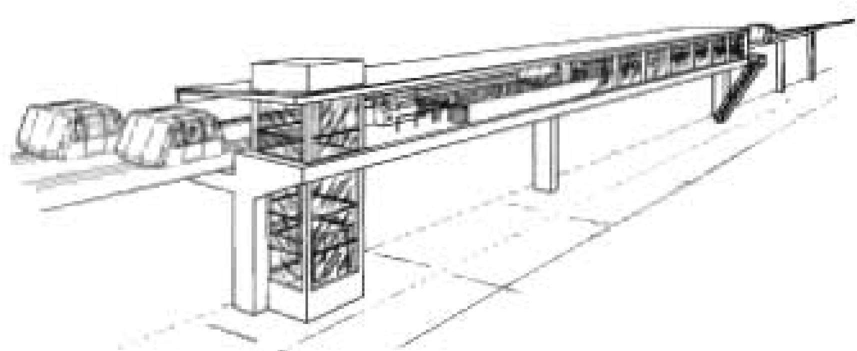


Fig. 3-17. Activity-Center Station Installed Over Curb Line

We have already discussed the functions of much of the electronics equipment to be located at each station, including the trip-selection equipment, the vending machine for "cash travel cards," the closed circuit television for platform surveillance, and the card readers located at each berth to control vehicle access and to inform the vehicle of its destination and other pertinent information. In addition, a station microcomputer would control the movement of vehicles and the assignment of berths to departing parties. The computer would be built with sufficient redundancy to ensure a very high probability of dependable service. For certain types of control systems, one of the functions of the microcomputer would be to dispatch empty vehicles, in the proper ratio, to preassigned destinations. For example, during the morning rush hours a CBD station will send its empty vehicles to a selected set of suburban stations, but later in the morning the empty vehicles will be sent to "car barns" for storage. In addition to diurnal variation of the empty vehicle dispatching instructions, they may be altered from time to time by a central computer.

Chapter 4
CONTROL ALTERNATIVES

Jack H. Irving

4.1 OVERVIEW OF PRT OPERATIONS AND CONTROL

There are many aspects of PRT operations and control.

In this chapter and in Chapter 5 we discuss PRT normal operations and control while in Chapter 6 we consider safety and the emergency operations employed in response to an operational failure or other hazardous condition. But, in comparing the various control options for normal operations, we will need to glimpse ahead to consider their safety, their vulnerability to failure, and the ease with which system operation may continue following a failure, although degraded in quality.

Chapter 5 treats the subjects of vehicle routing and empty vehicle management. To a large extent, those subjects can be treated quite independently of the type of control system used. The routing problem is one of assigning routes for all trips to minimize trip times or trip costs without leading to capacity overloads. It is true that the capacities achievable with different types of control systems may vary, but the routing problem can be solved by treating allowed capacity as an assigned parameter; the methodology does not change. There is further elaboration on this point in Sec. 5.1 after we have defined the control options in Chapter 4.

One aspect of the overall control problem is that of "lateral control." In some AGT designs there are four-wheeled vehicles which are steered much as a street-driven vehicle is steered, but automatically; the control system for the steering is a part of lateral control. We shall not treat steering control further because another approach has been broadly adopted in PRT designs, one which requires no active control system. In those designs the vehicles are physically constrained in their lateral motion by being continuously in contact with the sides of the guideway, although the vehicle usually is shock-mounted through appropriate springs and dampers to partially isolate passengers from being laterally buffeted by irregularities in the guideway wall. In Chapter 7 we will discuss vehicle suspension, including the lateral constraints.

Another facet of lateral control is vehicle switching. In Sec. 4.6.6 we will briefly discuss how switching can be controlled and in Chapter 7 we will discuss the design of switching mechanisms.

Any control concept must include the longitudinal control of vehicles along stretches of guideway, vehicle control at intersections and merges, and station operations and control. We already have covered the subject of station operations and control, so stations will be touched on only lightly in this chapter.

If a PRT system is to achieve the high capacities required in accordance with the arguments of Sec. 1.4, it must operate at very short headways; i.e., with small separations between vehicles. Section 4.2 treats the choice of minimum headway. This choice is not so much dependent on normal operations of the PRT system as it is on the question of passenger safety at the time of a vehicle failure. Thus the minimum allowable headways will depend on the safety policy adopted, the type and frequency of failures that can occur, the response times, the levels of emergency braking available, and on the use made of compressible bumpers and passenger constraints. Questions related to safety are discussed more fully in Chapter 6, but they will be touched on in Sec. 4.2 to show how they affect the choice of minimum headway and its possible dependence on line speed.

Then, in Sec. 4.3 through 4.5 we describe three of the more prevalent control concepts — synchronous, quasi-synchronous, and asynchronous control. Each of these concepts embodies a large number of characteristics. It is often assumed, incorrectly, that the characteristics of each must be grouped together and that a PRT (or GRT) must operate throughout with the same characteristics. Because of this impression, these three concepts have become stereotypes. Now we have come to understand that the characteristics can be admixed to give a very broad spectrum of control systems, and we also understand that the control characteristics can vary from one network element to another. In Sec. 4.6 we will discuss the spectrum of choices available. Nevertheless, it is still useful to first describe the three stereotypes, as a point of departure for the variations and hybrids.

4.2 THE CHOICE OF MINIMUM HEADWAY

It is important that PRT operations be very safe. In this section we shall examine some of the well-known safety criteria affecting the choice of minimum headway. If the adoption of criteria is capricious or based on an unreasoned standing tradition, it may rule out the possibility of short headways and therefore make PRT infeasible for certain applications. Criteria should be based on a realistic analysis of failure modes and other hazardous conditions and of their consequences on passenger safety. Chapter 6 presents such an analysis for some of the more important facets of safety.

One traditional criterion is the so-called "brick wall" approach which assumes that a failing vehicle stops instantaneously and that vehicles must be separated by a distance sufficient to allow the following vehicle to apply brakes and come to a stop before colliding with the disabled vehicle. The ratio of the separation to the stopping distance is known as k. The brick wall criterion corresponds to having $k > 1$. If the following vehicle, after a delay of 0.2 sec, were to decelerate at 0.7 g (the maximum attainable in standard automobiles), the initial separation would have to be 89 ft to avoid collision at an initial speed of 40 mi/hr.

Fortunately, a vehicle does not stop instantaneously when it malfunctions. Even in the extreme case where all wheels lock, a vehicle will traverse quite a distance while sliding to a stop. For example, at 40 mi/hr and with a 0.7 g deceleration rate, it will slide 77 ft. Even though vehicles do not stop instantaneously, the brick wall criterion has been adopted into regulations for conventional rail in many nations, and is now interpreted in many places as applying to all AGT systems. As a better understanding develops of the real safety issues and as systems are proven out on experimental test tracks, the old regulations will give way to more realistic ones.

Later we will return to a safety criterion which is closely related to the "brick wall" stop. This is where the sudden stop is not caused by a vehicle malfunction but by the striking of a massive object on the guideway. For the moment, however, let us continue our discussion of vehicle failures leading to inadvertent decelerations.

The approach at Aerospace (discussed in Chapter 6) is to have the failing vehicle measure its own inadvertent deceleration and report the measurement, together with other diagnostics, to a local computer which has control jurisdiction in the segment of the network where the failure occurs. (Other normal operational functions of the local computer will be discussed in later sections of this chapter.) If the local computer decides that the failing vehicle can be pushed, then the following vehicle is instructed to make a soft engagement with the failing vehicle, reaccelerate to line speed, and push the disabled vehicle to an emergency siding where a spare vehicle will be available. If the local computer decides that the failing vehicle cannot be pushed, then the following vehicle(s) are brought to a stop. At Aerospace we have chosen the headway to avoid impact during this emergency stop.

In Chapter 6 we shall demonstrate that if the deceleration of the following vehicle is about 15% greater than that of the failing vehicle, and if the onset of its braking is not delayed more than 0.2 sec after the onset of failure in the leading vehicle, then a 5-ft separation is more than adequate to ensure that vehicles do not collide. At first the vehicle separation will decrease, but as the velocity of the following vehicle drops below that of the failing vehicle, the separation reaches a minimum and starts to increase again. The total encroachment (maximum decrease in separation) is less than 4 ft. Had the delay been only 0.1 sec, the encroachment would be less than 1 ft. Only if there were a multiple failure, such as the failure of the second vehicle's brakes simultaneously with the locking of the first vehicle's wheels, would there be a collision. Thus, we have set the headway criterion in the Aerospace design so that no collisions will occur with "single-point" failures.

If the failing vehicle has locked its wheels, its rate of deceleration in g's will be equal to the coefficient of sliding friction between the wheels and guideway. Obviously, if the second vehicle's braking deceleration is to be 15% greater than the failing vehicle's rate of deceleration, then the second vehicle cannot rely on traction brakes. The primary mode for braking in the Aerospace design does not depend on traction; it is a linear motor used both for propulsion and braking. In the Federal Republic of Germany the Cabintaxi design is also independent of traction; it uses a different kind of linear motor for propulsion and uses eddy current braking. The Japanese CVS design uses traction brakes for normal braking and clamps the guideway for high-level 2 g emergency braking.

When braking is not dependent on traction, the guideway and wheels should be designed to minimize the coefficient of sliding friction. This has the effect of lowering the locked-wheel deceleration rate and thereby lowering the braking deceleration rate required in the following vehicle.

For a system that does use traction braking, let us first assume that braking deceleration on the following vehicle is 0.7 g (22.5 ft/sec^2), and that this exactly matches the deceleration of the failing vehicle with locked wheels. After a delay of 0.2 sec, there is a closing speed of 4.5 ft/sec, and subsequently this remains constant until the vehicles collide or the failing vehicle comes to a stop. Thus, an alternate policy to the one Aerospace adopted is to permit a collision velocity of about 4.5 ft/sec (3 mi/hr) for a single-point failure, rather than requiring that there be no collision.[1]

The problem with traction braking occurs when the following

[1] Even in the Aerospace design, if the reason for the inadvertent deceleration is that the leading vehicle has accidentally applied its brakes at their maximum rate of deceleration, then when the following vehicle matches this rate after a 0.2 sec delay, there will be a collision at somewhere around 4 to 5 ft/sec, depending on the maximum braking rate used.

vehicle has smooth tires and cannot develop as large a deceleration as the failing vehicle. If, for example, the following vehicle can only develop a 0.6 g deceleration rate (in contrast to 0.7 g in the failing vehicle), then the closing velocity will increase by 3.2 ft/sec for each additional second before impact. If the vehicles were initially only 5 ft apart, they will impact at a closing speed of 7.0 ft/sec (0.79 sec after the following vehicle applies its brakes). But if the vehicles are initially 30 ft apart and the line speed is at least 75 ft/sec (51.1 mi/hr), they will impact at 14.5 ft/sec (3.11 sec after the following vehicle applies brakes).

Thus far we have pointed out that with 0.2 sec for brake application, with 5 ft separation, and with nontraction brakes, no collision need occur when a single-point failure leads to inadvertent deceleration of a vehicle. (If braking response times can be brought down to around 0.1 sec, still shorter separations could be used.) Alternatively, if traction brakes are used, the impact velocity would normally be only about 4.5 ft/sec (but could be three to four times higher if the following vehicle has smooth tires and the vehicles are further separated). A separation of around 5 ft is more than adequate to manage the merging of vehicles.

Now let us return to the situation where a "brick wall" stop can occur, and that is the rare occasion where a massive object, such as a tree, has fallen across the guideway. Then, if no warning has occurred, the first vehicle that strikes the massive object will strike it at line speed, regardless of the headway. To protect the passengers in that vehicle, there must be such protective devices as compressible bumpers and passenger restraints (e.g., air bags). These are discussed in Chapter 6. As the striking vehicle rapidly decelerates, the following vehicle is warned and starts to brake.

Here is where the safety policy is involved. If the policy is that the second vehicle should avoid hitting the first, then the system must operate with $k > 1$ (i.e., according to the "brick wall" criterion). If the second vehicle is allowed to hit the first, then at what collision velocity may it strike the first? For the Cabintaxi system, operating at a line speed of 10 m/sec (32.8 ft/sec or about 22 mi/hr), the second vehicle was initially allowed to strike the first at 4 m/sec (13.1 ft/sec). The designers have considered increasing the allowed impact velocity up to 8 m/sec when shorter headways are required. At Aerospace our studies, reported in Chapter 6, have shown that with the proper design of the vehicle body structure, bumpers, and passenger constraints, the passengers can be well protected with "brick wall" collisions up to at least 75 ft/sec (about 50 mi/hr).

The minimum separation between vehicles which can be used, corresponding to any allowed impact velocity, is given by

$$S = V\tau + (V^2 - V_c{}^2)/2a_B, \qquad (4.1)$$

where

S = minimum allowed separation distance,

V = line speed,

V_c = allowed impact velocity between second vehicle and first after first vehicle has been stopped by brick-wall collision with massive object,

a_B = braking acceleration of second vehicle,

τ = effective delay time between collision of first vehicle and the effective[2] onset of braking of second vehicle.

Equation (4.1) is plotted in Fig. 4-1. The solid curves are for an effective delay of 0.2 sec and a braking deceleration of 0.8 g. They are given for values of allowed impact velocity ranging from 0 to 90 ft/sec. The dashed curves are based on limiting braking deceleration to 0.5 g.

