Practical Limits of Single-Track Light Rail Transit Operation

DUNCAN W. ALLEN

Increasing urban traffic congestion continues to stimulate interest in exclusive rights-of-way for new transit projects. Today's increased concern with cost-effectiveness, however, has focused attention on light rail transit (LRT) and other transitway technologies that are less capital-intensive than traditional heavy rail rapid transit. For these systems, planners have sometimes turned to single-track operation. Such operations have their limits, however. A planning-level method can identify whether single-tracking is appropriate for a particular application. The spacing and length of passing tracks depends on a number of factors, primarily the scheduled headway and the variability in vehicle travel time. Generalized design conditions, analogous to some levels of service can be considered in terms of maximum running times over single-track sections. For situations in which single-track operation is found to be feasible, the effects of additional practical considerations can be explored. Guidelines can help determine whether these practical considerations are likely to invalidate a solution originally determined to be feasible. Modern LRT and traditional street railways can also be compared in terms of the defined conditions.

Increasing urban traffic congestion continues to stimulate interest in exclusive rights-of-way for new transit projects, including high-speed commuter rail and rapid transit systems. Today's increased concern with cost-effectiveness, however, has focused attention on transitway technologies that are less capital-intensive than traditional forms. These include light rail transit (LRT), other guideway-based technologies, and busways. When the available right-of-way width is constrained by cost, physical obstacles, or other factors, planners have turned to single-lane or single-track operation to avoid the constraints. Single-track operation has its limits, however. A planning-level method can identify whether single-tracking is appropriate for a particular application.

SINGLE-TRACK APPLICABILITY AND PRINCIPLES

It should be noted at the outset that the use of railroad terminology is not intended to suggest that the techniques proposed are appropriate exclusively for rail vehicles. The terms are generally transferable to other guideway-based systems. The techniques are intended to be applicable to busways as well; when each "railroad" term is first used, a substitute term applicable to busways is either shown in parentheses, or a definition of the term applicable to bus operation is presented.

When traffic density is low enough, a single-track (lane) main line with appropriately located passing tracks (sections of two-lane roadway or two dedicated parallel single-lane roadways) can accommodate bidirectional operation with little or no delay. Vehicles running in opposing directions pass each other on double-track sections; these "meets" may require one or both vehicles to reduce speed or stop. As the frequency of operation increases, delays increase as well up to a point at which they become unacceptable, and a full double-track system is warranted.

Safe operation on single-track sections requires positive control of access. Railroad signaling technology has provided such control for decades, using the proven techniques of block signaling and interlocking logic. For a single main-line with passing tracks, control points at each end of each passing track are established to control access to the single track. The operation of signals and track switches is often electrically interlocked to ensure safe operation (e.g., to prevent the simultaneous display of signals to trains in opposing directions). Control points so equipped are generally referred to as interlockings.

The nearest common highway analogue to such operation is the use of traffic signals to control access to single-lane bridges, underpasses, and temporary work zones; in these cases, a single traffic signal controller ensures that conflicting signals are not displayed. Unlike railroad interlockings, however, these systems rely on the passage of time from the beginning of a red signal as the basis for an assumption that the single-lane section has cleared. The lack of a positive indication of block occupancy limits the applicability of this approach to sections that are entirely within line of sight from both control points.

The recent trend toward cost-effective rapid transit, however, has led to the development of bus presence detection technologies that can effectively function as signal systems. Elements of these technologies are already in service in a joint bus-LRT exclusive transit tunnel in Pittsburgh and in Germany. Busway planners should not, therefore, necessarily avoid single-lane sections.

ANALYTICAL MODEL OF SINGLE-TRACK OPERATION BETWEEN MEET POINTS

Techniques can be used to examine characteristics of a single-track section between specific meet points or to assess route-wide requirements.

Parsons De Leuw, Inc., Prudential Center, Boston, Mass. 02199.

Major Assumptions

The location and length of passing tracks depends on a number of factors. On North American railroads these generally include acceleration and braking characteristics; horizontal and vertical alignment; differences in train operation by direction; type of signal system; acceptable delays; train priorities; and train frequency. In a transit application, a number of simplifying circumstances are generally present, and were incorporated in this analytical model. They are as follows:

1. Vehicles have similar performance on successive trips in the same direction.
2. Service is scheduled on a fixed headway (i.e., the time interval between successive vehicles is constant).
3. All vehicles have the same priority.
4. Signal systems are optimized for the particular vehicles operating on the line.
5. Use of the single-track sections is on an alternating basis (i.e., successive occupancies of the block are by vehicles traveling in opposing directions).

