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.

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

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

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.

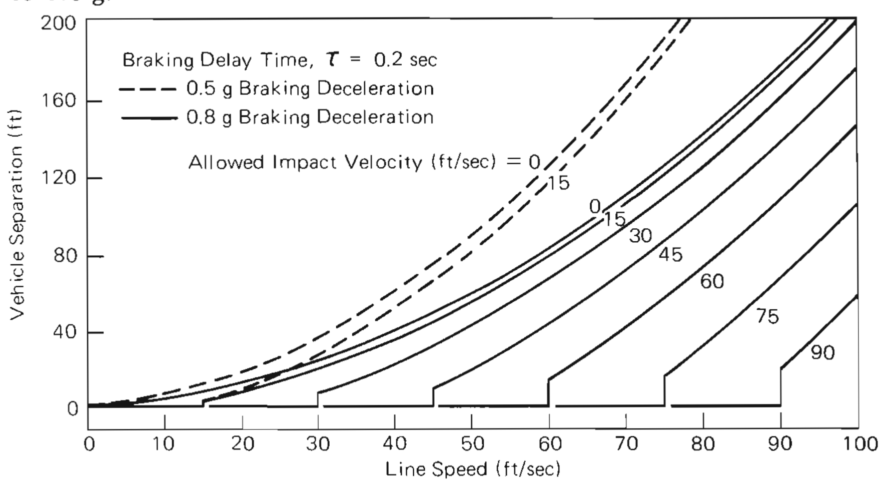


Fig. 4-1. Separation Required Between Two Vehicles if the First is Stopped Instantaneously by Hitting a Massive Object and the Second Brakes to Reduce its Impact Velocity to a Specified Value

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

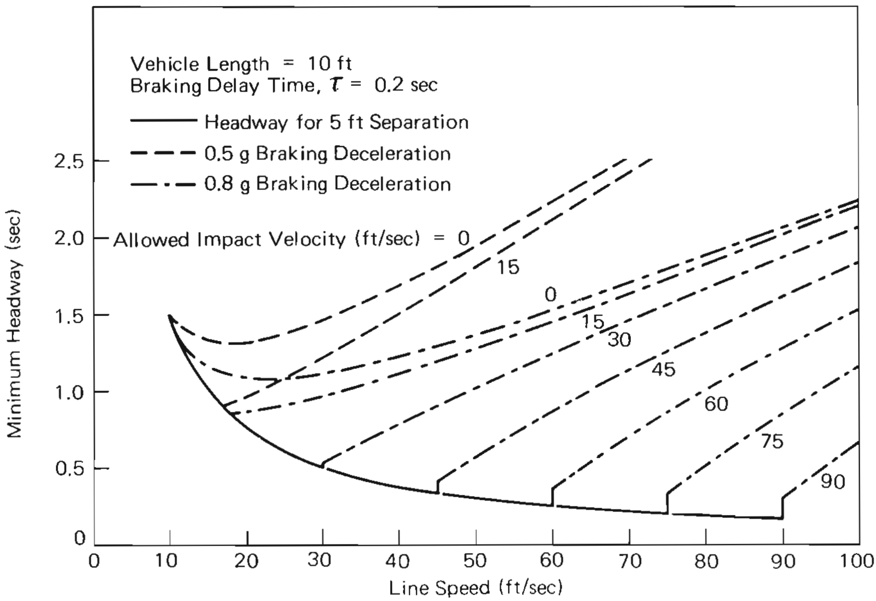


Fig. 4-2. Required Headway if Vehicle Separation is that Specified in Fig. 4-1 but Not Less Than 5 ft

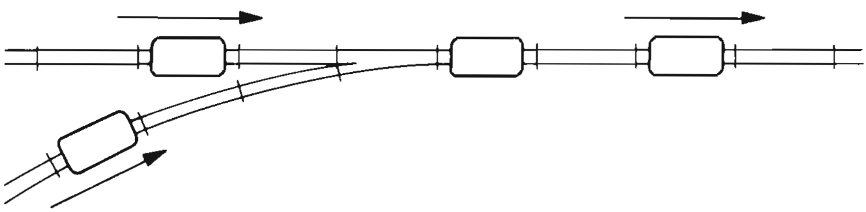
high impact velocity, like 75 or 90 ft/sec, then the system, in general, will operate on the solid curve; i.e., with 5-ft separation. This is the case with the Aerospace Corporation design. In that event, the higher the speed the shorter the headway. But, if the vehicle's protective devices and the adopted 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.