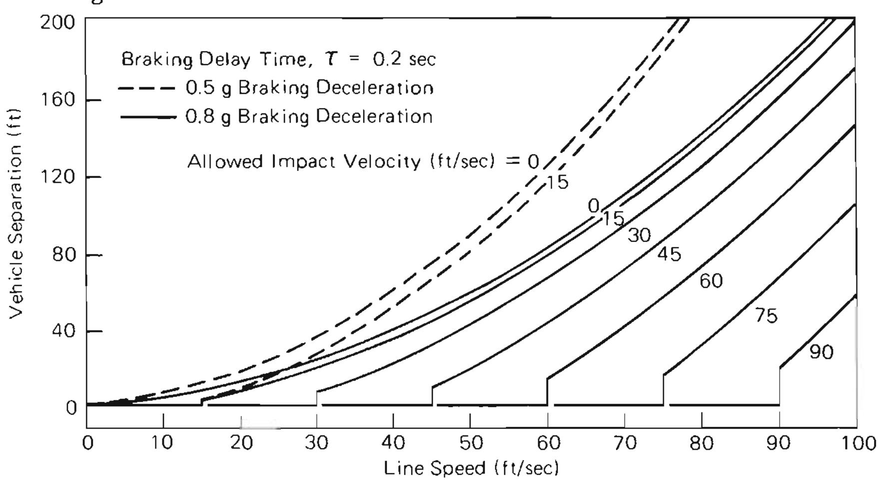


Fig. 4-1. Separation Required Between Two Vehicles if the First is Stopped Instantaneously by Hitting a Massive Object and the Second Brakes to Reduce its Impact Velocity to a Specified Value

To understand the sudden jump that appears in each curve, consider the case of limiting impact velocity to 60 ft/sec. If the line speed is 60.1 ft/sec, and there is a delay of 0.2 sec in braking, then the

[2] If one assumes a delay t_0 before brakes are applied, followed by a jerk duration t_J while the braking acceleration is being brought up to the value a_B, then "the effective onset of braking" is halfway through the jerk period; i.e., $\tau = t_0 + 0.5\, t_J$.

second vehicle will travel 12 ft before braking, and thus the separation must be at least 12 ft. But, if the line speed is only 59.9 ft/sec, no braking is required to keep impact velocity below 60 ft/sec, and the minimum headway could be zero if this were the only safety criterion.

It is not clear how seriously the separation criterion given in Fig. 4-1 should be taken. First, the scenario is predicated on a massive object that can instantaneously stop the first vehicle. With proper design the guideway would be protected from such objects, and even a heavy branch of a tree is not so massive that it would not be pushed some distance. Second, since passengers in the first vehicles have no warning of the foreign object, the danger to which they are exposed is not related in any way to the separation between vehicles. Third, if passengers in the first vehicle are to be adequately protected (by compressible bumpers and passenger constraints), then passengers in the following vehicle(s), will have at least the same protection. Fourth, the maximum exposure to this threat occurs only on the highest speed portions of the network and only when vehicles are following at minimum headway.

For all of these reasons we have not considered the separation criterion of Fig. 4-1 as being of primary significance in our work at The Aerospace Corporation. Rather, we have placed primary emphasis on inadvertent failure and on the ease of merging and therefore have planned on minimum separations of approximately 5 ft. As stated earlier, we believe that the data on passenger safety indicates that a brick wall collision of up to 50 mi/hr will cause no serious injury and consequently the 5-ft separation is quite adequate up to these speeds. To be conservative, one might lengthen the separation on lines with characteristic speeds above 50 mi/hr, but the advisability of doing this would depend on additional study. The determination will require more detailed design considerations, additional data relative to passenger injury at higher speeds, and an evaluation of the frequency of the rare occasions which might require additional separation between vehicles to further protect passengers in the second vehicle.

With the above caveats, Fig. 4-2 shows the minimum headway which would result from accepting the separations of Fig. 4-1, but limiting the minimum separation to be no shorter than 5 ft. The solid curve represents the headway in seconds which corresponds to using a 5-ft separation between vehicles. To illustrate how the figure is used, consider an allowed impact velocity of 60 ft/sec. For line speeds below 60 ft/sec, the minimal headway is that indicated by the solid curve; for line speeds above 60 ft/sec, the minimum headway is that given by the dash-dot curve labeled 60.

If passenger protection has been provided which allows some

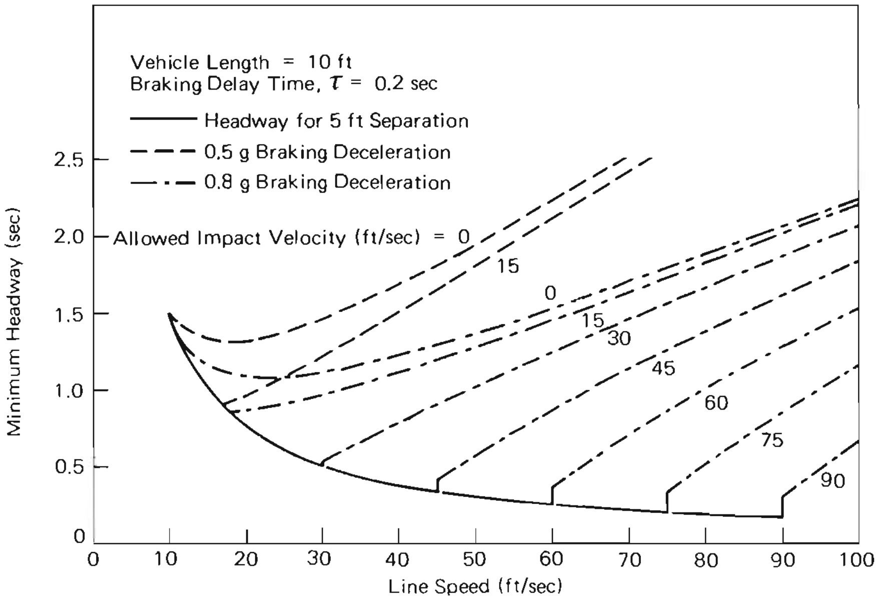


Fig. 4-2. Required Headway if Vehicle Separation is that Specified in Fig. 4-1 but Not Less Than 5 ft

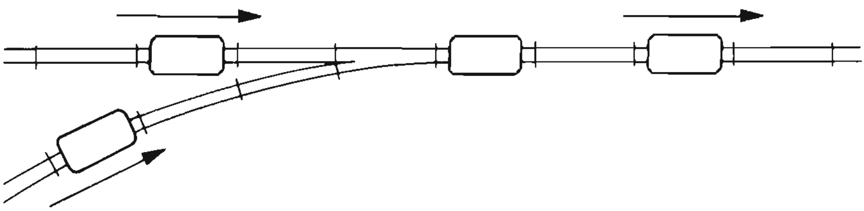
high impact velocity, like 75 or 90 ft/sec, then the system, in general, will operate on the solid curve; i.e., with 5-ft separation. This is the case with the Aerospace Corporation design. In that event, the higher the speed the shorter the headway. But, if the vehicle's protective devices and the adopted safety policy limit the impact speed between second and first vehicles to some low value, like 15 ft/sec, then the system will operate on the appropriate dashed or dash-dot curve and there is a critical trade-off that needs to be made between headway and line speed. It is because of this type of trade-off that the Cabintaxi line speed has been limited to 10 m/sec (about 22 mi/hr).

In summary, we have seen how sensitive the choice of a minimum safe headway can be to the adoption of a suitable safety policy and the response times, the level of emergency braking available, and the impact velocity that can be absorbed without injury to passengers. Because different control systems will be characterized by different response times to inadvertent deceleration, they may vary somewhat in the minimum headways achievable. In addition, capacity is dependent not only on minimum headway but also on the amount of space that must be left vacant on a line to permit the entry of vehicles coming from other lines or station sidings. Since different control systems may have different effectivity in using the available space for the merging, this too will influence the practical capacities attainable. These questions will be treated later in this chapter as a part of our comparison of different control alternatives.

4.3 SYNCHRONOUS CONTROL

Because of its complexity, and a number of other shortcomings to be discussed below, strict synchronous control is not taken very seriously today by most investigators. Yet, as did others, Aerospace started its investigation of PRT control by at first focusing on synchronous control. By discussing it first, we introduce some concepts which carry over to quasi-synchronous control.

Synchronous control is based on the concept of a moving "slot" which is a space of specified length moving along a guideway. Sometimes the "slot" is referred to as a "moving block." Either the slot is vacant or it is occupied by a vehicle centered in it. At a point of merging, the slots on the two merging lines are so synchronized that they exactly coincide on the merged line.



Slots may accelerate, but in doing so they stretch. (Likewise, during deceleration, slots shrink.) To understand this, consider a string of vehicles centered in adjacent slots 15 ft long and traveling at 30 ft/sec. The vehicle headway is 0.5 sec. When the vehicles pass a given point they start to accelerate up to a speed of 60 ft/sec. After reaching this speed, they still have a headway of 0.5 sec, but now the slot surrounding each one is 30 ft long.

It should be made clear that the slot is imaginary, not something physical; it is a useful concept to explain the allowed locations of moving vehicles. An equivalent concept is that of equally spaced points moving along a guideway, with each vehicle with its nose at one of the points, although not all points will have vehicles at them. The longitudinal control problem is to keep each vehicle centered in its slot, or, what is equivalent, following its point. For this reason such longitudinal control systems sometimes are called "point followers," although the term "point followers" would also include following points not equally spaced.

The longitudinal control is accomplished by observing the vehicle's position as a function of time, comparing that position with where it should be, and introducing speed adjustments to correct the position. The measurements and the determination of the correction needed can be made either from the vehicle itself or from the wayside; i.e., by instrumentation mounted on the guideway. These alternatives will be discussed further in Sec. 4.6.7.

The principal challenge for any control system is to avoid conflicts at merges and at stations. A conflict at a station occurs if a vehicle arrives at a station but finds it cannot enter the siding because there is no room for it. A conflict at a merge occurs if two merging vehicles are trying to occupy the same space (i.e., the same slot) on the merged line.

The essential idea for "synchronous control" is to set up a reservation system under the control of a large central computer, and not to allow a passenger to depart from his origin station until reservations for his whole trip are confirmed in advance. Here is how it works in its simplest form. When the passenger requests his trip, the request is transmitted to the central computer. There, the route to the destination station is looked up, and the exact time, measured from the instant of departure, past every merge point en route and to the destination siding is also looked up or computed. These times are very precise because of the synchronous slot motion.

A departure time is postulated, well enough in advance to ensure that the passenger(s) will have completed boarding at that time. Based on the postulated departure time, the time of arrival at the destination station is determined. If, as a result of previously confirmed reservations, the destination station is "booked to capacity," the process will be repeated either with a different route or with a new (later) postulated departure time. When the destination station is found to have available capacity at the calculated time of arrival, the next step is to check the availability of slots on each link of the route.

A "link" is here defined to mean the section of guideway from one merge point to the next. Slot availability is confirmed by checking a table of slot reservations. It is not enough to confirm that a slot is available where the vehicle turns onto a specified line, because that same slot could be reserved for another vehicle which will be merging into the slot as it passes a downstream intersection or as it passes a merge point with a siding from a station. That is why it is necessary to reserve the slot for every link along the way. If slots are not available, a new (and still later) departure time is postulated and the entire process is repeated, including checking both destination station and slot availability en route.

On a busy network it is extremely difficult to find available slots for the entire trip. For this reason, all of those who have worked with synchronous control have introduced a degree of flexibility by allowing the vehicle to move to neighboring slots on the main line and/or by allowing it to maneuver at an intersection to gain access to one of several slots after completion of the turn. One variation which uses slot changing at intersections and also allows flexibility in routing is referred to as "Trans-Synchronous."[3]

In some approaches the slots are thought of as being grouped into larger moving blocks. If the time of passage of a block were equal to the average interval at which the destination station can safely accept vehicles, then one (and only one) vehicle going to that destination station can be assigned to a block, but it could be in any slot of the block. The reservation of slots en route is facilitated by the freedom to move vehicles within the block, even though the order of the vehicles cannot be changed.

When, for some postulated time of departure, both destination station and slots are available, the new reservations are recorded and the ticket might be magnetically encoded with the planned departure time. If that time is some minutes away, the patron is informed that he must wait and he is not allowed to board until shortly before his scheduled departure. Alternatively, he can be allowed to board at once and the vehicle held in a holding area. In either case the station must be so designed as to allow a vehicle to depart precisely on schedule without being held up by others. This might be accomplished by the moving belt station described in Sec. 3.1.4, providing the departing party gets into the right vehicle and providing the belt doesn't need to be stopped for slow boarders. The docking station is another possibility.

The initial appeal of synchronous control is the general principle that the more information that exists on the state of the system and the totality of trips to be processed, the closer the control system can come to achieving some theoretical optimum operation. But in practice, synchronous control has a number of serious shortcomings:

a. The system requires a large computer to process and store reservations. Because failure of the computer would be catastrophic, two or more may be needed for redundancy.

b. The system is dependent on relatively long communication distances which makes communication vulnerable.

c. Destination stations would have to operate well below their capacity to assure that reserved time would be available. Departure areas would have to be designed to assure that departing vehicles could leave on schedule without interference from others. The station must provide a holding space for vehicles and/or an area for passengers waiting to board. Altogether, the station will have grown in size, cost, and complexity.

[3] "The Manhattan Project — A Cost Oriented Control System for a Large Personal Rapid Transit Network," R. Morse Wade, IBM Corporation, published in *Personal Rapid Transit-II*, University of Minnesota, Dec. 1973.

d. Should a vehicle fail, decelerating to a stop, it will cause all of the vehicles behind it to lose synchronization. Then other vehicles scheduled to turn onto that line will not be able to do so and must continue going straight. But the slot in which such a vehicle continues might be reserved after the next crossing, and so a conflict could be created. At the very least, a large number of vehicles would have to be reprogrammed en route with a new route and a new set of reservations, and possibly the desynchronization would propagate throughout the network.

To accommodate such failures more gracefully, it has been suggested that a certain fraction of all slots be left vacant for emergency use only. Then the vehicle forced to move straight ahead because it could not make its turn would adjust its position into one of the emergency slots and thus avoid conflict (except, perhaps, with another which had taken an emergency slot). Although this probably can be made to work, the effect under normal operations of not using emergency slots is to degrade the normal line capacity.

In summary, we do not favor synchronous control.

4.4 QUASI-SYNCHRONOUS CONTROL

Most of the work at The Aerospace Corporation has been devoted to quasi-synchronous control, including some of its variations which are discussed in Sec. 4.6. In Sec. 4.4.1 we describe the general concept of quasi-synchronous control and in 4.4.2 we consider in more detail the design and operation of intersections.

