Simulation of Merge Junctions in a Dynamically Entrained Automated Guideway Transit System

Steven E. Shladover

Systems Control, Inc. (Vt)
Palo Alto, California

Abstract

Merge junctions and intersections are the principal capacity limiters and sources of delay in automated guideway transit (AGT) networks. The capacity and delay performance of the merge junctions must be thoroughly understood before an AGT network can be designed. This paper describes the modeling and event-structured Monte Carlo simulation of a single merge junction, having two input lanes and one output lane, operating in a quasi-synchronous network. The simulation described here differs from previous merging models by permitting slots to be of variable length to accommodate trains of different lengths, and allowing unlimited maneuvering of trains (if needed). It has the unique ability to represent the operation of a special, newly devised concatenating merge, which permits trains arriving on separate input lanes to combine themselves into a longer train at the merge.

The simulation logic which is used to represent special features of merge performance (flow compressibility, upstream propagation of disturbances, etc.) is explained, and the implementation using GASP IIA is described. The statistical considerations underlying the experimental design are discussed and some sample diagnostic outputs displayed.

1. INTRODUCTION

1.1 Automated Guideway Transit Systems

Automated guideway transit (AGT) systems are designed to provide urban public transportation using driverless, automatically controlled vehicles operating on exclusive guideways. In the simplest AGT systems, all vehicles travel the same routes and make the same stops (at all stations). The more sophisticated AGT systems permit vehicles to operate on a variety of routes, stopping at different combinations of stations (which may be located off line). The routing of vehicles must then be controlled by wayside computers using one of several different operating policies (to be described in Section 2.)

The performance of the merge junctions and intersections which connect the links of an AGT network must be thoroughly understood because they have a dominant influence on the system capacity and level of service (delays to passengers). Although the line-haul capacity of a simple guideway lane is easily evaluated by kinematic analysis [1], that capacity is impossible to achieve in practice because some gaps must be provided in the flows entering a merge to enable vehicles to resolve potential merge conflicts. The proportion of gaps to vehicles needed to ensure acceptable merge performance must be determined before the effective capacity of the AGT system can be estimated.

1.2 Dynamically Entrained AGT Systems

There appear to be numerous advantages to operating AGT vehicles in dynamically reconfigured trains or platoons, as well as individually [2]. Platoons are defined to be noncontacting sequences of vehicles which maintain very close spacings among themselves by use of a vehicle-follower control system, while trains consist of mechanically coupled vehicles. Dynamic reconfiguration refers to the ability of a vehicle to operate either individually or in very close formation with others and to make the transition between these operating modes at any time, including while traveling at cruise speed.

The principal reasons for entraining AGT vehicles are the increase in the passenger-carrying capacity for each unit of costly infrastructure and the increase in reliability should one of the entrained vehicles suffer a failure. In addition, an entrained AGT system can be designed to have smaller, less costly and less visually obtrusive guideways and vehicles, while providing unique operational flexibility and simplifying network-level control problems. A straightfoward kinematic analysis has demonstrated the influence of entrained operation on AGT lane capacity [1]. However, merge junctions are the principal capacity limiters in a network and as such require careful study. The effect on merge performance of trains (or platoons) of varying

CH1437-3/79/0527-0539$00.75
1979 Winter Simulation Conference

lengths in the incoming traffic flow could not be evaluated using any previously developed model of merging, necessitating this new study. The model developed here has been used to show how the dynamic entrainment capability makes it possible to operate merge junctions at much higher traffic levels than in more traditional systems.

1.3 Goals of This Study

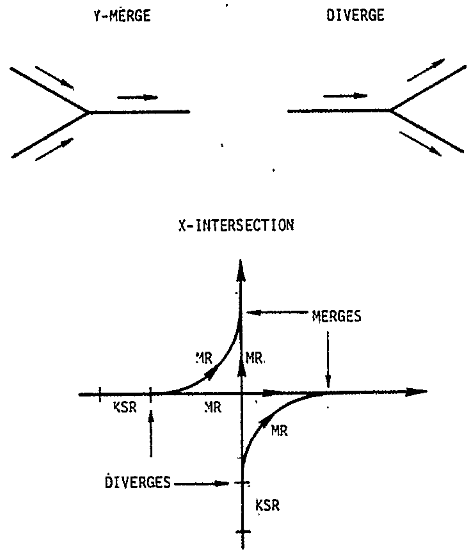
Straightforward kinematic analyses indicate that increasing train lengths produce increased lane capacity. However, the longer trains are likely to impair merge performance to some extent, attenuating the capacity gains. This study seeks to provide a realistic estimate of the fraction of the theoretical lane capacity which can be effectively used when merge conflicts must be resolved. A similar study for single-vehicle systems was conducted by Godfrey [3]. His work demonstrated the virtual impossibility of applying purely analytical methods to a simpler problem than that considered here (with no variety in train lengths), in the presence of congestion. As a result, an event-structured Monte Carlo simulation has been used to evaluate merge performance in the present study. The simulation results are intended to help produce realistic estimates of the tradeoff characteristics among line-haul capacity, level of service to passengers (delays) and infrastructure costs (maneuver ramp lengths) in a merge junction.

2. IMPORTANT DEFINITIONS

The links which form a guideway network can be connected in several different ways, as shown in Fig. 1. The Y-merge consolidates two lanes of traffic into one, while the diverge (or demerge) splits one lane of traffic into two. The X-intersection contains two merges and two diverges arranged so that traffic entering on either of the unidirectional guideway lanes can leave on either lane. Upstream of the diverges are "known status regions" (labeled KSR) in which trains are scheduled to maneuver, and downstream of the diverges are "maneuver regions" (labeled MR) in which trains maneuver so that they can proceed through the merges without interference.

Different network-level operating policies impose differing requirements for control of intersections. In a synchronous system, a large central control computer determines the precise routing of each vehicle before allowing the vehicle to depart from its origin station. All delays are then experienced at the origin station, while the computer searches for an available routing to the destination station. Once the vehicle departs, its trip should proceed at constant speed, without interruption. In a

Figure 1
MERGE DEFINITIONS

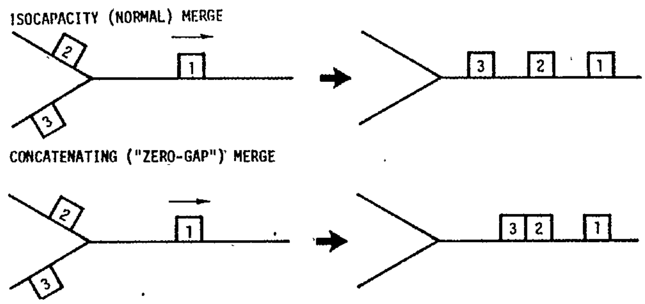


quasi-synchronous or asynchronous system, local computers at each intersection command vehicles to move behind or ahead of their original trajectories (slot slip or slot advance, respectively, in the quasi-synchronous systems) in order to pass through merge junctions without conflict.

The traffic intensity of a guideway lane is defined to be the ratio of the actual flow rate of trains on that lane to the maximum feasible flow rate of trains of the same mean length on that lane. The qualification based on mean train length is essential because the feasible flow rate is a function of train length. The effective traffic intensity at a merge junction is defined to be the sum of the traffic intensities on the upstream (incoming) lanes, ρ_{sum}.