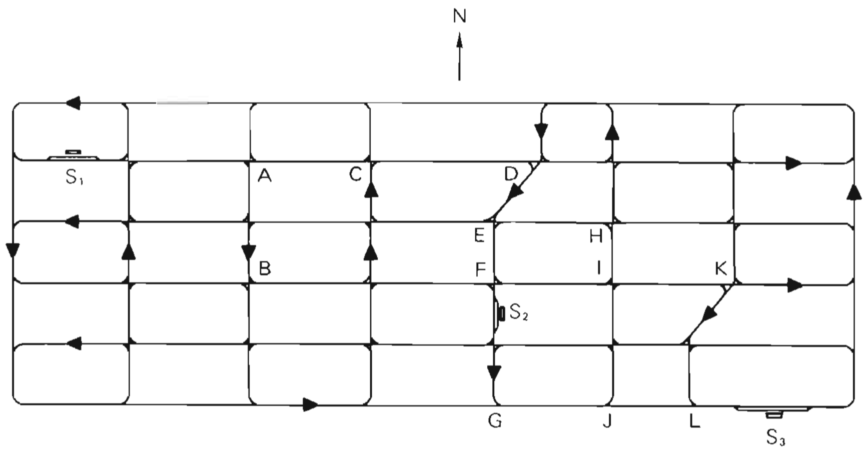


Fig. 4-3. One-Way Network Illustrating Alternate Paths

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.

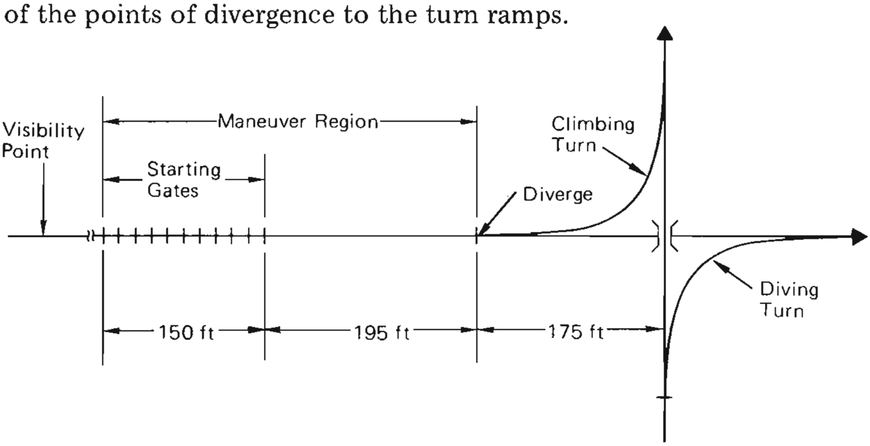


Fig. 4-4. Single-Stream Intersection for a Line Speed of 30 ft/sec

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.