4.4.1 General Description of Quasi-Synchronous Control

As in synchronous control, quasi-synchronous control uses the concept of imaginary slots moving in a synchronous manner along the guideway. Again, on most of the guideway between intersections, either a slot is empty or there is a vehicle centered in it. But, in the vicinity of an intersection, vehicles may be instructed to advance slots or to slip slots to resolve conflicts on merging.

The principal difference between this and synchronous control is that there is no reservation system. When a vehicle is boarded, it moves into an output queue on the siding, as described in Chapter 3. Then vehicles in this queue are merged into slots on the main line as soon as possible.

Conflict resolution at an intersection is under the control of a local microcomputer which, assuming a one-way network, has a jurisdiction extending back along both incoming lines to the first upstream merge points. At the entrance to its jurisdiction area (or even before), there are wayside sensors to determine which slots are empty and which have vehicles in them. A vehicle passing the sensor reports the number of its destination station and whether it is empty or occupied by passengers. Then the local computer refers to a routing table to see whether the nominal route to the destination is one requiring the vehicle to turn or to go straight ahead.

The "nominal route" will usually be the fastest route, although it could be the shortest or the one consuming the least energy, or some "least-cost" combination of these. More important, if all vehicles took the fastest (or least-cost) route, certain parts of the network might become overloaded; i.e., the assigned traffic could exceed the physical capacity. To avoid this situation, not all trips will be assigned fastest (or least-cost) routes, but some will be assigned slightly slower (or more costly) routes to "balance the traffic." In particular, empty vehicles may be sent on slower routes to allow occupied vehicles to be routed the fastest way. Thus, each local computer may have two routing tables, one for occupied vehicles and one for empty. It is also obvious that different routing tables should be used for different times of the day. During the nonpeak traffic, for example, fastest or "least-cost" routes could be used for all trips. How to set up routing tables to minimize trip times or "costs" consistent with avoiding overloads is presented in Chapter 5.

Once the local computer knows for both incoming lines which slots have vehicles and which of these vehicles should turn, it goes through a set of computations (algorithms) to determine which vehicles should maneuver (advance or slip slots).

The location of the maneuvering will depend on the geometry of the intersection. If the maneuvering takes place before the switch point, it is a "single-stream" intersection; if after, it is a "split-stream" intersection. The performance of these two types of intersections is discussed in Sec. 4.4.2 where it is found that for reasonable traffic densities, especially for the split stream, almost all conflicts can be resolved.

Now, occasionally it will be impossible to accomplish all planned turns without slowing down traffic on one of the lines upstream of the computer's jurisdictional region. (This could occur, for example, if all slots within the jurisdiction on one line were occupied, no vehicles were turning off of that line, but some wanted to turn onto it.) To avoid such an occurrence, the local computer has the authority to deny a turn and require the would-be turner to go straight. In giving the local computer this authority, maneuvers can be restricted to a stipulated region entirely within the computer's jurisdictional area, and each computer can act autonomously without interfering with the actions of its neighbors.

The vehicle which is denied its turn will move straight ahead and will be routed to its destination station by the local computers at downstream intersections. This situation is illustrated in Fig. 4-3 where a vehicle leaving Station S_1 is destined for Station S_2. The shortest path requires turns at A, B, and F. But, if there is heavy traffic coming from the north at A and there are many vehicles coming from the west and trying to turn south at A, then the vehicle destined for S_2 may be denied its turn. In that event it would proceed straight. As it approached intersection C, the local computer there would continue it straight ahead. The computer at D would cause it to turn, after which it will proceed south to Station S_2. The distance and time penalty for the "detour" is quite insignificant.

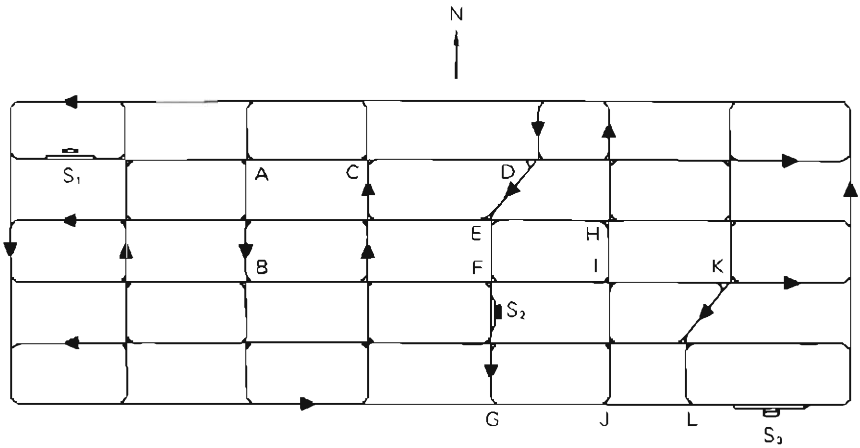


Fig. 4-3. One-Way Network Illustrating Alternate Paths

On the other hand, if the vehicle proceeds along its nominal path ABF and then is denied the turn at F, it will have to circle the block and reach S_2 along the path FIHEFS_2. This would add several minutes to its trip. To avoid this rather severe time penalty, the computer at F will choose instead to deny the turn to another vehicle going to S_3, since the path FIKLS_3 is only slightly longer than the nominal path FGJLS_3. Thus, a priority system is used in denying turns; the vehicles which would be most delayed by the denial will be the last to be denied.

When using quasi-synchronous control, the options for keeping vehicles centered within their assigned slots are the same as those for synchronous control. Measurements of time and position, and hence position error, can be done from the vehicle or the wayside. Moreover, if the vehicle is equipped to make the measurements, then, when the intersection computer requires the vehicle to carry out a maneuver to resolve a conflict, the computer need only command the vehicle to advance or slip a prescribed number of slots, beginning at a specified time (or position); the vehicle can program details of the maneuver. This is the approach which we used on our scale model test track. If measurements are made from the wayside, then the wayside computer must control the maneuver. This is the approach used in Japan's Computer-Controlled Vehicle System. Both the Aerospace and CVS measurement techniques are discussed in Sec. 4.6.7.

Switch actuation at a branch point can either be under the control of a local wayside computer or of the vehicle. In the Aerospace design, described in Chapter 7, we utilize electromagnetic switches on the guideway under the control of the local computer. In other designs the switch is on board the vehicle.

In addition to the many local computers, a quasi-synchronous system also employs a large central computer which is used for strategic and administrative functions, but not for the tactical control of individual vehicles. One of its strategic functions is the balancing of network traffic under exceptional circumstances. It accomplishes this by sending to the various intersection microcomputers appropriate tables for traffic routing. If, for example, a section of guideway were blocked, it would send out an emergency set of routing tables which would cause the intersection computers to route traffic around the blocked area. When a large sporting event was about to let out, it would send to the nearby intersection computers routing tables which would cause them to route through-traffic around the stadium area to minimize congestion in that area. It could also send to station computers instructions to dispatch their surplus empty vehicles to the stadium to meet the extraordinary demand.

Among its administrative functions would be validating travel cards when a trip was being ordered to make sure that the card had not expired and had not been reported as lost or stolen. Another would be customer billing. Still another would be sending each vehicle, perhaps once a day, to a facility where it would be automatically cleaned and checked out for incipient malfunctions.

One of the virtues of the quasi-synchronous approach is that it is relatively invulnerable to failure, and when it does fail, it fails gracefully. This subject is discussed at length in Chapter 6, but briefly here are the reasons:

a. If a vehicle or other object blocks the guideway, the central computer will be notified and new routing tables will be sent out so that little additional traffic will enter the affected area. The local computers will then clear the area, except for the blocking vehicle or other obstruction which must be manually removed.
b. If a local computer fails, or rather a redundant set of such

computers fails, then all intersection switches are set to the "straight ahead" position and there is no danger of collision, but routes will be somewhat longer.

c. Since the central computer is not involved in direct control of traffic, its failure will at most cause a degradation of service because of its unavailability for rebalancing the traffic for special situations. The unbalanced traffic would merely mean that a larger number of vehicles would be detoured by intersection computers. During the outage, all travel cards will be accepted as valid.

4.4.2 Quasi-Synchronous Intersection Control

The first intersection geometry investigated at Aerospace was the "single-stream" intersection.[4] An example is shown in Fig. 4-4 for a line speed of 30 ft/sec. The figure is based on the use of climbing and diving turn ramps, which, of course, must have double curvature. If double curvature is not used, then the divergence section, the climb (or dive), the turn, and the convergence section must all be distinct. This would move each divergence point about 200 ft farther away from the point of guideway crossing. Using a single-stream intersection, maneuvering (i.e., slot changing[5]) is accomplished upstream of the points of divergence to the turn ramps.

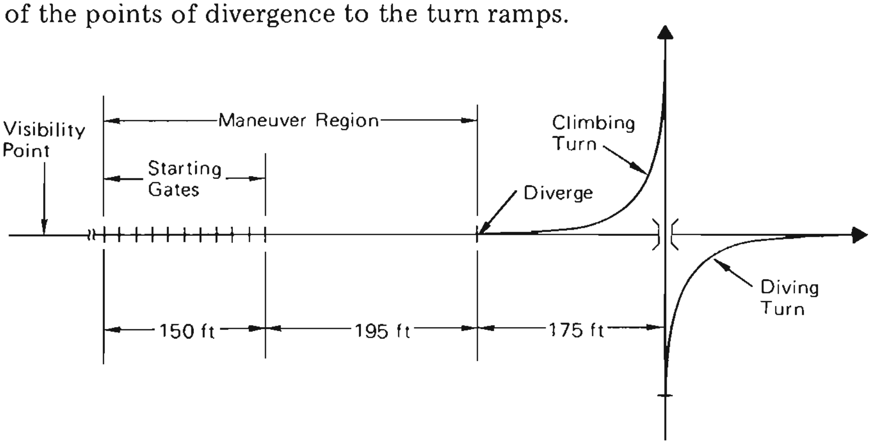


Fig. 4-4. Single-Stream Intersection for a Line Speed of 30 ft/sec

4 "Quasi-Synchronous Control of High-Capacity PRT Networks," A.V. Munson, Jr., et al., The Aerospace Corporation, published in *Personal Rapid Transit*, University of Minnesota, 1972.

5 Slot changing can mean either slot slipping or slot advancing. During slot advancing the vehicle temporarily accelerates to a higher speed and then returns to line speed, and during slot slipping it temporarily reduces its speed. The maneuvers assumed are limited both in acceleration (or deceleration) and jerk (rate of change of acceleration), and are discussed in Appendix A, Sec. A.2.

Within the broad framework of completing the maneuvers upstream of the divergence point, and not allowing traffic to "back up" beyond the jurisdictional area of the local computer, there are still many strategies which could be adopted. No attempt was made to find optimal performance strategies but rather we sought heuristic approaches which would be easy to implement and which would be able to handle the large majority of the tractable cases. The rules that were finally adopted for our simulation studies are the following:

a. At a line speed of 30 ft/sec, all maneuvers are carried out over a distance of 195 ft (thirteen 15-ft slots). This distance is adequate for the vehicle to come to a comfortable stop halfway, wait if necessary, and then accelerate up to line speed over the second half. This permits anywhere from one to an infinite number of slots to be slipped. Moreover, 195 ft is also an adequate distance to comfortably advance one or two slots. Had the distance been less than 180 ft, two-slot advances would not be possible.

   If maneuvers were all based on using the same acceleration or deceleration and the same jerk, they would take different guideway lengths, depending on the number of slots to be gained or slipped. We reasoned, however, that if the space had to be there anyhow for the more severe maneuvers, one might as well use all of it to make the less extreme maneuvers more comfortable; i.e., to use lower accelerations and jerks for them. Figure 4-5 is a plot of the maximum acceleration and jerks encountered as a function of the number of slots to be advanced or slipped.

b. To resolve an intersection conflict two types of maneuvers are allowed. Either the turning vehicle can advance or slip slots to a point where it can merge into a vacant slot on the other line (forcing others on its line to move if necessary), or a vehicle or string of vehicles on the other line can be advanced or retarded to make a slot available for the turner. Slot advances are preferred over slot slipping and, for each of these, moving the turning vehicle is given preference over moving the vehicle in conflict with it. In no case, however, is a turning vehicle already aligned for merging forced to move out of alignment to accommodate a would-be turner not yet aligned. The region upstream of the visibility point is assumed not to have any gaps or turners.

c. When a group of adjacent vehicles must all slip a slot, or several, then they must all start their decelerations simultaneously if they are not to encroach upon one another. Assuming 15-ft slot length, this means that the vehicles will be 15 ft apart when they start their maneuvers. These starting positions are referred to as "gates" and, as Fig. 4-4 illustrates, there might typically be 10 gates, although the number of gates was taken as a variable. If there are insufficient gates to provide starting positions for slipping all vehicles in a string, then the maneuver must not be allowed, and the would-be turner must be denied his turn.

   More generally, the starting gate to be used by any particular vehicle depends upon the number of slots that it is going to slip, the number of slots to be slipped by the vehicle ahead of it, the gate used by the vehicle ahead of it, and the number of vacant slots between the two vehicles. Each vehicle moves as far forward as it safely can before starting its maneuver; this provides more room for the vehicles behind it. Under some circumstances, such as when the vehicle ahead is advancing, the next vehicle may always move forward to the front gate before starting its maneuver.