Analytical Framework and Definitions

The analytical technique of the model is built around the concept of a "design early vehicle" and a "design late vehicle." The technique allows for estimation of passing track lengths for three basic design conditions. These are intended to approximate levels of service in the sense popularized by the 1965 and 1985 editions of the *Highway Capacity Manual.* These design conditions are as follows:

1. Condition B, under which there would be little or no delay to vehicles in either direction under normal operating conditions;
2. Condition C, under which some vehicles leaving double-track sections would be delayed, but few would be required to come to a complete stop; and
3. Condition E, under which all or most vehicles would experience delays waiting to enter single-track sections, but it is still possible to move the required traffic.

In the author's opinion, it is inappropriate to identify a design Condition A for systems employing single-track sections. Operating conditions analogous to highway Level of Service A (e.g., "individual users are virtually unaffected by the presence of others" and "extremely high freedom to maneuver" (*1*, pp. 1–3) can only be approached by an entirely double-track system.

Given the assumptions above, it is possible to analyze a single-track operation to determine what the primary factors are controlling the maximum single-track length. Figure 1 is a diagram of a typical bidirectional transit operation in what is usually called "stringline" form. With time on the horizontal axis and location on the vertical, the trajectories of individual vehicles in motion appear as lines or curves with non-zero slope; station dwell times have zero slope.

Figure 1 shows the on-time operation of two succeeding vehicles in each direction; the vehicles have been designated

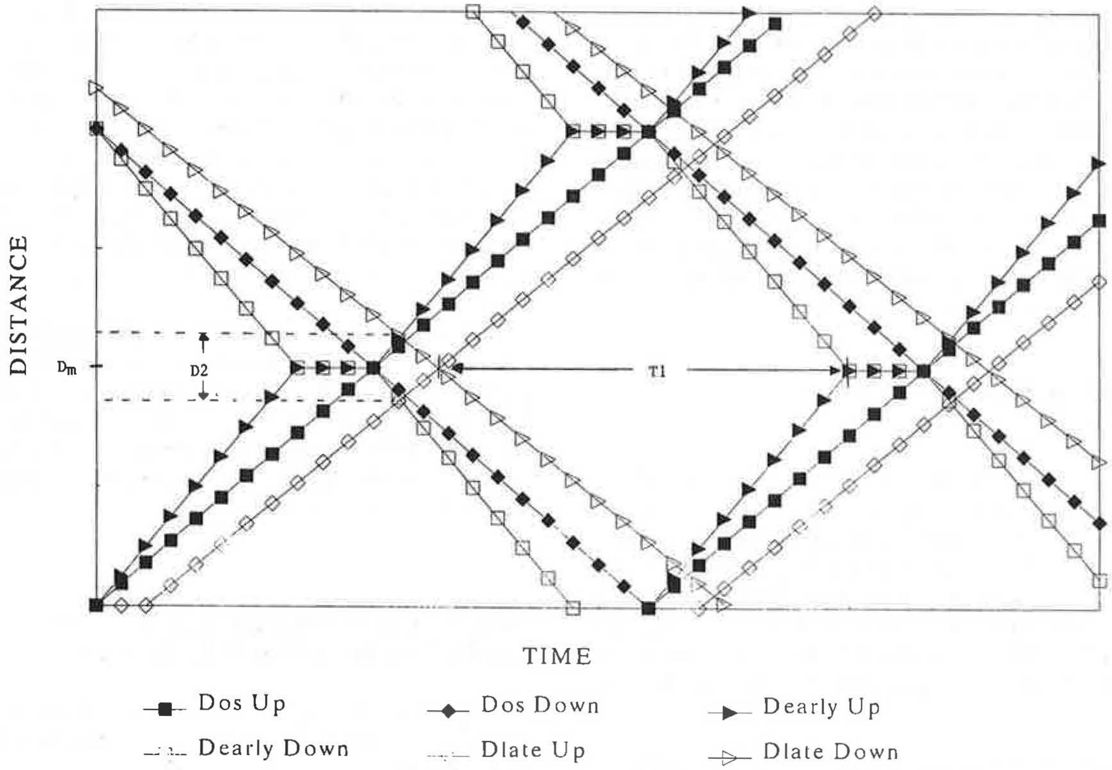

FIGURE 1 Stringline diagram.