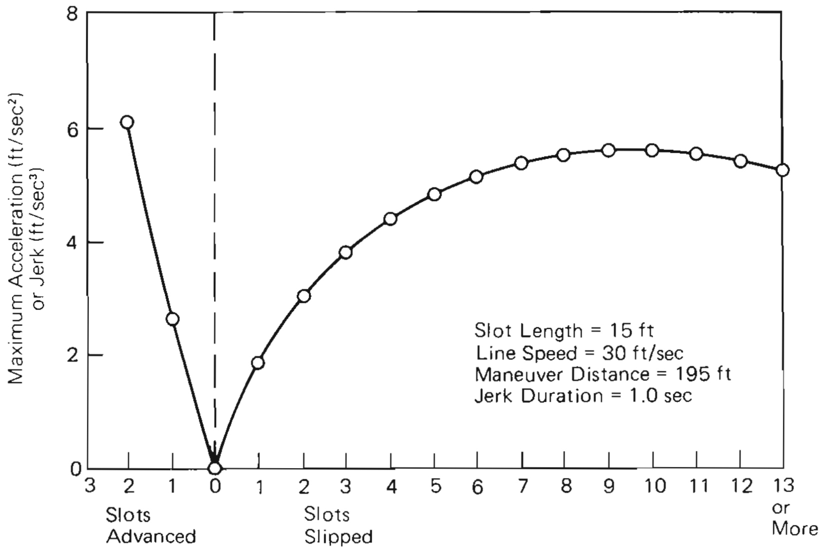


Fig. 4-5. Maximum Acceleration and Jerk for Various Slot Changes

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.

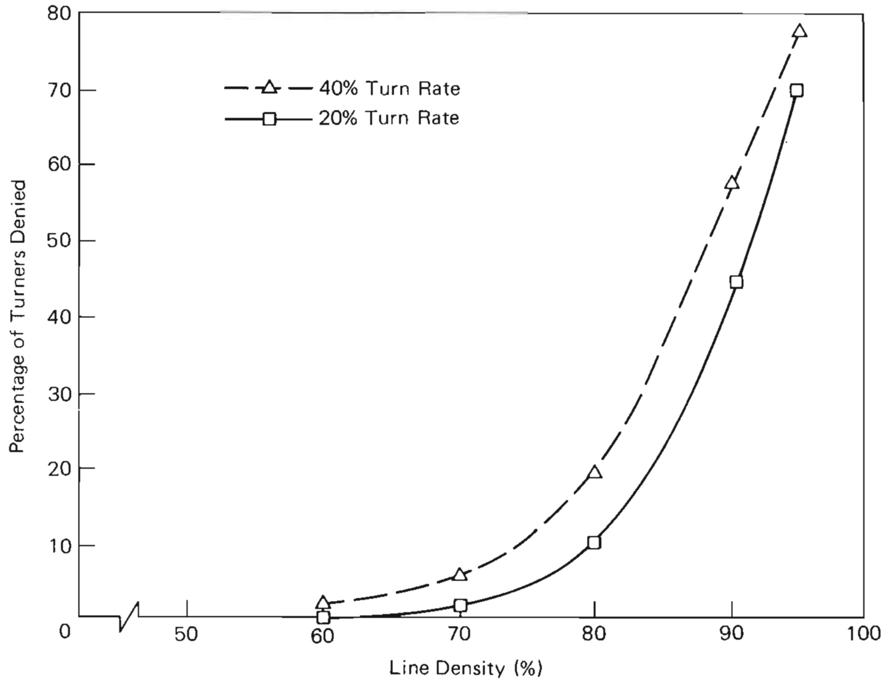


Fig. 4-6. Performance of a Single-Stream Intersection

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.

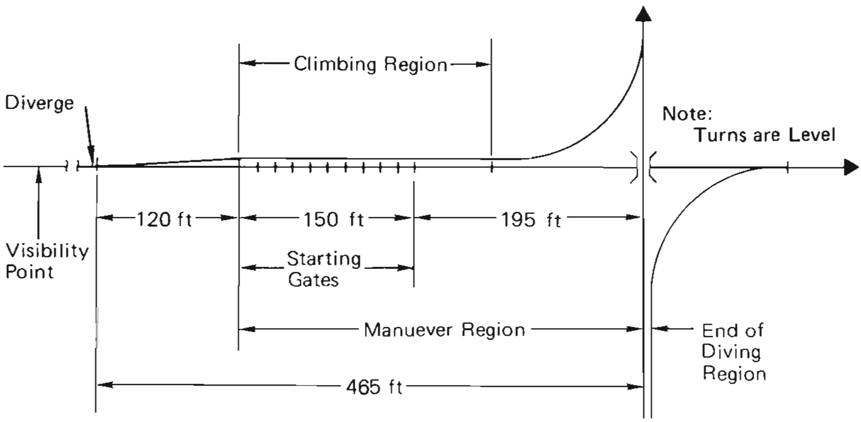


Fig. 4-7. Split-Stream Intersection for Line Speed of 30 ft/sec

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.