Two fundamentally different approaches to the operation of a merge in a dynamically entrained AGT system are pictured in Fig. 2. The first, referred to as an isocapacity (or normal) merge,

Figure 2
ISOCAPACITY AND CONCATENATING MERGES



is typical of all merge junctions reported in the literature, in that trains which enter the merge separately leave it the same way, separated by gaps of at least the minimum safe inter-train spacing. The number of trains leaving is the same as the number entering, and the downstream traffic intensity is the sum of the traffic intensities on the two incoming lanes. The second approach, the concatenating (or "zero gap") merge, is unique to this study of entrained AGT, and is specifically designed for use with dynamic entrainment. Under the concatenating merge concept, independent trains entering from the opposite input lanes can be combined into a single longer train at the merge (if that represents the minimum-maneuver-distance alternative). The number of trains leaving the merge is therefore less than the number entering, and the downstream traffic intensity is less than the sum of the traffic intensities on the incoming lanes. This provides the merge junction with a greatly improved tolerance for heavy traffic, and particularly for transients in which the sum of the traffic intensities on the two incoming lanes, ρ_{sum}, may exceed unity.

The system-level implications of the choice between isocapacity and concatenating merges are discussed at length in Refs. 4 and 5.

3. REVIEW OF PREVIOUS STUDIES OF MERGE AND INTERSECTION PERFORMANCE

Many studies of merges and intersections for AGT systems have been reported within the past decade. All have been concerned with single vehicles and fixed-length slots, rather than variable-length trains and slots, and none has included the concatenating merge.

The pioneering fundamental studies on merging, from which all subsequent studies have been derived, were performed by Godfrey [3], Athans [6], and Athans and Levine [7]. Godfrey established the capabilities of the simple Y-merge in a quasi-synchronous system, and demonstrated the dominance of the first-in, first-out (FIFO) merge control discipline. While Godfrey's study has served as the prototype for all the discrete, slot-based studies to follow, Athans has provided the prototype for the continuous, asynchronous merge by using optimal longitudinal control coupled with dynamic programming to choose the merge sequence.

Brown [8,9] designed an asynchronous Y-merge control strategy ("adaptive merging") using performance measures which were more relevant to AGT system performance than Athans'. His study demonstrated the difficulty of characterizing asynchronous merge performance concisely and in a way which permits comparisons with other studies. The papers by Whitney [10] and Sarachik and Chu [11] represent two attempts to combine the best features of the continuous and discrete approaches to merge control in hybrid merges.

The only fully synchronous (centrally controlled) systems which need consider merge performance explicitly are those which use cycle preprogramming to schedule vehicles [12,13]. In these systems, merge performance capabilities determine the lengths of the cycles which can be operated and the lengths of the maneuver ramps which are needed. The combination of dynamic entrainment at merges (the concatenating merge) and cycle preprogramming appears to offer a promising application for synchronous network control, and deserves further study.

Complex X-intersections have been studied by several investigators, beginning with Aerospace [14-16], who considered quasi-synchronous systems in which vehicles could maneuver either upstream or downstream of the diverges. Downstream maneuvering, which partially decouples the traffic streams destined for the two different output lanes, produced a major reduction in missed turns with no change in maneuver distance. Brown [17] developed a "chain-forming" algorithm to coordinate the maneuvers of the vehicles entering both sides of an intersection in a quasi-synchronous system. Some of the limitations of Brown's method were removed by McGinley [18,19], who demonstrated substantially improved performance with his more complicated "loop-former" algorithms.

4. MODELING OF SINGLE Y-MERGE

4.1 Baseline Merge Configuration

The baseline merge for the present study has been chosen to be a single Y-configuration merge, in a quasi-synchronous network, with effectively unlimited available maneuver distances. The essential features of merge performance with variable-length trains can be studied this way, without the complications introduced by a complete X-intersection, in which the two Y-merges and two diverges must be be coordinated with each other. It is expected that for eventual implementation in a guideway system, the X-intersections would be used, and would be under the control of "loop-former" algorithms such as those of McGinley [18, 19]. The single Y-merges to be studied here could eventually be coupled together by use of a "loop former" to produce a complete intersection. By providing no limits on maneuver distance, the problem of merge failures is avoided in the present study and the maneuver distance (or ramp length) which is needed can be studied as an output of the simulation. In this way, the results of a single simulation can be used to indicate how the merge failure rate would change with respect to changes in the available ramp lengths.

The quasi-synchronous, variable-slot-length control system used in this study is not typical of any standard system described in the literature, but has been developed specifically for use with the variable-length trains which will be studied here. A slot-oriented system was chosen to provide commonality with most existing AGT research, and to avoid the complications which arise in a completely asynchronous system (requiring a mixed discrete and continuous simulation incorporating explicit modeling of the performance of the vehicle longitudinal control system). The results of such an asynchronous merge study would have been awkward to characterize compactly (like those of Brown [8,9]) and difficult to compare with the results produced by most previous studies.

Conventional slot systems, in which the slot is sized for the vehicle length plus the needed safe gap between vehicles, are inappropriate when variable-length trains are used because the slots would have to be sized for the longest allowable trains, and any shorter trains would cause significant capacity losses. Much of the discretization loss inherent in such a fixed-slot-length system is eliminated here by using variable-length slots to provide most of the capacity advantages and flexibility of the asynchronous system without the added complications. The basic slot-length unit equals one vehicle length, and the minimum allowable spacing between trains is fixed at an integral number of vehicle lengths, M. Therefore, a train containing N vehicles occupies a large slot which is, in fact, composed of N+M vehicle-size slots. Merge operations and logic remain discrete and simple, but very little effective capacity is lost through this kind of fine-grain discretization. Slight modifications have been made to the baseline simulation code so that it can be used to describe a more conventional fixed-slot-length system as well. Sample runs using the modified code have demonstrated that the fixed slot lengths cause a major deterioration in merge performance, even when only single vehicles are used.

The Y-merge simulated here is different from most of those which have been previously analyzed because of the lack of limitation on maneuvering. Maneuvers are assumed to be initiated as far upstream of the merge as necessary in order to resolve potential conflicts before the merge itself is reached. The failure rate in a merge having a finite maneuver distance corresponds to the percentage of the maneuvers which are found to require a longer maneuver distance in this simulation. If an X-intersection is designed with ramps long enough that the merge failure rate is very low under all operating conditions, the performance of that intersection will be very close to the performance of the individual Y-merges of which it is composed.

The dominant merge priority discipline, as found by Godfrey [3] and others for single vehicles, is FIFO. For the present study, two different FIFO disciplines have been used, one based on arrival of the front of the train and the other on the arrival of the rear. The FIFO-rear strategy, which serves as the baseline case for this study, is designed to reduce the maximum maneuver distances and ramp lengths needed at merges. It has been found to have that effect for the concatenating merge, but is virtually indistinguishable from the FIFO-front discipline for the isocapacity merge.

The simulation code has been designed for maximum flexibility, including eventual use as the merge element of a full network simulation. The explicit representation of the entire sequence of merge events (rather than only the scheduling) permits later modifications to accommodate a variety of merge operations, such as slot advancing and rescheduling of maneuvers.

