3. Train Control and Signaling

3.1 INTRODUCTION

Signaling has been a feature of urban rail transit from the earliest days. Its function is to safely separate trains from each other. This includes both a separation between following trains and the protection of specific paths through junctions and cross-overs. The facilities that create and protect these paths or routes are known as interlockings.

Additional functions have been added to basic signaling, starting, again from a very early date, with automatic train stops. These apply the brakes should a train run through a stop signal. Speed control can also be added, usually to protect approaches to junctions (turnouts), sharp curves between stations and approaches to terminal stations where tracks end at a solid wall. Automatic trains stops are in universal use. Speed control is a more recent and less common application, often introduced in conjunction with automatic train control or to meet specific safety concerns.

Rail transit signaling is a very conservative field maintaining high levels of safety based on brick-wall stops and fail-safe principles. A brick-wall stop means that the signaling separation protects a train even if it were to stop dead, an unlikely though possible event should a train derail and strike a structure. This protection allows for a) the following train's failure to observe a stop signal, b) driver and equipment reaction time, and c) some impairment in the braking rate.

Fail-safe design principles ensure that failure of single—and often multiple—components should never allow an unsafe event. Traditionally in North America this involves the use of heavy railroad style relays that open by gravity and have nonwelding carbon contacts. Compact, spring opening, European-style relays or solid state (electronic or computer controlled) interlockings are now being accepted. Here equivalent safety is provided by additional logic, duplicate contacts or multiple polling processors.

The rigor with which fail-safe principles have been applied to rail transit has resulted in an exceptional safety record. However, the safety principles do not protect against all possibilities—for example, a derailed train could interfere with the safe passage of a train on an adjacent parallel track. Nor do they protect against all possible human errors whether caused by a signal maintainer, dispatcher or train driver. An increasing inability to control the human element—responsible for three-quarters of rail transit accidents or incidents[1]—has resulted in new train control systems using technology or automation to reduce or remove the possibility of human error.

Train control, or more properly automatic train control, adds further features to basic signaling. Automatic train control is an ill-defined term but usually encompasses three levels:

[1] PARKINSON, TOM, Safety Issues Associated with the Implementation of ATCS-Type Systems, *Transportation Development Centre, Transport Canada, August 1989.*

- Automatic train protection (ATP)
- Automatic train control[2] (ATC or ATO)
- Automatic train supervision (ATS)

Automatic train protection is the basic separation of trains and protection at interlockings. In other words, the signaling system as described above.

Automatic train control adds speed control and often automatic train operation. This can extend to automatically driven trains but more commonly includes a driver, operator or attendant who controls the train doors and observes the track ahead.

Automatic train supervision attempts to regulate train service. It can be an integral feature of automatic train control or an addon system. The capabilities of automatic train supervision vary widely from little more than a system that reports the location of trains to a central control office, to an intelligent system that automatically adjusts the performance and stop times of trains to maintain either a timetable or an even headway spacing.

Automatic train protection and automatic train control maintain the fail-safe principles of signaling and are referred to as *vital* or *safety critical* systems. Automatic train supervision cannot override the safety features of these two systems, and so it is not a vital system.

This chapter describes and compares the separation capabilities of various train control systems used on or being developed for rail transit. It is applicable to the main rail transit grouping of electrically propelled, multiple-unit, grade-separated systems. Specific details of train control for commuter rail and light rail modes are contained in the chapters dealing with these modes.

These descriptions cannot include all the complexities and nuances of train control and signaling but are limited to their effect on capacity. More details can be found in the references and in the bibliography. All urban rail transit train control systems are based on dividing the track into blocks and ensuring that trains are separated by a suitable and safe number of blocks. Train control systems are then broken down into fixed-block and moving-block signaling systems.

3.2 FIXED-BLOCK SYSTEMS

In a fixed-block system, trains are detected by the wheels and axles of a train shorting a low-voltage current inserted into the rails. The rails are electrically divided into blocks. Originally this required a rail to be cut and an insulating joint inserted. Only one rail is so divided. The other rail remains continuous to handle the traction power return.

[2] Sometimes termed automatic train operation to avoid confusion with the overall term automatic train control.

By moving from direct current to alternating current circuits,[3] the blocks can be divided by an inductive shunt[4] connected across the rails, avoiding the need for insulated joints. These are called jointless track circuits and both rails are then available for traction power return. A track circuit can be any reasonable length. Each circuit is expensive so lines use the minimum required for appropriate headways. Circuits will be short where trains must be close together, for example in a station approach, and can be longer between stations where trains operate at speed.

The signaling system knows the position of a train only by the relatively coarse measure of block occupancy. It does not know the position of the train within the block; it may have only a fraction of the train, front or rear, within the block. At block boundaries, the train will occupy two blocks simultaneously for a short time.

In the simplest two-aspect block system, the signals display only stop (red) or go (green). A minimum of two empty blocks must separate trains, and these blocks must be long enough for the braking distance plus a safety distance. The safety distance can include several components, including sighting distances, driver and equipment reaction times, and an allowance for partial brake failure, i.e. a lower braking rate.

Automatic train stops have long been a feature of rail transit (almost from the turn of the century). These prevent a train running through a red signal by automatically applying the emergency brakes should the driver ignore a signal. Called a *trip stop*, the system consists of a short mechanical arm beside the outer running rail that is pneumatically or electrically raised when the adjacent signal shows a stop aspect. If a train runs through this signal, the raised arm strikes and actuates a trip cock on the train that evacuates the main air brake pipe. Full emergency braking is then applied along the length of the train. To reset the trip cock the driver must usually climb down to track side and manually close the air valve.[5]

A two-aspect signaling system does not provide the capacity normally required on busy rail transit lines—those with trains an hour or better. Increased capacity can be obtained from multiple aspects where intermediate signals advise the driver of the condition of the signal ahead, so allowing a speed reduction before approaching a stop signal. Block lengths can be reduced relative to the lower speed, providing increased capacity.

The increased number of blocks, and their associated relay controls and color-light signals, is expensive. There is a diminishing capacity return from increasing the number of blocks and aspects as shown in Figure 3.1. This figure also shows that there is an optimal speed to maximize capacity. Between stations the line capacity is greatest with maximum running speeds of between 40 km/h (25 mph) with three aspects to 55 km/h (34 mph) with 10 aspects. At the station entry—invariably the critical point for maximum throughput—optimal approach speeds are from 25 km/h (15 mph) to 35 km/h (22 mph).

In North America, the most common block signaling arrangement uses three aspects. In Europe and Japan, a small number of systems extend to four or five aspects.

Optimizing a fixed-block system is a fine art, with respect both to block lengths and to boundaries. Block lengths are also influenced by grades because a train's braking distance increases on a down grade and vice-versa. Grades down into a station and curves or special work with significant speed restrictions, below the optimal levels given above, will reduce throughput and so reduce capacity. Fortuitously, one useful design feature of below-grade systems is a gravity-assisted profile. Here the stations are higher than the general level of the running tunnel. Trains use gravity to reduce their braking requirements in the station approach and to assist them accelerating away from the stations. This not only reduces energy consumption, equipment wear and tear and tunnel heating, but also reduces station costs because they are closer to the surface, allowing escalators and elevators to be shorter. More important to this study, it increases train throughput—altogether a good thing.

Requiring a train operator to control a train's speed and commence braking according to multiple aspect color-light signaling requires considerable precision to maximize throughput. Coupled with the expense of increasing the number of aspects an improvement has been developed over the past three decades—cab signaling.

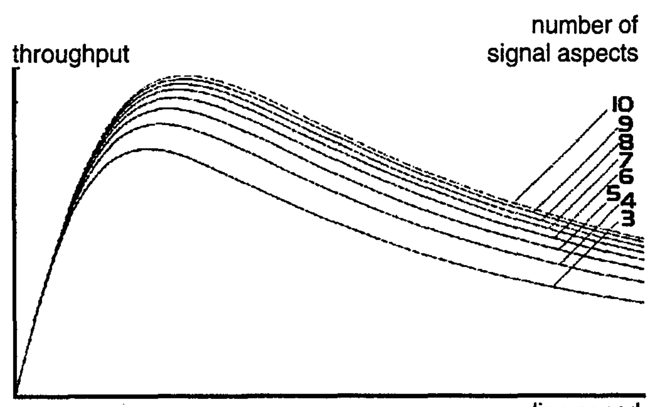


Figure 3.1 Throughput versus number of signal aspects(R26)

[3] Alternating current track circuits use different frequencies, combinations of frequencies or modulated frequencies. In all cases care must be taken to avoid interference from on-board vehicle equipment. Modern high power chopper and VVVF (variable voltage, variable frequency) three phase ac motor control equipment can emit considerable levels of EMI (electro magnetic interference). The systems engineering to coordinate and avoid such interference is difficult and complex and is beyond the scope of this report.

[4] In essence, the shunt shorts the small alternating current track circuits while presenting a low resistance to the high direct currents.

[5] Resetting the trip cock is understandably an unpopular task and consumes time. Consequently drivers may approach a trip cock cautiously at less than the optimal speed, particularly when closely following another train. In this case they expect the signal aspect to change as they approach but cannot be certain. Automatically driven trains will typically operate closer to the optimal speeds and braking rates and so can increase throughput.

There are times when it is operationally desirable to operate through a stop signal and its associated automatic train stop, particularly when the train ahead is delayed in a station and following trains wish to close up to expedite their subsequent entry to the station. The process is commonly called *key by* from an arrangement where the driver must lean out of the cab and insert a key in an adjacent electrical switch. However, the most common arrangement no longer involves a key, merely a slow movement of the train into the next block, which lowers the trip stop before it is struck by the train. The train must then proceed on visual rules toward the train ahead. In recent years an increase in the number of incidents caused by this useful, time saving, but not fail-safe, procedure has caused several systems to prohibit or restrict its use.

3.2.1 CAB SIGNALING

Cab signaling uses a.c. track circuits such that a code is inserted into each circuit and detected by an antenna on each train. The code specifies the maximum allowable speed for the block occupied and may be termed the *reference* or *authorized* speed. This speed is displayed in the driver's cab—typically on a dual concentric speedometer, or a bar graph where the authorized speed and actual speed can be seen together.

The authorized speed can change while a train is in a block as the train ahead proceeds. Compared to color-light signals, the driver can more easily adjust train speed close to the optimum and has less concern about overrunning a trip stop. Problems with signal visibility on curves and in inclement weather are reduced or eliminated.

Cab signaling avoids much of the high capital and maintenance costs of multiple-aspect color-light signals, although it is prudent and usual to leave signals at interlockings and occasionally on the final approach to and exit from each station. In some situations, dwarf color-light signals can be used. In this way trains or maintenance vehicles that are not equipped with cab signaling—or trains with defective cab signaling—can continue to operate, albeit at reduced throughput.

Reducing the number of color-light signals makes it economically feasible to increase the number of aspects and it is typical, although not universal, to have the equivalent of five aspects on a cab-signaling system. A typical selection of *reference* speeds would be 80, 70, 50, 35 and 0 km/h (50, 43, 31, 22 and 0 mph).

Signal engineers may argue over the merits of block-signaling and cab-signaling equipment from various manufacturers—particularly with respect to capital and maintenance costs, modular designs, plug versus hard-wired connections and the computer simulation available from each maker to optimize system design. However, for a given specification, the throughput capabilities vary little provided that—the signaling is optimized as to block length, boundary positioning and, when applicable, the selection of *reference* speeds. Consequently a listing or description of different systems is not relevant to capacity determination.

3.3 MOVING-BLOCK SIGNALING SYSTEMS

Moving-block signaling systems are also called transmission-based or communication-based signaling systems—potentially misleading because cab signaling is also transmission based.

A moving-block signaling system can be likened to a fixed-block system with very small blocks and a large number of aspects. Several analytic approaches to moving-block systems use this analogy. However a moving-block signaling system has neither blocks nor aspects. The system is based on a continuous or frequent calculation of the clear (safe) distance ahead of each train and then relaying the appropriate speed, braking or acceleration rate to each train.

This requires a continuous or frequent two-way communication with each train, and a precise knowledge of a train's location, speed and length; and fixed details of the line—curves, grades, interlockings and stations. These may be contained in a table that allows changes to be made without the normal full rigor required for changes to safety-critical software. Temporary changes can be easily made to add speed restrictions or close off a section of track for maintenance work.

Based on this information, a computer can calculate the next stopping point of each train—often referred to as the target point—and command the train to brake, accelerate or coast accordingly. The target point will be based on the normal braking distance for that train plus a safety distance.

Safety Distance Braking distance is a readily determined or calculated figure for any system. The safety distance is less tangible because it includes a calculated component adjusted by agency policy. In certain systems this distance is fixed; however, the maximum throughput is obtained by varying the safety distance with speed and location—and, where different types of equipment are operated, by equipment type.

In theory, the safety distance is the maximum distance a train can travel after it has failed to act on a brake command before automatic override (or overspeed) systems implement emergency braking. Factors in this calculation include

- system reaction time;
- brake actuation time;
- speed;
- train load (mass)—including any ice and snow load;
- grade;
- maximum tail winds (if applicable);
- emergency braking rate;
- normal braking rate;
- train to track adhesion; and
- an allowance for partial failure of the braking system.

The safety distance is frequently referred to as the "worst-case" braking distance, but this terminology is misleading. The truly worst case would be a total braking failure. Worst case implies reasonable failure situations, and total brake failure is not regarded as a realistic scenario on modern rail transit equipment that has multiple braking systems. A typical interpretation of the safety distance assumes that the braking system is three-quarters effective.

Train Position and Communication Without track circuits to determine block occupancy, a moving-block signaling system must have an independent method to accurately locate the position of the front of a train, then use look-up tables to calculate its end position from the length associated with that particular train's identification. The first moving-block systems, developed in Germany, France and the United States, all used the same principle—a wire laid alongside or between the running rails periodically transposed from side-to-side, the zigzag or Grecian square arrangement. The wire also serves to transmit signals to and from antennas on the train.

The wayside wires are arranged in loops so that each train entering a loop has a precise position. Within the loop, the control system counts the number of transpositions traversed, each a fixed distance apart— m (82 ft) is typical although much shorter distances have been used. Between the transpositions, distance is measured with a tachometer.[6]

The resultant positioning accuracy can be in the order of centimeters and with frequent braking rate feedback can result in station stop accuracy within ± 20 cm (8 in.) or better.

The use of exposed wayside wires is abhorred by maintenance-of-way engineers, and recent developments portend changes to existing systems and for the many moving-block signaling systems now under development. Inert transponders can be located periodically along the track. These require neither power nor communication wiring. They are interrogated by a radio signal from each train and return a discrete location code. Positioning between transponders again relies on the use of a tachometer. Moving-block signaling systems already have significantly lower costs for wayside equipment than do fixed-block systems, and this arrangement further reduces this cost as well as the occupancy time required to install or retrofit the equipment—an often critical factor in resignaling existing systems.

Removing the positioning and communicating wire from the wayside requires an alternate communication system. This can most economically be provided by a radio system using over-the-air transmission, wayside radiating cables, intermittent beacons or a combination thereof.

As with any radio system, interruption or interference with communications can occur and must be accommodated. After the central control computer has determined any control action, it will transmit instructions to a specific train using the identification number of the train's communication system. It is clearly vital that these instructions are received by and only by the train they were determined for.

There are numerous protocols and/or procedures that provide a high level of security on communication systems. The data transmission can contain both destination codes and error codes. A transmission can be received and repeated back to the source to verify both correct reception and correct destination, a similar process to radio train order dispatching. If a train does not receive a correctly coded confirmation or command within a set time, the emergency brakes will be automatically applied. The distance a train may travel in this time interval—typically less than 3 sec—is a factor in the safety distance.

Data Processing The computers that calculate and control a moving-block signaling system can be located on each train, at a central control office, dispersed along the wayside or a combination of these. The most common arrangement is a combination of on-board and central control office locations.

The first moving-block signaling systems used mainframe computers with a complex interconnection system that provided high levels of reliability. There is now a move toward the use of much less expensive and space-consuming personal computers (PCs).

PCs and their local area networks (LANs) have been regarded as less robust than mainframe systems, and as suspect for use in safety-critical applications. The first major application occurred in Vancouver in 1994 when, after 10 years of mainframe operation, the entire SkyTrain train control system was changed to operating on PCs with Intel 486 CPUs. Reliability has increased in the subsequent 15 months of operation. However, it is not possible to attribute this improvement solely to the new hardware because new software was also required by the change in operating systems. The proprietary computers and software on each train were not changed.

Safety Issues Safety on rail transit is a relative matter. It encompasses all aspects of design, maintenance and operations. In fixed-block signaling, electrical interlockings, switch and signal setting are controlled by relay logic. A rigorous discipline has been built around this long established technology which the use of processor-based controls is now infiltrating.

A moving-block signaling system is inherently processor controlled. Processor-based train control systems intrinsically cannot meet the fail-safe conventions of traditional signaling. Computers, microprocessors and solid-state components have multiple failure opportunities and cannot be analyzed and tested in the same way as conventional equipment.

Instead, an equivalent level of safety is provided on the basis of statistical failure modes of the equipment. Failure analysis is not an exact science. Although not all failure modes can be determined, the statistical probability of an unsafe event[7] can be predicted.

Determining failure probability is part of a safety assurance plan—a systematic and integrated series of performance, verification, audit, and review activities, including operations, maintenance and management activities that are implemented to assure safe and satisfactory performance. The plan can cover a specific area, such as software, or can encompass the entire system, where software would be but one aspect. Such a plan will usually include a fault tree analysis.

The typical goal in designing processor-based systems is a mean time between unsafe failures of 10^9 hours, or some 114,000 years.[8] After due allowance for statistical errors and the incorporation of a large safety margin, this is deemed to be equivalent to or better than the so-called fail-safe conventional equipment.

The possibility of even a low incidence of unsafe failure may give cause for concern and the acceptance of processor-based signaling, particularly moving-block systems, has been slow. However the safety of conventional rail transit signaling is not as absolute as is often made out. Minor maintenance errors can cause unsafe events. An estimated three-quarters of rail transit accidents are attributed to human error.[9]

Two methods are used to achieve the high levels of safety on processor-based control systems. One is based on redundancy, where two or more computers operate with the same software. The output of both or the output of at least two out of three

[6] Tachometer accuracy is helped by the ability for continual on-the-fly calibrations as the distance between each transposition is fixed and known. This fully compensates for wheel wear but not for slip or slide. Errors so caused, while small, can be minimized by the use of current sophisticated slip-slide control or, where feasible, placing the tachometer on an unmotored axle.

[7] An unsafe event may be referred to as a *wrong-side* failure.

[8] PARKINSON, TOM, Safety Issues Associated with the Implementation of ATCS-Type Systems, *Transportation Development Centre, Transport Canada August 1989.*

[9] Ibid.

must coincide before a comparator circuit transmits a command. Thereafter, the safety consequences of the output can be considered in a conventional fashion. This method is a hardware-intensive solution.

The other method is based on diversity. Two sets of software, created and verified by independent teams, are run on the same or separate computers. Again their output must agree before any commands are executed. This is a software-intensive solution.

Because software development can account for over half the cost of a moving-block signaling system, and with hardware costs declining—particularly with the use of PCs—the hardware-intensive approach to redundancy is invariably the most economic. However, the relative cost of software development, testing, commissioning and safety assessment is expected to drop with the introduction of modular code blocks—safety critical portions of software that remain unchanged from system to system.

In some regards, software-based systems, once fully tested and commissioned, are less prone to unsafe errors created during equipment installation and maintenance. However there are three major remaining areas of concern.

1. Revisions to software may be required from time to time and can escape the full rigor of a safety assurance plan.
2. Removing track circuits also removes broken rail detection. While no specific data for rail transit have been found, the Southern Pacific Railroad found that fewer than 2 percent[10] of broken rails were detected *in advance* by track circuits—it appears that most breaks occur from the stress of a train passing. Nevertheless, some moving-block signaling systems have long track circuits added to detect broken rails.
3. Removing track circuits also eliminates the detection of any and all vehicles whose wheels and axles short across the rails. A major hazard exists if maintenance vehicles, or a train with a defective train control system, enter into or remain in an area where automatically controlled trains are run. This requires a rigorous application of operating rules and requires the defect correction and reentry into the control system or removal of an automatic train protection failed train, before service can resume in the occupied area.

   This potential hazard can be reduced by adding axle counters at various locations. These count entry and exit into a specified track section. In conjunction with appropriate software, they will prevent an automated train from following an unequipped train at an unsafe distance. However, an unequipped train is not so protected but depends on the driver obeying rules, whether using line-of-sight operation, or depending on any remaining wayside signals.

Hybrid Systems There are times when an urban rail transit system shares tracks with other services, such as long distance trains, whose equipment is impractical or uneconomic to equip with the moving-block signaling system. Use of axle counters for the safety of unequipped rolling stock substantially reduces capacity. To avoid this reduction while still obtaining the close headway of the moving-block system for the urban or short distance trains requires a hybrid design.

The SACEM system developed by Matra is employed in Paris and Mexico City[11] The SACEM combines a fixed-block system with a transmission based system. Conventional blocks are subdivided into smaller increments that permit those trains, equipped with a continuous communication system, to operate on closer headways. Unequipped trains continue to be protected by the basic block system. As equipped trains operate through some signals displaying red an additional aspect must be added to such signals—indicating that the signal is not applicable to that specific train.

SACEM has a throughput capability between fixed-block and moving-block signaling systems that depends on the mix of equipped and unequipped trains. The manufacturer claims an increase in capacity up to 25%, which is comparable to the general 30% increase of moving-block over fixed-block signaling systems—all else being equal. The two equipped rail transit lines in Mexico City do not have any unequipped long distance trains with their longer braking distances and so should obtain the maximum capacity improvement.

While classed as a hybrid system, SACEM does not use moving-blocks and is really an overlay system. Shorter blocks—applicable to certain trains only—are overlaid onto a conventional fixed-block system.

Moving-block signaling systems have been installed by the SEL (Standard Electrik Lorenz) of Stuttgart, Germany, and its Canadian subsidiary SEL Canada. Both are now part of the Alcatel group, a French consortium.

The Alcatel SelTrac®™ system has evolved through five generations over two decades. There are some 20 worldwide installations of which five are in North America: Vancouver, Toronto, Detroit, San Francisco and Orlando (Disneyworld monorail).

The SelTrac system uses an inductive loop to both communicate with trains and, through the loop transpositions, to determine positioning. Processing power is centralized with the on-board computers limited to processing signals and controlling the vehicle subsystems. The use of Intel x86 processors to control critical train movements was introduced in 1994. Transponder positioning has been developed to reduce hardware costs and improve failure management. In addition, SelTrac includes an integrated automatic train supervision subsystem.