"up" and "down," which will appear as superscripts to distinguish terms applicable to each direction. The stringline trajectories for the leading vehicles shown on this figure are defined as follows:

$Dos^{up}(t)$ = location of head (front end) of an on-time upbound vehicle at time t, and

$Dos^{down}(t)$ = location of head of an on-time downbound vehicle at time t.

The trajectories for the following trains are shown separated by the scheduled headway, H. Points at which the scheduled upbound and downbound trajectories intersect are the meet points, or centers of double-track sections. These are normally separated in time by one-half the scheduled headway (i.e., $H/2$).

Figure 1 also shows trajectories for design early and design late vehicles in each direction. These are the trajectories that govern the extent of double-tracking that should be provided. The double-track length, $D2$, around a meet point must be sufficient for the design early vehicles in each direction to pass each other on double track, and for the design late vehicles in each direction to do likewise. The scheduled running time over the single-track section, $T1$, is also shown.

To simplify the model, the following assumptions are made:

1. Trains arriving early at meet points will depart on time. Although this is not strictly true in all cases, most transit operations do have time points at which vehicles may be held to regain schedule. Some railroad operating rules actually prohibit early departure from stations.

2. The design late vehicle will accumulate lateness from the beginning of its trip to the meet point. Unlike an early vehicle, which can be returned to schedule by holding, it is relatively difficult for a late vehicle to recover schedule.

Single-Track Limitations—Conditions B and C

In Figure 1, the following trajectories are indicated:

$D^{up}_{early}(t)$ = design early trajectory for the head end of upbound vehicles,

$D^{up}_{late}(t)$ = design late trajectory for the head end of upbound vehicles,

$D^{down}_{early}(t)$ = design early trajectory for the head end of downbound vehicles, and

$D^{down}_{late}(t)$ = design late trajectory for the head end of downbound vehicles.

From Figure 1 it is now possible to describe the maximum allowable scheduled single-track running time. This is evaluated at the meet point (Dm on the vertical axis). The governing late vehicle is the one which passes the point later in time. The design lateness, Tlate, at the meet point is as follows:

$$T^{meet}_{late} = \max\,(T^{down}_{late},\ T^{up}_{late}) \tag{1}$$

where

$$D^{down}_{late}\,(T^{down}_{late}) = Dm \tag{2}$$

and

$$D^{up}_{late}\,(T^{up}_{late}) = Dm \tag{3}$$

Through a similar process, the design early time can be identified:

$$T^{meet}_{early} = \max\,(T^{down}_{early},\ T^{up}_{early}) \tag{4}$$

where

$$D^{down}_{early}\,(T^{down}_{early}) = Dm \tag{5}$$

and

$$D^{up}_{early}\,(T^{up}_{early}) = Dm \tag{6}$$

The above quantities are critical to the analysis. For convenience, their sum is designated as the critical time, T_{crit}:

$$T_{crit} = T^{meet}_{early} + T^{meet}_{late} \tag{7}$$

Two additional steps must be taken before defining solutions. First, because the trajectories defined are for the head ends of vehicles, a deduction from the remaining time must be made for passage of the early train:

$$T_{pass} = L/V \tag{8}$$

where L is the vehicle length in meters and V is the operating speed through the control points in meters per second.

A second allowance, T_{clear}, is required for operation of the signal system protecting the single-track section. The system must recognize that the block is clear, perform any conflict checks required, and display the signal for the opposing direction. If the operation is dispatched manually, an allowance must also be included for dispatchers who may not immediately recognize that display of a signal is required; this is usually only true for commuter or intercity railroad operation.

With the above adjustments, the time value for Condition C is obtained by subtraction as follows:

$$T^{C}_{1} = (H/2) - T_{crit} - T_{pass} - T_{clear} \tag{9}$$

Derivation of a value for Condition B requires an additional allowance of time for the following vehicle to decelerate to a stop from the authorized speed; this is denoted as T_{stop}. Therefore

$$T^{B}_{1} = (H/2) - T_{crit} - T_{pass} - T_{clear} - T_{stop} \tag{10}$$

The maximum single-track time T_1 can now be seen to be dependent primarily on both the scheduled trajectory of vehicles, the headway H, and the inherent variability in operating time, as represented by T_{crit}.