4.2 Compressibility of the String of Trains

During all the time that the separate trains on an approach lane are maneuvering, they must maintain at least the minimum safe spacing determined by a kinematic capacity analysis. When a train decelerates to perform a slot-slip, the one following it may also have to slow down to avoid approaching too close. The need to maintain the safe spacing can cause disturbances to propagate far to the upstream side of a merge unless preventive measures are taken. This phenomenon was recognized by Godfrey [20], who suggested that the normal spacing between vehicles be somewhat larger than the minimum allowable in order to permit some compression of the vehicle string during maneuvers, and to avoid propagating disturbances upstream. In the single-vehicle case, provision of a small extra headway margin could guarantee that no disturbances would propagate upstream, permitting the following vehicle to begin its deceleration at the same guideway location as the decelerating leader.

For purposes of the present simulation, it was found to be useful to include some compressibility allowance, but with long trains it was not possible to completely avoid propagating disturbances upstream. For the baseline case, with trains cruising at 13.4 m/s and a nominal spacing of 33.5 m, the practical minimum headway is 2.5 s. Assume that the preceding vehicle suddenly decelerates at its service braking rate of 0.15 g. After 2.5 seconds, it will be about 4.6 m closer to the follower than it was before decelerating (6.1 m closer if deceleration is at

0.2 g). A direct simulation of such a maneuver using the longitudinal control system described in Refs. 2 and 4 predicted a spacing change of about 3 m, or approximately the length of one slot unit (one vehicle length). Therefore, if that single extra slot unit of headway were provided between the vehicles under nominal cruise conditions, it would be possible for the follower to begin its deceleration when located 10 slot units (33.5 m) closer to the merge than it was when the leader began to decelerate. The baseline simulation condition for this study includes a safe spacing between trains of 11 slot units (10 for safety plus the one for compressibility), and a forward displacement of the start of maneuvers (variable KOMPRS) of 10 slots.

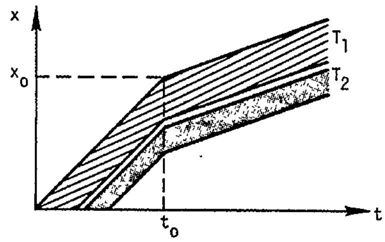
4.3 Upstream Propagation of Maneuvers

In spite of the provision of an extra headway margin, as described in Section 4.2, some slot-slip maneuvers will be displaced upstream so that they are completed prior to the merge junction, rather than right at the merge. This type of displacement is most often produced when the preceding vehicle is required to slip a large number of slots. For example, consider the case of two trains approaching a merge at the minimum allowable headway, with both required to slip a substantial number of slots, as indicated in the diagram of Fig. 3a. The second train, T_2, must decelerate at virtually the same time as T_1, which means that it must begin to decelerate at a location further upstream from the merge. The compressibility effect does allow the deceleration of T_2 to be delayed by KOMPRS slot units (generally 10 for the baseline case), but that length can be easily dominated by train length, particularly when the nominal spacing between trains is 11 slot units.

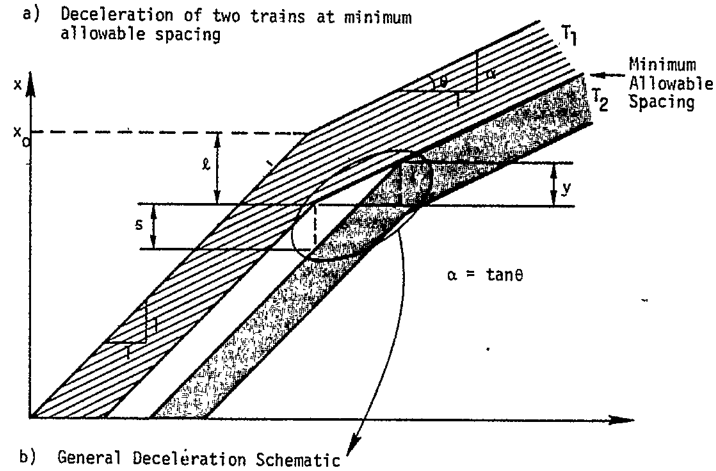
Even if the trains are not very close together when the first must decelerate, it may be necessary for the deceleration to propagate upstream. A simple analysis of this phenomenon, ignoring the finite deceleration rates of both vehicles and considering only their speed changes (because they begin far enough apart) follows, based on the diagrams of Figs. 3b and 3c. The train length, ℓ, is assumed to include both the length of the train and the minimum safe spacing, NSPACE. The analysis considers the limiting case in which the trains are at the minimum allowable spacing once both have decelerated to the minimum speed used for slot slipping.

The goal of the analysis is to find x, the location at which the front of the following train must begin to decelerate when the preceding train begins to decelerate at location x_o. The initial spacing between the trains (prior to the speed reduction) is s and the intermediate variable y is defined by the trigonometry shown in Fig. 3c. Notice that the magnitude of the two

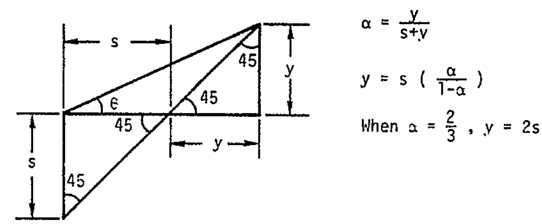
Figure 3
ANALYSIS OF UPSTREAM PROPAGATION OF MANEUVERS



a) Deceleration of two trains at minimum allowable spacing



b) General Deceleration Schematic



c) Detailed Trigonometry from b)

speeds does not matter, but only the percentage of speed reduction which occurs in the slot slip. In general,

$$x = x_0 - \ell + y + \text{KOMPRS} \qquad (1)$$

For the baseline case, the cruise speed is 50 km/h and the minimum speed in slot slipping is 33.3 km/h or 2/3 of the cruise speed. When the speed ratio, α, is 2/3 it follows from the equation shown on the figure that $y = 2s$. Therefore, for the simulation results reported here, the follower begins its slot slip at a location x when the leader begins its slip at x_0:

$$x = x_0 + 2s + \text{KOMPRS} + \text{NSPACE} - (\text{length of first train}) \qquad (2)$$

Note that for the worst-case condition shown in Fig. 3a, with $s = 0$, the follower begins its slip only KOMPRS slots beyond where it was when the leader began to slip.

4.4 Slot Slip — Maneuver Distance Relationship

The number of slots a train must slip (or advance) to avoid a conflict with an approaching train on the opposite input lane is calculated within the simulation code using information about the relative arrival times of the trains and the number of slots the earlier of the two trains was previously scheduled to slip. The slip and advance maneuvers must be scheduled so that they are completed prior to arrival at the merge itself, and must therefore start a proper distance upstream of the merge. The amount of time and maneuver distance needed to accomplish a slip or advance of any number of slots is computed in a subroutine which uses a trilinear curve fit to summarize the performance of the vehicle longitudinal control system previously described in Refs. 2 and 4. By changing the values of some coefficients, the same subroutine is used to represent kinematically optimal slot slipping and advancing as well.

4.5 Input Parameters and Output Measures

The merge simulation has been designed to be as flexible as possible, permitting a large range of input conditions to be varied, so that the user can perform a broad-based study of merge performance. The parameters which are available for variation are:

— Merge priority to be first-in, first-out (FIFO) based on either arrival of front or arrival of rear of the train.

— Average length of incoming trains, LENGTH.