The second manufacturer with a system in service is also French. Service started on Line D of the Lyon metro in 1992 using Matra Transport's Maggaly®™ system. The Maggaly system uses inductive transmission with positioning transponders and places the bulk of the processing power on-board. Line data are stored on-board with the wayside equipment limited to system management and providing the location of a leading train to its immediate follower.

The advantages of moving-block signaling systems are considerable. Beyond the capacity increase of interest to this study, the concept offers the potential for lower capital and maintenance costs, flexibility, comprehensive system management capabilities and inherent bi-directional operation. The

[10] Ibid.

[11] Line A and Line 8.

slow acceptance of processor based train control systems may explain why most conventional train control suppliers have stayed away from this concept until the recent selection of moving-block systems by London Transport and New York City Transit, together with several smaller systems. This selection is not necessarily based on the capacity increases but as much on the economics and relative ease of installing the system on top of a conventional signaling system on existing lines that must remain in operation throughout the conversion, modernization or replacement.

Subsequent to the London and New York decisions, many manufacturers have announced the development of moving-block signaling systems.

General Railway Signal is developing its ATLAS®™ system. This is a modular based concept that allows various forms of vehicle location and communication systems. A feature is a vital stored database and low requirements for the vehicle-wayside data communication flow.

Union Switch & Signal is developing its MicroBlok®™ which shares some similarity with Matra's SACEM, overlaying "virtual" software based blocks on a conventional fixed block system. With radio based communications and vital logic distributed on the wayside, the system uses some concepts developed for the Los Angeles Green Line which entered service in August 1995.

AEG Transportation System's Flexiblok®™ shares some features with MicroBlok and SACEM. It is a radio-based system designed for both standalone use and for incrementally adding capacity and features to traditional train control systems. Operational and safety responsibilities are distributed through the system, which incorporates nonproprietary interfaces conforming to Open System Interconnect protocol standards.[12] AEG's US division, previously Westinghouse Electric Transportation Systems, is developing a transmission-based train control system tailored to the North American market.

Harmon Industries' UltraBlock®™ system is radio based with transponder positioning technology. Line profile information is stored on-board. Vital processing is distributed along the wayside.

Siemens Transportation Systems is developing a moving-block system based on its Dortmund University people mover, an under-hanging cabin system that has been in service since 1984.

CMW (Odebretch Group, Brazil) is supplying a radio-based overlay system to the São Paulo metro with distributed processing. The system is claimed to reduce headways from 90 to 66 sec. As section 4.7 of this chapter shows, such close headways are only possible with tightly controlled station dwells which are rarely achievable at heavy volume stations.

Morrison Knudsen (with Hughes and BART) is developing a moving-block signaling system based on military communication technology. The system uses beacon-based, ranging spread spectrum, radio communications which are less susceptible to interference and can tolerate the failure or loss of one or more beacons.

[12] The proprietary nature of many moving-block signaling systems is a concern to potential customers who are then captive to a particular supplier. Traditional train control systems in theory allow many components from different manufacturers to be mixed and matched. However, particularly with the introduction of solid state interlockings, this is not always feasible.

NOTE: The above discussion represents the best information available to the researchers at the time this report was written. Other suppliers may exist and omissions were inadvertent. This discussion is not intended to endorse specific products or manufacturers.

All moving-block systems that base train separation on a continually adjusted distance to the next stop or train ahead (plus a safety distance) should have substantially similar train throughput capabilities. Capacity for a generic moving-block signaling system is developed in section 3.8 of this chapter, based on information from existing systems (Alcatel and Matra).

Those systems under development (above) that succeed in the market can reasonably be expected to have comparable capacities. However, there is insufficient information to confirm this.

3.4 AUTOMATIC TRAIN OPERATION

Automatic acceleration has long been a feature of rail transit. A driver no longer has to cautiously advance the control handle from notch to notch to avoid pulling too much current and so tripping the line breaker. Rather, relays, and more recently micro-processors, control the rate of acceleration smoothly from the initial start to maximum speed.

Cab signaling and moving-block signaling systems transfer speed commands to the train and it was a modest step to link these to the automatic acceleration features, and comparable controlled braking, to create full automatic train operation (ATO). The first North America application occurred in 1962 on NYCTA's Times Square Shuttle, followed in 1967 by Montreal's Expo Express, then, in short order by PATCO's Lindenwold line and San Francisco's BART. Most new rail transit systems have incorporated ATO since this innovative period.

The driver's or attendant's role is not necessarily limited to closing the doors, pressing a train start button and observing the line ahead. Drivers are usually trained in, and rolling stock is provided with, manual operating capabilities. PATCO pioneered the concept of having drivers take over manual control from time to time to retain familiarity with operations. Manual driving under cab controls, limited color-light signaling or radio dispatching is routine, if infrequent, on many ATO-equipped systems when there is a train control failure or to provide signaling maintenance time.

Dispensing entirely with a driver or attendant is controversial. In 1965 the driverless Transit Expressway was first operated in a controlled environment in Pittsburgh. This Automated Guideway Transit (AGT) system, and similar designs, have gained wide-spread acceptance in nontransit usage as driverless people movers in airports, amusement parks and institutional settings. Morgantown's AGT was the first public transit operation to gain acceptance for driverless operation when it opened in 1968. After a long gap Miami's downtown people mover opened in 1985 with the Detroit People Mover and the full-scale urban rail transit

SkyTrain system in Vancouver starting the following year. Driverless public transport is now well established in these cities but no subsequent operations have chosen to follow, despite their record of safety, reliability and lower operating costs. Fundamental concerns with driverless automatic train operation clearly remain.

Automatic train operation, with or without attendants or drivers, allows a train to more closely follow the optimum speed envelope and commence braking for the final station approach at the last possible moment. This reduces station to station travel times, and more important from the point of capacity, it minimizes the critical station close-in time—the time from when one train starts to leave a station until the following train is berthed in that station.

In the literature Klopotov(R32) makes claims of capacity improvements of up to 15% with ATO. Bardaji(R10) claims a 5% capacity increase with automatic regulation. Other reports allude to increases without specific figures. None of the reports substantiate any claims. Attempts to quantify time improvements between manual and automatic driving for this study were unsuccessful. Any differences were overshadowed by other variations between systems.

Intuitively there should be an improvement in the order of 5 to 10% in the station approach time. As this time represents approximately 40% of station headway, the increase in capacity should be from 2 to 4%.

The calculations used to determine the minimum station headway assume optimal driving but insert a time for a drivers sighting and reaction time—in addition to the equipment reaction time. The calculations in this report compensate for ATO by removing the reaction times associated with manual driving.

3.5 AUTOMATIC TRAIN SUPERVISION

Automatic Train Supervision (ATS) encompasses a wide variety of options. It is generally not a safety-critical aspect of the train control system and may not need the rigor of design and testing to its hardware and software that characterizes other areas of train control. At its simplest it does little more than display the location of trains on a mimic board or video screen in the central control or dispatcher's office.

One step up in sophistication provides an indication of on-time performance with varying degrees of lateness designated for each train, possibly grouped by a color code or with a digital display of the time a train is behind schedule. In either case corrective action is in the hands of the variously named controller, dispatcher or trainmaster.

Urban rail transit in North America is generally run to a timetable. Those systems in Europe that consistently operate at the closest headways (down to 90 sec) generally use headway regulation that attempts to ensure even spacing of trains rather than adhere strictly to a timetable. Although it appears that keeping even headways reliably provides more capacity, this is an issue of tradition, operating rules and safety[13] that is beyond the scope of this study.

In more advanced systems where there is ATO, computer algorithms are used to attempt to automatically correct lateness. These are rare in North America and are generally associated with the newer moving-block signaling systems.

Corrective action can include eliminating coasting, increasing line speed, moving to higher rates of acceleration and braking and adjusting dwell times—usually only where these are preprogrammed. Such corrective action supposes that the system does not normally work flat out.

The Vancouver system is an example of unusually comprehensive ATS strategies. Here trains have a normal maximum line speed of 80 km/h (50 mph) which ATS can increase to 90 km/h as a catch up measure—where civil speed restrictions so permit. Similarly acceleration and braking can be adjusted upwards14[14] or downwards by 10%.

In normal operation trains use less than their full performance which reduces energy consumption and maintenance, and leaves a small leeway for on-time corrective action. Together, these strategies can pick up 2 to 3 min in an hour.

Correcting greater degrees of lateness or irregularity generally involves manual intervention using short turn strategies or removing slow-performing or defective trains from service.[15] This is difficult to implement in the peak period and common practice is to let the service run as best it can and wait to make corrections to the timetable until after the peak period.

A further level of ATS strategies is possible—predictive control. Although discussed as a possibility, this level is not known to be used in North America. In predictive control a computer looks ahead to possible conflicts, for example a merge of two branches at a junction. The computer can then adjust terminal departures, dwell times and train performance to ensure that trains merge evenly without holds, or are appropriately spaced to optimize turn-arounds at any common terminal.

The nonvital ATS system can also be the host for other features such as on-board system diagnostics and the control of station and on-board information through visual and audio messages—including those required by ADA.

Summary ATS has the potential to improve service regularity and so help maximize capacity. However, the strategies to correct irregular service on rail transit are limited unless there is close integration with ATO and the possibilities of adjusting train performance and station dwells. Without such strategies, ATS allows dispatchers to see problems but remain unable to address them until the peak period is over. In Chapter Six, *Operating*

[13] Certain Russian systems that maintain remarkably even 90-sec headways require drivers to close doors and depart even if passenger flow is incomplete.

[14] A train's performance is limited by motor heating characteristics. Corrective actions that increase performance also increase heating. Depending on ambient temperature this can only be carried out for a limited period before the train's diagnostic equipment will detect over-heating and either cut one or more motors out or force a drop to a lower performance rate.

[15] One North American system is known to use a skip-stop strategy for seriously late trains, that is running through a station where the train would normally stop. Akin to the bus corrective strategy of "set downs only, no pick-ups," this is both unusual and can be difficult for passengers to accept.

Issues, an operational allowance to compensate for irregular operation is developed. A sophisticated ATS system in conjunction with a range of feasible corrective actions can reduce the desired amount of operating margin time.

3.6 FIXED-BLOCK THROUGHPUT

Determining the throughput of any rail transit train control system relies on the repetitive nature of rail transit operation. In normal operation trains follow each other at regular intervals traveling at the same speed over the same section of track.

All modern trains have very comparable performance. All low-performance equipment in North America is believed to have been retired. Should a line operate with equipment with different performance and/or trains of different length, then the maximum throughput rates developed in this section should be based on the longest train of the lowest performing rolling stock.

Trains operating on an open line with signaling protection but without station stops have a high throughput. This throughput is defined as *line* or *way capacity*. This capacity will be calculated later in this section although it has little relevance to achievable capacity except for systems with off-line stations. Only Automated Guideway Transit, or some very high capacity lines in Japan, can support off-line stations.

Stations are the principal limitation on the maximum train throughput—and hence maximum capacity—although limitations may also be due to turn-back and junction constraints. The project survey of operating agencies indicated that the station close-in plus dwell time was the capacity limitation in 79% of cases, turnback constraints in 15%, and junctions in 5% of cases. Further inquiry found that several turnback and junction constraints were self-imposed due to operating practices and that stations were by far the dominant limitation on throughput.

In a well-designed and operated system, junction or turnback constrictions or bottlenecks should not occur. A flat junction can theoretically handle trains with a consolidated headway approaching 2 min. However, delays may occur and systems designed for such close headways will invariably incorporate grade-separated (flying) junctions. Moving-block signaling systems provide even greater throughput at flat junctions as discussed in section 3.10.

A two-track terminal station with either a forward or rear scissors cross-over can also support headways below 2 min unless the cross-overs are long, spaced away from the terminal platform, or heavy passenger movements or operating practices when the train crew changes ends (reverses the train) result in long dwells. The latter two problems can be resolved by multiple-platform terminal stations, such as PATH's Manhattan and Hoboken terminals and Mexico City's Indios Verdes station, or by establishing set-back procedures for train crews.[16]

[16] Set back procedures require the train crew or operator to leave the train at a terminal and walk to the end of the platform where they board the next entering train which can be immediately checked and made ready for departure. On a system with typical close headways of two minutes this requires an extra crew every 30 trains and increases crewing costs by some 3%—less if only needed in peak periods. The practice is unpopular with staff as they must carry their possessions with them and cannot enjoy settling into a single location for the duration of their shift.

In this chapter the limitations on headway will be calculated for all three possible bottlenecks: station stops, junctions and turnbacks.

Nine reports in the literature survey provide detailed methods to calculate the throughput of fixed-block rail transit signaling systems:

- AUER, J.H., Rail-Transit People-Mover Headway Comparison[R9]
- BARWELL, F. T., Automation and Control in Transport[R11]
- BERGMANN, DIETRICH R., Generalized Expressions for the Minimum Time Interval between Consecutive Arrivals at an Idealized Railway Station[R13]
- DELAWARE RIVER PORT AUTHORITY, 90 Seconds Headway Feasibility Study, Lindenwold Line[R21]
- GILL, D.C., and GOODMAN C.J., Computer-based optimisation techniques for mass transit railway signalling design[R26]
- JANELLE, A., POLIS, M.P., Interactive Hybrid Computer Design of a Signaling System for a Metro Network[R31]
- LANG, A SCHEFFER, and SOBERMAN, RICHARD M., Urban Rail Transit Its Economics and Technology[R39]
- VUCHIC, VUKAN R., Urban Public Transportation Systems and Technology[R71]
- WEISS, DAVID M., and FIALKOFF, DAVID R., Analytic Approach to Railway Signal Block Design[R73]

All the reports deal with station stops as the principal limitations on capacity and use Newton's equations of motion to calculate the minimum train separation, adding a variety of nuances to accommodate safety distances, jerk limitations, braking system and drivers' reaction times plus any operating allowance or recovery margin. In the following section a classical approach is examined, followed by a recommended practical approach derived from the work of Auer[R09] in combination with information from several other authors. Then an examination is made of the sensitivity of the results to several system variables.

3.6.1 STATION CLOSE-IN TIME

The time between a train pulling out of a station and the next train entering—referred to as *close-in*—is the main constraining factor on rail transit lines. This time is primarily a function of the train control system, train length, approach speed and vehicle performance. Close-in time, when added to the dwell time and an operating margin, determines the minimum possible headway achievable without regular schedule adherence impacts—referred to as the *noninterference headway.*

When interference occurs, trains may be held at approaches to stations and interlockings. This requires the train to start from stop and so increases the close-in time, or time to traverse and clear an interlocking, reducing the throughput. With throughput decreased and headways becoming erratic, the number of passengers accumulated at a specific station will increase and so increase the dwell time. This is a classic example of the maxim that when things go wrong they get worse.

The minimum headway is composed of three components:

- the safe separation (close-in time),
- the dwell time in the station, and
- an operating margin.

Station dwells are discussed in Chapter Four, *Station Dwells*, recovery margins are discussed in Chapter Six, *Operating Issues.*

3.6.2 COMPUTER SIMULATION

The best method to determine the close-in time is from the specifications of the system being considered[17], from existing experience of operating at or close to capacity or from a simulation. It is common in designing and specifying new rail transit systems, or modernizing existing systems, to run a variety of computer simulation models. These models are used to determine running times, to optimize the design of track work, of signaling systems and of the power supply system. Where the results of these models are available they can provide an accurate indication of the critical headway limitation—whether a station close-in maneuver, at a junction or at a turnback.

Such models can be calibrated to produce accurate results. In particular, many simulation models will adjust train performance for voltage fluctuations in the power supply—a variant that cannot be otherwise be easily calculated. However caution should be exercised in using the output from simulations. Simulations can be subject to poor design, poor execution or erroneous data entry. In particular, increments of analysis are important. The model will calculate the voltage, performance, movement and position of the front and rear of each train in small increments of time, and occasionally in increments of distance or speed. Such increments should approach one tenth of a second to produce accurate close-in times.

Simulation programs are also often proprietary to a specific consultant or train control, traction substation or vehicle supplier. They require considerable detailed site and equipment data. As such, they may not be practical or available for determining achievable capacity, making it necessary to calculate the throughput of the particular train control system by more general methods.

If the minimum headway is not available from the system designers or from a simulation, then straightforward methods are available to calculate the time. Here train separation is based on a *line clear* basis—successive green signals governing the following train. The minimum line headway is determined by the critical line condition, such as the close-in at the maximum load point station plus an operating margin. The entire stretch of line between junctions and turnbacks, where train density is physically constant, is controlled by this one critical time.

The classical expression for the minimum headway of the typical rail transit three-aspect block-signal system is

$$H(t) = \frac{\frac{2BL}{v_{ap}} + D_w + \sqrt{\frac{2L}{a}}}{1 - M} \qquad \text{Equation 3-1}$$

The block length must be greater than or equal to the service stopping distance.[18]

$$BL \geq SD = \frac{V_{ap}^2}{2d} \qquad \text{Equation 3-2}$$

where
$H(t)$ = headway *in seconds*
BL = block length approaching station *(m)*
D_w = station dwell time *in seconds*
SD = service stopping distance *(m)*
L = length of the longest train *(m)*
v_{ap} = maximum approach speed *(m/s)*
a = average acceleration rate through the station platform clear-out *(m/s^2)*
d = braking rate *(m/s^2)*
M = headway adjustment combining operational tolerance and dwell time variance *(constant)*

Although the headway adjustment factor, *M*, can encompass a variety of items, it is difficult to encompass all the variables that can affect headway. These include

- any distance between the front of the train and the start of the station exit block,[19] particularly if the train is not berthed at the end of the platform;
- control system reaction time;
- on manually driven trains, the train operator sighting and reaction time;
- the brake system reaction time;[20]
- an allowance for jerk limitation;[21]
- speed restrictions on station approaches and exits whether due to speed control for special work or curves; and
- grades approaching and leaving a station.

In addition, the length of the approach block and the approach speed are not readily obtainable quantities. Consequently this traditional method is not recommended and an alternate approach will be developed, based, in part, on the work of Auer. This uses more readily available data accommodating many of the above variables. This approach encompasses both manually and automatically driven trains, multiple command cab controls, and, by decreasing block length, a moving-block system.

Even so, it should be borne in mind that not all variables can be included, and assumptions and approximations are still needed. This approach, while more comprehensive than many in the literature, is not as good as using information from signaling

[17] The train control design engineers will be aiming to minimize the close-in time and information from this source, particularly if the result of an accurate simulation, is invariably the most accurate way to determine practical capacity.

[18] On close headway systems block lengths may be less than the service stopping distance. New York has approach blocks down to 60m (200’) and lengths as short as 15m (50’) occur on some systems—particularly automated guideway transit systems.

[19] This allows for blocks that do not start at the end of the platform—at the headwall—or shorter trains that are berthed away from the headwall.

[20] Older equipment may have air brakes applied by releasing air from a brake control pipe running the length of the train (train-lined). There is a considerable delay as this command passes down the train and brakes are applied sequentially on cars. Newer equipment uses electrical commands to control the air, hydraulic or electric brakes on each car and response is more rapid.

[21] Limitations applied to the start and end of braking and the start of acceleration to limit the rate of change of acceleration—commonly, if somewhat erroneously called jerk.

engineers, based on actual block positions, or from a comprehensive and well-calibrated simulation.

3.6.3 CALCULATING LINE HEADWAY

On a level, tangent (straight) section of track with no disturbances the line headway *H(l)* is given by:

$$H(l) = \frac{L + S_{min}}{v_l} \qquad \text{Equation 3-3}$$

where
- $H(l)$ = line headway in seconds
- S_{min} = minimum train separation in meters
- L = length of the longest train in meters
- v_l = line speed in m/s[22]

The minimum train separation corresponds to the sum of the operating margin and safe separation distance shown in Figure 3.2. It can therefore be further subdivided: *(all in meters)*

$$S_{min} = S_{sbd} + S_{td} + S_{om} \qquad \text{Equation 3-4}$$

where
- S_{min} = mininimum train separation distance
- S_{sbd} = safe braking distance
- S_{td} = train detection uncertainty distance
- S_{om} = operating margin distance[23]

The safe braking distance is based on the rail transit assumption of brick-wall stops using a degraded service braking rate.[24] The train detection uncertainty reflects either the block length or the distance covered in the polling time increments of a movingblock signaling system. The operating margin distance is the distance covered in this time allowance. This will be omitted from further consideration in this section. It is developed in Chapter Six, *Operating Issues*, and added into the headway calculation by mode in Chapters Seven through Ten.

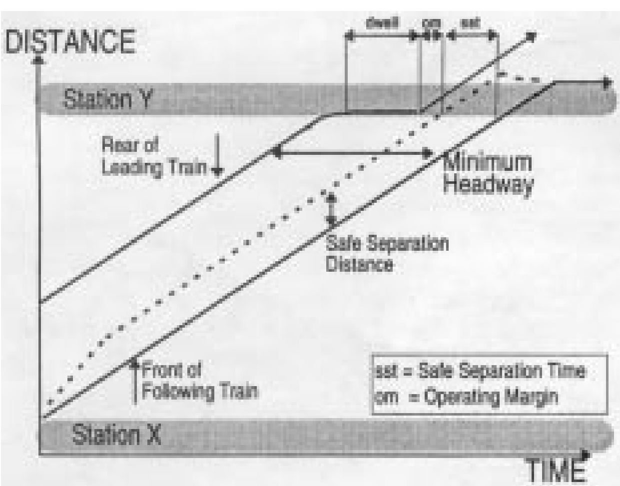


Figure 3.2 Distance-time plot of two consecutive trains (acceleration and braking curves omitted for clarity)

Substituting for S_{min} and removing S_{om} produces

$$H(l) = \frac{L + S_{td} + S_{sbd}}{v_l} \qquad \text{Equation 3-5}$$

There are several components in the safe braking time. The largest is the time to brake to a stop, using the service brake. A constant *K* is added to assume less than full braking efficiency or reduced adhesion—75% of the normal braking is an appropriate factor. There is also the distance covered during driver sighting and reaction time on manually driven trains, and on automatically driven trains brake equipment reaction time and a safety allowance for control failure. This overspeed allowance assumes a worst case situation whereby the failure occurs as the braking command is issued with the train in full acceleration mode. This is often termed *runaway propulsion.* The train continues to accelerate for a period of time t_{os} until a speed governor detects the overspeed and applies the brakes.[25]

$$S_{sbd} = \frac{100}{K} S_{bd} + S_{br} + S_{os} \qquad \text{Equation 3-6}$$

where
- S_{bd} = safe breaking distance *in meters*
- S_{bd} = service braking distance *in meters*
- K = braking safety factor
- S_{br} = train operator sighting and reaction distance and/or braking system reaction distance *in meters*
- S_{os} = overspeed travel distance *in meters*

The distance to a full stop from speed V_l at the constant service braking, deceleration or retardation rate is given by:

$$S_{bd} = \frac{v_l^2}{2d_s} \qquad \text{Equation 3-7}$$

where d_s = service deceleration rate *in m/s*^{2}

To be rigorous, the safe braking distance should also take into account grades, train load—passenger quantities and any snow and ice load and, in open line sections, any tail wind. These add complexities beyond the scope of this study and, except for downgrades, contribute a very minor increment to the result. Consequently they have been omitted. The effect of grades will be examined in the sensitivity analysis at the end of this section.