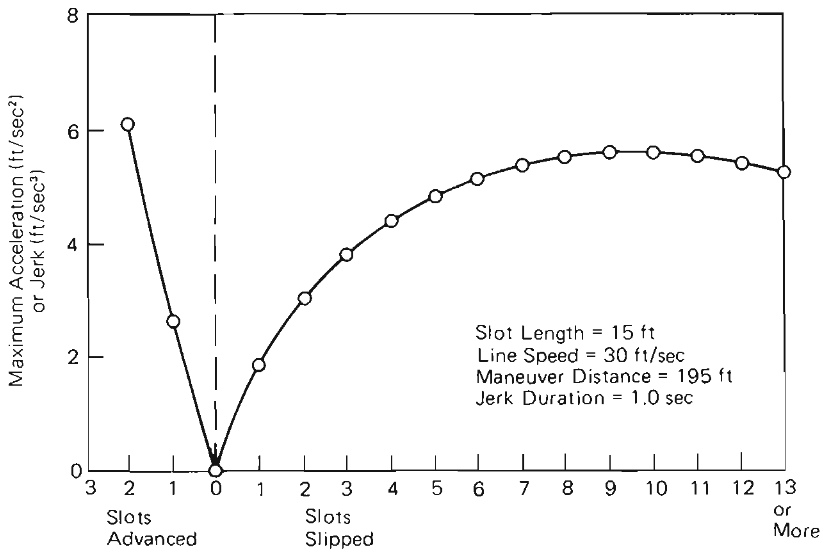


Fig. 4-5. Maximum Acceleration and Jerk for Various Slot Changes

The measure of performance is, of course, to keep the percentage of turns denied as small as possible. The results of the simulation are shown in Fig. 4-6 for 20% and 40% of the vehicles trying to turn. We found that for 60% line density, i.e., with each incoming slot having a 60% chance of being occupied, less than 1% of the turns were denied. However, at line densities over 70 or 75%, the percentage of turns denied increases rapidly with increases in line density. Thus, although the single-stream intersection gives very satisfactory performance at the lower line densities, its performance at the higher densities would tend to limit practical line capacities to less than 3/4 of their theoretical limit.

Minor improvements might be effected by increasing the number of gates or using a more sophisticated strategy for resolving conflicts, but there is a far more basic difficulty. It stems from the mutual interference in the maneuvering region between vehicles that should turn and vehicles that should go straight. A vehicle scheduled to turn may not be able to maneuver without forcing other vehicles to maneuver also. If any of these is scheduled to turn and is already aligned with its target slot, its alignment would be disturbed. A similar situation exists if the vehicle blocking the turn cannot vacate its slot without causing a vehicle turning off of its line to lose alignment.

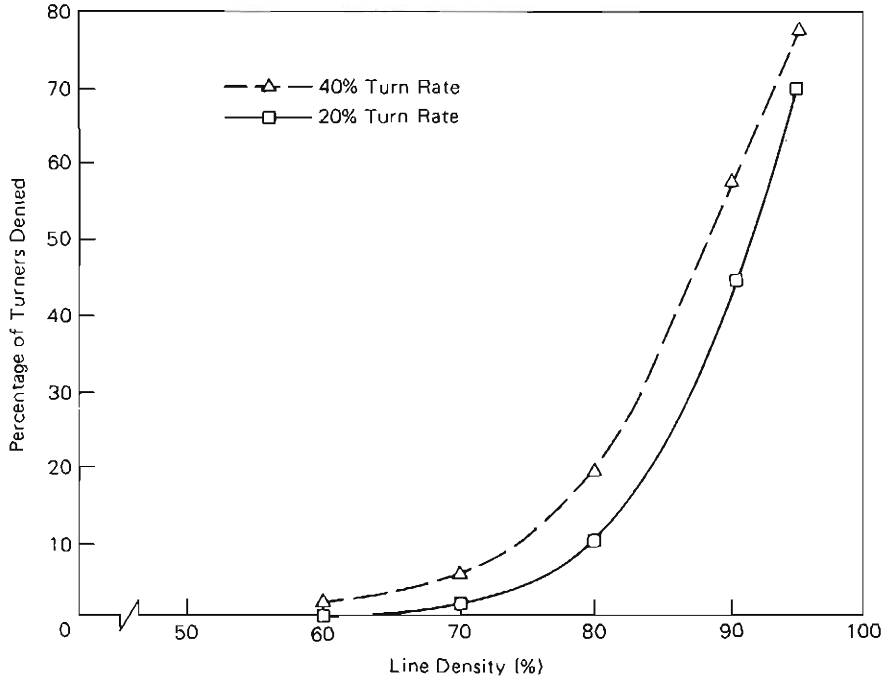


Fig. 4-6. Performance of a Single-Stream Intersection

The solution is to separate the vehicles intending to turn from those intending to go straight before they reach the maneuver zone. This has the effect of creating additional slot vacancies, and hence maneuvering flexibility, in the stream being maneuvered. Additionally, the turning and nonturning streams can now be maneuvered independently.

Therefore, a split-stream intersection geometry was defined (Fig. 4-7). This geometry necessitates additional guideway length on the turn ramp to accommodate the maneuver region between the divergence point and the intersection crossing.[6] For the split-stream intersection, the turning vehicles first go through an altitude change and then a moderately banked turn of small radius. Of course, on the turn, the lateral component of gravity will balance centrifugal force at only one speed; therefore, it was decided for purposes of initial simulation studies that slot changing maneuvers might best be performed only by nonturning vehicles. Again, both position advancing

[6] The reason for completing the maneuver before the crossing is related to the safety issue. If, because of malfunction, a vehicle moves into the wrong slot, there may be a conflict on merging. This situation will be detected by wayside sensors at the point of crossing, and there is still sufficient time to stop one or both of the conflicting vehicles before they reach the merge point.

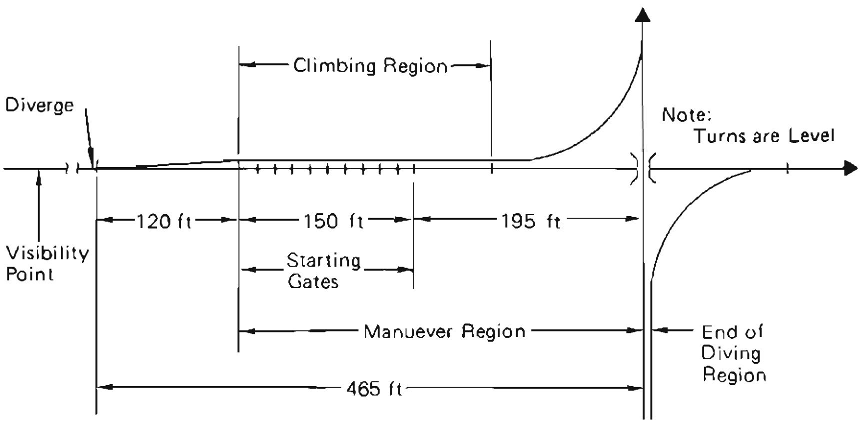


Fig. 4-7. Split-Stream Intersection for Line Speed of 30 ft/sec

and retarding maneuvers are permitted, with preference being given to advancing maneuvers for a maximum of two slots. The rules stated as (a) and (c) for the single-stream intersection still hold for the split-stream.

The results of the split-stream intersection simulation are shown in Fig. 4-8, where for ease of comparison the single-stream results are repeated. Split-stream clearly achieves significant improvement at the higher line densities.

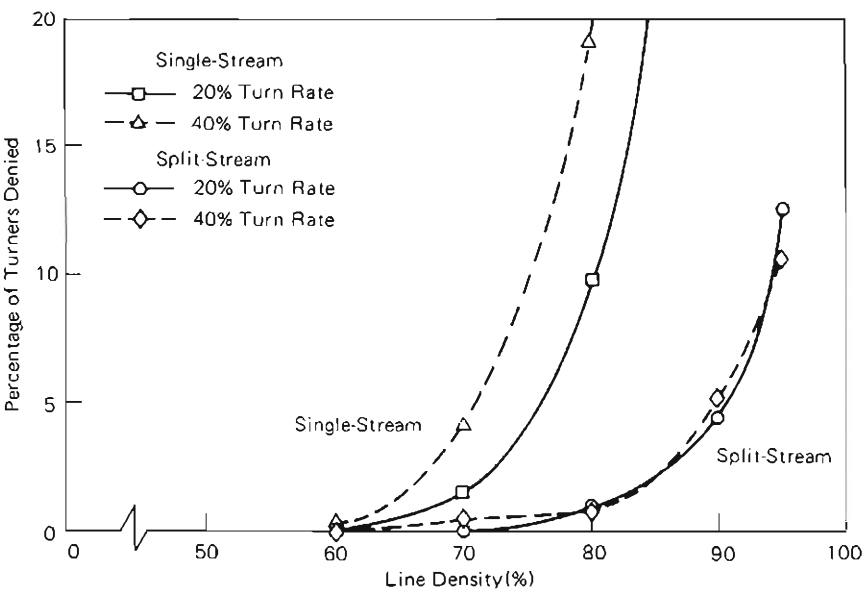


Fig. 4-8. Comparison of Single-Stream and Split-Stream Performance

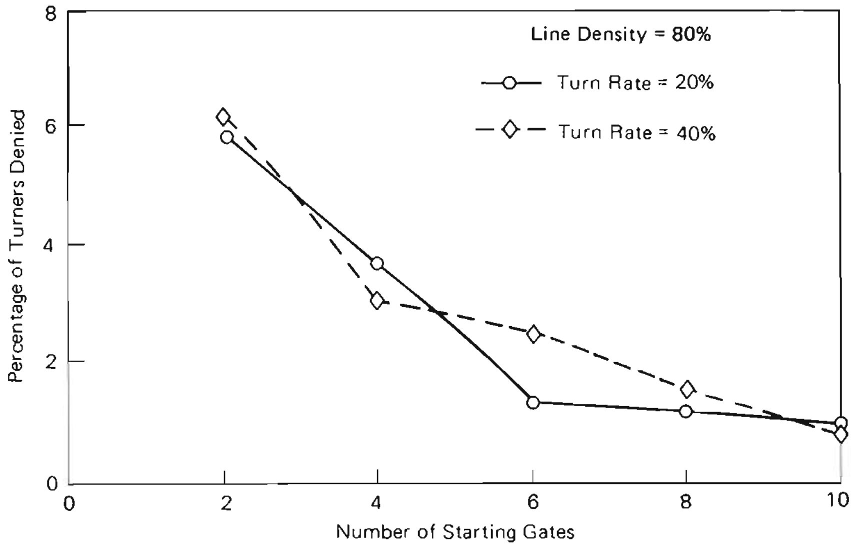


Fig. 4-9. Dependence of Turn Denial on Number of Starting Gates in a Split-Stream Intersection

Figure 4-9 shows how the percentage of turns denied depends on the number of starting gates in the split-stream intersection maneuver zone. It is a bit surprising how few starting gates can be used without serious degradation in performance, and the results are almost independent of turn rate.

When the space for starting gates is limited, further improvement in intersection performance may be achieved by

a. allowing the separation between vehicles in the maneuver region to temporarily fall below the nominal separation (of about 5 ft), especially when the speed of both vehicles is reduced below 30 ft/sec, and/or

b. to allow maneuvers to start at any arbitrary point (not fixed gates), as far forward as possible, consistent with satisfying the minimum separation criterion throughout the maneuver.

We recently wrote a computer program which computes these starting positions, but it has not yet been integrated into the intersection simulation program.

4.5 ASYNCHRONOUS CONTROL

Asynchronous control is not based on a principle of synchronous slot motion along the guideways, but rather on maintaining at least the minimum allowable headway between adjacent vehicles. Often the minimum separation between vehicles is considered a function of speed, with the separation between vehicles shortening at lower speeds as it does with automobiles on a highway. Whether the minimum separation is indeed a function of speed, and, if so, how it varies with speed is dependent on the headway policy adopted (Sec. 4.2). The traditional asynchronous system uses equipment on board the vehicle to measure the separation between it and the vehicle ahead. Thus, in contrast to the "point-follower" systems we have just been discussing, asynchronous systems are usually "car followers."

In the usual car-follower system, to eliminate the need for communication between vehicles, a vehicle knows only the location of the vehicle immediately ahead of it, and nothing about the locations of the vehicles ahead of that one. Thus, a vehicle has no knowledge of the sudden stopping of a downstream vehicle until the one immediately ahead of it has started to brake. As a result, the braking response propagates back along the guideway in a wavelike manner. Moreover, since the usual measurements are of separation and possibly relative velocity, detection of an inadvertent deceleration of the vehicle ahead may be delayed from when it would have been detected had the vehicle reported its own anomalous deceleration. As a result of these two types of delay, minimum headways need to be somewhat longer for the stereotypical asynchronous control than for quasi-synchronous control, although this has less influence on headway than does the safety policy discussed in Sec. 4.2.

Although a definitive comparison of headways for asynchronous and quasi-synchronous control can only be carried out once there is a detailed design of each system, it still will be instructive to illustrate by hypothetical although reasonable numerical examples. First let us compare the two types of system when the headway policy is one of avoiding a collision when the vehicle ahead has inadvertently locked its wheels and is decelerating at 0.7 g (22.5 ft/sec^2). Let us assume further that the vehicles are equipped with brakes capable of producing a deceleration of 0.8 g. For the quasi-synchronous control system, assume that it takes 0.1 sec for the accelerometer aboard the failed vehicle to detect the inadvertent deceleration, for a report to be made to the local computer, and for the local computer to order the succeeding vehicle to apply brakes. Assume further that the succeeding vehicle takes 0.2 sec to build up its braking deceleration to 0.8 g with a constant jerk rate during this build-up. The effective delay, τ, between the onset of inadvertent deceleration and the "effective time of braking" is then 0.2 sec (the delay of 0.1 sec plus one-half the jerk period). The maximum encroachment of the succeeding vehicle on the failed vehicle is 3.6 ft, which implies that an initial vehicle separation of 5 ft would be more than adequate to avoid collision.

For an asynchronous control system in which there is no communication between vehicles, it is highly unlikely that the inadvertent deceleration would be detected in 0.1 sec, for in that time the failing vehicle would have been displaced only 0.1 ft from the position that it would have occupied had there been no failure. Let us assume that the inadvertent deceleration is only detected after 0.2 sec when the displacement is 0.4 ft. The total effective delay, τ, will now be 0.3 sec, including one-half the jerk period. This leads to an encroachment of 8.1 ft, requiring an initial separation of about 10 ft. Thus, with a 10-ft long vehicle the minimum space allocated to a vehicle must be 20 ft, in contrast to 15 ft for the quasi-synchronous control system. As a result, for any characteristic line speed, headways will have to be 33% longer.

The assumption of 0.2-sec jerk time to build braking deceleration is compatible with the use of fast-acting mechanical brakes. However, for the primary braking methodology described in Sec. 7.3 (reversing current in the pulsed dc linear motor used for propulsion), full braking force can be reached in less than 0.002 sec. An asynchronous system would then require only a 5-ft separation to avoid collision, compared with about 1 or 2 ft for quasi-synchronous. But, a separation of, say, 3 ft would be required in any event to manage merging. Thus, slot size for asynchronous would be 15 ft compared with 13 ft for quasi-synchronous, which corresponds to a 15% increase in headway.