— Maximum length of incoming trains, LMAX.

— Distribution of train lengths (all same, or uniform from one to LMAX, or triangular from one to LMAX, with average value of LENGTH).

— Combined mean traffic intensity (summation of two input lanes), defined on basis of LENGTH and NSPACE, RHO_{sum}.

— Distribution of traffic between the two input lanes.

— Minimum safe spacing between trains on same lane, NSPACE.

— Minimum safe spacing between trains merging from opposite input lanes, MRGSPC (when MRGSPC=0, concatenating merge strategy is used, combining incoming trains into longer trains)

— Choice of primal or antithetic* random-number sequences for defining distributions of gaps between incoming trains and lengths of those trains (all combinations of primal and antithetic for both gaps and lengths).

— Seeds for four random-number sequences.

— Number of trains to simulate during the run.

— Slot advance capability in isocapacity merge (maximum number of slots which can be gained).

— Maneuver capabilities for concatenating merging:

- slot slipping only
- slot advance only
- choose slot slipping or slot advancing based on whichever requires the least maneuver distance.

— Percentage of incoming trains traveling at minimum allowable spacing (degree of bunchiness of input flow), modifying exponential distribution of gaps.

The output measures by which performance can be evaluated are also quite numerous. Within the main body of the simulation, the four moments and histograms of frequency of occurrence of eight variables are collected:

— Gaps on input lanes 1 and 2 (separately).

— Output lane gaps.

— Number of slots slipped per train.

— Number of slots slipped per vehicle (passenger delays).

— Maneuver distance used on input lanes 1 and 2 (separately).

— Output train lengths (important for concatenating merge, in which output trains are longer than input trains).

Additional output measures, chiefly plots of cumulative distributions and printouts of selected percentiles of the above eight variables, are provided in optional post-processing routines. The tail percentiles are particularly important for this study because of their impact on maneuver ramp design and usefulness for estimating merge failure rates when maneuver ramp lengths are limited.

5. IMPLEMENTATION OF THE MERGE SIMULATION

5.1 Computer and Simulation Language

The merge simulation was developed as part of a sponsored research project in the Dept. of Mechanical Engineering at the Massachusetts Institute of Technology. The most economical and readily accessible computer to use for this study was the Interdata Model 80 at the Joint Computer Facility shared by Mechanical Engineering and two other M.I.T. departments. This 64K-byte minicomputer accepted bulk input for initialization, but then operated in a user-interactive mode, in which the user could monitor the progress of the simulation at a CRT console display and could alter its course using data switches and typed inputs. Graphical outputs could be generated (at the user's option) quickly on a CRT plotter,

* The antithetic random-number sequence is the complement of the primal sequence; i.e., if the random-number generator produces a primal sequence of (0-1) variables, x_i, the antithetic variables are defined to be $y_i = 1-x_i$.

and hard copies were provided on either an electrostatic printer/plotter or a photographic raster-scan hard copy unit.

The minicomputer was less costly to use and more convenient to access at any hour of the day or night than the large, multi-user mainframe which was also available. The job control language for the minicomputer was much simpler and more user-oriented than that for the mainframe, and graphical outputs were also much more readily obtainable from the smaller system.

No simulation-oriented language was implemented on the Interdata 80 when this study of merge-junction performance was begun, but the availability of FORTRAN on this system strongly favored the use of GASP, which is a package of FORTRAN programs. The GASP IIA version which was available was not ideal for use in this simulation, and as a result it had to be modified in several ways. This version of GASP only had provisions for a single random-number generator, and because the merge simulation needed four separate random-number sequences, new code had to be provided to keep track of the four separate seeds. In addition, the filing system had the awkward feature that each file only had a single set of pointers, and could only be ordered on the basis of the values of a single attribute. This limitation required that some extra (redundant) files be set up, making inefficient use of the available storage space.

5.2 Simulation Events and Files

The simulation of a single Y-merge includes nine separate events, each represented by its own FORTRAN subroutine(s).

(1) and (2) Entry of trains into known status regions upstream of maneuver regions on lanes 1 and 2, respectively. These events each call two subroutines, the first, called SKEDLR, to perform the scheduling of merge maneuvers and collect most performance statistics; and the second, called GENTRN, to generate the next train to enter on that lane.

(3) Warm-up completion. Because only steady-state performance statistics are desired, this event returns all accumulated statistics to zero after the starting transient has passed (at a time chosen by the user, and punched on a data card for each simulation run).

(4) Entrance of vehicle into maneuver region. At any time after this event, trains may maneuver to sequence themselves through the merge properly. Train records are transferred into the maneuver-region files, and trains are scheduled for slot slipping (if required according to the calculations done in SKEDLR) or for passage through the merge if they are not to slip or advance at all.

(5) Initiate slot slip. Train record is transferred to the slot-slipping file for its lane, and end of slot slip is scheduled.

(6) End of slot slip. Train is removed from slot-slipping file and scheduled for passage through the merge. In most cases, the merge follows immediately (the slot slip ends at the merge), but under certain conditions of heavy traffic, when the preceding trains on the same lane must undergo large slot slips, this may occur somewhat upstream of the merge.

(7) Pass through merge. Collect statistics on gaps between trains (and on output train length if concatenating merge is used), and remove trains from filing system.

(8) End of simulation. This is defined as an explicit event so that no stray trains are left in the system at the end, and so that the statistics collected upstream of the merge in SKEDLR apply to all the same vehicles as those collected at the merge in event number (7). The end of simulation event is scheduled after the desired number of trains (as read off a data card) have been simulated.

(9) Interim statistical reporting. The GASP subroutine SUMRY, which prints out moments and histograms, can be scheduled to be called at any time during the simulation (again using data cards). This has been used to study the stationarity of the simulation record, and to determine how the output statistics change with respect to simulation run length.

The main GASP filing array has been used to hold seven files, the first being the event file which contains all scheduled events which have not yet occurred. The next two files contain the trains which occupy the two known status regions upstream of the maneuvering regions. The trains are entered into these files when generated (in GENTRN) and are removed when they enter the maneuver region (event number (4)). Another two files are for the trains in the maneuver regions which are not in the process of slot slipping. Trains enter these files on entering the maneuver region, but are transferred to the final two files during the time they are slot slipping. If the slot-slip maneuver ends before the merge is reached, these trains are then returned to the maneuver region file.

5.3 Practical Limitations

The 64K-byte memory of the available computer imposed a strong constraint on this simulation model. The event subroutines and GASP mainline and subroutines, together with the filing arrays needed to accommodate most merge conditions, occupied all of the available memory to within a handful of bytes. The simulation was coded with conservation of memory space as a prime design criterion. For example, integer variables were used to represent everything except simulation time (which would, in many cases, have exceeded the largest possible integer representation on the Interdata 80) because integers occupied only 16 bits, as compared to 32 bits for floating-point variables. Some highly congested operating conditions, in which many trains had to be retained in the files at once, could not be simulated because of the space constraint.

Sophisticated analyses of the simulation data and graphical displays of the results had to be performed by post-processing routines separate from the simulation for the same reason. The vital performance parameters for each train were written onto direct-access disk files during the simulation and read later. Three files were used: one each for the queues on the incoming lanes and a third for the departing trains (only used with concatenating merge). The tight memory size constraint did not permit buffering of the inputs to the direct-access disk files, imposing a significant run-time penalty. The majority of the execution time for each simulation run was devoted to disk operations: searching for the appropriate one out of two or three randomly ordered files in which to write the entry for each train.