Modern rail transit equipment uses a combination of friction and electrical braking,[26] in combination with slip-slide controls, to maintain an even braking rate. An allowance can be added for the jerk limiting features that taper the braking rate at the beginning and end of the brake application.

---

[22] Can be worked in feet with speed in feet per second. 10 mph=14.67 ft/sec, 10 km/h = 2.78 m/s

[23] Auer used the term *service control buffer distance.*

[24] Some workers use the emergency braking rate. As this is highly variable depending on location, equipment, and wheel to rail adhesion, it is not recommended.

[25] As the braking so applied is usually at the emergency rate, a case can be made that this component may be discounted or reduced.

[26] Electrical braking is both dynamic—with recovered energy burned by resistors on each car, or regenerative braking with recovered energy fed back into the line—here it feeds the hotel load of the braking train, adjacent trains, is fed back to the power utility via bi-directional substations or is burned by resistors in the substation. The latter two modes are rare. Regenerative braking was common in the early days of electric traction. It then fell out of use when the low cost of electricity failed to justify the additional equipment costs and maintenance. With increased energy costs and the ease of accommodating regeneration on modern electronic power conversion units, regeneration is now becoming a standard feature. Regeneration is sometimes termed recuperation.

The distance an automatically operated train moves until the overspeed governor operates can be expressed as

$$S_{os} = v_l t_{os} + \frac{a_l t_{os}^2}{2} \qquad \text{Equation 3-8}$$

where
- S_{os} = overspeed distance
- t_s = time for overspeed governor to operate
- a_l = line acceleration rate in m/s^2 at v_l
- v_l = line speed

Substituting Equations 3-6, 3-7, and 3-8 in Equation 3-5 and adding a jerk limiting allowance produces

$$H(l) = \frac{L + S_{td}}{v_l} + \frac{100}{K}\left(\frac{v_l}{2d_s}\right) + \frac{a_l t_{os}^2}{2v_l} + t_{os} + t_{jl} + t_{br} \qquad \text{Equation 3-9}$$

where
- t_{br} = train operator sighting and reaction time and/or braking system reaction time
- t_{jl} = jerk limiting time allowance

Service acceleration is said to be following the motor curve as it reduces from the initial controlled rate to zero at the top, maximum, or balancing speed of the equipment. The acceleration rate at a specific speed may not be readily available and an approximation is appropriate for this item—a small component of the total line headway time. On equipment with a balancing speed of 80 km/h, the initial acceleration is maintained until speeds reach 10-20 km/h then tapers off, approximately linearly until speeds of 50-60 km/h, then approximately exponentially until it is zero. At line speeds appropriate to this analysis the line acceleration rate can be assumed to be approximate to the inverse of speed so that for intermediate speeds

$$a_l \cong a_s\left(1 - \frac{v_l}{v_{max}}\right) \qquad \text{Equation 3-10}$$

where
- v_l = line speed in m/s
- v_{max} = maximum train speed in m/s
- a_l = line acceleration rate in m/s^2
- a_s = initial service acceleration rate in m/s^2

The train detection uncertainty distance is not readily available but can be approximated as either the block length(s)—again not easily obtained—or the braking distance plus some leeway as a surrogate for block lengths on a system designed for maximum throughput. This quantity is particularly useful as a simple method to adjust for the differences between the traditional three-aspect signaling system, cab controls with multiple aspects (command speeds) and moving-block signaling systems.

$$S_{td} \cong B\left(\frac{v_l^2}{2d_s}\right) \qquad \text{Equation 3-11}$$

where *B* is a constant representing the increments or percentage of the braking distance—or number of blocks—that must separate trains according to the type of train control system. A *B*-value of 1.2 is recommended for multiple command cab controls. A value of 2.4 is appropriate for three-aspect signaling systems where there is always a minimum of two clear blocks between trains.[27] The value of *B* for moving-block signaling systems can be equal to or less than unity and is developed in the next section.

Accepting these approximations and substituting Equations 3-10 and 3-11 in Equation 3-9 produces

$$H(l) = \frac{L}{v_l} + \left(\frac{100}{K} + B\right)\left(\frac{v_l}{2d_s}\right) + \frac{a_s t_{os}^2}{2v_l}\left(1 - \frac{v_l}{v_{max}}\right) + t_{os} + t_{jl} + t_{br} \qquad \text{Equation 3-12}$$

where
- $H(l)$ = line headway *in seconds*
- L = length of the longest train *in meters*
- v_l = line speed in *m/s*
- K = braking safety factor—worst case service braking is K% of specified normal rate—typically 75%
- B = separation safety factor—equivalent to the number of braking distances (surrogate for blocks) that separate trains
- t_{os} = overspeed governor operating time[28] (*s*)
- t_{jl} = time lost to braking jerk limitation *(s)*
- t_{br} = operator & brake system reaction time *(s)*
- a_l = line acceleration rate *in* m/s^2
- d_s = service deceleration rate *in* m/s^2

North American rail transit traction equipment tends to have very similar performance derived from the work of the Presidents' Conference Committee (PCC) in the mid 1930s. The chief engineer, Hirschfeld,[29] placed subjects on a moving platform and determined the acceleration rate at which they lost their balance or became uncomfortable. A wide variety of subjects were tested including people who were pregnant, inebriated or holding packages. From this pioneering work, the PCC streetcar evolved and with it rates of acceleration and deceleration (and associated jerk[30]) that have become industry standards. The recommended maximum rate is 3.0 mphps (1.3 m/s^2) for both acceleration and deceleration.

Attempts have been made to increase these rates, specifically on the rubber tired metros in Montreal and Mexico City, but subsequently these were reduced close to the industry standard. Except for locomotive hauled commuter rail, almost all rail transit in North America operates with these rates. The main difference in equipment performance is the maximum speed. Most urban rail systems with closer station spacing have a maximum speed of 50-60 mph (80-95 km/h), light rail typically has a maximum speed of 50 mph (80 km/h),[31] while streetcars have a maximum in the range of 40-50 mph (65-80 km/h). The few suburban type rail rapid transit systems have a higher maximum of 70-80 mph (110-130 km/h)—BART in San Francisco and PATCO in Philadelphia are the principal examples.

[27] On existing systems the results can be calibrated to actual performance by adjusting the value of "*B*".

[28] t_{os}+ t_{jl}+ t_{br} may be simpified by treating as a single value—typically 5 sec for systems with ATO, slightly longer with manual driving.

[29] HIRSCHFELD, C.F., Bulletins Nos. 1-5, Electric Railway Presidents' Conference Committee (PCC), New York, 1931-1933.

[30] jerk—rate of change of acceleration.

[31] SEPTA's Norristown line is a higher speed exception.

The higher gearing rates required for these higher speeds result in either a reduced initial acceleration rate or, more typically, an acceleration rate that more rapidly reduces (follows the motor curve) as speed increases.

Braking rates are invariably uniform. Emergency braking rates vary widely and are significantly higher and more sustainable on equipment fitted with magnetic track brakes—all streetcars, most light rail and the urban rail transit systems in Chicago and Vancouver.

This relative uniformity of rates allows a typical solution of Equation 3.11 using the following data for a cab control system with electrically controlled braking and a train of the maximum length in North American rail transit.

The results of applying typical rail transit data to Equation 3-9 are shown in Figure 3.3 using the data values of Table 3.1.

Table 3.1 Data values for line headway

TERM	VALUE	
Train length	200 m	660 ft
Overspeed governor time[32]	3 sec	3 sec
Braking safety factor ***K***	75%	75%
Separation safety factor ***B***	1.2	1.2
Initial service acceleration rate	1.3 m/s²	3.0 mphps
Service deceleration rate	1.3 m/s²	3.0 mphps

[32] The 3-sec figure is conservative. For automatically driven trains, a time of 1 sec is appropriate and can drop as low as 0.2 sec on AGT systems. The higher figure is useful on cab control systems. When the overspeed detection occurs, and alarm is sounded in the cab to allow the driver to apply service braking and so cancel the automatic application of emergency brakes—avoiding wheel flats and passenger discomfort or loss of balance. The delay time is then based on typical manual reaction times of 2 to 3 sec. With entirely manual operation this term becomes a surrogate for driver sighting and reaction time. Values of 2 to 5 sec have been quoted in the literature. 3 sec is an appropriate value.

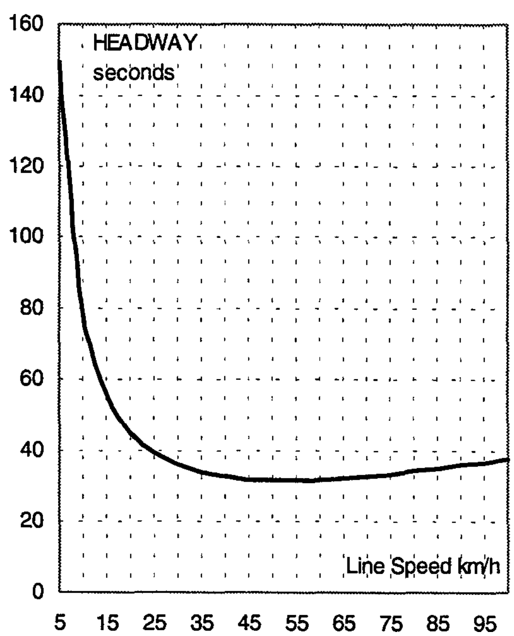


Figure 3.3 Line headway versus speed

Table 3.2 Breakdown of line headway time components

Speed km/h	Time to travel length	Time to brake x(100/K+B)	Over-speed accel. time	Over-speed time[33]	Jerk allow-ance	Line head-way
COLUMN	1	2	3	4	5	6
10	72.00	2.70	0.40	3	0.5	78.61
20	36.00	5.41	0.20	3	0.5	45.10
30	24.00	8.11	0.13	3	0.5	35.74
40	18.00	10.81	0.09	3	0.5	32.40
50	14.40	13.51	0.07	3	0.5	31.49
60	12.00	16.22	0.06	3	0.5	31.78
70	10.29	18.92	0.05	3	0.5	32.75
80	9.00	21.62	0.04	3	0.5	34.16
90	8.00	24.33	0.03	3	0.5	35.86
100	7.20	27.03	0.03	3	0.5	37.76

[33] Overspeed time is applicable to automatically driven trains.

These are somewhat theoretical, showing headways down to 31.5 seconds—120 trains per hour. There is a clear minimum at 50 km/h (31 mph). Obviously restricting train line speed to so low a value would be uneconomic, requiring a larger number of cars to meet a given demand—which would, in any event, diminish because of the slow travel times deterring passengers.

The equation and results will be applied in Chapter 10 for automated guideway transit with off-line stations and will be used as a basis for determining realistic headways with station stops.

To this end it is useful to examine the value of the components in the line headway, shown in Table 3.2 with all figures in seconds. Columns one through five in this table represent, consecutively, the first five terms of Equation 3-12. The time to travel the length of train and the factored braking time predominate. No value has been assigned to the brake system reaction time. The time associated with the runaway acceleration is small. Equation 3-12, adjusted to compensate for grades and line voltage variations, is included in the spreadsheet on the computer diskette. For manual calculations, the equation can be simplified to:

$$H(l) = 4 + \frac{L}{v_l} + 1.3\left(\frac{v_l}{d_s}\right) \qquad \text{Equation 3-13}$$

where the constant 4 is approximately the rounded up sum of columns 3, 4 and 5 plus a small allowance for brake reaction time. This should be increased to 7 for manually driven systems to add the train operator sighting and reaction time.

The next step is to accommodate station stops. Reference to the literature will show numerous ways to calculate the station headway. This approach is based on adapting the line headway equation.

3.6.4 CALCULATING STATION HEADWAY

Station headway, the time for one train to replace another at the maximum load point station, is by far the most common capacity limitation. Having derived an expression for line headway that uses readily available information with as few approximations as possible, it is possible to adapt this to station headway by

- changing line speed to approach speed and solving for this speed,
- adding a component for the time a train takes to clear the platform,
- adding the station dwell, and
- adding an operating margin.

The time for a train to clear the platform is

$$t_c = \sqrt{\frac{2(L+D)}{a_s}} \qquad \text{Equation 3-14}$$

Adding Equation 3-14 to 3-12 plus components for dwell and an operating margin produces the station headway

$$H(s) = \sqrt{\frac{2(L+D)}{a_s}} + \frac{L}{v_a} + \left(\frac{100}{K} + B\right)\left(\frac{v_a}{2d_s}\right) + \frac{a_s t_{os}^2}{2v_a}\left(1 - \frac{v_a}{v_{max}}\right) + t_{os} + t_{jl} + t_{br} + t_d + t_{om}$$

Equation 3-15

where
$H(s)$ = station headway *in seconds*
L = length of the longest train *in meters*
D = distance from front of stopped train to start of station exit block *in meters*
v_a = station approach speed *in m/s*
v_{max} = maximum line speed *in m/s*
K = braking safety factor—worst case service braking is K% of specified normal rate—typically 75%
B = separation safety factor—equivalent to number of braking distances plus a margin, (surrogate for blocks) that separate trains
t_{os} = time for overspeed governor to operate
t_{jl} = time lost to braking jerk limitation—(seconds) typically 0.5 seconds
t_{br} = operator and brake system reaction time
t_d = dwell time *(seconds)*
t_{om} = operating margin *(seconds)*
a_s = initial service acceleration rate *in* m/s^2
d_s = service deceleration rate *in* m/s^2

Typical values will be used and this equation solved for the approach speed under two circumstances:

1. three-aspect signaling system ($B = 2.4$)
2. multiple command speed cab controls ($B = 1.2$)

A 45-sec dwell time is used—typical of the busiest stations on rail transit lines operating at capacity—together with an operating margin time of 20 sec. The brake system reaction time will use a moderate level of 1.5 sec—this should be higher for old air brake equipment, lower for modern electronic control, particularly with hydraulically actuated disk brakes. Other factors remain at the levels used in the line headway analysis. (See Table 3.3.) The results of solving Equation 3.15 for minimum headway in Table 3.4 show a distinct optimum approach speed for fixed-block systems. Moving-block signaling systems, which adjust their separation according to speed, are discussed in the next section. The values are calculated in Table 3.5 with different values of dwell and operating margin times. Speeds are rounded to the nearest km/h or mph reflecting the approximations used in their derivation. As Figure 3.4 deals with maximum length trains, running at minimum headways, at the longest dwell[35] station, dwell times of 30 sec may not be possible and the lower values of H(s) are unlikely. The above calculations do not take into account any speed restriction in the station approach. Reference to Figure 3.4 shows a rapid fall off in throughput as the approach speed decreases. Speed restrictions may be due to curves, special work, or speed controls approaching a terminal station. The Figure 3.5 shows the speed of a braking train against

Table 3.3 Data values for station headway

TERM	VALUE
Train length	200 m (660')
Front of train distance	10 m (33')
Overspeed governor time	3 sec
Jerk limitation time	0.5 sec
Brake system reaction time	1.5 sec
Controlling dwell time	45 sec
Operating margin time	20 sec
Braking safety factor K	75%
Separation safety factor B	1.2 or 2.4[34]
Initial acceleration rate	1.3 m/s²
Service braking rate	1.3 m/s²

[34] $B = 1.2$ for cab control, 2.4 for 3 aspect signaling

Table 3.4 Optimum approach speeds

	Three-aspect signaling		Multi-code cab signaling	
Optimum approach speed	47 km/h	29 mph	52 km/h	32 mph

Table 3.5 Headways with dwell and operating margins

TIMES		MINIMUM HEADWAY	
Dwell time	Operating margin	Three-aspect signaling	Multi-code cab signaling
45 sec	25 sec	127 sec	121 sec
45 sec	15 sec	117 sec	111 sec
30 sec	25 sec	103 sec	96 sec
30 sec	15 sec	93 sec	86 sec

[35] The longest dwell station is usually at the maximum load point station and is so assumed through this report. Reference to Chapter Four, *Station Dwells* shows that a high-volume mixed-flow station could have a longer dwell than the higher volume maximum load point station.

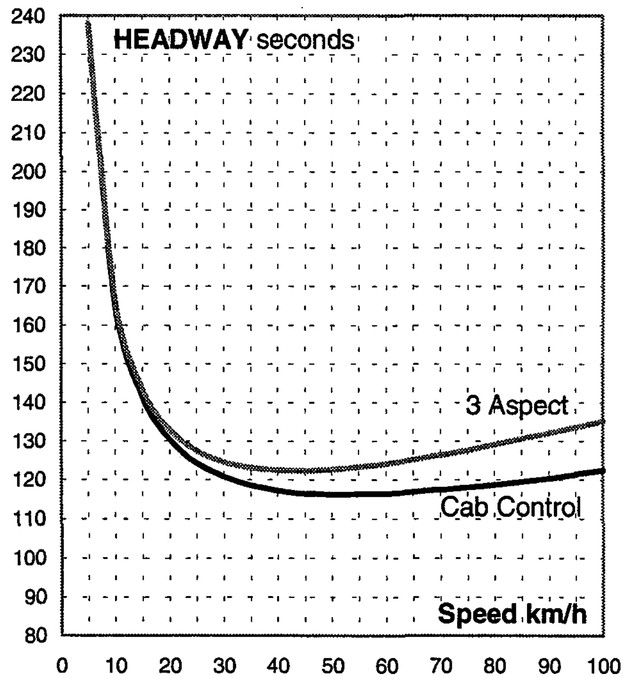


Figure 3.4 Station headway for lines at capacity

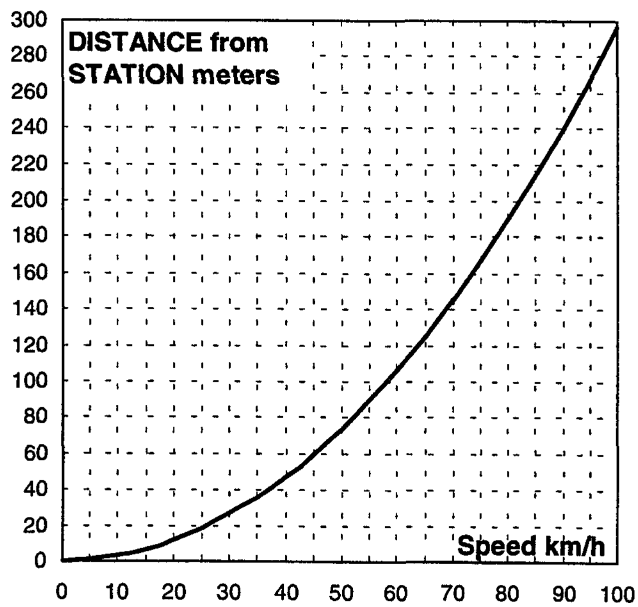


Figure 3.5 Distance—Speed chart

distance—using the performance data of Table 3.3. If a more restrictive speed limit is within the distance for a given approach speed—plus the length of the train—then that more restrictive limit should be used in Equation 3-15 to calculate the minimum headway.

On existing systems speed limits are usually posted on the wayside and included in the rule book. On new systems where speed limits are not known they can be approximated from

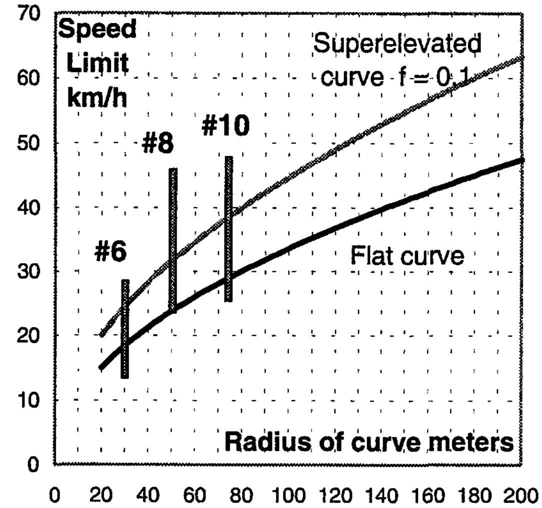


Figure 3.6 Speed limits on curves and switches

$$v_{sl} = (87R(e + f))^{1/2} \qquad \text{Equation 3-16}$$

where
- v_{sl} = speed limit in km/h
- R = radius of curvature in meters
- e = superelevation ratio (height the outer rail is raised divided by track gauge) usually not greater than 0.10
- f = comfort factor (ratio of radial force to gravitational force—0.13 is the maximum used in rail transit with some systems using as low as 0.05)

In U.S. customary units, mph and feet, the speed limit is

$$v_{sl} = (15R(e + f))^{1/2} \qquad \text{Equation 3-17}$$

The results of speed limits due to curves are plotted below for both flat curves and curves superelevated with the maximum radial force (e = 0.10). Transition spirals are not taken into account in Figure 3.6. The vertical bars show the AREA[36] recommended speed limit range for lateral and equilateral level turnouts of size #6, #8 and #10. Note that many operators have their own speed limits for turnouts that may differ from those shown.

3.7 SENSITIVITY

Two factors have not been taken into account in the determination of minimum headways in the preceding section—grades and fluctuations in traction voltage.

3.7.1 GRADES

The principal effect of grades is where downgrades into stations increase the braking distance[37] and the distance associated with

[36] American Railway Engineering Association.

[37] Certain modern equipment uses accelerometers to adjust propulsion and braking to constant levels—independent of train load or grades. In this case grade need not be taken into account—up to the point that wheel-rail adhesion becomes inadequate—an unlikely event.