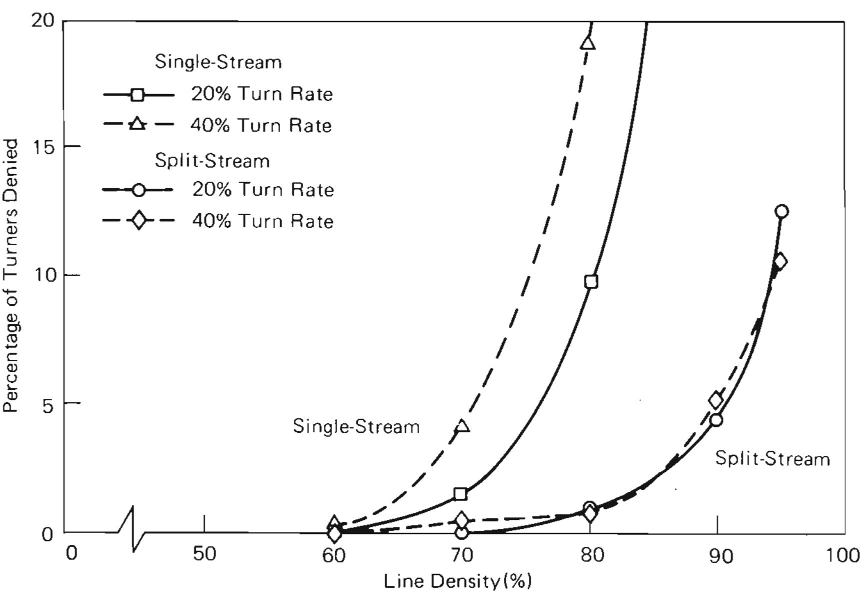


Fig. 4-8. Comparison of Single-Stream and Split-Stream Performance

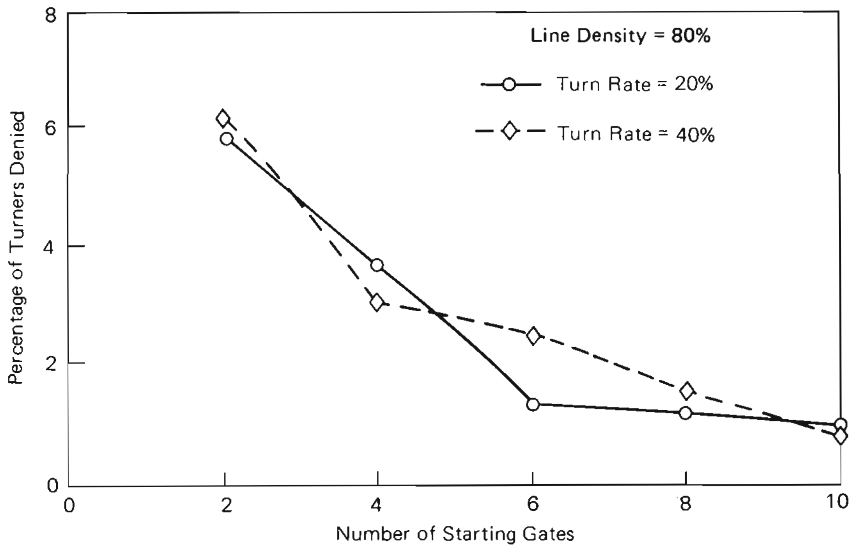


Fig. 4-9. Dependence of Turn Denial on Number of Starting Gates in a Split-Stream Intersection

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.