5.4 Post-Processing Routines

At the conclusion of each simulation (normally indicated by the passage of the number of trains specified on an input data card, but also possibly triggered by toggling a data switch or exceeding a specified period of simulated merge operation), a series of post-processor routines, chained together in successive core loads, was invoked. After reading the direct-access disk files, these post-processors performed the following functions:

– Calculation of key percentiles (50, 80, 90, 95, 98, 99, 99.5) of slot slips, maneuver distances and passenger delay times on each input lane, as well as output gaps and (for concatenating merge) output train lengths.

– Plots of cumulative distributions of slot slips and maneuver distances on both input lanes, on either a linear or normal probability scale.

– Plots of slot slips and maneuver distances on each input lane with respect to simulation time.

– Cross-plots of slots slipped on two input lanes at fixed intervals of time.

– Cross-plots of maneuver distances on the two input lanes at fixed intervals of time.

– Correlations of slot slips and maneuver distances on the two input lanes.

– Cross-plots of maneuver distances vs. slots slipped on each input lane.

– Plots of spectral density of slot–slip or maneuver-distance data on each input lane (either linear or logarithmic frequency scale).

– Plots of autocorrelations of slot slips or maneuver distances on each input lane.

Because of the computer's limited memory size, the post-processing occurred in several different programs. One program computed the percentiles and provided linear-scale plots of the cumulative distributions. A separate program produced the time plots, cross plots, and correlations, and also included the option to produce the normal-probability scale cumulative distribution plots. This program was chained together (in successive overlays) with three additional programs which did the spectral analysis and autocorrelations, all maintaining the same COMMON block.

6. STATISTICAL CONSIDERATIONS IN THE MERGE SIMULATION EXPERIMENT

The experimental design used for this study does not fit any standard formulation. Factorial experiments would have been much too extensive for the number of parameters to be investigated, and other methods were not well suited to this multiple-parameter search. As a result, the experiment was conducted as a series of essentially independent tests, looking for trends in different directions independently.

In order to avoid unnecessary random variation among simulation runs, which would have confounded the performance trends of interest, common random-number sequences were used for all simulations. Either two or four separate sequences were used for each simulation (depending on whether train length was variable). The seeds for these separate sequences were chosen to be far enough apart that there would normally be no overlapping among them.

Antithetic sampling of the random-number sequence was built into the simulation as an option, but was not very extensively used. A set of pilot simulations was run to evaluate the efficiency of antithetic sampling. It was found that the use of antithetics for the train-length sampling was disadvantageous, producing an efficiency of less than unity (i.e., the negative correlation between the antithetic pair of runs was too small to produce a more efficient estimate of mean values than a single simulation run as long as the two antithetic runs combined). Antithetic sampling of the inter-arrival times produced an efficiency of about 1.2 at most, indicating that the queuing process at the merge was very effectively filtering out the differences in the arriving sequences of trains. Even though this efficiency estimate indicated some advantage to using antithetic sampling, the advantage was not significant enough to overcome the additional effort which would be involved in processing the results of twice as many separate simulations. The economics of the computer facility which was used strongly discouraged use of antithetics because of the fixed charge per log-in, the fixed charge per printout of a simulation (regardless of how long the simulation lasts, the same number of lines are printed), and the decreasing charge per unit of execution time as length of job increases. The efficiency of the antithetic method would have had to approach 3.0 to justify its use for this project.

The antithetic sampling method was used to help choose the length of simulation to run for each test case. Some pilot runs of 10,000 trains were simulated, with statistical outputs reported at regular intervals in the course of these simulations. The antithetic pairs of runs had very different output statistics after 2000 trains, but by the time 6000 had been simulated, they were extremely close to each other. Convergence of the results from the antithetic pair of simulations provided some assurance of the long-term stationarity of the simulation. It was apparent that even if other statistical considerations were satisfied for simulations of 2000 trains, the nonstationarity of the results would rule against the general use of such short simulations.

The output samples from a simulation of a congested merge junction are strongly autocorrelated, and for purposes of hypothesis testing, may represent only a relatively limited number of independent samples. The confidence intervals around the output measures of a simulation of 5000 trains may have to be determined on the basis of significantly less than 5000 independent samples if the correlation between consecutive samples is strong. The problems inherent in evaluating such simulation results were described at length in the paper by Fishman and Kiviat [21]. Fishman and Kiviat discussed the use of spectral analysis of the simulation record as a way of determining the "correlation time," a characteristic interval at which samples could be said to be independent. They preferred use of the spectral density function to the autocorrelation, but had to make strong assumptions about the nature of the correlation process to estimate a correlation time from the spectral density.

For purposes of the present study, a spectral analysis post-processor was developed to analyze the outputs of short (2000-train) simulations in an effort to estimate a correlation time. The simulation record was sampled at a frequency faster than the fastest change expected in merge conditions (to avoid aliasing), and was passed through a cosine "window" to eliminate spurious side lobes in the spectrum, which was calculated using a standard Fast Fourier Transform (FFT) method. The autocorrelation of each simulation record was calculated from the spectrum using the method outlined by Bendat and Piersol [22]. These autocorrelation functions were used to generate rough estimates of the correlation times of the processes, based on the lag time needed for the autocorrelation to approach zero. Some care was needed in performing this analysis, because the autocorrelation function is sensitive to the sample interval chosen.

The autocorrelation tests were performed for four baseline simulation cases, all using a triangular distribution of train lengths having a mean value of six vehicles and a maximum of 11. For the isocapacity merge, at a traffic intensity ρ_{sum} of 0.9, the correlation time was about 400 slot-passage units. This corresponds to about 1.5 times the maximum maneuver distance used, or five times the mean maneuver distance. For the same input conditions, but with the concatenating merge, the correlation time ranged from 50 to 85 slot passages, depending on sample intervals. These correspond roughly to the mean and maximum maneuver distances respectively. When the concatenating merge was forced to operate at a combined input traffic intensity ρ_{sum} of 1.8, correlation time increased to about 120. This was about 1.5 times the mean maneuver distance, or approximately at the 90 percentile level of maneuver distance. When the isocapacity merge was operated in light traffic, at a ρ_{sum} of 0.5, the correlation time was in the range of 100 to 150, again comparable to the maximum maneuver distance, or about twice the mean maneuver distance.

As expected, the more congested the merge the longer the correlation time, corresponding to longer "busy periods" during which delays to one train are propagated to the next train in the input sequence. The concatenating merges exhibit weaker correlation than the isocapacity merges by an amount comparable to their shorter maneuver distances at the same traffic intensities. The results of the correlation study reported here can be used to estimate the number of independent samples which can be assumed to be included in a 5000-train simulation under each of the baseline conditions, producing the results shown in Table 1.

Table 1
NUMBER OF INDEPENDENT SAMPLES IN BASELINE SIMULATIONS OF 5000 TRAINS

Merge Category	Combined Input Traffic Intensity	Independent Samples
Isocapacity	0.5	1200 – 1700
Isocapacity	0.9	235
Concatenating	0.9	1100 – 1900
Concatenating	1.8	400

The increased difficulty of simulating the more congested conditions is obvious.