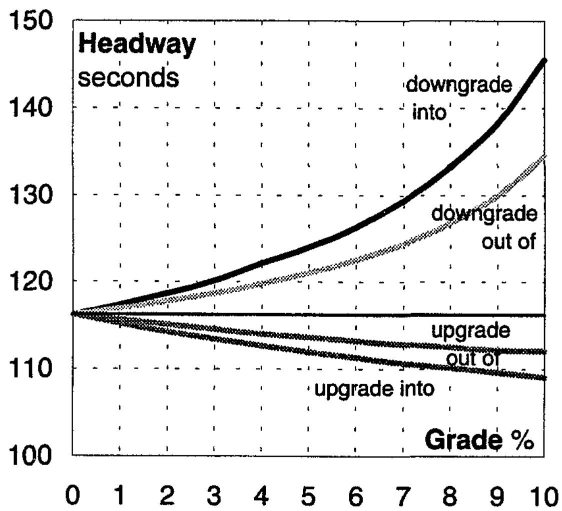


Figure 3.7 Effect of grade on station headway
(cab signals, dwell = 45, margin = 20 secs)

Table 3.6 Result of 4% station grades on headway
(cab signals, dwell = 45, margin = 20 secs)

4% grade	down in	up into	down out	up out
change	+5.9 secs	-3.6 secs	+3.5 secs	-2.2 secs
%	+7.3%	-3.1%	+3.0%	-1.9%

the runaway propulsion factor. A simple method to compensate for grades is to adjust the service braking and acceleration rates in Equation 3-15 while holding the component of the equation that relates to the time for a train to exit a platform constant. The acceleration due to gravity is 9.807 m/s^2. Thus each 1% in downgrade reduces the braking rate by 0.098 m/s^2. The results are shown in Figure 3.7. Note that most rail transit systems have design standards that limit grades to 3 or 4%, a few extend to 6% and the occasional light rail grade can extend to 10%. The impact of grades is greater into a station. The greatest impact is a downgrade into a station which increases the braking and so the safe separation distance. Block lengths must be longer to compensate for the longer braking distances. The absolute and percentage changes are tabulated in Table 3.6 for the typical heavy rail maximum grade of 4%.

3.7.2 LINE VOLTAGE

Rail transit in North America is supplied by direct current power at a potential of 600 to 750 volts with the occasional 1,500-volt system. As more power is drawn through the substations, feeders and third rail or overhead catenary, the voltage drops. Voltage is higher in the vicinity of substation feeders and drops off with distance. Voltage is said to be regulated within a system specification that is typically +20% to -30%.[38] The lowest

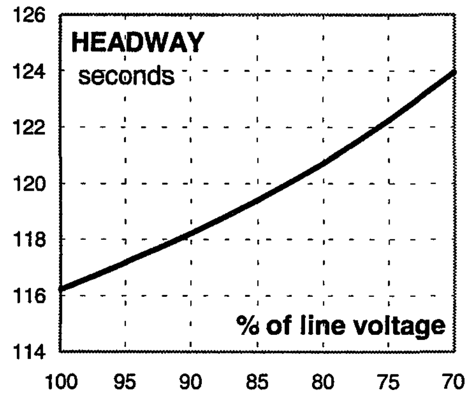


Figure 3.8 Headway changes with voltage

voltage occurs at locations most remote from sub-stations in the peak hour when the maximum number of trains are in service. The lower voltage reduces train performance—at a time when the heavy passenger load is doing likewise. Both acceleration and balancing speed are reduced; braking is not affected.

The acceleration of a train is approximately proportional to the power applied to the motors, which in turn is proportional to the square of the supply voltage. This is particularly true for older equipment with switched resistor controls[39], less so with modern electronically controlled equipment.[40] Consequently, for older equipment without on-board motor voltage feedback and control, the common 10% reduction in voltage will reduce acceleration to 81% of normal, the very rare 30% drop will reduce acceleration to 49% of normal.

Reduced acceleration affects the platform clear out component of the headway calculation. The resultant headway sensitivity to voltage is shown below. At a typical 15% drop in voltage (85% in Figure 3.8), headway increases by 3.2 seconds, a 2.7% change. It is not possible to calculate line voltage at any instance of time without a complete train performance and traction supply system simulation. This will automatically occur if a simulation is used to determine the minimum headway. Otherwise it is uncertain whether a manual adjustment should be made based on the above chart—with certain designs of modern rolling stock the effect of voltage drop can be less than shown.

3.7.3 ACCELERATION

Changes in acceleration affect the time required for a train to clear the platform and make minor adjustments to the runaway

[38] Certain newer rail systems have purchased vehicles with electronic motor controls that are intolerant of voltage drops. Consequently the traction supply voltage has to be regulated to closer tolerances.

[39] Estimated to be used on about three quarters of the rolling stock in North America, including all NYCT cars except prototypes.

[40] Modern electronically controlled equipment may use accelerometers which will command the vehicle's power conversion unit to compensate for reduced voltage. Similar feedback systems may attempt to regulate motor voltage—ven with reduced line voltage. However such corrective action defeats the self regulating effect of the reduced line voltage—a rationing of power when demand from the trains exceeds the capability of the power supply — and so increases the likelihood that the power supply system will trip (disconnect) due to overload. On manually driven systems lower line voltage is immediately apparent to the driver and serves as an advisory to reduce demand or, when trains are lined up due to a delay, to start up in sequential order rather than simultaneously. Consequently, providing full correction for drops in line voltage is unwise.

propulsion safety factor. Headways for a cab signal train control system are shown with acceleration adjusted to 50%, 75% and 125% of the normal value—1.3 m/s^2(3.0 mphps). (See Figure 3.9).

3.7.4 BRAKING

Changes in braking rate affect both the braking time and the safe separation time. Headways for a cab signal train control system are shown with braking adjusted to 50%, 75% and 125% of the normal value in Figure 3.10. Changes in the braking rate have a greater effect on headway than those of acceleration. Note that the optimum approach speed increases with the braking rate. The normal rate (100%) is 1.3 m/s^2 (3.0 mphps).

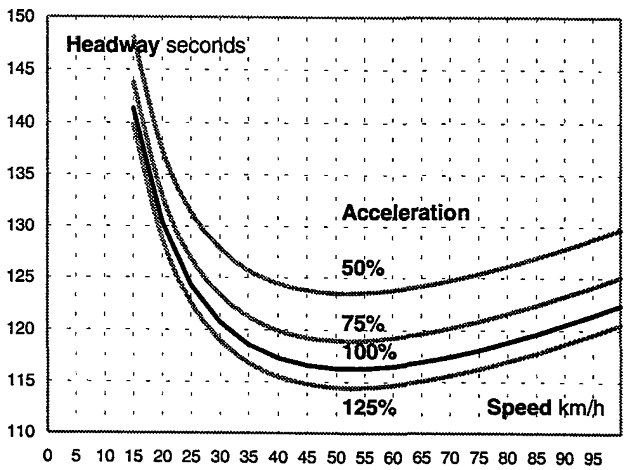


Figure 3.9 Headway changes with the acceleration rate

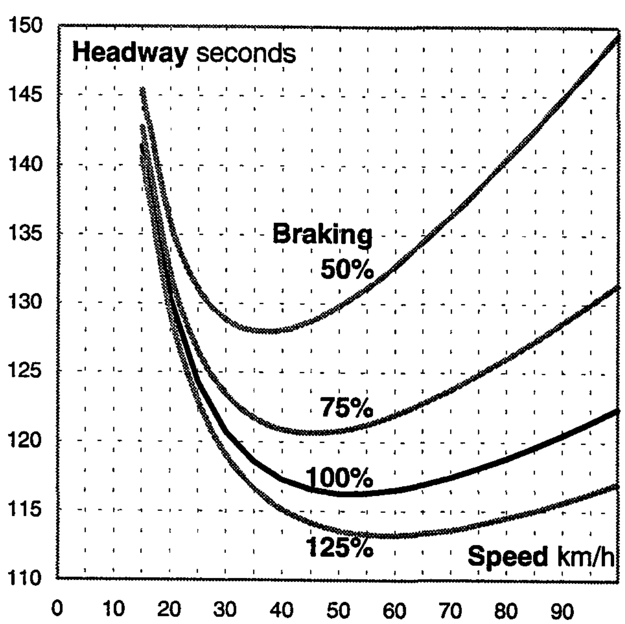


Figure 3.10 Headway changes with the braking rate

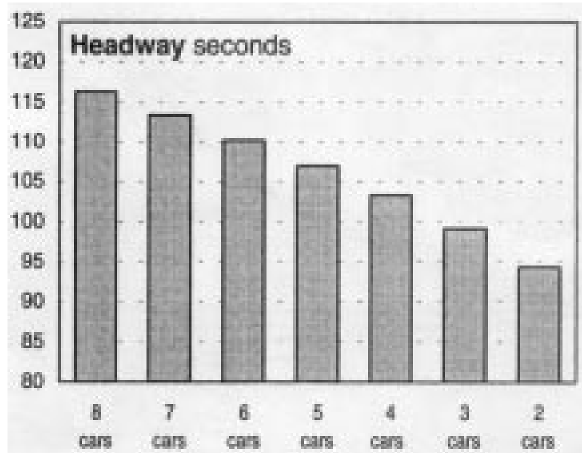


Figure 3.11 Headway changes with train length

3.7.5 TRAIN LENGTH

All previous work in this section has used a maximum train length of 200 m (660 ft). Shorter trains will permit closer train spacing as shown in Figure 3.11.

3.8 MOVING-BLOCK THROUGHPUT

Moving-block signaling systems can use a fixed safety separation distance, plus the calculated braking distance, to separate trains, or a safety distance that is continually adjusted with speed and grades. In this section both approaches will be developed and compared.

3.8.1 FIXED SAFETY DISTANCE

The minimum station headway for the close-in operation is expressed in Equation 3-15. For a moving-block signaling system there is no requirement for a train to travel its own length and vacate the station platform before freeing up a block for the following train. Rather, the moment a train starts from a platform the distance so freed is added to that available for the following train to proceed.

The term for the time to clear the platform block can be removed. The safety separation constant *B*—a surrogate for the number of blocks between trains can be set to zero. The fixed safety distance can be added to the train length to produce a term that represents the time to travel both the train length plus the fixed safety distance. The overspeed acceleration time equivalent and time constant terms can be removed—allowance for runaway propulsion is included in the fixed safety distance. The overspeed time can similarly be deleted.

The other factors in the equation should remain. The braking reaction time can be adjusted for the specific equipment. The station headway Equation 3-15 is shown below with the main components identified

$$H(s) = \sqrt{\frac{2(L+D)}{a_s}} + \frac{L}{v_a} + \left(\frac{100}{K} + B\right)\left(\frac{v_a}{2d_s}\right) + \frac{a_s t_{os}^2}{2v_a}\left(1 - \frac{v_a}{v_{\max}}\right) + t_{os} + t_{jl} + t_{br} + t_d + t_{om}$$

Station close-in head-way	Time to vacate platform	Time to travel own length	Brake system level %	Safe separa-tion factor	Brak-ing time	Time equivalent of overspeed acceleration	Over-speed time	Jerk limit time	Brake system reac-tion time	Station dwell time	Operat-ing margin time

where
- $H(s)$ = station headway *in seconds*
- L = length of the longest train *in meters*
- D = distance from front of stopped train to start of station exit block in meters
- v_a = station approach speed *in m/s*
- v_{max} = maximum line speed *in m/s*
- K = braking safety factor—worst case service braking is K% of specified normal rate—typically 75%
- B = separation safety factor—equivalent to number of braking distances (surrogate for blocks) that separate trains
- t_{os} = time for overspeed governor to operate on automatic systems—to be replaced with driver sighting and reaction times on manual systems (seconds)
- t_{jl} = time lost to braking jerk limitation—typically 0.5 seconds
- t_{br} = brake system reaction time—*older air brake equipment only (seconds)*
- t_d = dwell time *(seconds)*
- t_{om} = operating margin *(seconds)*
- a_s = initial service acceleration rate *in m/s²*
- d_s = service deceleration rate *in m/s²*

The final four time constants can be abbreviated so that

$$\Sigma t = t_{jl} + t_{br} + t_d + t_{om} \qquad \text{Equation 3-18}$$

The adaptation of Equation 3-15 for a moving-block signaling system with fixed safety separation becomes

$$H(s) = \frac{L + S_{mb}}{v_a} + \frac{100}{K}\left(\frac{v_a}{2d_s}\right) + \sum t \qquad \text{Equation 3-19}$$

where S_{mb} = moving-block safety distance

The calculation of the appropriate safety distance is described by Motz[(R47)]. The process is complicated and requires judgment calls on how to represent the worst case situation. The final figure may involve compromises involving decisions of the appropriate government regulatory body (if any) and/or the rail transit system executive.

The Vancouver SkyTrain moving-block signaling system uses a short safety distance of 50 m (165 ft), reflecting the short trains and high levels of assured braking from magnetic track brakes and motor braking—both independent of traction power. The resultant throughput is high and becomes limited by station dwells, junctions and issues of operational allowances.

Safety distances for more conventional equipment are triple or quadruple, particularly if there are significant grades. In these circumstances a variable safety distance will increase the throughput.

This alternate approach develops an approximation for a safety distance that adjusts with circumstances. In this case the assumption is made that the safety distance comprises the braking distance (i.e., B = 1) plus the runaway propulsion components and a positioning error distance—all adjusted for any downgrade into the *headway critical station.*

Discounting grades for the moment the station headway can be represented by:

$$H(s) = \frac{L + P_e}{v_a} + \left(\frac{100}{K} + B\right)\left(\frac{v_a}{2d_s}\right) + \frac{a_s t_{os}^2}{2v_a}\left(1 - \frac{v_a}{v_{max}}\right) + t_{os} + \sum t$$

Equation 3-20

where P_e = positioning error
B = 1

Adjusting for a the grade into a headway critical station, the service acceleration should be increased by one hundredth of the force of gravity for each percentage of grade, and the service braking rate reduced similarly. Thus the acceleration rate is multiplied by (1 - gG/100) where g is the acceleration due to gravity (9.807 m/s2) and G is the percentage grade—negative for downgrades. This adjustment approximates to (1 - 0.1G). The result becomes

$$H(s) = \frac{L + P_e}{v_a} + \left(\frac{100}{K} + B\right)\left(\frac{v_a}{2d_s\,(1 + 0.1G)}\right) + \frac{a_s\,(1 - 0.1G)t_{os}^2}{2v_a}\left(1 - \frac{v_a}{v_{max}}\right) + t_{os} + \sum t$$

Equation 3-21

The results of this equation are shown in Figure 3.12 using data from Table 3.3 with B = 1 and a positioning error of 6.25 m (21 ft). The resultant minimum headway of 97 sec occurs at an approach speed of 56 km/h (mph). The respective curves for a conventional three-aspect signaling system and a cab control system are included for comparison. As would be expected, a

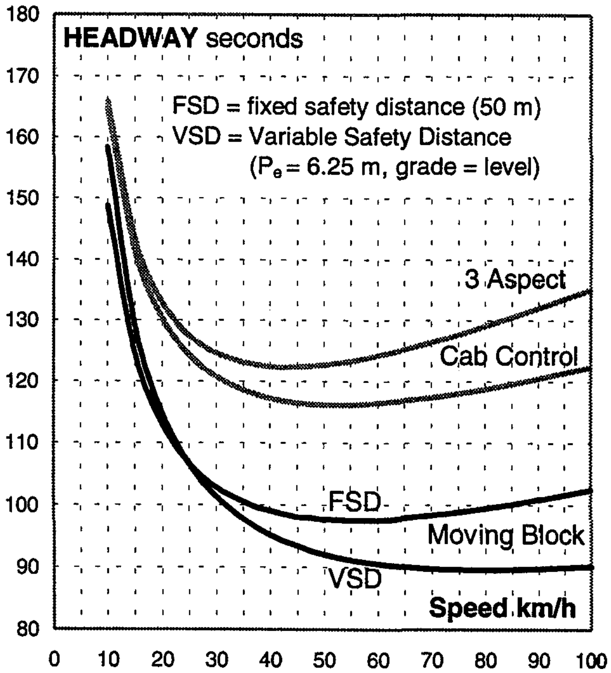


Figure 3.12 Moving-block headways with 45-sec dwell and 20-sec operating margin compared with conventional fixed-block systems

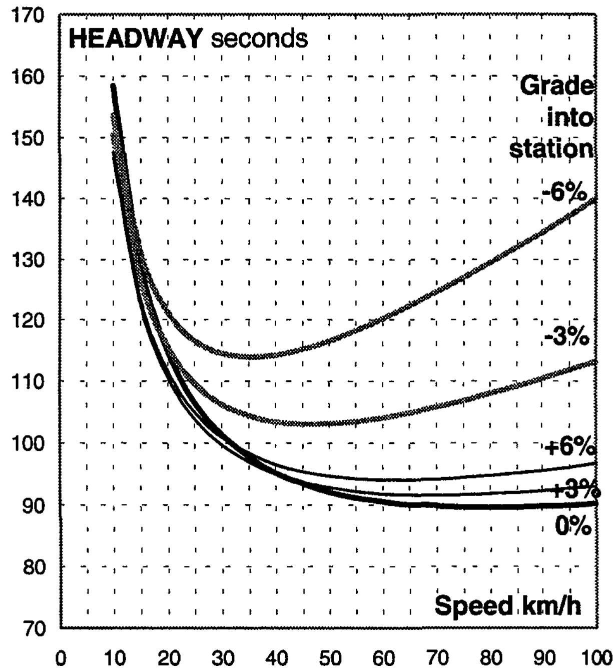


Figure 3.13 Effect of grades on a moving-block signaling system with variable safety distance

moving-block system with a speed variable safety distance shows the lowest overall headway. The difference between the two methods of determining the safety distance represents an eight second difference in the minimum headway—pointing out the importance of selecting the best method when a close headway is required.

The elasticity of moving-block headways with respect to voltage fluctuations will be negligible as the time to clear the plat-form is not a component in calculating the moving-block signaling system headway. The effect of grades is shown in Figure 3.13.

Downgrades (negative) into a station significantly reduce the minimum headway while positive grades have little effect.

3.9 TURN-BACK THROUGHPUT

Correctly designed and operated turn-backs should not be a constraint on capacity. A typical minimal terminal station arrangement with the preferred[41] center (island) platform is shown in Figure 3.14. The worst case is based on the arriving

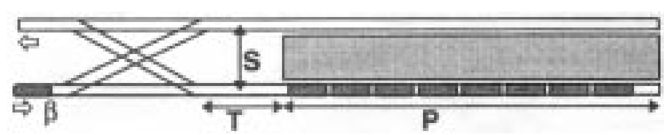


Figure 3.14 Terminal station track layout[42]

train (lower left) being held at the cross-over approach signal while a train departs. It must, moving from a stop, traverse the cross-over and be fully berthed in the station before the next exiting train (lower right) can leave. The distance involved is

$$D_a = P + T + CS \qquad \text{Equation 3-22}$$

where
- D_a = approach distance
- P = platform length
- T = distance from cross-over to platform
- S = track separation (≅ platform width +1.6m)
- C = switch angle factor
  - 5.77 for #6 switch
  - 6.41 for #8 switch
  - 9.62 for #10 switch

The time for this maneuver is expressed as

$$t_a = 2\sqrt{\frac{2D_a}{a_s + d_s}} = 2\sqrt{\frac{2(P + T + CS)}{a_s + d_s}} \qquad \text{Equation 3-23}$$

where
- t_a = approach time
- a_s = initial service acceleration rate in m/s^2
- d_s = service deceleration rate in m/s^2

[41] While side platforms reduce the track to track centers and so reduce the maneuver time, they require passengers to be directed to the correct platform for the next departing train. This is inherently undesirable and becomes more so when a train cannot depart because of a defect or incident and passengers must be redirected to the other platform.

[42] The diagram shows no run-on space beyond the station platform. Where there is little or no such space, mechanical or hydraulic bumpers should be provided.

The distance to exit the station, a straight run, is shorter but the initial acceleration rate will start to taper off. Leaving the travel distance the same to compensate for this, the time for the exiting train to clear the cross-over can be approximated as:

$$t_e = \sqrt{\frac{2(P + T + CS)}{a_s}} \qquad \text{Equation 3-24}$$

In between these two travel times is the terminal time that includes the dwell for alighting and boarding passengers, the time for the train operator to change ends and conduct any necessary inspections and brake tests, the time for the crossover switches to move and lock plus any desired schedule recovery time.

With two terminal tracks, the headway restriction is half the sum of these time components, expressed as:

$$H(t) \geq \frac{t_t + t_e + t_a}{2} + t_s \qquad \text{Equation 3-25}$$

where $H(t)$ = terminal headway time
t_a = terminal approach time
t_e = terminal exit time
t_t = terminal layover time
t_s = switch throw and lock time
(all in seconds)

Determining the terminal layover time is difficult. An approach is to look at the maximum terminal layover time for a given headway by transposing Equation 3-24.

$$t_t \leq 2(H(t) - t_s) - t_e - t_a \qquad \text{Equation 3-26}$$

The maximum terminal layover time can then be calculated. With the following typical *worst case* parameters:

where the headway = 120 sec
train length = 200 m
track separation = 10 m
distance from cross-over to platform = 20 m
initial service acceleration rate = 1.3 m/s^2
service deceleration rate = 1.3 m/s^2
switch is #10
switch throw and lock time is 6 sec

the terminal time $t_t \leq 175$ sec. This would increase by 9 sec if the incoming train did not stop before traversing the cross-over. While this is not a generous amount of time, particularly to contain a schedule recovery allowance, many systems maintain such close headways with minimal delays.

This maximum permitted terminal time can be calculated for the specific system and terminal parameters. Where the time is insufficient there are numerous corrective possibilities. These include moving the cross-over as close to the platform as possible— note that structures can restrict the cross-over location in subways.

The full terminal layover time is available for station dwell. If passenger movement time is a limiting factor then this can be reduced with the use of dual-faced platforms. At terminals with exceptionally heavy passenger loading, multiple track layouts may be needed. An atypical alternative, used at SEPTA's 69th Street; PATH's World Trade Center termini; and the Howard, Desplaines, and 54th St. CTA Stations is the use of loops—with the exception of several examples in Paris this is rare for rail transit.

Crew turnaround time can be expedited with set-back crewing. At a leisurely walking pace of 1 m/s, it would take 200 sec for a driver to walk the length of a 200 m train, more if the driver were expected to check the interior of each car for left objects or passengers. Obviously this could not be accommodated reliably in a 175-sec terminal layover time.

Terminal arrangements should accommodate some common delays. An example would be the typical problems of a train held in a terminal for a door-sticking problem; waiting for police to remove an intoxicated passenger—euphemistically termed a *sleeper*; or for a cleaning crew. Alternately one track may be preempted to store a bad order train. On these occasions the terminal is temporarily restricted to a single track and the maximum terminal layover time is reduced to 61 sec with the above parameters (70 sec without an approach stop). This may be sufficient for the passenger dwell but cannot accommodate changing ends on a long train and totally eliminates any schedule recovery allowance.

More expensive ways to improve turn-backs include extending tracks beyond the station and providing cross-overs at both ends of the station. This permits a storage track or tracks for spare and disabled trains—a useful, if not essential, failure management facility. With cross-overs at both ends of the station, on-time trains can turn-back beyond the station with late trains turning in front of the station—providing a valuable recovery time of some 90 sec at the price of additional equipment to serve a given passenger demand.