Estimation of T_{crit} for Conditions B and C

The model's technique for estimating T_{crit} is based on several underlying premises:

1. For any particular type of vehicle, there is a minimum "dead time" at stations, designated T_0, which is not available for loading passengers. This includes the time required for door operation, brake release, and so forth. This is supported both by data collected by the author (*2*) and the "lag time" referenced in the 1985 *Highway Capacity Manual* discussion of transit capacity (*1*, pp. 12–20).
2. The passenger service time is proportional to passengers boarding and alighting.
3. Passenger boarding times for individual passengers are statistically independent.
4. Passenger boarding times at successive stations are statistically independent.
5. Successive vehicle run times between stations are not statistically independent. At least two major mechanisms for dependence can be identified: first, operators are generally aware of whether they are early or late and can often take measures to compensate; second, transit schedulers often build in schedule recovery time to account for variations.
6. For successive trains, the distributions of both run times between stations and of station passenger service service time exhibit a characteristic asymmetrical probability distribution function. The following equation was estimated by a least squares fit to a cubic polynomial to generate simulated run or dwell times:

$$T_{sim} = T_{min} + T_{var} * (3.15*R - 6.2*R**2 + 5.75*R**3) \qquad (11)$$

where

T_{sim} = a randomly occurring value of a time to be simulated,
T_{min} = a minimum observed or possible value for the time to be simulated, presumed to be invariant (e.g., T_0),
T_{var} = a variable component of time, computed as the difference between the mean value of T_{sim} and the value of T_{min}, and
R = a randomly generated number with a uniform probability distribution, ranging between 0 and 1.0.

Based on the above premises, an expression for estimating the earliest likely arrival relative to the timetable, assuming an on-time departure from a point from which the timetable shows a travel time of T_{ref}, was constructed:

$$T_{early}(T_{ref}) = \{6.25*[Fd*T_{ref} - N*T_0] + 0.0004*[(1 - Fd)*T_{ref}]**2\}**0.5 \qquad (12)$$

where

T_{ref} = the scheduled travel time upstream of the meet point. For late vehicles, this is the total travel time from the upstream terminal; for early vehicles, it is one-half the headway (i.e., $H/2$),
N = number of station stops within scheduled travel time T_{ref} upstream of the meet point,
Fd = fraction of T_{ref} that is scheduled station dwell time, and
T_0 = station stop dead time for the particular equipment and station configuration under analysis.

Equation 12 implies that, for planning purposes, an amount equal to 1.7 times the early allowance would account for late operation not resulting from equipment failure or other major occurrences. Therefore

$$T_{late}(T_{ref}) = 1.7 * T_{early}(T_{ref}) \qquad (13)$$

Development of Basic Input Parameters

Application of the model described above for any proposed meet point requires the following information:

- H, the design headway,
- T_{clear}, the signal clearance time,
- T_{stop}, the vehicle stopping time (for Condition B),
- Fd, the scheduled station dwell time as a fraction of total scheduled time,
- T_0, the station stop dead time, and
- T_{pass}, the time required for an entire vehicle to pass a control point.

The design headway is strictly a user input. The values for T_{clear}, T_{stop}, T_0, and T_{pass} should, where possible, be derived from the actual operating characteristics (e.g., speed, braking characteristics, and vehicle lengths) of the proposed transit service. For planning-level feasibility assessments, however, an assumed value may be desired. Table 1 provides suggested typical values for these parameters, assuming commonly found operating characteristics for each of several modes. It is strongly recommended that values for *Fd* be specifically estimated for the particular operation under analysis; in the case of bus operations, considerable guidance is available from the *Highway Capacity Manual* (*1*, pp. 12–19); in principle, these techniques are also probably valid for LRT systems using low-platform stations. For high-platform stations and railroad op-

TABLE 1 Typical Model Parameters for Various Transit Modes

Parameter	Busway	LRT	HRT(1)	Railroad(2)
Tstop (seconds)	25	25	30	45
Tclear (seconds), manual dispatching	25	25	25	35
Tclear (seconds), automatic dispatching	6	8	8	10
To (seconds) "dead time"	3	5	8	15
Tpass (seconds)	1	3	10	10
Fd, ratio of dwell time to travel time (3)	0.25	0.25	0.30	0.20

(1) "heavy" rail rapid transit
(2) Diesel locomotive with passenger coaches
(3) Derivation of application-specific values for Fd is strongly recommended.

erations, dwell time is highly dependent on vehicle door configuration and other factors; the values for these modes in Table 1 represent approximate averages only, and should be used with caution.