Now we consider another numerical example for the case where passengers are not protected by air bags and the headway policy is that indicated in Fig. 4-2. Let us assume that the maximum braking deceleration available is 0.5 g and that the policy is that when one vehicle has struck a large immovable object (such as a fallen tree), the vehicle behind it will be allowed to impact the first vehicle at any speed up to 15 ft/sec. Let us assume a line speed of 60 ft/sec. If, as above, the quasi-synchronous control is characterized by a τ of 0.2 sec, Fig. 4-2 shows that the headway would be 2.1 sec. In considering an asynchronous system with a τ of 0.3 sec, the first term in Eq. (4.1) for vehicle separation would be increased by 6 ft and, as a result, the headway would be increased by 0.1 sec. This represents only a 5% increase in headway.

In summary, when relatively longer headways are being used, the additional delays of a car-follower system detecting an inadvertent deceleration are not significant, but if there is an attempt to maximize capacity with the use of very short headways, then the extra delays can be quite significant, depending on the jerk time for emergency braking. When the extra delay for asynchronous control is significant, it may be possible to avoid that extra delay by having the failed vehicle measure its own inadvertent deceleration and report it directly to the vehicle behind.

The Cabintaxi system under development in the Federal Republic of Germany is an example of a system which uses car-follower techniques. Each vehicle broadcasts a 100-kHz signal into a lossy line; the signal propagates backward along the guideway. The next vehicle back detects this signal and can determine the separation by the amplitude of the received signal. By using two separated transmitters on each vehicle it is possible to cancel out the forward-moving signal and reinforce the backward-moving signal so that the net signal propagates only backwards along the guideway. Also, each vehicle transmits backwards a signal which just cancels the backward-moving signal from the vehicle ahead; this keeps the signals from propagating back to more than one vehicle. Except in the vicinity of a merge, this whole car-follower system is redundant with two lossy lines, one on each side of the guideway.

The difficult problem in asynchronous control is that of merging, because there is no direct way for a vehicle to compare its distance from the merge point with that of a vehicle on the other guideway with which it may be in conflict. As a result, there needs to be some means for letting a vehicle know the location of the potentially conflicting vehicle. In the Cabintaxi system, on each guideway upstream of a merge, the inside lossy line (i.e., the one closer to the merging guideway) is broken into segments. The vehicles no longer transmit into the broken line, but each segment carries a signal brought to it by an electrical connection from the corresponding point on the outside lossy line of the other guideway. Thus, each vehicle measures the separation from the actual vehicle ahead of it on the outside lossy line, and on the inside line it measures the separation to a "ghost" vehicle which is the same distance from the merge point as the real conflicting vehicle on the other guideway. To avoid overreaction to the ghost (jamming on the brakes) when it first appears and there is still a long way to the merge, the signals coming into the first few segments of the broken line are attenuated to make the ghost appear farther away. As the merge point is approached, the amount of attenuation is gradually decreased to zero so that the true distance to the ghost can be measured.

One of the features which distinguishes a stereotype asynchronous system from the stereotype quasi-synchronous system is the response that takes place to conflicts at intersections. We noted above that in quasi-synchronous control a turn is denied rather than forcing traffic to slow down upstream of the jurisdictional area of an intersection computer. This was necessary to provide each local intersection computer with autonomy. In the stereotypical asynchronous system, turns are not denied. Therefore, if conflicts develop, incoming traffic is slowed down and this slowdown can propagate back to upstream intersections and merges, much as automobile traffic "backs up" on a busy highway. In Sec. 4.6.3 we shall show how this stereotypical approach might be improved upon.

As a result of the somewhat larger minimum headway and the less efficient merging which comes from not knowing the make-up of both merging streams, line capacities on an asynchronous system are somewhat lower than on a corresponding quasi-synchronous system. As pointed out, however, these differences are not as significant as those that might arise from differences in safety policy (Sec. 4.2).

Routing on an asynchronous system could be quite similar to that on a quasi-synchronous system where at each intersection there would be a local computer to look up whether the nominal route requires the vehicle to turn. Again, these routing tables could be varied from time to time as necessary to balance the traffic. This function, as before, would be carried out by a central computer.

Asynchronous control shares with quasi-synchronous control the virtue of failing gracefully. Failures of the central computer or the local routing computers have substantially the same impact as their failures on a quasi-synchronous system. If the guideway were blocked, it would be necessary for a local computer to supervise the line-clearing procedures, as indeed was the case with quasi-synchronous control.

Since a car-follower system has no need to depend on a wayside computer or a communications link to maintain separation, it might seem to be safer than a quasi-synchronous control system. However, one should be cautious with such arguments because the maintenance of separation still depends upon the proper functioning of certain equipments. Again taking the Cabintaxi system as an example, the avoidance of collision between two vehicles on the same guideway is dependent on the proper functioning of the transmitters of the vehicle ahead and of the receivers of the following vehicle because a loss of signal would be interpreted as an infinite separation. A loss of signal is especially critical when approaching a merge point because in those regions there is no longer redundancy. At a merge there is also dependence on the transmitters of the conflicting vehicle on the other guideway. Less serious are breaks in the continuity of the lossy line along the vehicle's own guideway or a break in the connection to one of the segments near a merge, because these cause only transient errors. Before one can reach any firm conclusions about relative safety, it is necessary to look very deeply into the design, the degree and kind of redundancy, and the consequences of the failure.

Before leaving the subject of asynchronous control we should briefly describe a novel PRT system, Aramis, under development by Engins Matra in France. In that system, optical ranging is used to keep vehicles traveling 1 ft apart in platoons or "trains," although the trains are separated from each other by headways of about 1 minute. If one of the vehicles in a platoon should decelerate suddenly, the vehicle behind it makes contact so soon that very little relative velocity will have developed.[7]

As a train passes a station siding, some of the vehicles will leave the train and enter the siding. The remaining vehicles will close ranks as soon as possible, again reducing separations to 1 ft. A vehicle leaving a station siding does not try to merge into a train, but rather waits until the train has gone by and merges into the very large space between trains. It then accelerates to catch up to the train ahead and becomes the last vehicle in that train.

The Aramis is very effective in a line-haul configuration but is not intended for use in a network with many closely spaced crossing lines. The problem is in turning from one line to another. Vehicles that need to turn might have to be queued for some time to wait for a train to go by. If there were many vehicles waiting to turn, there might not be adequate space for storing them without building an off-line storage area.

Vehicles arriving at the intersection when a train was not going by would be able to turn without delay, but then it might take them a long time to catch up with the last train to pass. While they were catching up, the train might have passed other intersections and stations. Thus the problem is introduced as to how to merge vehicles from these downstream intersections and stations into the stream of vehicles already trying to catch up with the train. To the best of our knowledge, Engins Matra, the developers of Aramis, have not tackled this problem, since they envision Aramis as a line-haul system.

4.6 THE SPECTRUM OF CONTROL OPTIONS

In Secs. 4.3 through 4.5 we have described synchronous, quasi-synchronous, and asynchronous control. Each had a number of characteristics with similarities in some areas and dissimilarities in others. Now we shall try to get to the root of these characteristics.

The principal characteristics represent the system designer's choice as he makes the critical decisions which will define the control concept for his system. (After the major decisions are made there still are many possible design implementations of any chosen control strategy.) Although there is no unique way to list the critical decisions,

[7] For example, if the failing vehicle decelerates at 0.7 g (22.5 ft/sec^2) and the following vehicle does not brake, the impact velocity will be 6.7 ft/sec.

the following may be regarded as a representative list of the questions that need to be addressed:

a. Which control functions should be centralized and which decentralized?
b. What kind of a reservation system should there be, if any?
c. What uses should be made of "wait-to-merge" and "wave-on" strategies for handling excessive traffic at merges and intersections? How does this affect network design?
d. Should sequencing of vehicles at a merge or intersection be under the control of a local computer?
e. Should a car-follower or point-follower system be employed?
f. How should switching be controlled?
g. For a point-follower system, should position and speed be measured by the vehicle or from the wayside? How should the position be controlled?
h. For a point-follower system, should discrete or continuous positions be used? Is systemwide synchronization desirable?

We shall now discuss these critical decision areas and some of the viable control options available.

4.6.1 Centralization versus Decentralization

Which functions should be centralized and which decentralized?

By this time the reader will understand that we believe that those functions which are vital to the continuing operation of the system should be decentralized as much as possible. In particular, the functions of headway maintenance, merging, switching, intersection control, and station control should be decentralized. They may be under the control of small "local computers," working perhaps in cooperation with small computers on board the vehicles. If these functions were centralized, then a failure of the central computer might paralyze the entire network. In contrast, the failure of a local computer might at most disable a single station or cause all turns to be denied at a single intersection.

There are two aspects of routing and empty-vehicle management — the tactical and the strategic aspects. The tactical aspect is how to control the routing of individual vehicles and when and where to dispatch individual empty vehicles. We envision these as decentralized functions. For example, routing may be accomplished by having a local computer at each intersection (or shared by a small group of intersections) interrogate the vehicle to determine its destination and then refer to a table of turn instructions to find out whether the vehicle should turn or not. Dispatching of empty vehicles from any station should be under the control of the station's local computer, which first determines which vehicles are surplus and then refers to a list of stations in need of empty vehicles to determine where the next one should be sent.

The strategic function, which must be carried out centrally to have any meaning, is to modify the intersection local computers' turn instruction tables or the station computers' dispatching lists to better balance the traffic or to serve special needs. A failure of the central computer will, at most, degrade the quality of service; it will not leave the vehicles bereft of turn instructions, and surplus empties will have somewhere to go. The strategic "override" by the central computer must always "pass through channels" and never go directly to the vehicle, for otherwise the vehicle may be receiving conflicting orders and/or the local computer might not know that its orders are being countermanded.

It is also valid to think of the central computer as carrying out certain administrative functions listed in Sec. 4.4.1, which it can carry out efficiently and which are not vital to safety or service dependability.

4.6.2 Reservations

What kind of a reservation system should there be, if any?

In discussing synchronous control we considered the reservation of both stations and slots. Our conclusions were that such a reservation system is unnecessarily complex and does not fail gracefully. Of course, it must be acknowledged that in principle a centrally controlled reservation system could use very sophisticated algorithms to optimize the vehicle flow; but, if a much simpler approach will work almost as well, then there is very little incentive to introduce the full-blown reservation system. Indeed, we have shown that very high line densities are feasible with quasi-synchronous control, and in Chapter 5 we shall show how nominal routes may be chosen to keep average line densities safely within practical limits. Thus slot reservation would certainly seem unnecessary.

There may, however, be some virtue in having a station reservation system or, as an alternative, a "station delay warning" system. Either system could be superimposed on quasi-synchronous control or asynchronous control.

Here is how a station reservation system might work. When a patron inserts his travel card into the trip selection equipment and enters the number of his destination station, the information will be transmitted to the central computer. The central computer predicts the time of arrival at the destination station, assuming that the patron and his party proceed at once to the boarding platform and that their vehicle is routed along its nominal route. The prediction is only within crude tolerances of about ± 1/2 minute at best. The computer then looks up previously confirmed reservations to find the average rate of arrival at the predicted arrival time. If the destination station is not saturated, the reservation is confirmed and the travel card magnetically encoded in the usual way. If the destination station is saturated at the predicted arrival time, the computer searches forward through the record of confirmed reservations until it finds a period when the average arrival rate is below the station's capacity. The patron is then informed of how long he must delay his boarding.

At the same time he is informed of the delay, the patron may be shown a map of his destination area which would display not only his requested station but the neighboring stations as well. Each of these can be marked with the delay, if any, associated with it. After examining the map, the patron either confirms his original selection or he may change his request to one of the neighboring stations. His travel card is appropriately encoded with the number of his requested station, but it also carries encoded information on when he may be allowed to board. He is also informed directly of the time he may board. Until that time his card will not open the boarding gates.

A "station delay warning" system operates in a quite similar manner except the passengers are not delayed in boarding. They are allowed to board at once and the delays refer to how long they will have to "circle the block" around the destination station. Another difference is that, in a reservation system, precedence is given to those who request their trips first; in a delay warning system, precedence is given to those who arrive at the destination station first. If a vehicle has circled the block, it will be given priority in entering the station siding over neighboring vehicles which have not yet circled; one that has circled the block twice will be given priority over one that has circled once, etc.

In a station delay warning system, the central computer estimates the arrival time and predicts the number of vehicles that will then be circling the block with precedence over the new patron's vehicle. It will thus be able to predict the number of circlings for the new patron and hence his delay. In making this prediction it must include all vehicles which will arrive ahead of the patron's vehicle, even vehicles for trips not yet requested from origin stations close to the destination station. The latter can be projected on the basis of normal demand, possibly discounted if very long delays are encountered.

The station delay warning system, as the reservation system, presents the patron a map showing neighboring stations and their delays (projected circling times). This gives him the opportunity to confirm his original request or to change it. Then his travel card is encoded with the number of his selected station, but no delay times are recorded. His party proceeds at once to the boarding platform and boards.

Are the benefits of a station reservation system or station delay warning system worth the cost and added complexity? If all stations were sized properly to meet their demand, such systems would be completely unnecessary. But, if the demand at a station is badly underestimated, or if the demand increases suddenly, then there is a problem. Of course, if the high demand persists, in many cases the station can be enlarged to satisfy and probably exceed the demand. But, until the enlargement is completed, station reservations or delay warning could improve the service. Moreover, there will be occasions when the station cannot be enlarged, either because funds are not available, or because there is no room for a larger station with its longer siding. The latter situation is most likely to occur in a CBD or other activity center where stations are close together. Under those conditions, the patron would find it particularly useful to know that there is a long delay to his requested station but that there is no delay to its nearest neighbor, a block or two away.