The above analysis has assumed that any speed restrictions in the terminal approach and exit are below the speed a train would reach in the calculated movements—approximately 21 km/h (13 mph) on a stop-to-stop approach, 29 km/h (18 mph) as the end of the train leaves the interlocking on exit. For safety reasons, some operators have imposed very low entry speeds, occasionally enforced with speed control signaling.

Slow terminal approaches are common on manually driven rail transit systems in the United States. In some cases this approach could be a greater restriction than the start from stop at the approach cross-over represented in Equation 3-24. If an approach speed restriction exists that is less than $(t_a . a_s/2)$ *(m/sec)* then the above methodology should not be used.

3.10 JUNCTION THROUGHPUT

Correctly designed junctions should not be a constraint on capacity. Where a system is expected to operate at close headways, high use junctions will invariably be grade separated. At such *flying junctions*, the merging and diverging movements can all be made without conflict and the only impact on capacity is the addition of the switch throw and lock times, typically 3 to 6 sec. Speed limits, imposed in accordance with the radius of curvature and any superelevation, may reduce the schedule speed but should not raise the minimum headway—unless there is a tight curve close to a headway limiting station.

The capacity of a flat junction can be calculated in a similar manner to the terminal station approach. The junction

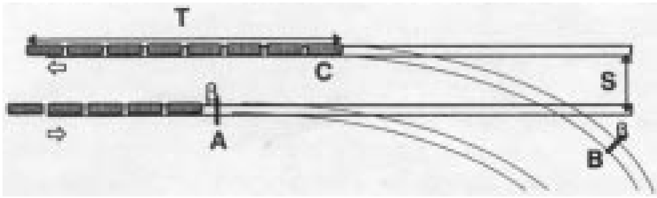


Figure 3.15 Flat junction track layout

arrangement is shown in Figure 3.15. The worst case is based on a train (lower left) held at signal "A" while a train of length "T" moves from signal "B" to clear the interlocking at "C". The minimum operable headway is the line headway of train "A" (from Figure 3.3) plus the time for the conflicting train to clear the interlocking plus the extra time for train "A" to brake to a stop and accelerate back to line speed. Ignoring specific block locations and transition spirals, this can be expressed approximately as:

$$H(j) = H(l) + \sqrt{\frac{2(T + 2CS)}{a_s}} + \frac{v_l}{a_s + d_s} + t_s + t_{om}$$

Equation 3-27

where $H(j)$ = limiting headway at junction *(seconds)*
$H(l)$ = line headway (Figure 3.3) *(seconds)*
T = train length *in meters*
S = track separation *in meters*
C = switch angle factor
5.77 for #6 switch
6.41 for #8 switch
9.62 for #10 switch
a_s = initial service acceleration rate *in m/s*^{2}
d_s = service deceleration rate *in m/s*^{2}
v_l = line speed *in m/s*
t_s = switch throw and lock time *(seconds)*
t_{om} = operating margin time *(seconds)*

The limiting headway at the junction can then be calculated with the following typical parameters:

where line headway = 32 sec
line speed = 100 km/h
train length = 200 m
track separation = 10 m
initial service acceleration rate = 1.3 m/s^2
service deceleration rate = 1.3 m/s^2
switch is #10
switch throw and lock time is 6 sec

The result is a junction limiting headway of 102 sec plus an operating margin. While in theory this should allow a 120-sec headway with a flat junction, it does not leave a significant operating margin and there is a probability of interference headways. General guidance in rail transit design is that junctions should be grade separated for headways below 150 to 180 sec.

An exception is with a moving-block signaling system incorporating an automatic train supervision system with the capability to look forward—and so adjust train performance and station dwells to avoid conflicts at the junction, i.e. trains will not have to stop or slow down at the junction—other than for the interlocking's civil speed limit. In this case, the junction interference headway drops to 63 sec, allowing 120 sec, or slightly lower, headways to be sustained on a flat junction—a potentially significant cost saving associated with a moving-block signaling system.

A real-life example of the restrictions created by junctions is contained in a NYCTA study.[43] This capacity analysis of NYCTA operations focused on the backbone of services in Queens—the Queens Boulevard line to 179th Street. The analysis determined headway constraints due to train performance, the signaling system, and station dwell times. An analysis of the partially flat junction at Nostrand Avenue indicated a throughput that was four trains per hour per single track lower than the 29 to 31 trains per hour that is typically the NYCTA maximum.

3.11 SUMMARY

Using as few approximations as possible, the minimum headway has been calculated for a range of train control systems with a wide number of variables. Table 3.7 summarizes the results including the raw minimum headway with the dwell and operating margin times stripped away.

The spreadsheets contained on the available disk allow the user to change most variables and obtain the minimum headway under a wide range of circumstances.

CAUTION This table and the spreadsheet make assumptions and approximations. The results are believed to be a reliable guide but are not a substitute for a full and careful simulation of the train control system in conjunction with a multiple train performance simulation. To these times approximately 6 seconds should be added for a 4% downgrade into the headway critical station. Three to four seconds can be added to allow for voltage drops at peak times on systems at full capacity—except for the moving-block signaling system.

The results of this chapter concur with field data and agree or are close to the calculations of most other headway determination

[43] As reported by panel member Herbert S. Levinson from the study: BOOZ ALLEN and HAMILTON INC., in association with Abrams-Cherwony; Ammann & Whitney; George Beetle: Merrill Stuart, Queens Transit Alternatives Technical Appendix, Part 5a, Operations/Capacity Analysis, NYCTA, New York, January 1981.

Table 3.7 Headway result summary in seconds with 200 m (660 ft) (8-10 cars) VSD = variable safety distance

Station dwell	0	30	45
Operating margin	0	15	25[44]
3 aspect system	57	102	122
Cab controls	51	96	116
Moving Block-VSD	32	77	102

[44] Perversely, the operating margin should be increased as the dwell time increases

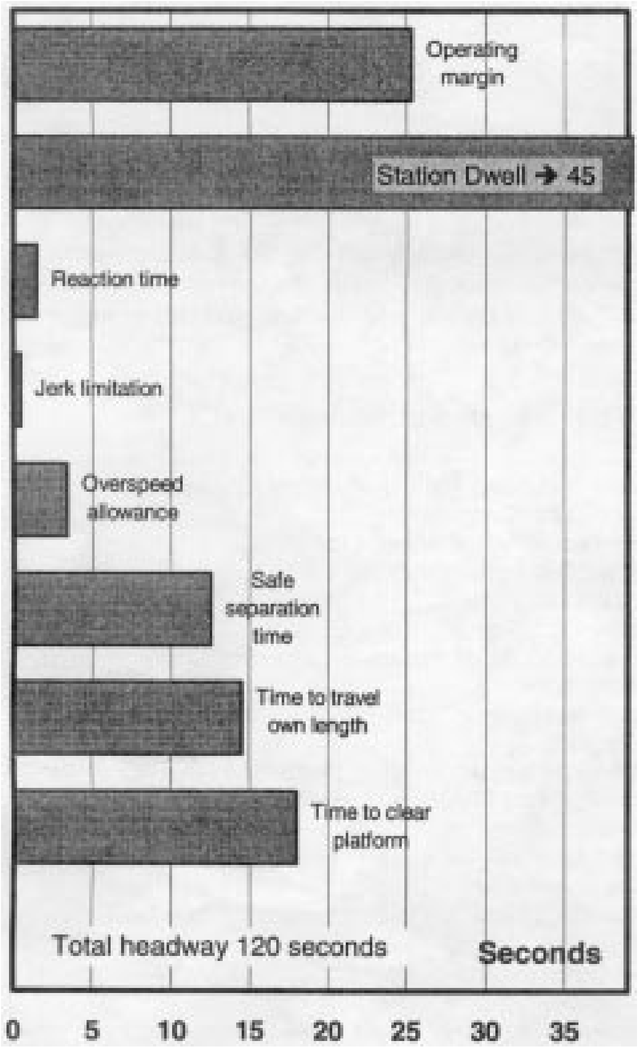


Figure 3.16 Headway components for cab control signaling that comprise the typical North American minimum headway of 120 sec

methods reviewed in Appendix One. Typical cited minimum headways, without dwell or operating margin times, are in the range of 50 to 60 sec for conventional train control—compared to the 51 to 57 sec in the above summary.

Auer[(R09)] estimates that a moving-block system should increase system capacity by 33% based on a 20-sec dwell[45] and 10-sec operating margin. With these quantities the headway of the VSD moving-block signaling systems is 62 sec—providing a capacity increase of 30% over the cab control signaling system value of 81 sec.

This reflects a slightly conservative approach in calculating the moving-block signaling system headway with the safety separation factor "B" set at a full braking distance. "B" can be reduced to less than one. Auer's capacity gain is achieved if "B" is set to 0.77.

The value of "B" can be adjusted for the three types of signaling to calibrate the equations of this chapter with actual field experience or system simulation.

The components of headway for the above mid range cabcontrol data are shown in the Figure 3.16 with a station dwell of 45 sec and operating margin of 25 sec.

The components are shown in the order of Equation 3.15 with terms running from the bottom upwards. Dwell is the dominant component and the subject of the next chapter.

[45] Note that many of the referenced headway analyses use a fixed dwell of 20 or 30 sec. This is rarely adequate. On heavy rail transit systems with long trains running at or below headways of 120 sec the dwell at the headway controlling stations will often reach into the range of 40-50 sec—and so become the largest headway component.

4. Station Dwells

4.1 INTRODUCTION

In Chapter Two, *Capacity Basics*, station dwells were introduced as one of three components of headway. Dwells are the major component of headways at close frequencies as shown in Figure 4.1—based on a heavy rail system at capacity, operating 180-m-long trains with a three-aspect signaling system. The best achievable headways under these circumstances are in the range of 110 to 125 sec.[1] In Chapter Two the concept of *controlling dwell* was also introduced. Controlling dwell is the combination of dwell time and a *reasonable* operating margin—the dwell time during a normal peak hour that controls the minimum regular headway. Controlling dwell takes into account routine perturbations in operations—but not major or irregular disruptions. The sum of controlling dwell and the train control system's *minimum train separation time* produces the maximum train throughput without headway interference.

In this chapter the components of dwell time are examined. The major component—passenger flow time—is analyzed. and methodologies developed for determining passenger flow times and dwell times.

4.2 LITERATURE REVIEW

The literature review produced 26 dwell time references listed in Table 4.1. The full listing is contained in Chapter Twelve, *Bibliography* and a summary of each reference is contained in Appendix One. These references can be divided into three categories. The largest category discussed dwell as a component in calculating train throughput.

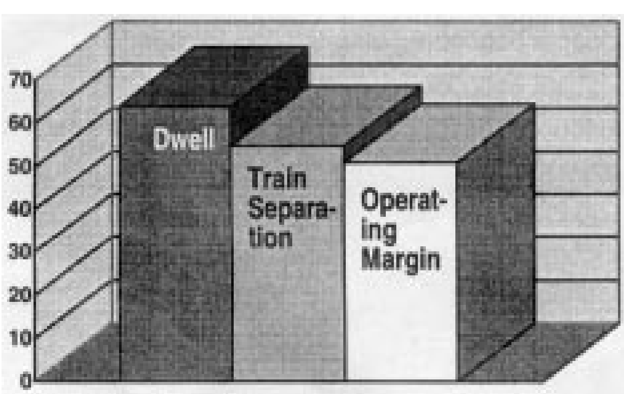


Figure 4.1 Typical headway components in seconds

The second category analyzed dwell time relative to the number of passengers boarding and alighting. This group concluded that linear regression provided the most suitable fit for both rapid transit and light rail with high- and low-level loading for specific systems.[2] Three references improved the data fit by including

Table 4.1 List of dwell time references

Alle, Improving Rail Transit Line Capacity Using Computer Graphics
Anderson, Transit Systems Theory
Auer, Rail-Transit People-Mover Headway Comparison
Barwell, Automation and Control in Transport
Canadian Urban Transit Association, Canadian Transit Handbook
Celniker, Trolley Priority on Signalized Arterials in San Diego
Chow, Hoboken Terminal: Pedestrian Planning
Gray, Public Transportation Planning, Operations and Management
Jacobs Transit Project--Estimate of Transit Supply Parameters
Janelle, Interactive Hybrid Computer Design of a Signaling System
Klopotov, Improving the Capacity of Metropolitan Railways
Koffman, Self-service Fare Collection on the San Diego Trolley
Kraff, Evaluation of Passenger Service Times
Levinson, Some Reflections on Transit Capacity
Levinson, ITE Transportation Planning Handbook Chapter 12
Levinson, Capacity Concepts for Street-Running Light Rail Transit
Lin, Dwell Time Relationships for Light Rail Systems
Miller, Simulation Model of Shared Streetcar Right-of-Way
Motz, Attainable Headways Using SELTRAC
Pushkarev, Urban Rail in America
Schumann, Status of North American LRT Systems
TRB, Collection and Application of Ridership Data on Rapid Transit
TRB, Highway Capacity Manual, Chapter 12
US DoT Characteristics of Urban Transportation Systems
Vuchic, Urban Public Transportation Systems and Technology
Walshaw, LRT On-Street Operations: The Calgary Experience

[1] Some European systems operate three or more aspect signaling systems with headways down to 90 sec by strict control of dwells—on occasion, closing doors before all passenger movements are complete. This is not an acceptable practice in North America.

[2] Lin and Wilson [R44] indicate that crowding may cause a non-linear increase in dwell time during congested periods. Koffman, Rhyner and Trexler,[R33] after testing a variety of variables, including various powers, exponentials, logarithms and interaction terms, conclude that a linear model produced the best results for the specific system studied.

the number of passengers on-board a car as a variable. One paper, by Koffman, Rhyner and Trexler(R33), evaluated a variable to account for passenger-actuated doors on the San Diego trolley.

In the third category, a single paper (Alle(R02)) answered two key questions: “How many trains can realistically pass a point in one hour?” and “What is the impact of station dwell times on this throughput?”

Using an at-capacity section of the MTA-NYCT E & F lines, Alle analyzed the actual peak-hour dwells at Queens Plaza Station in New York by trapping 85% of the area under the normal distribution curve. The upper control limit becomes the mean plus one standard deviation with a 95% confidence interval. The results determined that this specific single track, with the given set of dwells, can support trains every 130 sec—almost identical to the actual throughput of 29 trains per hour (124 sec).

Alle’s methodology is based on measurements of actual inservice dwell times, and so it is unsuitable for determining controlling dwells of new systems or new stations added to existing systems where such information would not be available.

With the above exception, the literature offers only methods to determine passenger flow times; no material was found that adjusts these flow times to either the full station dwell time or a controlling dwell time. Many reports, and even some simulations, use a manually input average dwell time, a worst case dwell time, or merely a typical dwell time—often quoted at 15 to 20 sec per station with 30 sec or more for major stations. These gross approximations usually produce a throughput of 40 to 50 trains an hour and so require applying one or more factors to adjust the resultant throughput to the actual North American maximum of 30 to 32 trains an hour.

This situation required the authors to make a fresh start at developing a methodology for calculating dwells. Much of the field data collection involved timing dwells and passenger flows.

4.3 DWELL CONSTITUENTS

Dwell is made up of the time passenger flow occurs, a further time before the doors are closed and then a time while waiting to depart with the doors closed. Figures 4.2 through 4.5 show these dwell components for the peak period of four selected systems. Each of the systems has a different operating philosophy. BART is automatically driven with door closure and departure performed manually; the latter subject to override by the automatic train control. NYCT is entirely manual, subject only to a permissive departure signal. BC Transit is an entirely automatic system with unattended cars; door closing and departure times are preprogrammed. Station dwells are contained in a nonvital table of the train control system and are adjusted by station, destination, time-of-day and day-of-week. The Toronto Transit Commission is also entirely manual but, unlike New York, has recently implemented a safety delay between door closure and train departure on the Yonge subway.

The data collection did not time any delays between a train stopping in a station and the doors opening. Although there were such minor delays, few were long enough to possibly annoy passengers. Delays do occur with passenger-actuated doors used on many light rail systems. These are discussed separately in section 4.4.2 of this chapter.

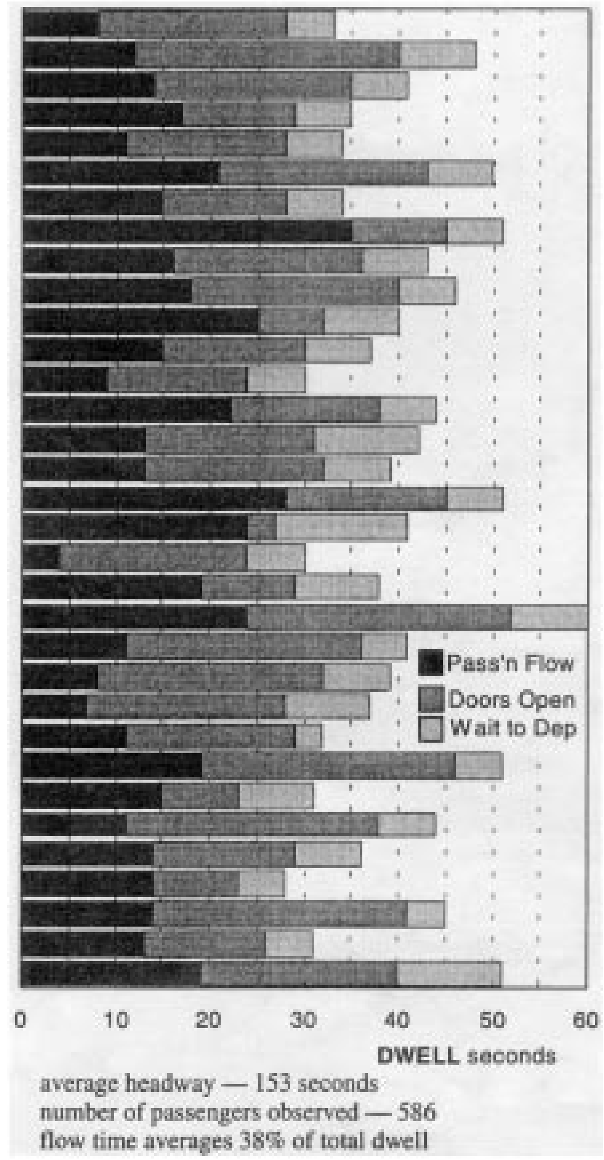


Figure 4.2 BART Montgomery Station dwell time components p.m. peak February 9, 1995

The preprogrammed nature of the BC Transit observations are very evident. There are two services in the data set. The short turn service has shorter dwells until it ends—just over halfway down the chart. Minor variants in the total dwell time for each service are due to observation errors. Data were collected at the heaviest used doorway(s) on the train. While it was not always possible to guarantee that this was selected, it is still surprising that the proportion of dwell time productively used for passenger movements is so small, ranging from 31 to 64% of the total dwell. Only New York fares well in this regard with a percentage of productive time double the other examples. However, there were major variations in the percentage of productive time between stations on the same system (See Table 4.3).

These four charts are representative of 61 data sets of door flows collected in early 1995 for those few systems operated at, or close to, the capacity of their respective train control systems.

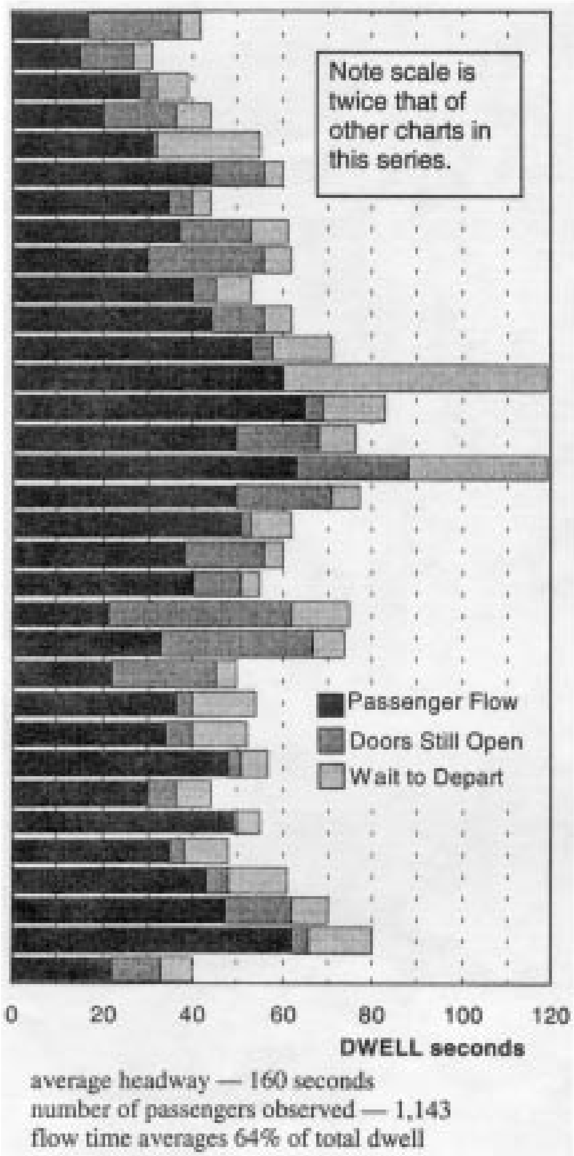


Figure 4.3 NYCT Grand Central Station dwell time components a.m. peak February 8, 1995

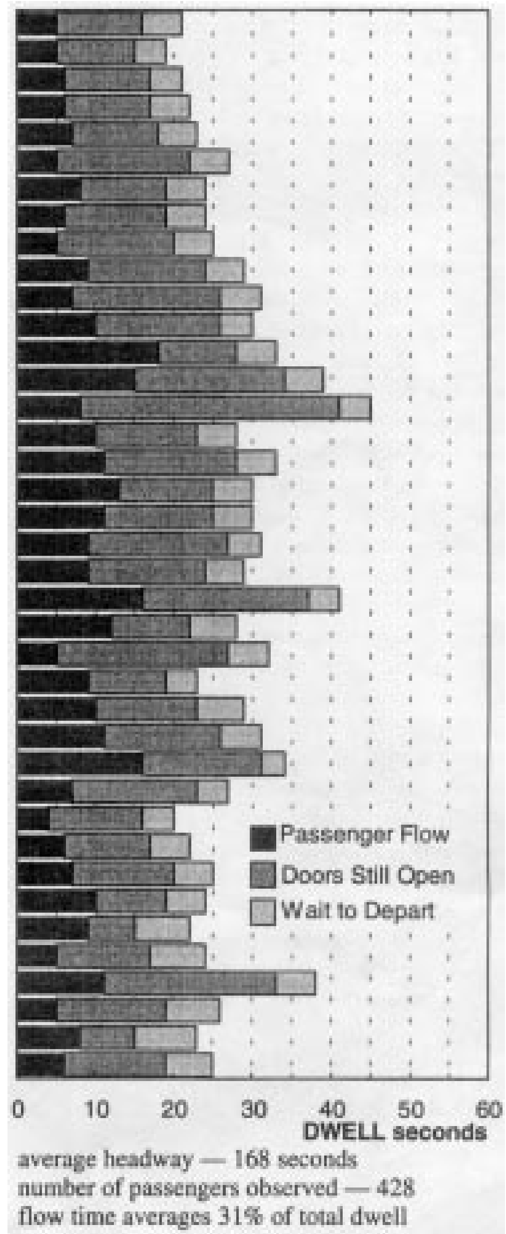


Figure 4.4 Toronto Transit Commission King Station S/B dwell time components: am peak February 6, 1995

The data represent the movement of 25,154 passengers over 56 peak periods, two base (inter-peak) and three special event times, at 27 locations on 10 systems. All data sets are contained on the computer disk. Table 4.2 summarizes the results. The low percentage of dwell time used for passenger flow at the heaviest use door presents a challenge in determining dwell times from the passenger volumes in section of this chapter.