Single Track Limitations—Condition E

In most cases it is possible to decrease the extent of double-tracking implied by Conditions B or C and continue to provide service on a fixed headway. There is, however, a maximum single-track occupancy time which can be scheduled for a given headway, assuming each trip through the single-track section follows one in the opposing direction. Values of T_1 above this limit, designated as Condition E, or capacity, will result in either accumulating delays or the need to "fleet" (i.e., dispatch more than one vehicle at a time in the same direction through single-track sections). This limit is as follows:

$$T_1^E = (H/2) - T_{\text{clear}} - T_{\text{pass}} \quad (14)$$

EXTENSION OF MODEL TO ROUTEWIDE ANALYSIS

Assumptions and Methods

The model described above can be extended to complete lines or routes by suitably defining the variables to be representative of an average or typical condition. To do this, the following assumptions were made:

1. For a given mode, the location-specific default values from Table 1 for the model parameters would apply.
2. The value of T_{crit} will, on average, grow as the square root of distance from the terminal. This implies that the value any point will be the larger of two values (see Equation 1), one proportional to the square root of the value from each terminal.
3. T_{ref} for design early trains will be $H/2$.

An electronic spreadsheet was developed to calculate location-specific values for T_{crit} according to Equations 1 through 7, and was then exercised for a wide range of relative locations along routes of various length for each set of modal parameters. Functions proportional to the square root of relative location were then fitted to the results via linear regression. Based on the coefficients of these functions, standard curves for various values of *Fd* were developed, as well as mode-specific adjustment factors.

Application of Routewide Technique

The most convenient form for using these results requires successive use of two sets of curves, or nomographs. The first set, appearing as Figure 2, represents a typical value for T_{late} (see Equation 1). This value depends primarily on the one-way scheduled travel time along the line (T_{ref}) and the routewide average dwell time ratio (*Fd*). A mode-specific adjustment factor is also provided. A second curve, appearing as Figure 3, allows the effect of the headway H on T_{early} (as specified in Equation 4) to be considered, and therefore the effect on T_{crit}. This curve provides a factor by which T_{late} can be multiplied to provide a value for T_{crit}.

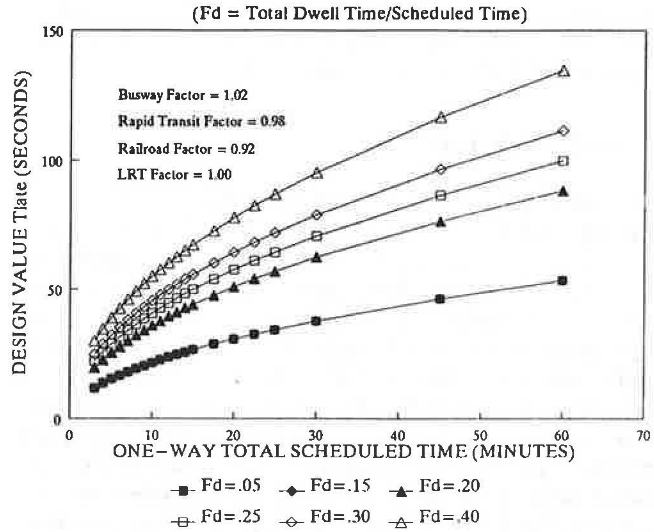

FIGURE 2 **Routewide assumptions for T_{late}.**

Once T_{crit} is determined, the equations governing T_1 for Conditions B and C (Equations 10 and 9, respectively) can be applied. Equation 14 remains applicable for Condition E, and does not require use of the nomographs.

VALIDATION AND COMPARISON WITH ACTUAL SYSTEMS

The techniques described have been compared with both present and historical North American rail systems. Table 2 com-

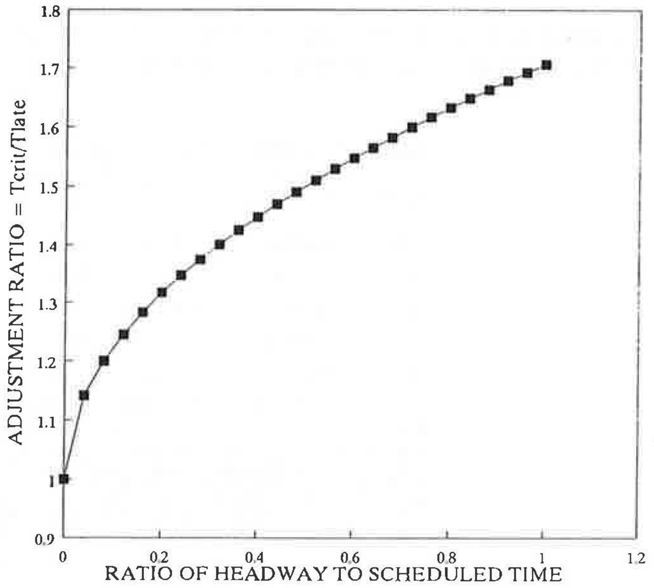

FIGURE 3 **Estimation of T_{crit} from T_{late}.**

pares key parameters and T_1 values, derived for selected current systems with single-track sections, using the routewide method described above, against actual values.