The estimates of the number of independent samples listed here must be used with care. It is tempting to use these estimates, in conjunction with the simulation outputs of standard deviations of slots slipped and maneuver distances, to derive confidence intervals about the estimated mean values. If the output variables are not normally distributed, but are strongly skewed, a confidence interval based on the standard normal distribution would be misleading. Sometimes the output measures are close to being Gaussian, and at other times they are not, so no general guideline can be established here. Determination of confidence intervals for the upper tails of the output distributions would have required some special data-collection methods [23-25] because of the strong autocorrelation of the merge queuing process under congested conditions.

7. DIAGNOSTIC OUTPUT OPTIONS

A simulation of a complicated process such as the operation of a merge junction cannot be debugged unless comprehensive diagnostic outputs, which completely describe the operation of the simulated system, are provided. The complete diagnostics are needed only for debugging, being much too cumbersome and costly to use in full-scale production runs. The merge simulation has been provided with a combination of printed and graphical outputs for diagnostic purposes.

7.1 Detailed Printout of Events

The most powerful diagnostic tool is a printout which describes each simulated event on a separate line, including all relevant information about the maneuvering of the train involved. A sample section of this diagnostic is shown in Fig. 4. Obviously, the volume of information in such a printout for a complete simulation would be unmanageable, but careful inspection of only a small section of printout can reveal whether the simulation is consistently reflecting the merge behavior it is designed to represent.

7.2 Graphical Outputs

Selected output plots can concisely represent the behavior of a large number of trains passing through a merge junction. This has proven useful for fast debugging, by immediately revealing inconsistent, illogical or physically unrealistic output data points. These plots have also helped promote a more thorough understanding of the merging process and of the distinctions between the isocapacity and concatenating merges.

A sample graphical description of an isocapacity merge operating at a combined traffic intensity of $\rho_{sum} = 0.9$ is shown in Fig. 5. The time plot of maneuver distance shows the variability of merge conditions with time, and clearly illustrates the busy periods of the merge. The cross-plot of slot slips demonstrates the degree of correlation between the conditions on the two incoming lanes. This correlation is much stronger for the isocapacity merge shown than it is for the concatenating merge. The maneuver distance cross-plot, Fig. 5c, demonstrates similar correlation characteristics but looks somewhat different because the minimum nonzero maneuver distance (for a single slot slip) is 26 vehicle lengths. The cross-plot of maneuver distances and slot slips on the same lane, Fig. 5d, is a very useful consistency check, and also provides a very clear demonstration of the upstream propagation of merge delays (represented by data points scattered above the minimum maneuver distance frontier defined by the performance of the vehicle longitudinal control system [2,4]).

The detailed diagnostic plots were not generated for most simulated cases because of their expense and complexity. Most output performance measures were derived from printouts of statistics and percentiles, and printer histograms. The only graphical outputs which were regularly generated were cumulative distribution plots of slot slips and maneuver distances, such as those shown in Fig. 6. Percentile information can be derived from these (within the limits of plotter resolution), and the shapes of the distributions, which differ significantly for the isocapacity and concatenating merges, can be clearly seen.

8. CONCLUSIONS

This paper provides an overview of a simulation which has been developed to investigate the performance of merge junctions in dynamically entrained AGT systems. This simulation is more complicated than previously developed merge simulations because of its representation of variable-length trains, upstream propagation of maneuvers, and the special concatenating merge. It has been used to investigate the influences of many different parameter changes and merge operating conditions on merge performance (results discussed in Ref. 5, with further detail in Ref. 4). For example, the results derived using this simulation model have been applied to estimating the tradeoffs between lane capacity and needed maneuver ramp length for trains of various lengths using the isocapacity and concatenating merges.

Figure 4
SAMPLE DIAGNOSTIC PRINTOUT

```
T= 1605. TRAIN #  60 ENDS SLIP ON LANE 2 SCHED   MERGE AT T= 1610.
T= 1610. TRAIN #  60 MERGES FROM LANE 2 AFTER GAP OF    0. SLOTS
T= 1613. TRAIN #  87 ENTERS KSR 2 WITH L= 6 SCHED TO SLIP 11 SLOTS FROM INITIAL GAP= -5 SLOTS AT T= 2103. USING 61 OF MR
T= 1616. TRAIN #  83 ENTER MR FROM LANE 1 SCHED  MERGE AT T= 2066.
T= 1627. TRAIN #  61 STARTS SLIP ON LANE 2 SCHED END SLIP AT T= 1701.
T= 1627. TRAIN #  59 ENDS SLIP ON LANE 1 SCHED   MERGE AT T= 1627.
T= 1627. TRAIN #  59 MERGES FROM LANE 1 AFTER GAP OF    0. SLOTS
T= 1634. TRAIN #  84 ENTER MR FROM LANE 1 SCHED  MERGE AT T= 2084.
T= 1644. TRAIN #  89 ENTERS KSR 2 WITH L= 8 SCHED TO SLIP  0 SLOTS FROM INITIAL GAP=  1 SLOTS AT T=    0. USING  0 OF MR
T= 1647. TRAIN #  88 ENTERS KSR 1 WITH L= 3 SCHED TO SLIP 11 SLOTS FROM INITIAL GAP=-11 SLOTS AT T= 2140. USING 58 OF MR
T= 1655. TRAIN #  82 ENTER MR FROM LANE 2 SCHED  MERGE AT T= 2105.
T= 1680. TRAIN #  62 MERGES FROM LANE 1 AFTER GAP OF   36. SLOTS
T= 1681. TRAIN #  86 ENTER MR FROM LANE 2 SCHED  MERGE AT T= 2131.
T= 1681. TRAIN #  90 ENTERS KSR 2 WITH L= 5 SCHED TO SLIP  0 SLOTS FROM INITIAL GAP=  7 SLOTS AT T=    0. USING  0 OF MR
T= 1683. TRAIN #  91 ENTERS KSR 1 WITH L= 6 SCHED TO SLIP 15 SLOTS FROM INITIAL GAP=-15 SLOTS AT T= 2165. USING 69 OF MR
T= 1695. TRAIN #  85 ENTER MR FROM LANE 1 SCHED  SLIP AT T= 2102.
T= 1696. TRAIN #  65 STARTS SLIP ON LANE 2 SCHED END SLIP AT T= 1767.
T= 1701. TRAIN #  61 ENDS SLIP ON LANE 2 SCHED   MERGE AT T= 1701.
T= 1701. TRAIN #  61 MERGES FROM LANE 2 AFTER GAP OF    0. SLOTS
T= 1704. TRAIN #  66 STARTS SLIP ON LANE 1 SCHED END SLIP AT T= 1783.
T= 1705. TRAIN #  92 ENTERS KSR 2 WITH L= 4 SCHED TO SLIP  8 SLOTS FROM INITIAL GAP=  7 SLOTS AT T= 2203. USING 53 OF MR
T= 1707. TRAIN #  87 ENTER MR FROM LANE 2 SCHED  SLIP AT T= 2103.
T= 1708. TRAIN #  63 STARTS SLIP ON LANE 1 SCHED END SLIP AT T= 1750.
T= 1725. TRAIN #  67 STARTS SLIP ON LANE 2 SCHED END SLIP AT T= 1818.
T= 1729. TRAIN #  93 ENTERS KSR 1 WITH L= 3 SCHED TO SLIP  0 SLOTS FROM INITIAL GAP=  2 SLOTS AT T=    0. USING  0 OF MR
T= 1735. TRAIN #  64 MERGES FROM LANE 2 AFTER GAP OF   14. SLOTS
T= 1736. TRAIN #  89 ENTER MR FROM LANE 2 SCHED  MERGE AT T= 2186.
T= 1744. TRAIN #  88 ENTER MR FROM LANE 1 SCHED  SLIP AT T= 2140.
T= 1748. TRAIN #  69 STARTS SLIP ON LANE 1 SCHED END SLIP AT T= 1853.
T= 1749. TRAIN #  94 ENTERS KSR 2 WITH L= 9 SCHED TO SLIP  0 SLOTS FROM INITIAL GAP=  0 SLOTS AT T=    0. USING  0 OF MR
T= 1750. TRAIN #  68 STARTS SLIP ON LANE 1 SCHED END SLIP AT T= 1792.
T= 1750. TRAIN #  63 ENDS SLIP ON LANE 1 SCHED   MERGE AT T= 1750.
T= 1750. TRAIN #  63 MERGES FROM LANE 1 AFTER GAP OF    0. SLOTS
T= 1759. TRAIN #  70 STARTS SLIP ON LANE 2 SCHED END SLIP AT T= 1830.
T= 1767. TRAIN #  65 ENDS SLIP ON LANE 2 SCHED   MERGE AT T= 1767.
T= 1767. TRAIN #  65 MERGES FROM LANE 2 AFTER GAP OF    0. SLOTS
T= 1770. TRAIN #  71 STARTS SLIP ON LANE 2 SCHED END SLIP AT T= 1886.
T= 1776. TRAIN #  90 ENTER MR FROM LANE 2 SCHED  MERGE AT T= 2226.
T= 1776. TRAIN #  96 ENTERS KSR 2 WITH L= 2 SCHED TO SLIP  0 SLOTS FROM INITIAL GAP= 14 SLOTS AT T=    0. USING  0 OF MR
T= 1777. TRAIN #  91 ENTER MR FROM LANE 1 SCHED  SLIP AT T= 2165.
T= 1779. TRAIN #  72 STARTS SLIP ON LANE 1 SCHED END SLIP AT T= 1864.
T= 1783. TRAIN #  66 ENDS SLIP ON LANE 1 SCHED   MERGE AT T= 1783.
T= 1783. TRAIN #  66 MERGES FROM LANE 1 AFTER GAP OF    0. SLOTS
T= 1789. TRAIN #  74 STARTS SLIP ON LANE 2 SCHED END SLIP AT T= 1896.
T= 1791. TRAIN #  95 ENTERS KSR 1 WITH L= 8 SCHED TO SLIP  4 SLOTS FROM INITIAL GAP= -4 SLOTS AT T= 2295. USING 48 OF MR
T= 1792. TRAIN #  68 ENDS SLIP ON LANE 1 SCHED   MERGE AT T= 1803.
T= 1801. TRAIN #  92 ENTER MR FROM LANE 2 SCHED  SLIP AT T= 2203.
T= 1802. TRAIN #  75 STARTS SLIP ON LANE 2 SCHED END SLIP AT T= 1907.
T= 1803. TRAIN #  68 MERGES FROM LANE 1 AFTER GAP OF    0. SLOTS
T= 1818. TRAIN #  67 ENDS SLIP ON LANE 2 SCHED   MERGE AT T= 1818.
```