In Chapter Three, *Train Control and Signaling*, it was suggested that automatic driving—when compared with manual driving—should permit a train to run closer to civil speed limits and not commence braking until the last moment, thus reducing train separation by 5 to 15% and increasing capacity by a like amount and improving regularity.

There was insufficient data to confirm this, although Figure 4.5, shows BC Transit's automated operation with a short-turn service integrated into two other services at a very consistent 90-sec separation.

However, the project observers, timing dwells and counting a total of over 25,000 passengers at various locations on 10 systems noted a wide variation in operating practices that ranged from efficient to languid, with automatically driven systems predominantly in the latter group. It would appear that any operating gains from automatic driving may be more than offset by time lost in station dwell practices.

Several light rail and heavy rail systems were notably more expeditious at station dwells than their counterparts, contributing to a faster—and so more economic and attractive—operation. Most automatically driven systems had longer station dwells extending beyond the passenger movement time.

This inefficiency is extending to some manually driven systems where safety concerns have resulted in the addition of an

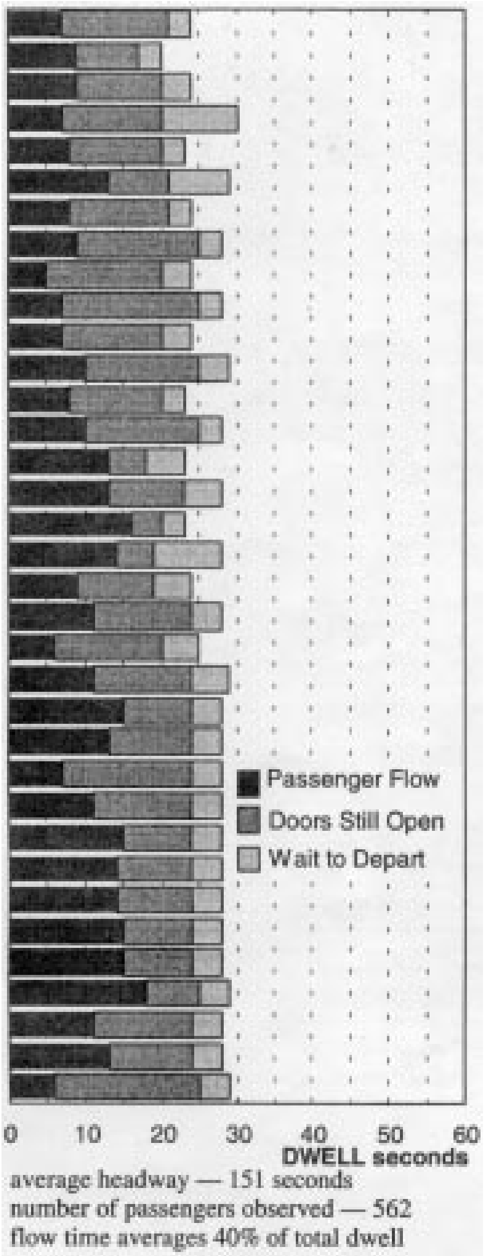


Figure 4.5 BC Transit SkyTrain Burrard Station inbound dwell time components am peak April 5, 1995

Table 4.2 Summary of door observations through one double-stream door during the peak period—four rail transit systems operating at or close to capacity (1995)

SYSTEM	Headway seconds	Total Pass.	% flow/ dwell
BART Montgomery	153	586	38%
NYCT Grand Central	160	1,143	64%
BC Transit Burrard	151	562	40%
TTC King Station	168	428	31%

Table 4.3 Summary of all door observations through a single double-stream door during the peak period (1995)

System Location	Pass.	Headway		% flow/ dwell	
		am	pm	am	pm
BART Montgomery	3400	02:43	02:43	38	37
BART Embarcadero	2298	03:22	02:50	58	43
BCT Burrard	562	02:31		40	
BCT Broadway	257		02:26		34
BCT Metrotown (off-peak)	263		04:01		35
CTS 1st St. SW	298	02:57		30	
CTS 3rd St. SW	339		03:01		26
CTS Heritage	100	05:54		38	
CTS City Hall	201		03:11		30
ETS Central	37	04:39		33	
ETS Churchill	103		04:53		33
NYCT Grand Central (4&5)[3]	3488	02:45	03:09	58	39
NYCT Queens Plaza (E&F)	401	02:15		34	
NYCT Queens Plaza (E&F)	634		02:37		25
PATH Journal Square	478	03:20		23	
PATH Exchange Place	525		01:56		36
Tri-Met 5th Ave. Mall	804	07:28		40	
Tri-Met Pioneer Sq. S.	471		08:22		28
SDT Civic Center	251	06:26		34	
SDT Imperial & 12th	20	07:31		20	
SDT City College	241	07:20	06:40	24	19
SF Muni Montgomery	2748	02:27	02:26	56	45
SF Muni Irving & Arguello	252	04:49		38	
SF Muni Duboce/Church	298		06:10		35
SF Muni 9th & Judah/Irving	176	04:32		37	
TTC King	1602	02:48	02:37	32	37
TTC Bloor	4907	02:42	02:38	52	58
TOTAL—AVERAGES	**25,154**			**43**	**35**

artificial delay between the time the doors have closed and the train starts to move from the platform.

A companion Transit Cooperative Research Program project A-3, TCRP Report 4, *Aids for Car Side Door Observation,* and its predecessor work, National Cooperative Transit Research & Development Program Report 13, *Conversion to One-Person Operation of Rapid-Transit Trains,* address some of these issues but do not examine overall door-platform interface safety or the wide differences in operating efficiency between various light and heavy rail systems. This issue is discussed further in Chapter Eleven, *Future Research.*[3]

4.4 DOORWAY FLOW TIMES

4.4.1 FLOW TIME HYPOTHESES

Flow time is the time in seconds for a single passenger to cross the threshold of the rail transit car doorway, entering or exiting, per single stream of doorway width.

[3] Dwells may be intentionally extended to enable cross-platform connections between local and express trains.

In the course of conducting this study, several interesting conjectures and educated guesses were encountered relating to flow times and rail transit vehicle loading levels. Certain of these suggest the attractiveness of air-conditioned cars on hot days may decrease both doorway flow times and increase the loading level. Similarly with warm cars in cold weather—with loading levels offset by the bulk of winter clothing. While there is some intuitive support for these hypotheses no data were obtained to support them.

Other hypotheses related to different flow times between old and new rail transit systems, for example, that after delays and under emergency operation passengers will load faster and accept higher loading levels. Similar circumstances apply when rail transit is used to and from special events—such as sporting venues.

4.4.2 FLOW TIME RESULTS

Part of the dwell time determination process involves passenger flow times through a train doorway. Data were collected from a representative set of high-use systems and categorized by the type of entry—level being the most common, then light rail with door stairwells, with and without fare collection at the entrance. These data sets were then partitioned into mainly boarding, mainly alighting and mixed flows. The results are summarized in Figure 4.6. The most interesting component of these data is that passengers enter high-floor light rail vehicles faster from street level than they exit. This remained consistent through several full peak period observations on different systems. Hypotheses include brisker movement going home than going to work, entering a warm, dry car from a cold, wet street and, in the Portland light rail case, caution alighting onto icy sidewalks. Balance may also be better when ascending steps than when descending.

The fastest flow time, 1.11 sec per passenger per single stream, was observed on PATH boarding empty trains at Journal Square station in the morning peak. These flow data are consolidated and summarized by type of flow in Figure 4.7. The results show that, in these averages, there is little difference between the high-volume, older East Coast rail rapid transit systems, and the medium-volume systems—newer light rail and rail rapid transit. Doorway steps approximately double times for all three categories: mixed flow, boarding and alighting. Light rail boarding up steps, with exact fare collection, adds an average of almost exactly 1 sec per passenger.[4]

While most field data collection on doorway flow times is from the peak periods, the opportunity was taken on BC Transit's rail rapid transit system to compare peak-hour with off-peak and special event flows, as summarized in Figure 4.8. Project resources prohibited significant data collection at special events and outside peak periods. However, four field trips were made to survey flows and loading levels on BC Transit. One was before a football game, the second before a rock concert. In both cases a single station handled 10,000 to 15,000 enthusiasts in less than an hour. The other data collection trips surveyed a busy

[4] No data were collected for light rail fare payment alighting down steps—a situation unique to Pittsburgh.

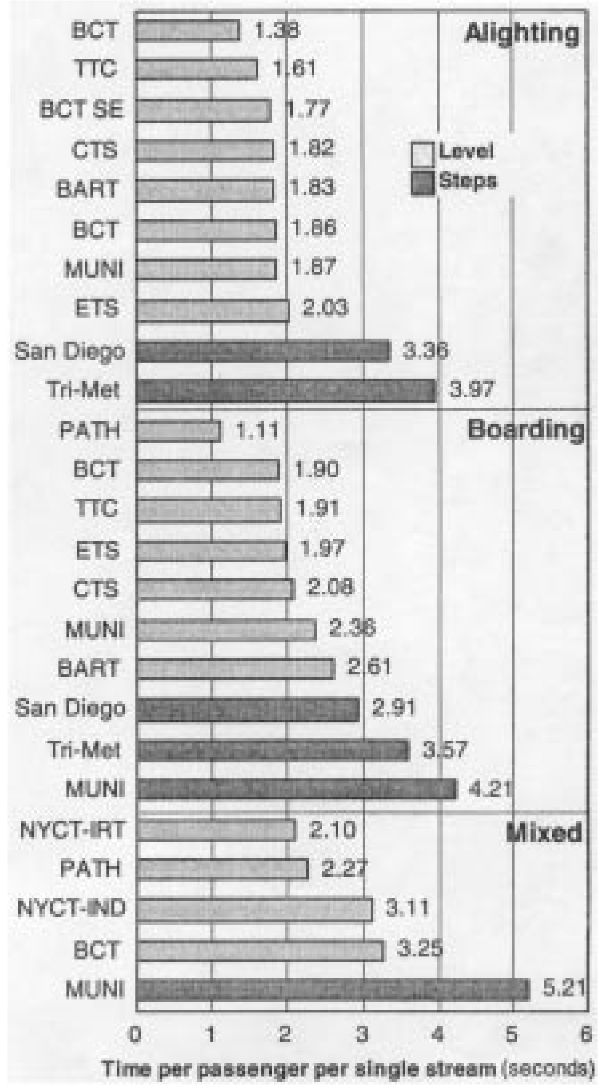


Figure 4.6 Selection of rail transit doorway flow times (1995)

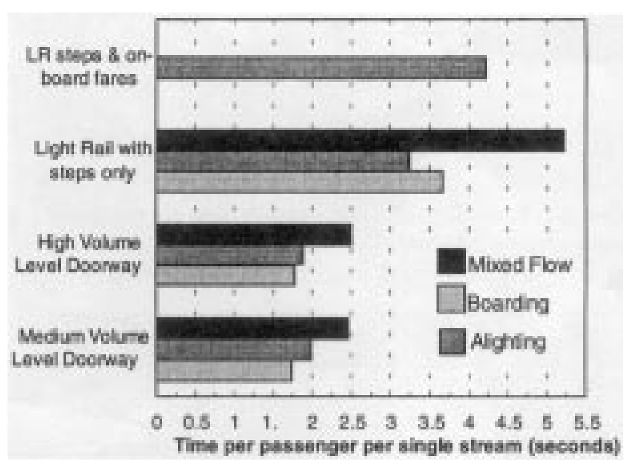


Figure 4.7 Summary of rail transit door average flow times

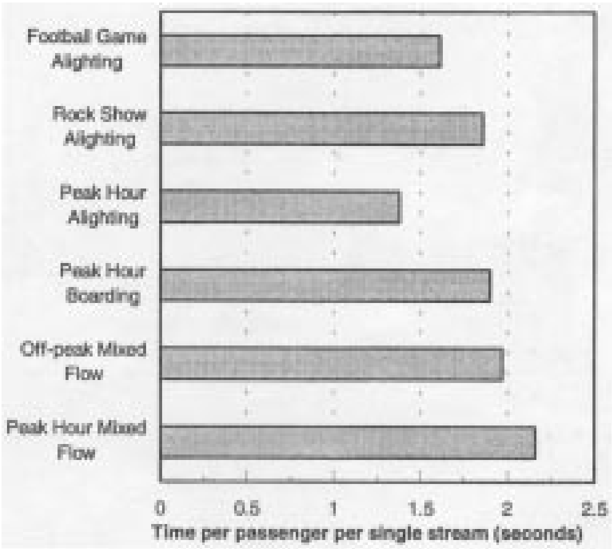


Figure 4.8 BC Transit doorway flow time comparisons (1994-5)

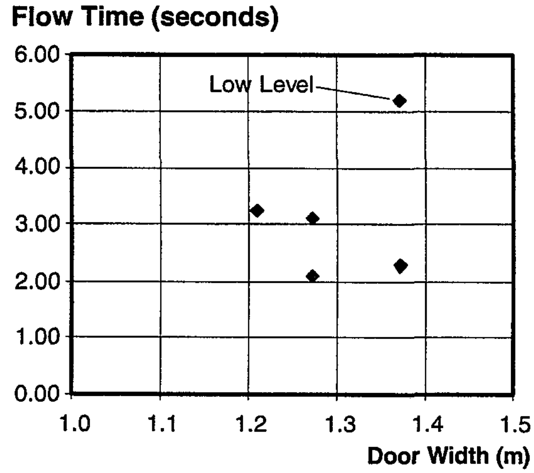


Figure 4.9 Mixed flow times versus door width

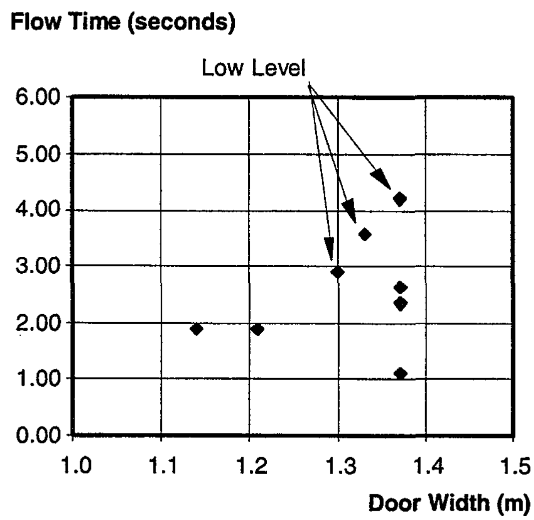


Figure 4.10 Boarding flow times versus door width

suburban station in the early afternoon base (inter-peak) period. The resultant data are contrary to the supposition that special event crowds move faster and that off-peak flows are slower than in the peak hour.

The results showed an increase in alighting flow times before special events. However, loading densities were 20 to 30% higher than during a normal peak hour. This higher level of crowding, together with the fact that many special event passengers are not regular riders, may account for the slower alighting time. Separate BC Transit analysis[R27] has measured car occupancy differences between normal peak-hour operation and after service delays. Standing density increased from a mean of 2.8 passengers per m^2 to 5 passengers per m^2. The equivalent standing space occupied declined from 0.36 m^2 per passenger to 0.2 m^2 per passenger (3.9 to 2.2 sq. ft. per passenger).

Off-peak flows are invariably mixed. The BC Transit off-peak data, an average of 21 trains over a 2-hour period, show faster movement than comparable peak hour mixed flows. However, these data are insufficient to draw firm conclusions.

4.4.3 EFFECT OF DOOR WIDTH ON PASSENGER FLOW TIMES

Figures 4.9, 4.10 and 4.11 plot the relationship between flow times in seconds per passenger per single stream against door width. A variety of statistical analyses failed to show any meaningful relationship between door width and flow time. The only conclusion can be that, within the range of door widths observed, all double-stream doors are essentially equal.

Field notes show that double-stream doors frequently revert to single-stream flows and very occasionally three passengers will move through the doorway simultaneously when one is in the middle and two move—essentially sideways—on either side. At some width below those surveyed a doorway will be effectively single stream. At a width above those surveyed a doorway will routinely handle triple streams. There are no singleor triple-stream doors on any modern North American rail transit vehicle although they exist on AGT and in other countries. JR East in Tokyo is experimenting with a quadruple-stream doorway — shown in Figure 4.12. Wide doors have been a characteristic of the AEG[5] C100 AGT used in many airports and on Miami's Metromover. This four-stream 2.4-m (8-ft) door is shown in Figure 4.13.

4.5 ANALYZING FLOW TIMES

Procedures must be developed that will translate station passenger volumes and flow times per passenger into total doorway use times and then into dwell times. Other work has developed

[5] Previously Westinghouse Electric Corporation.

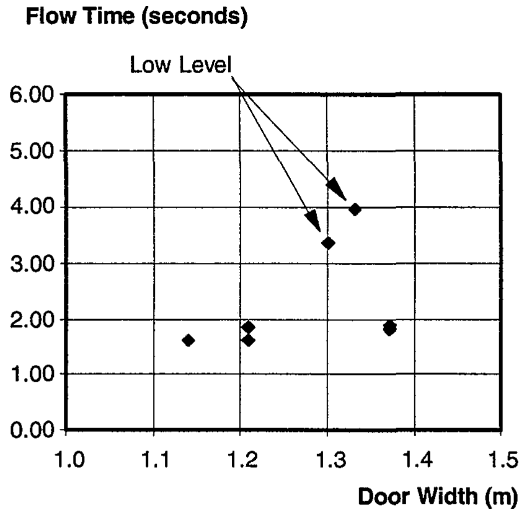


Figure 4.11 Alighting flow times versus door width

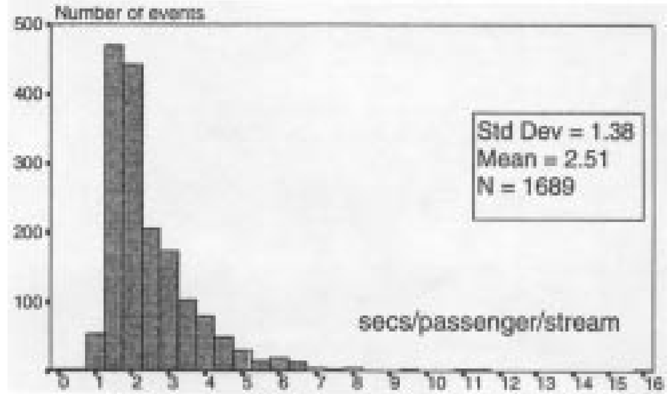


Figure 4.14 Histogram of flow time

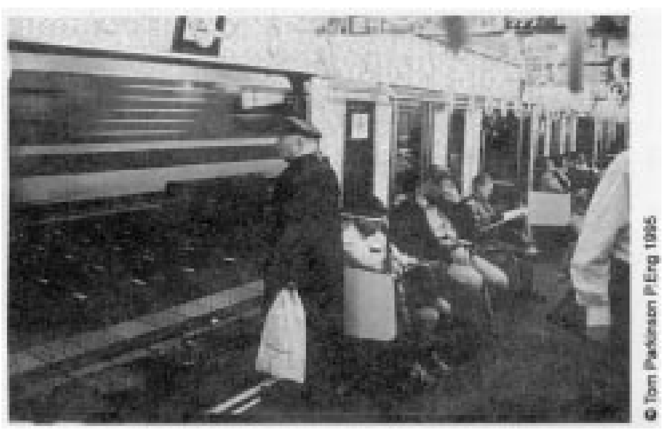


Figure 4.12 Quadruple-stream doorway in Tokyo

Figure 4.13 Quadruple-stream doorway, Miami Metromover

relatively simply linear regression formulae with slight improvements in fit using quadratic terms and the number of passengers remaining on-board—a relatively crude surrogate for the level of doorway congestion. Most work in this area has been restricted to limited amounts of data from a single system.

Linear regression would also be possible for the more extensive data collected during this project. However an examination of these data indicated that separate regression equations would be required for each system—and even for different stations and different modes, alighting, boarding and mixed, on a single system. This is undesirable and unsuitable for determining the capacity of new rail transit systems where regional transportation models provide an estimate of hourly passenger flow by station, from which dwell times must be estimated.

tThe project's statistical advisory team pursued the goal of a single regression formula for all systems with level loading, accepting the need for variations between mainly alighting, mainly boarding and mixed passenger flows. The result, in the following sections of this chapter, involves relatively erudite statistical analysis. The only satisfactory results required logarithmic transforms. Readers may elect to skip the remainder of this chapter. Section 7.5.3 in Chapter Seven offers simpler methods to estimate station dwell times and presents the results of the following work in a simplified manner. The computer spreadsheet allows the calculations to be carried out without any knowledge of the underlying methodologies.

4.5.1 DATA TRANSFORMATION

To assess the distribution of the flow time (seconds/passenger/single stream), the explicit outliers (5 zero times and one time of 36.0) were removed. The histogram in Figure 4.14 shows a clear skewing. In the next step logarithmic transformations were made of the flow times to obtain a normally distributed set of data.