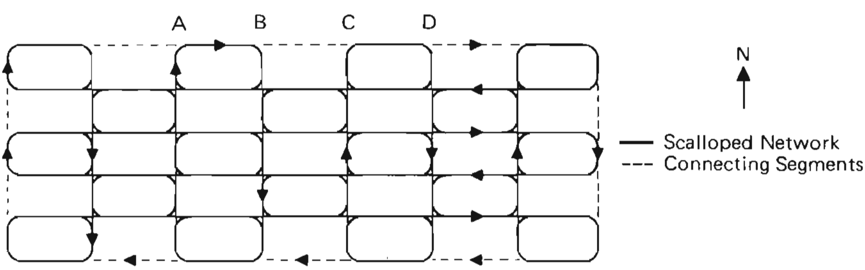


Fig. 4-10. Scalloped Network with Connecting Segments

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

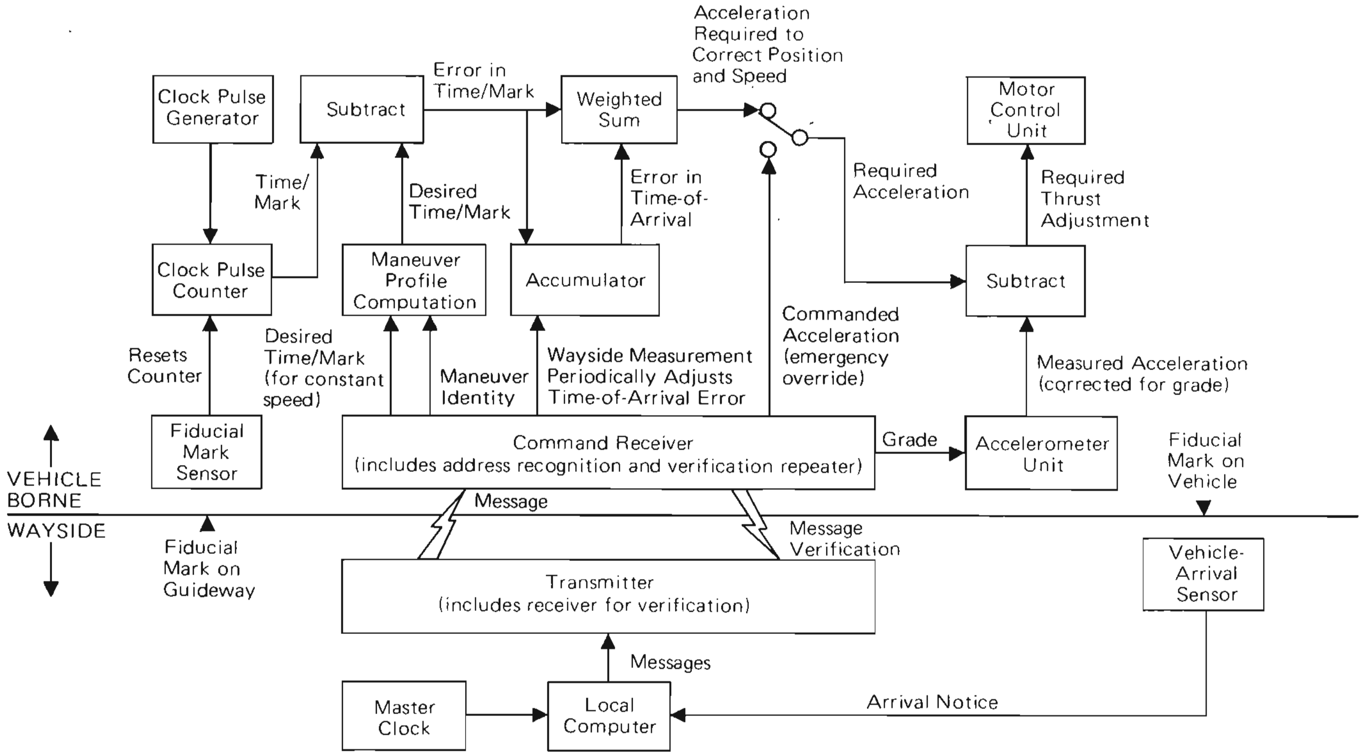


Fig. 4-11. Essentials of The Aerospace Corporation Approach to Longitudinal Control

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.