It may be argued that, even without a station reservation or delay warning system, the patrons will adjust their requests, through a learning process, to equalize the demand among neighboring stations. Certainly this will occur at the activity-center stations during the evening rush hours because, if patrons see long waiting lines at one station, but not at the next, many of them will walk to the station with the shorter lines. The learning process will be more difficult in the morning if no indication is given at the suburban origin station on delays to be encountered at the activity-center destination station. But, patrons will learn by experimentation and they will learn from friends. As they circle the block they may observe stations whose input queue is not full. Some may even push the "Next Station" button (Sec. 1.7.1) which would bring them into the first station approached with space available.

It is the author's belief that, if either can be justified, a station delay warning system is generally preferable to a station reservation system. As far as the patron is concerned, the two are about equivalent. Each warns him of delays and informs of the availability of neighboring stations. It is of little concern to him whether he is delayed at his origin station or by circling the destination station. The reservation system saves a little energy involved in circling, and a few vehicles, but if almost all stations have been properly sized to meet their demand, the savings would not be significant. The disadvantage of a reservation system, when compared to a delay warning system, is that it requires all stations to have a waiting area and means for keeping a party from boarding before the assigned time. This not only increases station cost but it may complicate the security problem at the station.

An exception may occur when a significant part of the PRT network is in a line-haul configuration, for then, if the station is "missed," there are no blocks to be circled. Under these circumstances, a station reservation system would be preferred. But, if there are only a few stations on the line-haul portion of the network, the best approach may be to overdesign these stations to virtually eliminate the possibility of a vehicle being forced to bypass one of them because the input queue was full; then the reservation system would be unnecessary.

4.6.3 Wait-to-Merge versus Wave-on

What uses should be made of "wait-to-merge" and "wave-on" strategies for handling excessive traffic at merges and intersections? How does this affect network design?

Both quasi-synchronous and asynchronous control systems can involve a certain amount of slowing down or waiting at intersections. In an asynchronous car-follower system, vehicles will slow down to allow merges from the other line, and in a quasi-synchronous system they may slip slots. Where the stereotype asynchronous system differs from the stereotype quasi-synchronous is in the means of handling excessively high traffic densities.

In the stereotypical quasi-synchronous system, maneuvering is confined to two regions within the jurisdictional area of a local computer, one on each of the two lines approaching the intersection. If a would-be turner cannot be accommodated by maneuvering vehicles within these maneuver regions, the would-be turner is "waved-on;" i.e., it is denied its turn. By this means there is no "backing up" of traffic congestion; congestion at one intersection will have no direct influence on upstream intersections.

In contrast, the stereotypical asynchronous system does not use "wave-on" tactics; rather, each vehicle will follow its predestined route no matter how long it must "wait to merge" or what impact this waiting may have on propagating congestion upstream.

Thus, it would appear that excessive traffic at intersections is managed by "wave-on" tactics when using stereotypical quasi-synchronous control and by "wait-to-merge" tactics for stereotypical asynchronous control. We shall now examine these stereotypical approaches to find their implications and we shall explore variations to improve overall system performance. We begin by examining quasi-synchronous control.

"Wave-on" is possible at an intersection because there are alternate routes to the destination. If the vehicle is denied its turn, it can go straight and still reach its destination. At a simple merge there are two incoming lines but only one outgoing line; wave-on has no meaning. Thus, there is an apparent implication that no simple merges can be used in a network under quasi-synchronous control because at a simple merge there is no way of avoiding the backing up of traffic when the traffic flow on the two incoming lines is greater than the capacity of the single outgoing line. One solution is indeed to design a network with no simple merges (except at station sidings), and we shall shortly illustrate how this can be done. But an alternate approach that does allow simple merges is to precede the simple merge by a branch point (point of divergence) under the control of the same local computer which controls the merge. In this way traffic may be diverted to keep from overloading the merge.

First let us illustrate how to design a network without simple merges. One natural location for simple merges is at the borders of a network. Referring to Fig. 4-3, the points G and L are such merge points. However, one way to avoid merges at the edge of a network is to use a "scalloped" network, as indicated in Fig. 4-10. The reader should at first ignore the dotted lines in this figure. The scalloped network consists of four loops indicated by the solid lines of the figure. Three of these loops are simple rectangles; two are predominantly north-south and the other is east-west. The fourth loop is around the perimeter and crosses itself in four places. It will be noted that this network has 24 intersections, but no merges.

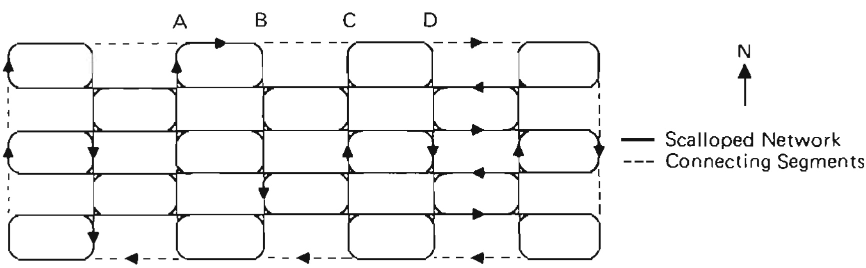


Fig. 4-10. Scalloped Network with Connecting Segments

Now imagine that we add the 10 connecting segments shown by the dotted lines. The network now has 10 merge points, one of which is marked C. Conflicts can be resolved at C by treating the segment BC much like a siding. If the traffic coming from the south at C plus that from the west does not exceed the capacity of the line segment CD, conflicts can be resolved by employing slot advancing and slipping maneuvers. But, if the densities are too high, then traffic coming from the south at C has precedence on the line CD over traffic coming from the dotted segment BC. In short, "wait-to-merge" is employed at merge point C with the waiting done by the vehicles on BC, just as though they were in the output queue on a station siding. Should the line segment BC become completely occupied, the traffic coming from A would be forced to turn south at point B. Thus, even though there might be a backing up of traffic from C, the backing up can go no further than the branch point B. This illustrates how one can manage traffic at a merge point (C) by diverting traffic at an upstream branch point (B).

We have thus shown an example of a network (the scalloped network) which has no merge points (except at station sidings), and we have shown at least one way of using wait-to-merge control at merge points without an uncontrolled backing up of traffic congestion. Another interesting case is provided by the Los Angeles network shown in Fig. 2-13. Although we envisioned the network operating under a control system that might generally be classified as quasi-synchronous, it had many simple merges. Therefore, there would be many line segments, predominantly on north-south lines, that would operate on the wait-to-merge principle.

Now let us consider a PRT system which uses stereotypical asynchronous car-follower control. The performance of the system might be improved by introducing a variation of wave-on at busy intersections. This is especially true if the system is operating with a safety policy which allows the vehicles to operate at minimum separations which cannot be significantly decreased when traffic slows down.[8] When, on the other hand, the separations can be significantly decreased, then slowing down may so *increase* the capacity of the slowed down line as to relieve traffic congestion at upstream merges. In such cases little can be gained by the wave-on variation.

Here is how the wave-on system works. When the average line densities on both outgoing lines at the intersection are within certain specified limits, then no turns are denied; all vehicles follow the turn instructions specified by the routing table for that time of day. (These turn instructions, it may be remembered, do not necessarily direct all vehicles along minimum time paths, but rather along paths as fast as possible, consistent with having the projected average traffic densities less than practical capacities throughout the network.) But,

[8] This is the case where passengers are well enough protected so that, in the rare event when the vehicle ahead has been stopped instantaneously by striking a massive object, the following vehicle can be permitted to strike the stopped vehicle at an impact speed higher or nearly as high as the line speed. Referring to Fig. 4-2, the line speed would be to the left of the minimum for the permitted impact speed (i.e., the system would be operating on the solid curve) or, at most, slightly to the right of the minimum.

when one of the outgoing lines would be too crowded for a short period of, say, 30 sec, as a result of an upward fluctuation of the number of vehicles coming in on that line and required to go straight and/or the number of vehicles coming in on the other line and required to turn, then the local computer has the authority to deny turns onto the crowded line. This authority avoids the backing up of traffic congestion which would result from the excessive slowdown of vehicles trying to merge. In determining which vehicles should be denied their turns, the computer must refer to a priority table and deny turns for those vehicles whose trip times will be the least penalized by the denial.

The reader will see that this wave-on strategy is identical in almost every way to that which we discussed under quasi-synchronous control (Sec. 4.4.1). The only difference is that for a car-follower control system the wave-on is called when projected densities averaged over some short period of time are too high, while for quasi-synchronous control the wave-on is invoked when the local computer's conflict resolution algorithms are unable to find a solution that limits maneuvers to prescribed maneuver regions.

In summary, we have seen that both wave-on and wait-to-merge strategies can be employed on a single network, regardless of whether the vehicles are otherwise controlled quasi-synchronously or asynchronously.

4.6.4 Sequencing of Vehicles at a Merge or Intersection

Should sequencing of vehicles at a merge or intersection be under the control of a local computer?

In quasi-synchronous control the sequencing of vehicles at a merge or intersection is a function under the control of a local computer. The local computer knows which vehicles should turn, if possible, and where gaps exist in the incoming traffic streams; it computes how the vehicles should maneuver to effectively use the available space. The maneuvers used can include a vehicle moving forward relative to the stream, or moving backward. Thus a vehicle on one of two merging lines might be closer initially to the merge point than several vehicles on the other line, but after the maneuvers it might be more distant from the merge point than the several. This could occur either by the vehicle dropping back a considerable distance and/or the several advancing. For simplicity we might call this "passing," even though the two merging lines might initially be perpendicular to each other.

In contrast, under typical car-follower control, like that of Cabintaxi (Sec. 4.5), there is no passing. It will be recalled that each vehicle measures the distance not only to the vehicle ahead of it on its own guideway, but to a "ghost" vehicle which is the same distance from the merge point as the conflicting real vehicle on the other line. A vehicle will slow down if it is too close to the ghost. There is never an attempt to accelerate and pass the ghost. Because no advancing maneuvers are used, and because there is no way for a vehicle to drop back to allow vehicles on the other line to pass it (they would only drop back further), there is a less efficient use of available space. To the best of our knowledge there have been no studies to date which have quantized this difference in efficiency.

The reader will note that the essential benefits in using a local computer are that advantage can be taken of a knowledge of the entire stream of vehicles and their longitudinal spacing, that vehicles can be instructed either to advance or fall back, and that vehicle "passing" is permitted. Whether or not the vehicles are nominally constrained to synchronized slots is immaterial to the argument. Nor does it matter how the measurements are carried out so long as they are made known to the local computer. For example, the measurements can be made by the vehicles or by wayside instrumentation. Finally, it does not matter how the maneuvers are controlled as long as there is a high degree of certainty that they will be carried out faithfully.

One reason often given for adopting a car-follower system is to keep the system implementation "simple." It is argued that with the car-follower approach no local computers are required, although some means are required to get information to each vehicle on the location of the conflicting vehicle on the merging line. But, isn't the local computer really needed for other functions, even when using a car-follower approach?

We have described asynchronous control in Sec. 4.5 as including a local computer to look up routing instructions, and in Sec. 4.6.3 we pointed out how with car-follower control, overall system performance might be improved by having the local computer monitor line densities and use the wave-on principle to avoid excessive backing up of traffic congestion. It might be argued that the latter function is not necessary and that the former could have been performed by routing tables at the departure station with turning instructions stored aboard the vehicle. However, there is another important function of the local computer — that of controlling emergency situations. In Chapter 6 we shall discuss car-pushing strategies where one vehicle makes a soft engagement with the vehicle ahead of it which is failing, and pushes that vehicle to an emergency siding. Making such a soft engagement would be contrary to the normal working of a car-follower control system and would have to involve an override from a local computer. If traffic has come to a stop because of guideway blockage, the local computer is required to "clear the lines." This line-clearing procedure may even involve moving some vehicles backward. All in all, the local computers carry out so many functions that it is difficult to see how a well-designed system could do without them. If, indeed, they are there and there are communication links to them, then, in the author's opinion, when line capacity is an important issue the local computers should be used to sequence vehicles at merges and intersections.

4.6.5 Car Follower versus Point Follower

Should a car-follower or point-follower system be employed?

We have already discussed a number of the possible disadvantages in car-follower systems. These will be reviewed briefly and then a few new points touched.

In Sec. 4.5 we pointed out that, because of delays encountered in detecting the inadvertent deceleration of downstream vehicles, car-follower systems require somewhat longer minimum headways, and therefore lower theoretical capacities, than systems under the supervision of a local computer. Under the local computer, the anomalous deceleration is sensed by the failing vehicle and reported to the local computer which warns the following vehicles to start braking at once. Without such a reporting system the sudden deceleration of a vehicle would not be detected by the vehicle immediately behind it until a measurable difference developed in the relative velocity or possibly even in the separation distance. A vehicle further back would not know of the failure until the chain of braking reactions propagated back to the vehicle immediately ahead of it.

This disadvantage of the typical car-follower system could be eliminated if each vehicle reported its anomalous deceleration to the vehicle behind it, and if this information were relayed back along the line. There would also need to be some way to report to vehicles on a merging line. All of this, of course, complicates the system mechanization.

In Sec. 4.6.4 we pointed out that when merges and intersections are under the control of a local computer, higher efficiencies can be achieved than when a car-follower control system is used. The local computer can take advantage of a knowledge of the two streams of incoming vehicles and the location of gaps; with a car-follower, each vehicle has knowledge only of the vehicle immediately ahead and of the conflicting vehicle on the other guideway. With a local computer, vehicles may be ordered to advance or slip back relative to the nominal traffic stream and "passing" vehicles on the other guideway is permitted; with a car-follower, advancing and "passing" are prohibited. When we say that higher efficiencies can be achieved at merges or intersections under the control of a local computer, we mean that the incoming lines can operate at a higher fraction of their theoretical capacity without causing serious overloads at the merges or intersections; i.e., without significant backing up of traffic congestion or, in the case of intersections using wave-on, without excessive detouring of vehicles from their nominal routes.

Thus, we have seen that the theoretical capacity of a car-follower system is somewhat lower than that of a system under the control of a local computer unless it is complicated by the addition of a reporting system which allows a vehicle to report its anomalous deceleration to other vehicles, and we have seen that a car-follower must operate at a lower fraction of its theoretical capacity to keep from overloading merges and intersections. Therefore, the car-follower approach is not indicated when the highest capacities must be achieved. (However, as pointed out earlier, safety policy is of far greater importance in achieving high capacity.)