This is achieved by a power transformation technique due to Box and Cox, which raises the flow time to a power determined by an algorithmic procedure. The procedure chooses the power to get a best fit (i.e., minimize the residual sum of squares due

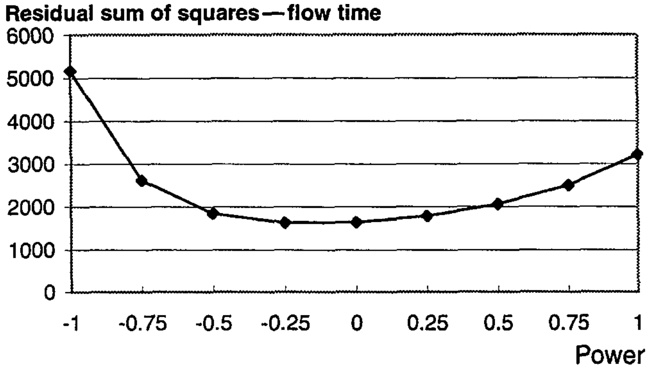


Figure 4.15 Residual sum of squares

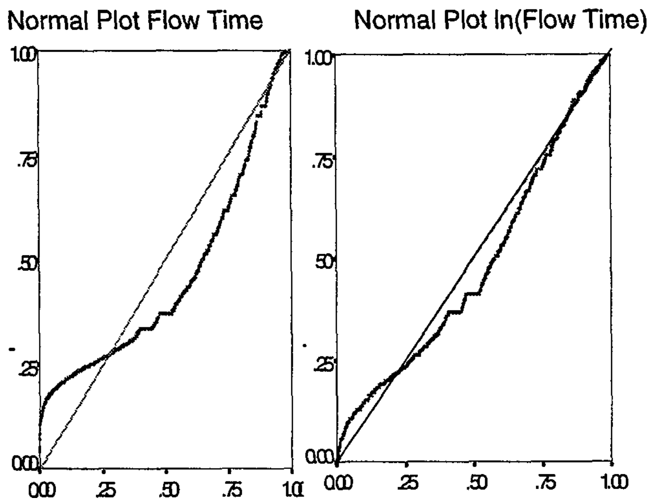


Figure 4.16 Expected flow time cumulative probabilities versus observed cumulative probabilities (abscissa)

to error) in a typical regression. The results of these calculations are shown in Figure 4.15.

This graph indicates that a power of -0.25 or 0 is appropriate. For ease of interpretation a power of zero, which corresponds to a natural logarithm (ln) transform, is preferable. Further calculation shows that this transformation is statistically warranted. Confirmation of this decision can be seen by comparing the normal probability plots obtained from regressions of flow time and ln(flow time) against time of day, shown in Figure 4.16.

4.5.2 COMPARISONS

Box plots are the easiest way to visually compare the natural log transformed flow time data between cities, time of day, loading levels and event types. These plots enable the researcher to quickly compare the central values (the mid box horizontal line is the median) and gauge the spread of the data (the box represents the interquartile range; i.e., the top is the 75th percentile and the bottom is the 25th percentile).

Analysis of variance is used to examine differences in the mean value of ln(flow time) between different levels of a variable (e.g., by system).

Table 4.4 Overall data set summary (seconds)

Variable	Mean	SD	Min	Max	No.
Flow Time	2.51	1.38	.08	16.00	1689
ln(Flow Time)	.81	.44	-2.48	2.77	1689

Table 4.5 System comparison summary ln(flow time (secs))

	Mean	Standard Deviation	Cases
BART	.6939	.4415	297
NYCT	.8893	.3640	254
PATH	.7277	.4345	128
Portland	1.2990	.2788	34
San Diego	1.1208	.2771	105
SF Muni	1.0042	.5346	393
Vancouver	.7437	.3333	155
TTC	.5449	.2178	323

There are highly significant differences between the cities ($p<0.0001$) which are enumerated in the following table. An 'x' indicates a difference significant at the 5 percent level between the cities.

Table 4.6 Significant differences between systems

#	System	Mean	1	2	3	4	5	6	7	8
1	TTC	0.54								
2	BART	0.69	x							
3	PATH	0.73	x							
4	Vancouver	0.74	x							
5	NYCT	0.89	x	x	x	x				
6	SF Muni	1.00	x	x	x	x	x			
7	San Diego	1.12	x	x	x	x	x	x		
8	Portland	1.30	x	x	x	x	x	x	x	

RESULTS

Overall Descriptive statistics for the overall data set are as follows: where SD or Std Dev = standard deviation, No. = Number of observations or Cases, ln = natural logarithm.

City/system comparison In this comparison all data are used and the descriptive statistics for the eight systems are as follows (Table 4.5): There are highly significant differences between the cities ($p<0.0001$), which are enumerated in the Table 4.6. An 'x' indicates a difference significant at the 5 percent level between the cities.

Alighting/boarding comparison All trains with greater than or equal to 70% boarding passengers were declared to be boarding

Table 4.7 Alighting/boarding comparison

	Mean	Standard Deviation	No.
Board	0.9021	0.4994	605
Alight	0.6806	0.3787	442

The mean natural log of the flow time was significantly (p<0.0001) less for alighting.

Table 4.8 Time of day comparison ln(flow time (secs))

	Mean		Cases
am	.8389	.4443	804
pm	.7891	.4431	885

The morning mean natural log of the flow time was mildly significantly (p=0.02) higher than that in the afternoon.

Table 4.9 Loading level comparison ln(flow time (secs))

	Mean	Standard Deviation	Cases
High	.6939	.4254	222
Low (1 door)	1.5944	.3516	29
Low (>1 door)	1.3688	.3596	142

There were significant differences in the mean natural log of the flow times between each pair of loading levels (p<0.05).

Table 4.10 Event time comparison ln(flow time (secs))

	Mean	Standard Deviation	Cases
Normal	.8823	.3275	91
Special	.5466	.2260	64

The special event log flow time was significantly (p<0.0001) lower than that during normal peak time

and similarly those with greater than or equal to 70% alighting passengers were declared to be alighting. This reduced the data set to 1047 cases with descriptive statistics as follows (Table 4.7): The mean natural log of the flow time was significantly (p<0.0001) less for alighting.

Time of day comparison All data were used in comparing am and pm natural log flow times. The descriptive statistics are as follows (Table 4.8): The morning mean natural log of the flow time was mildly significantly (p=0.02) higher than that in the afternoon.

Loading level comparison In order to have a homogeneous dataset for comparing the effect of boarding levels, attention was restricted to the SF Muni datasets. The following descriptive statistics were calculated (Table 4.9). There were significant differences in the mean natural log of the flow times between each pair of loading levels (p<0.05).

Event Time Comparison In order to have a homogeneous dataset for the comparison of the normal and special event times, attention was restricted to the Vancouver Sky Train (Table 4.10). The special event log flow time was significantly (p<0.0001) lower than that during normal peak times. Figures 4.17 through 4.21 show the comparison box plots with the following key.

4.5.3 PREDICTION OF DOOR MOVEMENT TIME USING BOARDING AND ALIGHTING

Preliminary regressions indicate that it is preferable to use the natural logarithm of the door movement (DM) time. This is illustrated in Figure 4.22, where the normal plot for the transformed DM time is much closer to the line of identity, that indicates normality. So, as with the flow time, the natural logarithm of the door movement time is modeled, and the resulting prediction is transformed back to the raw scale by exponentiation. There is evidence (p=0.02) that separate fits are warranted for mainly boarding (i.e. > 70% boarding), mainly alighting (i.e. > 70% alighting) and mixed.

A number of parameterizations and combinations of the two independent variables, number boarding (B) and number alighting (A) are possible. The coefficients of determination for the various models are shown in the following table. The coefficient represents the proportion of variation in the data that is explained by the model. In addition to these parameterizations, the natural logarithm of the numbers boarding and alighting were considered, and dummy variables were used to model the levels resulting from a discretization of the variables. However, these latter approaches did not provide better fits than those above and so were not considered further.

The models were applied to the overall dataset and the three mutually exclusive subsets of mainly boarding (i.e. > 70% boarding), mainly alighting (i.e. > 70% alighting) and mixed; results are shown in Table 4.11. From the table, it can be seen that there are gains of up to 16% in the proportion of variation explained by considering separate models for the subsets of mainly boarding, mainly alighting and mixed. The gains in considering more complex models than the simple additive linear model (Model 1) are less clear.

There is little gain from introducing a term for the interaction between the number boarding and the number alighting as in model 2. However, there is an approximate gain of 10 percent, resulting from the introduction of quadratic terms in model 3, but no further gain from adding an interaction to this as in model 4. Similarly, there is no gain from higher order terms and interactions, which also tend to make the prediction more unstable. Hence the quadratic model (Model 3) is chosen as the best fit, explaining 50% to 80% of the variation in the data.

Residual plots from the regression with this quadratic model show an inverse fanning indicating that the residuals are inversely proportional to the logarithms of the flow times. While this could be transformed toward an identical error structure, in the interests of parsimony, no reparameterization of the logarithm of the flow time is attempted. The Durbin-Watson statistic ranges between 1.3 and 1.6 indicating significant first-order positive auto correlation among the residuals and so standard errors for parameters and associated tests must be viewed with some caution.

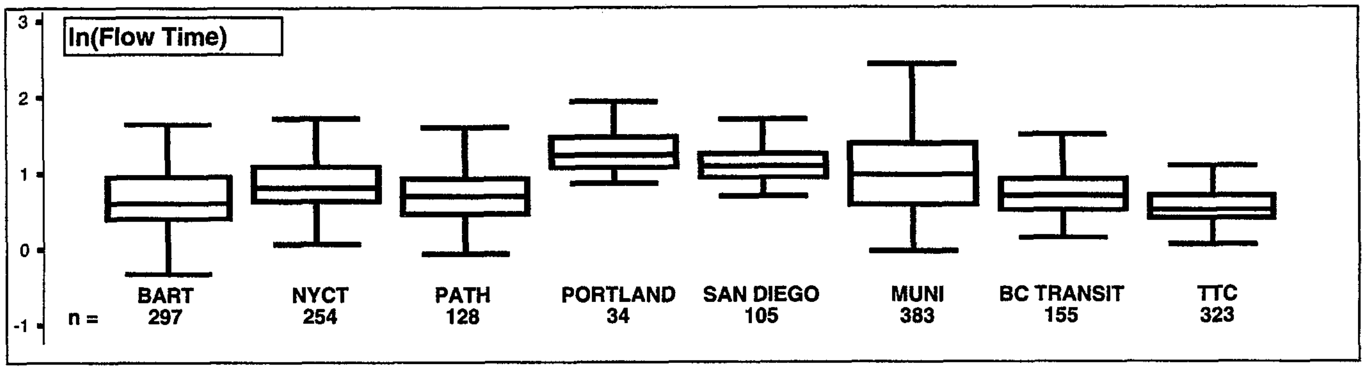
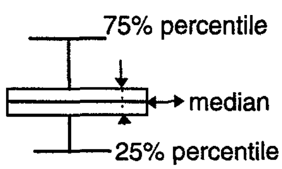


Figure 4.17 City/company comparison



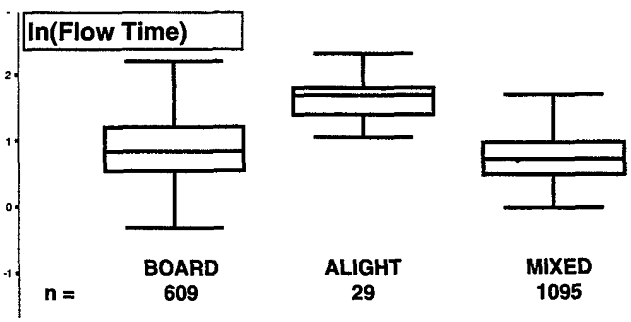


Figure 4.18 Alighting/boarding comparison

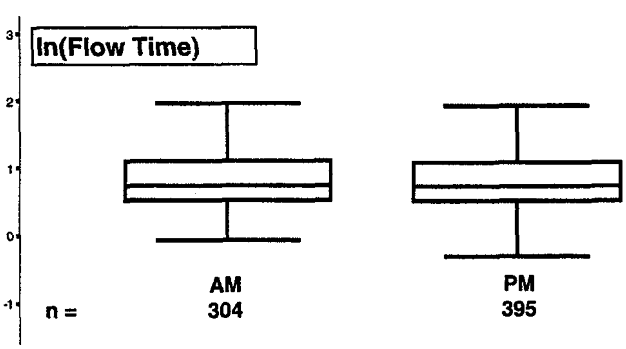


Figure 4.19 Time of day comparison

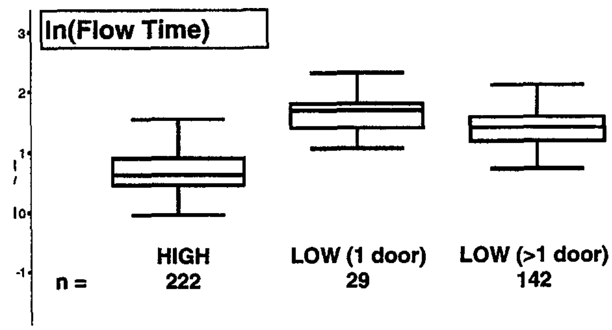


Figure 4.20 Loading level comparison

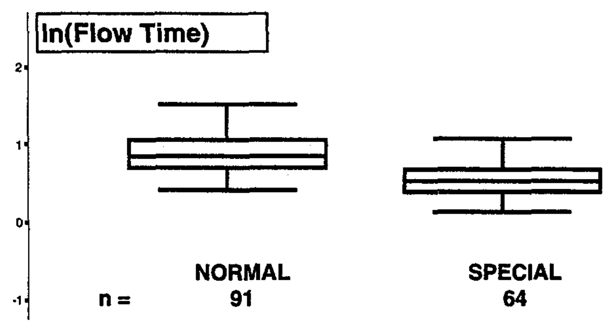


Figure 4.21 Event time comparison

KEY

FT = Flow Time = the time in seconds for a single passenger to move through a single-stream doorway

DM Time = Doorway Movement Time, the time in seconds a single doorway is used for all continuous passenger movements during a single dwell

A = number of passengers alighting and;

B = number of passengers boarding through a single stream level loading rail transit car doorway

SN = number of standing passengers on-board the surveyed car at the end of the dwell

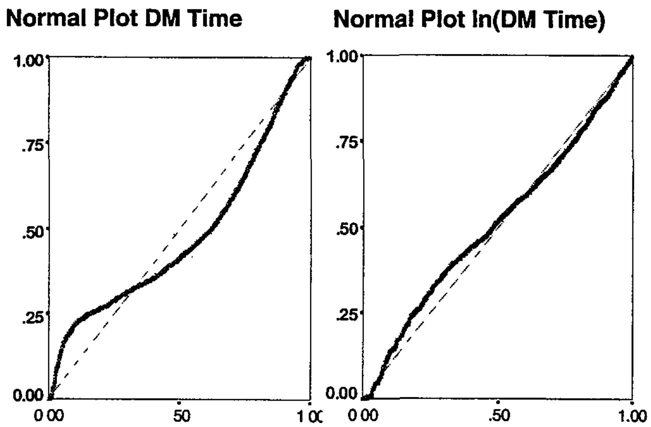


Figure 4.22 Expected cumulative probabilities versus observed cumulative probabilities of door movement time and ln(door movement time)

Table 4.11 R^2 data for tested models 1-4

Model number and terms	Overall	Mainly Boards	Mainly Alights	Mixed
	n=1749	n=667	n=425	n=657
1—B, A	.59	.43	.63	.71
2—B, A, B*A	.60	.43	.63	.75
3—B,A,B^2,A^2	.66	.55	.71	.78
4—B,A,B^2,A^2,B*A	.69	.56	.71	.79

Table 4.12 Flow time regression results for model 3

Model terms	Overall	Mainly Boards	Mainly Alights	Mixed
	n=1749	n=667	n=425	n=657
Constant	1.514	1.380	1.440	1.368
B	0.0987	0.124	0.0979	0.112
A	0.0776	0.0722	0.0922	0.0948
B^2	-0.00159	-0.00214	-0.00103	-0.00225
A^2	-0.000985	-0.000857	-0.00116	-0.00184

The final regression models are presented in Table 4.12. All coefficients are highly significant ($p<0.001$), except for A^2 in the mainly boarding dataset ($p=0.2$), and B^2 ($p=0.6$) in the mainly alighting dataset. Expressed as equations these are

$$\textit{ln(flow time overall)} = 1.514 + 0.0987B + 0.0776A - 0.00159B^2 - 0.000985A^2$$

$$\textit{ln(flow time mainly boarding)} = 1.380 + 0.124B + 0.0722A - 0.00214B^2 - 0.000857A^2$$

$$\textit{ln(flow time mainly alighting)} = 1.440 + 0.0979B + 0.0922A - 0.00103B^2 - 0.00116A^2$$

$$\textit{ln(flow time mixed)} = 1.368 + 0.112B + 0.0948A - 0.00225B^2 - 0.00184A^2$$

Table 4.13 R^2 data for tested models 5-7

Model number and terms	Over-all	Mainly Boards	Mainly Alights	Mixed
n=	963	249	178	531
5—B, A, B^2, A^2, SN	.78	.76	.64	.81
6 — B, A, B^2, A^2, SN, SN^2	.79	.76	.64	.81
7 — B, A, B^2, A^2, SN, SN^2, A*SN, B*SN	.80	.77	.64	.81

Table 4.14 Doorway movement regression results, model 5

Model terms	Overall	Mainly Boards	Mainly Alights	Mixed
	n=963	n=249	n=178	n=536
Constant	1.412	1.0724	1.302	1.363
B	0.0845	0.124	0.147	0.106
A	0.0890	0.104	0.105	0.0864
B^2	-0.00131	-0.00194	-0.00511	-0.00235
A^2	-0.00149	-0.00153	-0.00165	-0.00159
SN	0.0460	0.0782	0.653	0.0563

4.5.4 PREDICTION OF DOORWAY MOVEMENT TIME USING NUMBER BOARDING AND ALIGHTING PLUS THE NUMBER STANDING

The above quadratic model for the logarithm of the DM time was augmented with the number standing standardized for the floor area of the car (SN) to give model 5. Models 6 and 7 introduce quadratic terms in SN and its interactions with B & A.

Data from BART, MUNI and PATH were not used, thus reducing the car numbers to half of those in the previous section. Table 4.13 presents the coefficients of determination for these models. In comparing these models to model 3 of the previous section, there appear to be gains for the mainly boarding and mixed models. However, there is no point in considering more complex models than model 5 which is linear in SN. The residual analyses show similar characteristics to the model without the standardized number standing, so once again all standard errors must be viewed with some caution. The final regression models are presented in Table 4.14. All regression coefficients are highly significant ($p<0.001$) except for B ($p=0.006$), B^2 ($p=0.6$) and SN ($p=0.009$) in mainly alighting dataset. Expressed in equation form the models are

$$\textit{ln(flow time overall)} = 1.412 + 0.0845B + 0.0890A - 0.00131B^2 + 0.00149A^2 + 0.0460SN$$

$$\textit{ln(flow time mainly boarding)} = 1.0724 + 0.124B + 0.104A - 0.00194B^2 - 0.00153A^2 + 0.0782SN$$

ln(flow time mainly alighting) = 1.302 + 0.147*B* + 0.105*A* - 0.00511B^2 - 0.00165A^2 + 0.653*SN*

ln(flow time mixed) = 1.363 + 0.106*B* + 0.0864*A* - 0.00235B^2 - 0.00159B^2 + 0.0563*SN*

where *B* and *A* are the numbers boarding and alighting and *SN* is the number standing normalized for floor area.

This model, with examples, is demonstrated in the computer spreadsheet. The model has limitations and becomes inaccurate with values of A or B > 25.

4.5.5 PREDICTION OF DWELL TIME FROM DOORWAY MOVEMENT TIME

As shown in Figure 4.23 it is desirable to transform the dwell time using natural logarithms, since the normal plot is considerably straighter, indicating a progression toward normality. The dwell time is modeled using its natural logarithm and exponentiated back to the raw scale. Examination of interaction terms shows no evidence (p=0.5) of a need to consider separate predictions for the automatic systems (BART and Vancouver's Sky-Train). The coefficient of determination has a value of 0.34 with a linear model and there is no gain evident from considering quadratic terms.

Residual analysis indicates an inverse fanning that will not be corrected for so as to keep the model simple. However, the Durbin-Watson statistic is 1.2 indicating strong positive serial auto correlation, so that all standard errors and associated tests must be viewed with some caution. The final regression model for the natural logarithm of the dwell time is shown in Table 4.15. It is noted that this relationship is not as strong as those in the previous section. The association is displayed in the scatterplot of Figure 4.24. The mean dwell and DM times, together with their standard deviations, are displayed in Table 4.16.

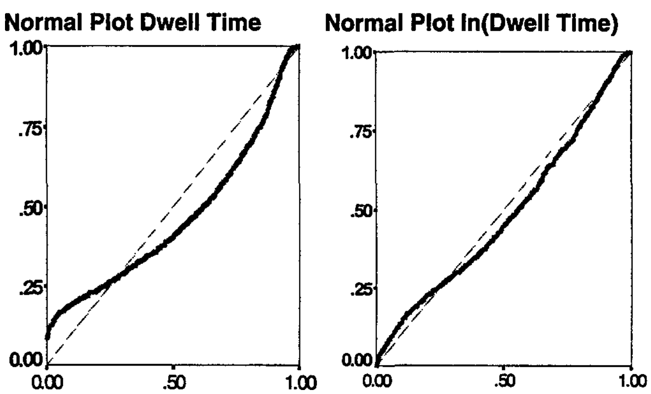


Figure 4.23 Expected cumulative probabilities (ordinates) versus observed cumulative probabilities (abscissa)

Table 4.15 Modeling dwell time on doorway movement time

Model terms	Overall
	n=1661
Constant	3.168
Flow time	0.0254

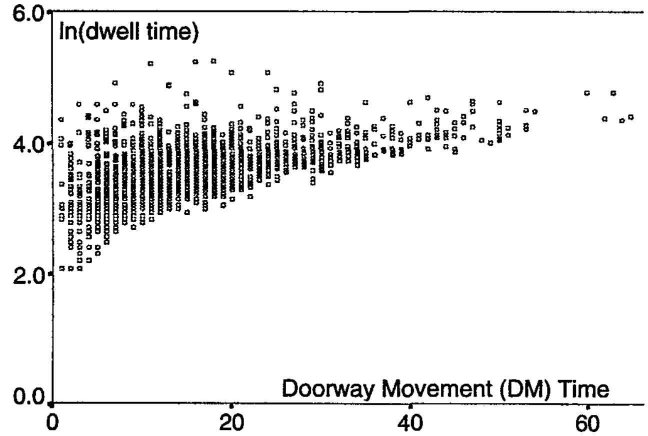


Figure 4.24 Scatterplot of ln(dwell time) versus DM time

Table 4.16 Mean doorway movement and dwell times (with standard deviations) for all data sets of selected systems (s)

	DM Time (secs)		Dwell Time (s)	
	mean	SD	mean	SD
BART	20.1	8.7	46.3	12.0
CTS	9.9	5.0	35.7	15.6
Edmonton	7.7	3.4	24.7	8.8
NYCT	14.5	8.8	30.7	20.9
PATH	20.2	13.5	51.3	22.9
Portland	8.8	4.9	32.0	19.4
San Diego	17.4	5.5	51.1	17.9
SF Muni	11.1	5.8	50.4	21.8
TTC	17.0	11.8	36.6	23.2
Vancouver	14.1	6.6	30.7	7.2

4.5.6 ESTIMATING THE CONTROLLING DWELL

It is usually the longest dwell time that limits the capacity of a rail transit system. This *controlling dwell* is determined at the most heavily used doorway on the peak-15-min train with the highest loading and is typically at the busiest station on the line being examined. Occasionally the controlling dwell may be at other than the busiest station on a line. This can be due to speed restrictions that increase the other headway components at this station or to congestion that increases the passenger doorway movement time—for example platform congestion due to inadequate platform exits, platform obstructions or, at stations with multiple routes, due to passengers waiting for other trains.