Figure 5
SAMPLE DIAGNOSTIC PLOTS FOR ISOCAPACITY MERGE AT $\rho_{sum} = 0.9$

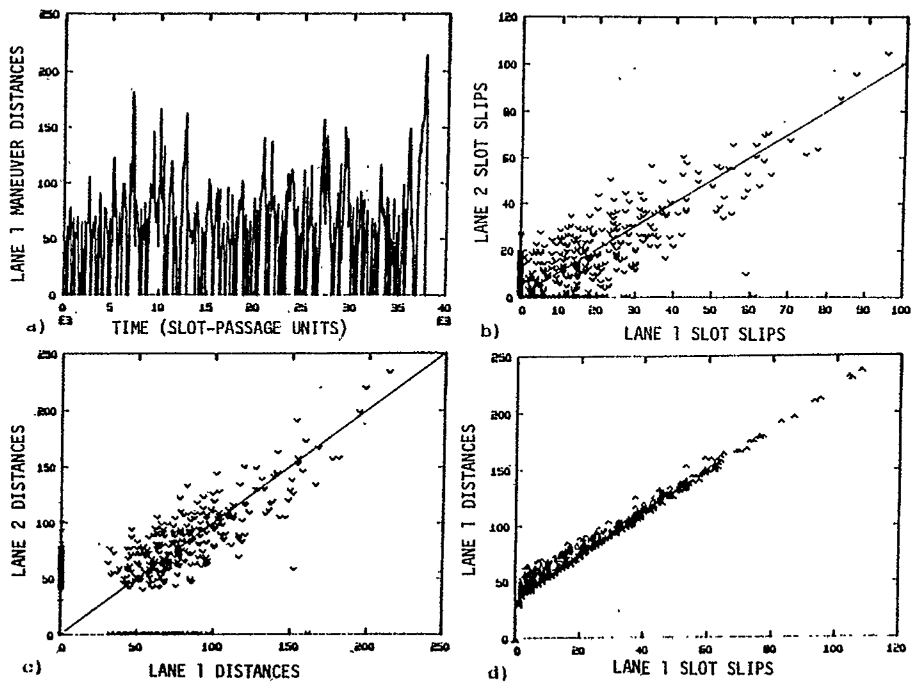


Figure 6
CUMULATIVE DISTRIBUTIONS OF SLOT SLIPS AND MANEUVER DISTANCES FOR CONGESTED ISOCAPACITY MERGE

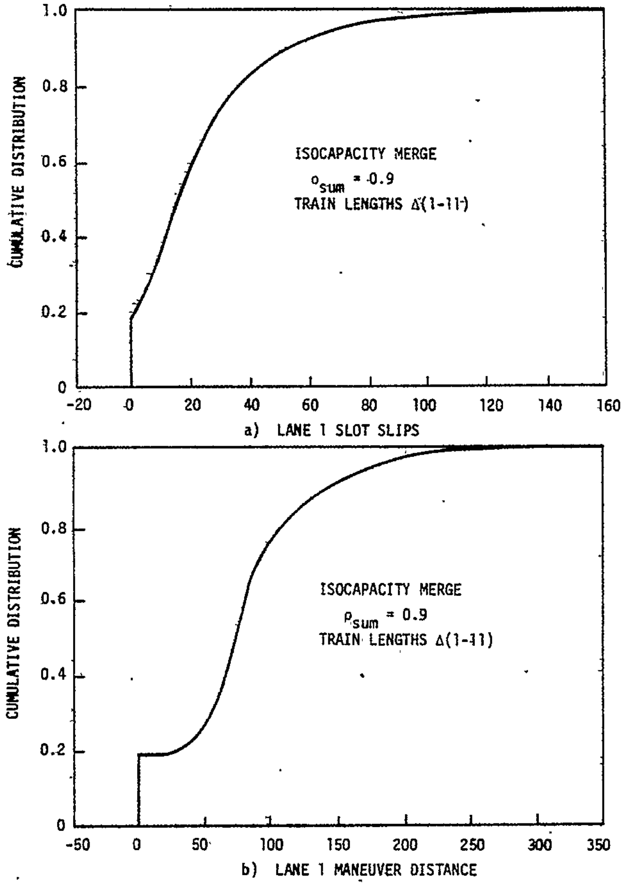
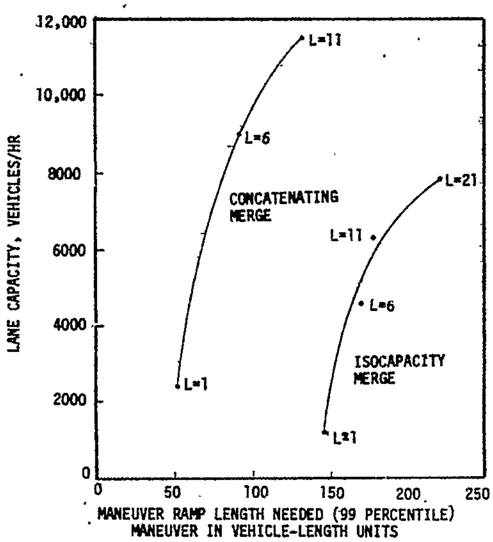