In addition to the questions of capacity just discussed, there are considerations of emergency operations such as car pushing and line clearing. It would seem that, regardless of the type of normal operations, the supervision of such emergency operations would have to be under the control of a local computer.

Up to this time we have been discussing the term "car follower" in its usual context of meaning a system where vehicle-borne equipment measures the distance to the vehicle ahead and then the following vehicle's speed is adjusted to maintain safe separation. However, there might be a second meaning of the term "car-follower," relating to the motion of one vehicle being adjusted to maintain the separation from the vehicle ahead, regardless of how that separation is measured. For example, if there were continuous or very frequent wayside measurements of the positions of all vehicles, as in the Japanese CVS system (see Sec. 4.6.7), then it might be possible to use control algorithms which adjust a vehicle's speed, not to keep it in a prescribed slot or to follow a designated "point," but rather to adjust its distance from the vehicle ahead. If, indeed, a system operating under these principles were under the control of a local computer, then there is no reason why such a system could not achieve capacities as great as those achievable by quasi-synchronous control. In fact, we alluded to the use of such car-following techniques in Sec. 4.4.2 when we spoke about improving intersection performance by allowing maneuvers to start at arbitrary points, rather than fixed gates, consistent with satisfying a criterion for minimum separation from the vehicle ahead.

4.6.6 Control of Switching

How should switching be controlled?

There are two stages in the control of switching. First, it must be decided which of two branches the vehicle should take. If merges and intersections are under the control of a local computer, the local computer will make that decision. Second, the decision should be communicated either to wayside equipment or to the vehicle so that the switching mechanism may be activated.

The switching cannot involve moving any massive parts of the guideway because if it did, short headways could not be maintained. For example, at a line speed of 75 ft/sec (about 50 mi/hr), if vehicles are separated by 5 ft the time between the passage of the tail of one vehicle and the nose of the next is only 1/15 sec. There are generally two ways of accomplishing switching in such a short time. One way is to have the switching mechanism on board the vehicle. In that event it can be activated well in advance of the vehicle reaching the point where guideways begin to diverge. For example, it could be a set of rollers on the vehicle which "grab" one side of the guideway. Another means for accomplishing switching rapidly is to use electromagnets mounted on the guideway to pull the vehicle onto one branch or the other. This is the approach Aerospace followed in its scale model development (Appendix B).

When the latter method is used, there is no need to communicate any switching instructions to the vehicle and there is no need to rely on proper operation of on-board switching mechanisms. The electromagnetic switching should be designed, however, so that in the event of a power failure the vehicle will automatically lock into one of the diverging lines. At an intersection, any vehicle not yet into its turn at the time of power failure would be locked to whichever side of the guideway would carry it straight through the intersection with no turns. One of our reasons for choosing electromagnets was related to the specific approach we had to propulsion and braking. This relationship is developed in Chapter 7.

4.6.7 Measurement and Longitudinal Control

For a point-follower system, should position and speed be measured by the vehicle or from the wayside? How should the position be controlled?

One approach to position measurement is to have this function performed by wayside equipment. The Japanese CVS design is an example of this approach. In CVS the wayside computer takes a poll of vehicles by addressing each one by its own unique identification code. The vehicle replies by broadcasting a signal through an antenna just a few centimeters away from a number of wire pairs running along the length of the guideway. Some of the pairs are twisted in the vicinity of the antenna and they cannot pick up the signal, but other pairs are separated in that vicinity and they will pick up the signal. The pattern of which wire pairs are twisted and which are separated changes about every 20 cm on the station sidings and somewhat less frequently on the main lines. The vehicle can be located by which of the wire pairs have picked up the signal. By interpolation, positions can be determined quite accurately. To the best of our knowledge, there is no direct wayside speed measurement in CVS. Speed can, of course, be determined quite accurately from successive position measurements, provided that such measurements are frequent and not too "noisy."

If position and speed are determined by very frequent wayside measurements, at most every few feet of travel, there are two generic alternatives for controlling the vehicles:

a. One alternative is to transmit to each vehicle the amount of acceleration or deceleration required to correct its position and speed errors. This commanded acceleration might be corrected to include the acceleration necessary to compensate for gravity when the vehicle is climbing or diving. If the vehicle had an accelerometer on board, it could adjust its motor thrust until the measured acceleration equaled that commanded by the wayside controller. This control method has the advantage of having tight (fast acting) feedback around an "inner loop" which adjusts the motor current promptly in response to gusts which might accelerate or decelerate the vehicle. There is no need to wait for updated position measurements.

b. Alternatively, if the vehicle were not equipped to measure acceleration, the wayside controller would transmit commanded vehicle thrust (or motor current) which should then include not only the thrust required for acceleration and grade, but also an estimate of the thrust required to overcome friction and air drag (including the effects of wind). This approach would depend on a motor calibration so the commanded thrust could be used to adjust motor current. If the estimated thrust (or current) was wrong, the vehicle would temporarily depart from the moving point it was to follow. This would be detected by the wayside position-measurement equipment and the commanded thrust would be altered. This alternative depends solely on feedback from the position-measurement "outer loop."

If vehicles are traveling at very small separations, then, from the safety standpoint, any sudden inadvertent deceleration must be detected on board and reported to the local computer as soon as possible. As discussed in Sec. 4.2, a 5-ft separation requires the following vehicle to effectively apply its braking force within approximately 0.2 sec after an extreme (locked wheels) inadvertent deceleration of the leading vehicle if collision is to be avoided. To be compatible with this figure, the measurement and reporting of the extreme anomalous deceleration might take about 0.1 sec. During that time the speed will have changed by only about 2 ft/sec and the position error by only about 0.1 ft. Thus, the wayside position measurement is of little use in the early detection and measurement of an extreme inadvertent deceleration (unless the position were measured at intervals substantially less than 0.1 sec and with a measurement error substantially less than 0.1 ft). If, indeed, an accelerometer is to be included for safety purposes, then there appears to be no reason for not using control alternative *a*, which, as pointed out earlier, should be more responsive to gusts than alternative *b*.

Since it seems worthwhile to measure acceleration on board, the question might arise as to whether speed too should be measured on board. If speed is measured on board, then wayside position measurements may be made far less frequently, as we shall soon demonstrate. But first we note that if wayside position measurements are infrequent, it becomes difficult, if not impossible, to deduce speed from these measurements because they give average speed between measurement sites, and not instantaneous speed. Therefore, if wayside position measurements are infrequent, then measuring speed on board is highly desirable.

Once the vehicle is at the "moving point" that it is tracking, then, in principle, it would never depart from this point if it could maintain the correct speed at all times. When speed is measured on board, the wayside controller (local computer) will specify desired speed, and the on-board longitudinal controls will attempt to keep the measured speed matching the commanded speed. Of course, there may be wind gusts or other disturbances that prevent the vehicle from maintaining an absolutely correct speed; but if the vehicle could measure precisely the deviation of its speed from that required, it could integrate this deviation to find its position error and subsequently could adjust its speed to eliminate the position error.

No analog measurement of speed is accurate enough to prevent the vehicle from "drifting off" from the moving point it should be following. As a result it is necessary to make a periodic position measurement to eliminate cumulative drift. If, for example, there were a 1% error in measuring speed, and if the vehicle is required to drift no further than 1.0 ft from its "moving point," then there must be an independent position measurement every 100 ft. Such position measurements could be made by wayside sensors, or, as we shall see later, they could be made from the vehicle.

In contrast to the relatively inaccurate analog measurement of speed, certain digital means for measuring speed could, in principle, maintain position indefinitely. This would require that all clocks throughout the system be perfectly synchronized (but not necessarily accurate). This is the approach that was used in The Aerospace Corporation's 1/10-scale model test track; timing pulses were sent to each vehicle from a master clock. Because of the danger that some of these pulses could be lost in transmission, it might be better to have clocks on each vehicle which would be periodically synchronized to a master clock. But even with perfect synchronization there is still a necessity to take some position fixes. For example, there is the problem (discussed below) of initialization after system shutdown. Also, it will be recalled that in our discussion of quasi-synchronous control we suggested a guideway-mounted vehicle sensor at the entrance to the jurisdictional area of each local computer. In Sec. 6.5.1 we will describe how such guideway-mounted sensors might also be used just upstream of a merge point to make certain that no two vehicles approaching the merge are in conflict. If guideway-mounted vehicle sensors are needed for other purposes, it would seem advisable to use them to obtain position fixes, and if this is done, then there is no necessity to synchronize the vehicle clocks with a master clock. For purposes of illustration, assume that a position fix is taken every 0.5 mi and that the vehicle clocks are only accurate to 1 part in 10,000 (about 9 sec per day). The drift would then be only 0.26 ft between position fixes.

Figure 4-11 illustrates how a digital system, depending upon such position fixes, might work. As described above, each vehicle would generate a continuous stream of clock pulses by having an oscillator on board whose frequency was controlled to about 1 part in 10,000. For purposes of illustration let us assume that the pulse rate is 10,800 pulses/sec. Along the guideway there are evenly spaced fiducial marks which, for purposes of illustration, we shall assume to be separated by 1.0 ft. These marks are detected by a fiducial mark sensor aboard the vehicle.

In the Aerospace design the propulsion and primary braking systems (Secs. 7.2 and 7.3) employ evenly spaced guideway-mounted ceramic magnets. Every second magnet has its north pole facing inward toward the vehicle-mounted motor primary, and the alternate magnets have their south poles facing inward. These magnets are 6 inches long and are spaced 6 inches apart full scale. Hall-effect

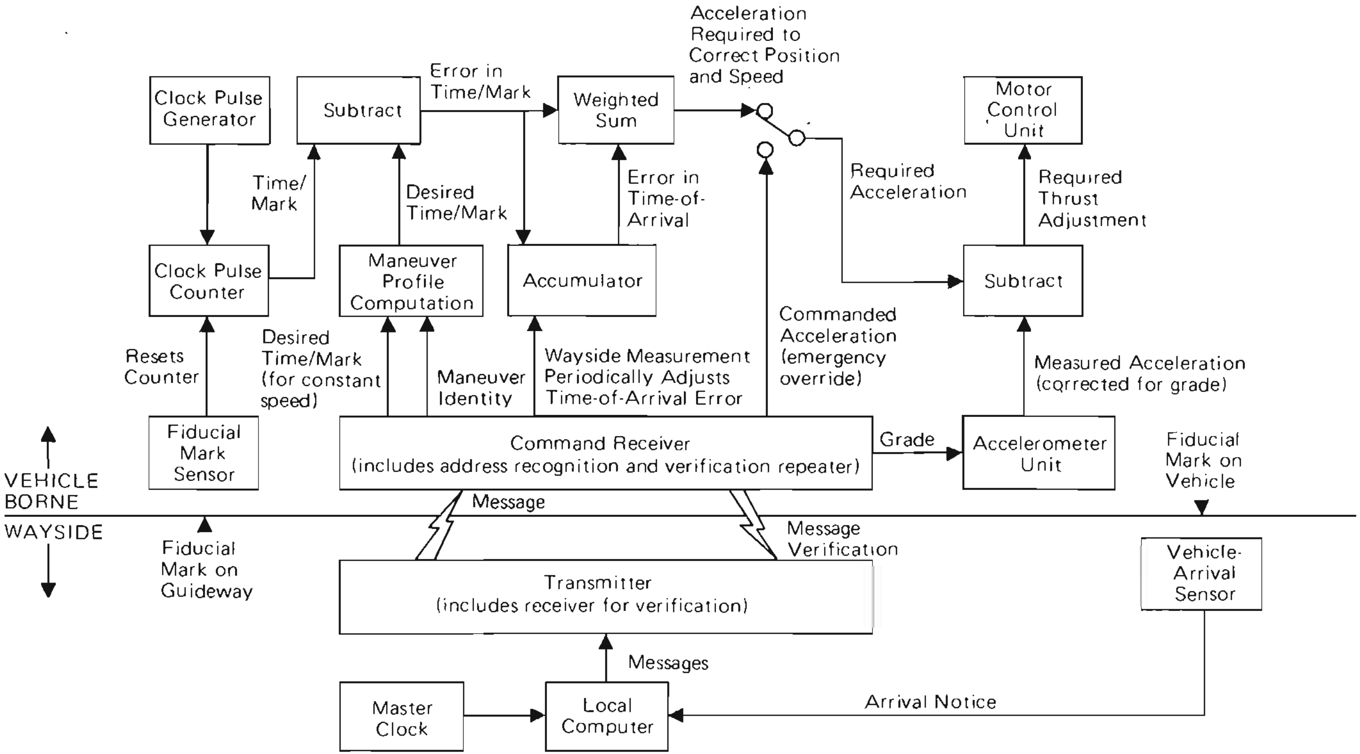


Fig. 4-11. Essentials of The Aerospace Corporation Approach to Longitudinal Control

detectors aboard the vehicle sense the leading edge of the magnets to commutate current among a number of primary coils. (The current in any coil is turned off when it is between magnets. When the current in a coil is on, the direction of the current will depend on the polarity of the magnet adjacent to the coil.) Since the magnet's leading edge must be sensed for purposes of commutating current, this sensing may also be used as the fiducial mark sensing required for speed control. Thus we regard the leading edge of the magnets as being the fiducial marks spaced 1.0 ft apart.

Let us continue our illustrative example. If the vehicle is to travel at a characteristic speed of 60 ft/sec, it is informed by the local computer that it should count 180 clock pulses between successive fiducial mark detections — 10,800 pulses/sec divided by 60 fiducial marks/sec. (In the figure this is shown as "Desired Time/ Mark" where the unit of time is the time between successive clock pulses.) If the count is higher than 180, the vehicle is moving too slowly; if the count is lower than 180, it is moving too fast.