Segment-Specific Data

The data in Table 2 suggest that the technique described in this paper is indicative of the actual performance of the routes or branches represented. Operation of the LRT starter line in Sacramento and the Needham commuter rail branch line are generally satisfactory for the scheduled headways; these systems are indicated as being in the range of Conditions B and C. The Media and Sharon Hill LRT lines (in metropolitan Philadelphia) were studied by Transportation and Distribution Associates (TAD), Inc., in 1987. TAD's report (*3*) concluded that the single-track operation on Sharon Hill (approximately Condition C according to the model) was acceptable, but that the Media Line (approximately Condition E) required improvement. Pittsburgh's single-track 2-km Drake extension, which is similar to the Philadelphia lines in that the single-track section lies entirely at the outer end of a route, is scheduled at less than capacity. Of some interest is the Overbrook segment of Pittsburgh's South Hills LRT, which operates a section of single-track midroute with frequently spaced short passing tracks. As of 1987 the Port Authority Transit (PAT) was actually operating this segment over its capacity (i.e., Condition E) by "fleeting" two or three trains at a time in one direction.

System or Routewide Data

A second indication of the general validity of the model is provided in Figure 4. In this graph, the vertical axis represents the ratio of physical track kilometers to route kilometers for a particular light rail system or route. The horizontal axis represents the actual or estimated average number of trains per hour per direction. Six sets of information are shown on Figure 4:

1. Points represent those North American LRT routes operating with at least some single-track as of 1989, according to a recent survey conducted by the Institute of Transportation Engineers (*4*). They are labeled individually.
2. Points represent *systemwide* average values for street railway companies operating in Massachusetts in 1916, according to a Public Service Commission report (*5*). Average frequencies were estimated based on reported fleet size and typical operating speeds.
3. Values represent Condition B, according to the model, from the default parameters assuming that the fraction of single-track length would be proportional to the fraction of single-track time.
4. Values represent Condition C under assumptions similar to No. 3 above.
5. Values represent Condition E under assumptions similar to No. 3 above.
6. A straight line represents a fit to the 1916 data.

Several interesting conclusions can be drawn from Figure 4: First, on a routewide basis, all the recently constructed LRT systems with single-track sections appear to be able to meet Condition B. Second, with the exception of PAT's Drake extension, all existing single-track operations appear to at least meet Condition C. Third, the points representing 1916 operations suggest that the private companies of that era designed to a fairly consistent practice lying somewhere between Conditions C and E. This might almost be regarded as a de facto condition D.

Condition D—Historical/Empirical

Based on the discussion in the previous two sections, it may be appropriate to add the concept of Condition D to the set of conditions that can be evaluated. This is probably a more realistic practical upper limit than E, even though it cannot be directly derived from the analytical method. Given the statistical fit to historical data, however, it can be estimated once the other T_1 values have been calculated according to the procedures described above:

$$T_1^D = 0.66 * T_1^C + 0.39 * T_1^E \tag{15}$$

TABLE 2 Comparison of Actual Designs and Model Results

System	Line	Tref	H	Fd	Tcrit	T1/"B"	T1/"C"	T1/"E"	Actual T1's
MBTA(1)	Needham	2400	1500	0.21	104	542	582	686	555
SEPTA(2)	Sharon Hill	1330*	900	0.20*	85	334	354	439	360
SEPTA(2)	Media	1440	900	0.22	90	329	349	439	420
SDTA(3)	Starter	3240	900	0.30	145	269	294	439	240,285,310,325,325
PAT(4)	Drake	2340	1260	0.25**	90	504	529	619	240
PAT(4)	Overbrook	2400	240	0.25**	100	(16)	9	109	N/A

(1) Based on field observations by author
(2) Data from Media/Sharon Hill Productivity Study, Transportation and Distribution Associates, Inc., for Southeastern Pennsylvania Transportation Authority (SEPTA), April 1987.
(3) From "Light Rail Transit X-T diagram" dated March 7, 1983, for Sacramento Transit Development Agency, by Foster Engineering, Inc.
(4) Port Authority Transit (Pittsburgh, PA) public timetables effective 1987.

* Estimated value based on station spacing vs. Media
** based on default values from Table 1.