Figure 7
TRADEOFF BETWEEN LANE CAPACITY AND LENGTH OF MANEUVER RAMPS NEEDED



These tradeoffs, shown in Fig. 7, demonstrate the clear dominance of the concatenating merge, achieved by use of dynamic entrainment of vehicles.

The simulation which has been described here is useful for evaluating the performance of a single, isolated, merge junction. Conclusions about the design and performance of a network system should be based on detailed simulations of the entire network, with each merge represented by a model similar to that presented here. This level of detail is necessary for representing the significant effects of changes in the distributions (not just the mean values) of the gaps between trains. In a network environment, with interactions among consecutive merges, the assumptions of "perfectly" random train arrivals are no longer valid. This means that the results obtained by simulating the isolated merge, with "perfectly" random arrivals, cannot be extended directly to apply to a merge in a network.

ACKNOWLEDGMENTS

This work was performed while the author was a Lecturer and Research Assistant in the Dept. of Mechanical Engineering at M.I.T. Financial support was provided by the Urban Mass Transportation Administration under Contract DOT-UT-80013. Valuable advice and suggestions were offered by Dr. Marvin A. Sirbu and Professors Herbert H. Richardson and Nigel H.M. Wilson.

REFERENCES

1. Shladover, S.E., "Dynamic Entrainment of Automated Guideway Transit Vehicles," High Speed Ground Transportation Journal, Vol. 12, No. 3, Fall, 1978, pp.87-113.
2. Shladover, S.E., "Longitudinal Controller Design for Dynamic Entrainment and Platooning of AGT Vehicles," Proc. of 1978 International Conf. on Advanced Transit and Urban Revitalization, Advanced Transit Assoc., Indianapolis, April 1978.
3. Godfrey, M.B., Merging in Automated Transportation Systems, Sc.D. Thesis, Dept. of Mechanical Engineering, M.I.T., 1968.
4. Shladover, S.E., Operation of Automated Guideway Transit Vehicles in Dynamically Reconfigured Trains and Platoons, Sc.D. Thesis, Dept. of Mechanical Engineering, M.I.T., September 1978.
5. Shladover, S.E., "Operation of Merge Junctions in a Dynamically Entrained Automated Guideway Transit System," accepted for publication in Transportation Research, 1979.
6. Athans, M., "A Unified Approach to the Vehicle Merging Problem," Transportation Research, Vol. 3, 1969, pp.123-133.
7. Athans, M. and Levine, W.S., Optimal Vehicle Control for the Merging Problem, M.I.T. Electronic Systems Laboratory Report ESL-R-327, November 1967, PB 176 921.

8. Brown, S.J., Adaptive Merging Under Car-Follower Control, Johns Hopkins University Applied Physics Laboratory Report CP 038/TRP 029, October 1974.
9. Brown, S.J., "Traffic Merging Under Vehicle-Follower Control," Personal Rapid Transit III, Audio Visual Library Service, University of Minnesota, 1976, pp.257-268.
10. Whitney, D.E., "A Finite State Approach to Vehicle Merging," Journal of Dynamic Systems, Measurement and Control, Vol. 94, No. 2, June 1972, pp.147-151.
11. Sarachik, P.E. and Chu, K-C., "Real-Time Merging of High Speed Vehicular Strings," Transportation Science, Vol. 9, No. 2, May 1975, pp.122-138.
12. Boyd, R.K. and Lukas, M.P., "How to Run an Automated Transportation System," IEEE Trans. on Systems, Man, and Cybernetics, Vol. SMC-2, No. 3, 1972, pp.331-341.
13. Stefanek, R.G., "Network Effects on the Interface Problem for an Automated Highway System," Proc. of the Ninth Annual Allerton Conf. on Circuit and System Theory, University of Illinois, October 1971, pp.547-558.
14. Munson, A.V. and Travis, T., "Development Simulation of an Urban Transit System," Proc. of 1971 Summer Computer Simulation Conf., July 1971, Boston, pp.512-521.
15. Munson, A.V. et al., "Quasi-Synchronous Control of High Capacity PRT Networks," Personal Rapid Transit, Institute of Technology, University of Minnesota, 1972, pp.325-350.
16. Irving, J.H. et al., "Vehicle Management in Large PRT Networks," Personal Rapid Transit III, Audio Visual Library Service, University of Minnesota, 1976, pp.345-368.
17. Brown, S.J., "Merge Control in Automated Transit System Networks," ASME Paper 73-ICT-109, Intersociety Conf. on Transportation, 1973.
18. McGinley, F.J., "An Intersection Control Strategy for a Short-Headway PRT Network," Transportation Planning and Technology, Vol. 3, No. 1, 1975, pp.45-53.
19. McGinley, F.J., "A Survey of Quasi-Synchronous PRT Interchange Control Algorithms," Transportation Planning and Technology, Vol. 3, No. 3, 1977, pp.233-245.
20. Godfrey, M.B., Headway and Switching Strategies for Automated Vehicular Ground Transportation Systems, M.I.T. Engineering Projects Laboratory Report DSR 76106-4, PB 173 644, 1966.
21. Fishman, G.S. and Kiviat, P.J., "The Analysis of Simulation-Generated Time Series," Management Science, Vol. 13, No. 7, March 1967, pp.525-557.
22. Bendat, J.S. and Piersol, A.G., Random Data: Analysis and Measurement Procedures, New York, Wiley-Interscience, 1971.
23. Crane, M.A. and Iglehart, D.L., "Simulating Stable Stochastic Systems — I: General Multiserver Queues," Journal of the Association for Computing Machinery, Vol. 21, No. 1, January 1974, pp.103-113.
24. Crane, M.A. and Iglehart, D.L., "Simulating Stable Stochastic Systems — II: Markov Chains," Journal of the Association for Computing Machinery, Vol. 21, No. 1, January 1974, pp.114-123.
25. Crane, M.A. and Iglehart, D.L., "Simulating Stable Stochastic Systems — III: Regenerative Processes and Discrete-Event Simulations," Operations Research, Vol. 23, No. 1, January — February 1975, pp.33-45.