We shall now show that for small errors the error in velocity is proportional to the time/mark error. The measured velocity is

$$V_M = \frac{D}{t_M}, \tag{4.2}$$

where D is the distance between fiducial marks and t_M is the measured time to traverse D (i.e., t_M is the measured "time/mark"). The desired velocity, V_D, is related to the desired time/mark, t_D, by the equation:

$$V_D = \frac{D}{t_D} \ . \tag{4.3}$$

Therefore, the velocity error is

$$\delta V = V_M - V_D = D\left(\frac{1}{t_M} - \frac{1}{t_D}\right) = -\frac{D}{t_M t_D}(t_M - t_D)$$

$$= -\frac{V_D}{t_M}(t_M - t_D) \approx -\frac{V_D}{t_D}(t_M - t_D). \tag{4.4}$$

In the last step we have replaced t_M by t_D since the error in time of passage between adjacent marks is assumed small compared with the time itself. Equation (4.4) shows that the error in velocity is proportional (but of the opposite sign) to the time/mark error, $t_M - t_D$.

In addition to determining the time (count) error in passing between successive fiducial marks, the on-board equipment also accumulates the time error. This cumulative time error at the instant of detecting a fiducial mark represents the error in the time of arrival at that fiducial mark. For small errors, the position error, δX, is merely $- V_D$ multiplied by the time-of-arrival error. Thus,

$$\delta X = - V_D \left[\sum_i (t_M - t_D)_i + \delta t_f \right], \tag{4.5}$$

where $(t_M - t_D)_i$ is the time error in passing the ith spatial interval between marks, starting with the site of the last position fix, and δt_f was the time-of-arrival error at the last position fix.

As the vehicle passes the guideway-mounted vehicle-arrival sensor, a fiducial mark on the vehicle is detected and the arrival event is reported to the wayside computer. This enables that computer to instruct the vehicle to adjust the time-of-arrival error at the output of the accumulator, thus eradicating any drift errors caused by the imperfect clock or due to missed counts or missed wayside fiducial marks.

Both velocity and position errors may be nulled by requiring an acceleration, a_R, given by

$$a_R = -\frac{1}{\tau_1} \left(\delta V + \frac{1}{\tau_2} \delta X \right), \tag{4.6}$$

where τ_1 and τ_2 are time constants which determine the dynamics with which position and velocity errors are eliminated. Substituting δV from Eq. (4.4) and δX from Eq. (4.5), a_R may be written

$$a_R = \frac{V_D}{\tau_1} \left[\frac{1}{t_D} (t_M - t_D) + \frac{1}{\tau_2} \sum_i (t_M - t_D)_i + \frac{1}{\tau_2} \delta t_f \right] \tag{4.7}$$

The required[9] acceleration, a_R, is then compared with the measured acceleration, a_M, to find the required thrust change. In finding a_M, grade information is used to correct the accelerometer measurement to compensate for the component of gravity which is measured if the vehicle is climbing or diving.

The digital longitudinal control system we have just described can also be used to control the advancement or slipping of slots (or fractions of slots). By way of illustration, let us assume that the vehicle is to slip back one 15-ft slot or 15 fiducial marks. To slip back it must temporarily reduce its speed and will get a higher than 180 count at the reduced speed. The total number of extra clock-pulse counts during the slot-slipping maneuver is 2,700 (15 fiducial marks x 180 pulses/fiducial mark). Thus, one way to control the slip would be to have the vehicle follow some deceleration profile until 1,350 extra clock pulses were counted and then to accelerate back

[9] Under certain emergency situations, as when the vehicle ahead is inadvertently decelerating, the local computer will specify the required acceleration (deceleration) as an override.

to line speed while another 1,350 extra pulses are counted. This can be done either by storing a sequence of the clock-pulse counts (between neighboring pairs of fiducial marks) desired during the maneuver or by using a formula to compute the sequence. For the 1/10-scale model, which only operated at one line speed, we stored a few sequences in the vehicle's digital control electronics, each sequence representing a different maneuver, i.e., a different number of slots gained or slipped. We now believe it might be better to compute the sequences because of the very large number of different maneuver profiles that might be used if the system employs many different line speeds.

Let us now turn to the question of initialization, as might be required during reestablishment of traffic flow following a power failure or line blockage.

We need to distinguish three cases:

a. The first is typified by the situation which might occur following the removal of a line blockage. There would be a string of stopped vehicles but the line ahead would be clear. In that event, the first vehicle of the string would be given an instruction to accelerate to the characteristic line speed; one second later the next vehicle would be so instructed, etc. Some distance after each vehicle had achieved line speed it would pass a wayside sensor[10] which would report its arrival to the local computer. This would allow the local computer to decide what slot should be assigned to the vehicle or what moving point the vehicle should track, and the computer would then transmit a time-of-arrival error to the vehicle. As explained above, this would reset the output of the on-board count-error accumulator which would then lead to a transient adjustment of speed to eliminate the time-of-arrival (or position) error.

   In this type of initialization it is not necessary to know the exact position of the stopped vehicles but only their identities and order so that they may be given the start instruction in the proper sequence.

b. The second case is typified by the situation that might occur following reestablishment of power after a systemwide power failure. Consider, for example, a single loop on which all vehicles had stopped. In this event, all vehicles on the main

[10] At the very least, such wayside sensors should be located a little upstream of intersections and merges to provide the local computer with accurate information for resolving potential conflicts.

line would simultaneously be instructed to accelerate up to line speed, using a standard acceleration profile. During the acceleration process and the subsequent cruise at line speed they would hold approximately to their initial spacing. Then, as each passed a wayside sensor, it would be given a time-of-arrival error which would allow it to correct its position.

This initialization approach should be satisfactory following a power failure because the vehicles would probably have received simultaneous instructions to come to a stop at the time of the failure so that their separations when stopped would be approximately equal to their previous running separations. However, to facilitate this type of initialization, wayside sensors should probably be as close as 500 to 1,000 ft, and certainly should be placed upstream of intersections and merges so that a vehicle may correct its position before it attempts to merge.

If the vehicles on the network shown in Fig. 4-10 were brought to a stop, then this method of initialization would be used for vehicles on the solid lines in that figure. It will be recalled that such lines consist of complete loops. Vehicles that had stopped on the dotted lines would not be started up until those on the loops were up to line speed, and they would be handled by initialization method c, discussed below. An exception would be those on the dotted line already committed to the merge. Those vehicles would be started simultaneously with those on the loops, since there clearly is space for them on the loops and otherwise they might block the vehicles on the loops.

c. The third method of initialization applies to vehicles stopped on a siding or on a main-line segment treated like a siding for merge control. Vehicles on the dotted line segments of Fig. 4-10 are an example of the latter. This method would also apply to certain line-clearing procedures, discussed in Sec. 6.3.1, where vehicles on a blocked line are waiting to merge into the traffic stream on a crossing line.

Here the technique is very similar to that described under a. above, with the vehicles accelerated one at a time. But, instead of starting them up at some regular intervals, they are started at such times as necessary to merge them into available spaces on the main line they are entering. This requires an approximate knowledge of the stopped vehicle's position. One way of obtaining this position is to have the vehicle count the number of fiducial marks it has passed since passing the last wayside sensor.[11] (The vehicle's fiducial mark counter would be reset to zero by an instruction from the local computer when a wayside sensor detects the vehicle's arrival.) When initialization is about to occur, the vehicles would be interrogated as to their fiducial-mark counts.

Thus far we have postulated the use of wayside sensors to take position fixes. An alternative would be to have each vehicle periodically measure its own position in absolute terms. There might be, for example, a number of identifiable master fiducial marks, say, every 1,000 ft. The vehicle could, on reaching such a mark, report the event to the local computer. The computer could then inform the vehicle of its time-of-arrival error which would reset the output of the count accumulator shown in Fig. 4-11, resulting in a position adjustment. This alternative approach for taking position fixes has the disadvantages of depending on each vehicle to correctly determine the identity of master fiducial marks and of requiring additional communication from the vehicle to the local computer. It has the advantage, however, that once the vehicles are equipped to read the identity of master fiducial marks, these marks may be placed closely together at negligible extra cost, and this may facilitate initialization.

The reader will see that there are many different approaches of approximately equal merit for accomplishing longitudinal measurement and control. Let us try to summarize what we have learned:

a. There must be absolute position fixes. This may be accomplished either by a wayside sensor which observes the arrival of each vehicle (or, more precisely, of an identifiable fiducial mark on the vehicle) and reports the event and the vehicle's identity to the local computer, or, alternatively, there can be identifiable master fiducial marks along the guideway and the vehicle can report its arrival at such a mark, together with the mark's identity, to the local computer. In either case, the local computer will become aware of all such events and will use the error in time of arrival to instruct the vehicle.

b. If the absolute position measurements are frequent, at most every few feet, the speed may be derived from the position measurements. Otherwise there must be an independent speed measurement, probably aboard the vehicle. Speed may be measured with considerable precision by using a digital technique to measure the time of passage between closely spaced

[11] An alternative would be to start each vehicle creeping along the guideway until it passes a wayside sensor just upstream of the turn or merge, and then, if no space is available for merging, to stop it there until a space comes along into which it can merge.

fiducial marks (fiducial marks not requiring an encoded identity). Because of the position fixes described in a. above, there is no need to have the vehicle clocks synchronized with a master clock.

c. Maneuvers to allow merging should be ordered by the local computer but can be carried out without wayside supervision (or with wayside supervision if the designer prefers).

d. It is desirable to have acceleration measured on board the vehicle, not only to minimize the time for detection of an inadvertent deceleration which might cause a safety hazard, but also to provide quick response to gusts.

e. After a system or partial system shutdown, there are several means for reinitializing traffic flow. These may require somewhat more closely spaced position fixes than would otherwise be necessary, and probably, in any event, will require a position fix just upstream of intersections and merges to enable the local computer to determine the necessary maneuvers for resolving conflicts. If position fixes are not very close together, it may be necessary in some circumstances for the vehicle to report its approximate position to the local computer so that it can be merged into traffic already in progress on a main line. This position could be obtained by counting fiducial marks from the last position fix.

It should be noted that the entire discussion of measurement and control in this section applies to point-follower systems, whether or not slots are used.

4.6.8 Discrete versus Continuous Positions – Synchronization

For a point-follower system, should discrete or continuous positions be used? Is systemwide synchronization desirable?

For both synchronous and quasi-synchronous control (Secs. 4.3 and 4.4), we described a system of moving imaginary slots absolutely synchronized throughout the system. Either a slot would be vacant or a vehicle would be centered in a slot, except near a merge or intersection where vehicles might be changing position from one discrete slot to another. In Sec. 4.6.4 we noted that the arguments for efficient merging and intersection control were dependent on supervision by a local computer but were not dependent on whether incoming and outgoing vehicles were indeed restricted to slot centers. We now reexamine the question more broadly to see whether slots really perform a useful function and whether there is any need to synchronize them throughout the system.

We know that to accomplish merges from station sidings or turn ramps there must be ample space on the main line available for the entering vehicles. Does it matter whether this available space is aggregated into vacant slots or scattered about? To be more specific, if 20% of the capacity of a line is to be left vacant (80% line density), does it matter whether four vehicles are spaced to leave one whole slot vacant, or two 1/2 slots, or three 1/3 slots? If a turning vehicle were aligned with the whole vacant slot it could merge directly into it without requiring any maneuvering of the four vehicles. But, more likely, they would have to shift slots to move the vacant slot into alignment with the merging vehicle. This being the case, it would be about as easy to maneuver to create a slot from the two 1/2 slots or the three 1/3 slots. Thus it would seem that the average line densities that can be used should be about the same whether or not the vehicles are confined to slots, so long as vehicle sequencing is under the control of a local computer with full knowledge of where the vacant space is available.

It will be recalled that in Sec. 4.4.2 we suggested that there not be a discrete set of maneuver starting gates in the maneuvering regions of an intersection. Rather, we recommended a continuum of maneuver starting points where each maneuver is started as far forward as possible consistent with adequate separation from the vehicle ahead being maintained throughout the maneuver. The precise starting point would depend on the maneuver to be performed (i.e., how much distance was to be gained or lost), the maneuver being performed by the vehicle ahead, the point where it started its maneuver, and the initial separation between the two vehicles. Although this prescription was intended for a quasi-synchronous system adhering to a slot-oriented approach, it clearly applies equally well if the vehicles neither start nor end their maneuvers centered in slots.

One argument which might be proposed for adhering to slots is the relative simplicity of implementation, especially if the fine control of maneuvers (not the choice of maneuver) is to be delegated to the vehicles. The local computer would merely instruct the vehicle to "drop back 3 slots" and the velocity profile for the 3-slot slip could be stored in the vehicle's computer. This is how we carried out maneuvers on the 1/10-scale model.[12] But, as pointed out in Sec. 4.6.7, it is probably better to compute the velocity profile because of the large number of possible maneuvers when several line speeds are used. If the maneuver is computed, it can easily be computed for an arbitrary distance to be gained or slipped.

[12] More accurately, we stored profiles of the number of timing pulses that should be counted between the passage of neighboring fiducial marks.

In summary, it would seem that there is no compelling reason for adhering to slots, but also there is no compelling reason for abandoning them, except in the maneuver region of a split-stream intersection where the length of double guideway may be somewhat shortened by using a continuum of maneuver starting points rather than discrete starting gates.

If a system uses the slot principle, there is still the question of whether the slots need to be synchronized throughout the system or only within the jurisdiction of a local computer. The reason for having slots synchronized within at least the area under the jurisdiction of a local computer is so that a vehicle coming in on one line can merge into a slot coming in on another without having to move vehicles fractional slot lengths.

If the synchronization is not universal, then, as vehicles leave one jurisdiction (where they were slot-centered) and enter another, they will no longer be centered in slots. Of course, it is not difficult to instruct them to shift a fraction of a slot to recenter themselves. The instruction must come from the local computer so that a vehicle which must be moved will not be moved into an occupied slot.

Thus, it would seem that systemwide synchronization is not necessary; on the other hand, it is not difficult to achieve. All that is required is that the local clocks be synchronized periodically with a systemwide master clock. When that is done, there is no longer the necessity for position adjustment for vehicles leaving one jurisdiction and entering another. But, if the master clock should fail or the communications to it break down, the operation can continue with each local clock responsible for local synchronization.