There are a number of possible methods for estimating the controlling dwell. In essence, all these methods seek to determine an upper bound for the dwell time below which the bulk of the population falls.

Examples of these methods, comparison with actual field data and suggestions of the most appropriate method to use in different circumstances are discussed in the application chapter: Chapter Seven, *Grade Separated Rail Capacity Determination,* Section 7.5.3 *Determining the Dwell Time.*

ALLE'S METHOD(R02)

This approach focuses on providing a prediction interval for the mean. In other words, in the long run all sample means should fall within these limits 95% of the time. However, it is really a prediction for a typical dwell time that is desired as this will provide the reference limit or bound that is required. As such, Alle's formula seems inappropriate. Moreover it is a nonstandard approach which consists of adding the 95% confidence widths for the distribution of the sample mean and the sample standard deviation. The rationale for adding the confidence width of the sample standard deviation is not clear.

The prediction interval for the sample mean is a random variable itself, and as such, it is possible to construct a confidence interval around it, which may have been the intent. If one were considering the limits for the dwell time of a typical new train, then the variance of the upper prediction limit is approximately $3s^2/n$ where s is the sample standard deviation and n is the sample size. As Alle's method considers a limit for the mean and not a typical unit, it is not considered further.

MEAN PLUS STANDARD DEVIATIONS

This is the traditional approach derived from control theory. It provides a prediction interval for a new train as opposed to one for the mean of all trains. Since it is maximum capacity that is the ultimate objective, only the upper limit is of interest.

A dwell based on the **statistical mean plus one standard deviation** ensures that 83% of the observed data would be equal to or less than this value. A dwell based on the **statistical mean plus two standard deviation** ensures that 97.5% of the observed data would be equal to or less than this value.

Both one and two standard deviations have been used in other work. In either case it is necessary to ensure that the calculated controlling dwell contains sufficient operating margin or allowance to compensate for minor irregularities in operation. With the addition of one standard deviation some additional allowance for operational irregularities is necessary. With two standard deviations the need for any additional allowance is minor or unnecessary

DWELL TIME PLUS AN OPERATIONAL ALLOWANCE OR MARGIN

In many situations, particularly new systems, sufficient data is not available to estimate the dwell standard deviation over a one

Table 4.17 Controlling dwell data limits (seconds)

System	mean	SD	n	m+SD		Operational	
			no. of	upper limit		Mar gin	
	secs	secs	sam-ples	ONE SD	TWO SD	+15 sec	+25 sec
BART	46.3	12.0	290	58.3	70.2	61.3	71.3
CTS	35.7	15.7	91	51.5	67.0	50.7	60.7
ETS	24.7	8.8	18	33.6	42.3	39.7	49.7
NYCT	30.7	20.9	380	51.6	72.6	45.7	55.7
PATH	51.3	23.0	252	64.3	97.3	66.3	76.3
Portland	32.0	19.4	118	51.4	70.8	47.0	57.0
S. Diego	51.1	17.9	34	69.1	86.8	66.1	76.1
MUNI	50.4	21.8	75	72.2	93.9	65.4	75.4
TTC	36.6	23.2	322	59.8	83.0	51.6	61.6
Vanc'ver	30.7	7.2	82	37.9	45.1	45.7	55.7

hour or even a 15 min peak period. In these cases or as an alternate approach an operational allowance or margin can be added to the estimated dwell time due to a specific volume of passenger movements. The figures for the controlling dwell are listed in Table 4.17 using both the mean plus one or two standard deviations and the mean plus operational allowances of 15 and 25 sec.

Chapter Six, *Operating Issues*, discusses the need for, and approaches to, estimating a reasonable operating margin. Application Chapter Seven, *Grade Separated Rail Capacity Determination,* Section 7.5.4, discusses how to select an operating margin in specific cases.

4.6 SUMMARY

The analysis in this chapter has produced methodologies whereby the passenger doorway flow time can be determined from four logarithmic models—overall, mainly boarding, mainly alighting and mixed flow—using as input the number of passenger movements, without reference to a specific mode, system or city.

A fifth model, also logarithmic, but considerably simpler, determines dwell time from passenger doorway flow time. Three alternative methods are then examined to convert the resultant dwell time to the controlling dwell time. The first two methods, traditional dwell plus two standard deviations, which most closely matched the field data, and Alle's method both require information on dwells over the peak hour. This information is not readily available when trying to estimate the capacity of new or modified rail transit systems, leaving the third method, adding an estimated operating margin to the calculated maximum dwell.

These methodologies are deployed in Chapter Seven, *Grade Separated Rail Capacity Determination* and in the spreadsheet as one of several complete methods to calculate system capacity.

5. Passenger Loading Levels

5.1 INTRODUCTION

Establishing the loading level of rail transit is usually the final step in determining capacity—and one of the most variable. After the maximum train throughput has been calculated from the inverse of the sum of signaling separation time, dwell time and operating margin, then capacity is based only on train length and loading level.

It is important to remember the feedback processes; that train length significantly changes the signaling separation time and that loading levels affect dwell times.

The existing loading levels on North American rail transit vary from the relaxed seating of premium service (club cars) operated on specific trains of a few commuter rail lines to the densest loading of an urban subway car in Mexico City—a range of 1.5 to 0.17 m^2 per passenger (16 to 1.8 sq ft).

This wide range is more than eight to one. A more normal loading level range, discounting Mexico City and commuter rail, is two or three to one. This range makes the precise determination of loading level difficult. The main factor is a policy issue, the question of relative comfort—heavily restrained by economic issues.

Notwithstanding Toronto's subway and PATCO's Lindenwold line, the first new rail transit network in North America in the last half century was BART. In the early 1960s, planning for this network—more a suburban railway than an inner-city subway—was based on the provision of a seat for every passenger. Subsequently economic reality has forced acceptance of standing passengers, particularly for shorter trips in San Francisco and through the Transbay tube. Nevertheless, BART remains an example of a system that was designed to, and succeeded in, attracting passengers from alternate modes.

More so now, entering the twenty-first century, than 30 years ago, rail transit is being planned as an alternative to the automobile. While additions to existing systems can be expected to follow existing standards, new systems have to determine their service standards. The principal standards include speed, frequency of service at peak and off-peak times—often termed *policy headways*—and loading levels. Schedule speed is fixed when the alignment, station spacing and equipment specifications are set; headways are usually closely tied to demand, although unmanned trains, as used on Vancouver's SkyTrain and Miami's Metromover, make short, frequent trains over much of the day more affordable. Loading level is the remaining variable. Loading levels and headways interact as more comfortable standards require either longer or more frequent trains.

Demery[(R22)] states:

> *Long before crowding levels......reached New York levels, prospective passengers would choose to travel by a different route, by a different mode, at a different time, or not at all.*

and

> *Outside the largest, most congested urban areas, the level of crowding that transit passengers appear willing to tolerate falls well short of theoretical "design" or "maximum" vehicle capacity.*

These are important issues to consider in establishing loading standards.

In the next section, existing loading standards are reviewed. The remainder of the chapter determines a range of loading standards that can be applied in specific circumstances for each mode.

It is possible to determine the interior dimensions of a rail transit vehicle; subtract the space taken up by cabs, equipment and, for low-loading light rail, stairwells; then assign the residual floor space to seated and standing passengers on the basis of selected densities. This approach is one of several followed in this chapter. However, the recommended method is simply to apply a passenger loading per unit of train length.

5.2 STANDARDS

A 1992 New York City Transit policy paper, *Rapid Transit Loading Guidelines*,[(R48)] gives the loading and service standards that have been applied, with minor modifications, to the New York subway system since 1987. The guidelines provide for slightly more space per passenger than those in effect until 1986. Modifications have allowed for a relaxation in the nonrush hour passenger loading guideline to allow for the operation of short trains.

The loading guidelines were established from test loadings of different car types, loading surveys of revenue service at the peak load point and comparisons with the policies of other rail transit operators. Additional concerns such as passenger comfort, dwell time effects, uneven loading within trains, and an allowance for *slack* capacity in the event of service irregularities and fluctuations in passenger demand were also considered. A rush hour standard of 3 sq ft per standing passenger (3.6 passengers per m^2) was generated from this work. The policy recognizes that this condition is only to be met at the maximum load point on a route and so is effective for only a short time and small portion of the overall route. For comparison, the agency's calculations of the maximum capacity of each car type are based on 6.6 - 6.8 passengers per m^2.

Figure 5.1 compares the loading standards of the older North American subway systems. NYCT standards for loading in the nonrush hours are more generous, with a seated load at the maximum load point being the general standard. If this would require headways of 4 min or less, or preclude operation of short trains, a standard of 125% of seated capacity applies. This

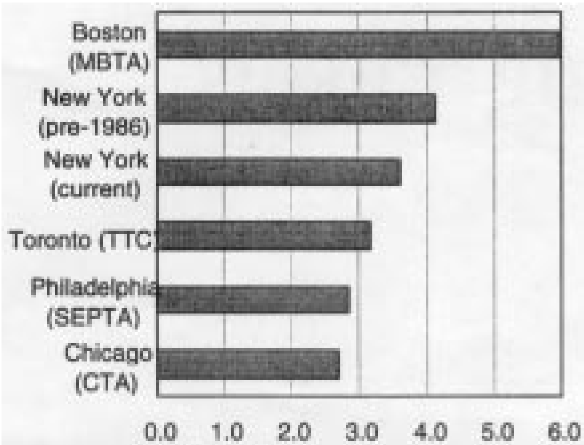


Figure 5.1 Scheduled loading guidelines (passengers/m^2)

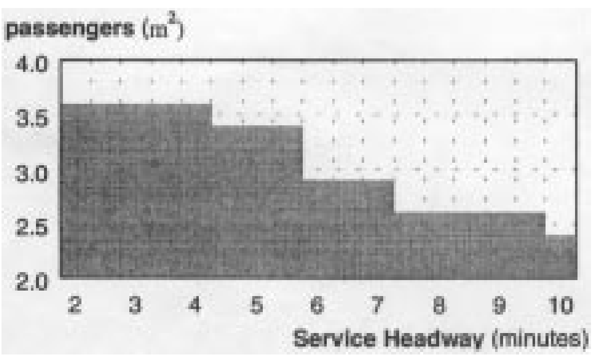


Figure 5.2 New York loading guidelines (passengers/m^2)

Table 5.1 New York policy service levels

Schedule	Time Period	Minimum Headway
Weekday	Rush hours	10 minutes
Weekday	Midday	10 minutes
Weekday	Evening	12 minutes
Saturday	Midday	10 minutes
Saturday	Evening	12 minutes
Sunday	All Day	12 minutes
All days	Midnight	20 minutes

consideration of passenger comfort also extends to rush hour service on lines where the headway is longer than 4 min. In these cases a sliding scale is used to ensure lower standing densities on routes with longer headways, as shown in Figure 5.2. Minimum headways for each day and service period are shown in Table 5.1. The NYCT standard of 3.6 passengers per m^2 can be compared with the average occupancy into the CBD over the peak period as shown in Table 5.2. Table 5.3 tabulates and compares daily and peak-hour ridership and passengers per vehicle for 19 New York CBD trunks for 1976 and 1991. This decrease in NYCT car loadings partly reflected the improvement

Table 5.2 Passenger space on selected US systems[R22]

City	Passengers/m^2 of Gross Floor Space
New York	2.6 into CBD
Chicago	1.5 into CBD
Philadelphia	1.3 into CBD
Boston	2.0 into CBD
San Francisco	1.2 - 1.9
Washington	0.9 - 2.0
Atlanta	1.4 - 1.6
Toronto	1.8 - 2.4
Montreal	2.6 - 3.2

Table 5.3 Changes in NYCT peak-hour car loading[R22]

NEW YORK LOCATION	Average Passengers per car in peak hour		Change 1976-1991
	1976	1991	15 years
IRT Lexington	155	138	-10.97%
IRT Lexington Loc	147	112	-23.81%
IRT Broad Exp.	152	125	-17.76%
IRT Broad Local	104	95	-8.65%
IRT Flushing	116	115	-0.86%
IND Queens	200	195	-2.50%
IND 8th Exp.	146	128	-12.33%
IND 8th Local	91	74	-18.68%
IND 6th Ave	91	99	8.79%
BMT Astoria	129	108	-16.28%
BMT Canarsie	138	113	-18.12%
BMT Jamaica	103	139	34.95%
BMT Man. Bridge	136	119	-12.50%
BMT Montague	106	101	-4.72%
PATH WTC	79	112	41.77%
PATH 33rd	91	91	0.00%
Average[2]	124.4	120.2	-3.30%
Median	129	115	-10.85%

[2] Average and Median include additional data sets.

in service standards of 1987, among other factors. Several trunks continue to operate at or near capacity.[1]

Care should be taken in comparing and applying the service standards with hourly average loadings. Service standards are usually based on the peak within the peak—15 min or less.

A loading diversity factor equating 15-min and peak-hour flows was introduced in Chapter One, *Rail Transit In North America.* Section 5.6 of this chapter discusses the issues of loading diversity, provides data on existing factors by system and mode, and recommends factors for use in capacity calculations. The loading diversity factor for New York trunk routes, shown

[1] Similar comparisons can be made for other cities and earlier years using data from this report and from the TRB's Highway Capacity Manual, Chapter 12 and appendices. Ridership and loading level information in the HCM are based on data to 1976 plus some historic data.[R67]

in Figure 5.3, ranges from 0.675 to 0.925 with an average of 0.817. This diversity must be taken into account to determine peak-hour capacity from a given service standard. NYCT's standard of 3.6 passengers per m^2 over the peak-within-the-peak becomes 3.6 × 0.82 or 2.95 (3.65 sq ft per passenger) on average, over the peak hour.

Outside New York the peak-within-the peak tends to be more pronounced and the peak-hour diversity factor is lower.[3] In part this is due to the long established Manhattan program to stagger work hours and the natural tendency of passengers to avoid the most crowded period—particularly on lines that are close to capacity.

Space occupancy during the peak period on other North American rail transit systems varies widely from below 0.3 passengers per m^2 (3.2 sq ft) to over 1.0 m^2 (11 sq ft) on some commuter rail lines, as shown in Figure 5.4. Note that the highest capacity entry (labeled NYCT) represents two tracks that combine local and express service.

In analyzing this data Pushkarev et al.[(R51)] suggest a standard of 0.5 m^2 (5.4 sq ft) per passenger. This will be discussed in the next section. In addition to standards or policies for the maximum loading on peak-within-the-peak trains and for minimum headways (*policy headways*) at off-peak times, some operators specify a maximum standing time. This is more often a goal rather than a specific standard—20 min is typical.

Commuter Rail Loading levels for commuter rail are unique and uniform. Although standing passengers may be accepted for short inner-city stretches or during times of service irregularities, the policy is to provide a seat for all passengers. Capacity is usually cited at 90 to 95% of the number of seats on the train.

5.3 SPACE REQUIREMENTS

The surveyed literature contains many references to passenger space requirements. The Batelle Institute[(R12)] recommends comfort levels for public transport vehicles. The passenger standing density recommendations are

- COMFORTABLE 2-3 passengers per m^2
- UNCOMFORTABLE 5 passengers per m^2
- UNACCEPTABLE >8 passengers per m^2

In contrast, Pushkarev et al.[(R51)], suggesting *gross vehicle floor area* as a readily available measure of car occupancy, recommends the following standards:

- ADEQUATE 0.5 m^2—provides comfortable capacity per passenger space
- TOLERABLE WITH DIFFICULTY 0.35 m^2—lower limit in North America with "some touching"
- TOTALLY INTOLERABLE 0.2 m^2—least amount of space that is occasionally accepted

Batelle[(R12)] also provides details of the projected body space of passengers in various situations. The most useful of these for

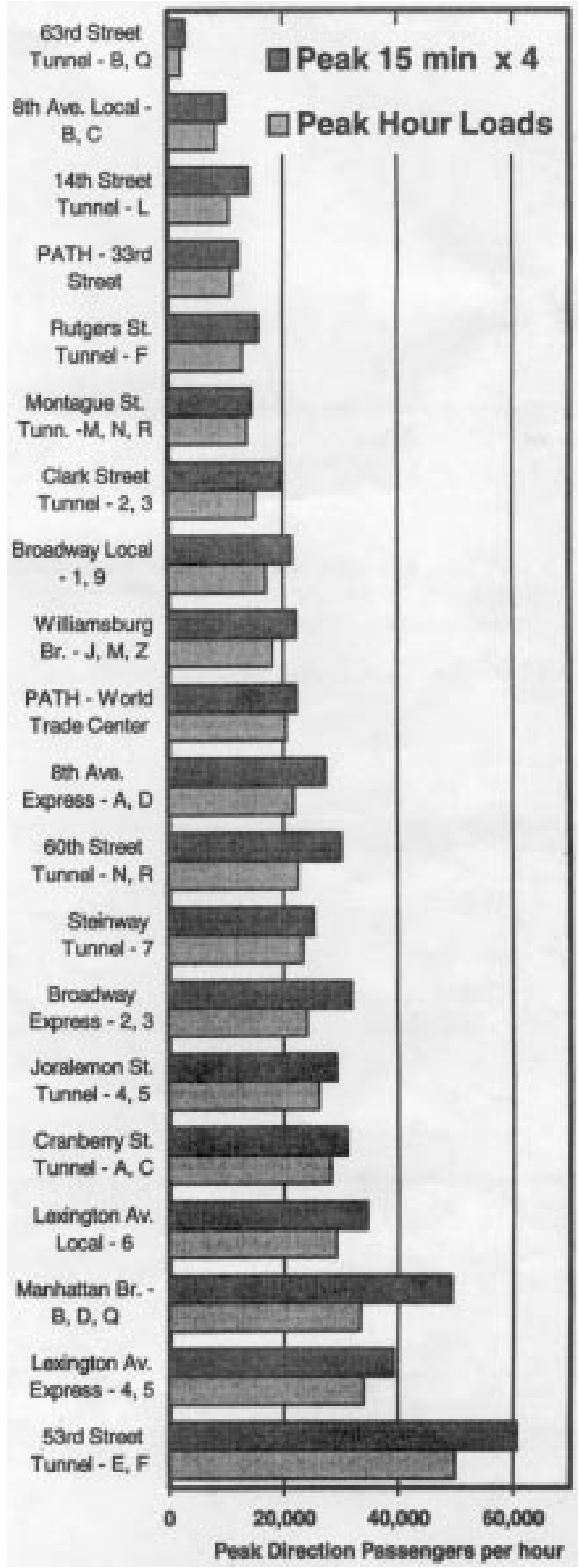


Figure 5.3 15-min peak-within-the-peak compared to full peak-hour ridership on New York subway trunks

[3] Shown in Chapter One, Figures 1.4 and 1.6.

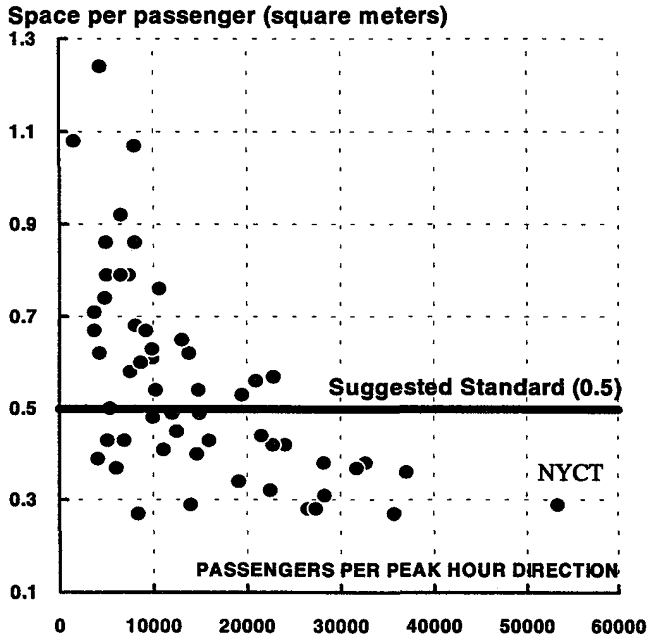


Figure 5.4 Peak-hour space occupancy—all U.S. systems[R51]

Table 5.4 Passenger space requirements [R12]

Situation	Projected Area m^2
Standing	0.13 to 0.16
Standing with briefcase	0.25 to 0.30
Holding on to stanchion	0.26
Minimum seated space	0.24 to 0.30
Tight double seat	0.36 per person
Comfortable seating	0.54 per person

Table 5.5 Passenger space requirements [R30]

CRITERIA	Passenger/ area	Mean space per passenger m^2
Max. practical (NY)	6.0 /m^2	0.17 (1.8 sq ft)
Typical rapid transit	2.2 - 3.6 /m^2	0.34 (3.7 sq ft)
Crush rapid transit	2.6 - 5.4 /m^2	0.26 (2.8 sq ft)
Design rapid transit	1.4 - 4.0 /m^2	0.38 (4.1 sq ft)
Design light rail	2.3 - 4.0 /m^2	0.30 (3.3 sq ft)
Actual light rail	2.9 - 5.7 /m^2	0.25 (2.7 sq ft)
To avoid contact	3.8 - 4.5 /m^2	0.24 (2.6 sq ft)
Unconstrained	1.2 - 2.7 /m^2	0.50 (5.4 sq ft)

rail transit capacity are shown in Table 5.4. The tight double seat corresponds closely to the North America transit seating minimum of 34- to 35-in.-wide double seats on a 27- to 33-in. pitch (0.88 m by 0.76 m)— 3.6 sq ft or 0.33 m^2 per seat.

Jacobs et al. [R30] contains a comprehensive section on vehicle space per passenger, stating that while 53% of U.S. rapid transit lines enjoyed rush hour loadings of 0.5 m^2 per passenger or better, the space requirements shown in Table 5.5 are recom-

Table 5.6 International transit space use [R30]

GROUP	pass/m^2
Some European and most North American	2.0 - 3.0
Some European systems and New York	3.1 - 5.0
Most European large cities	5.1 - 6.0
Large Soviet and Japanese systems	7.1 - 8.0

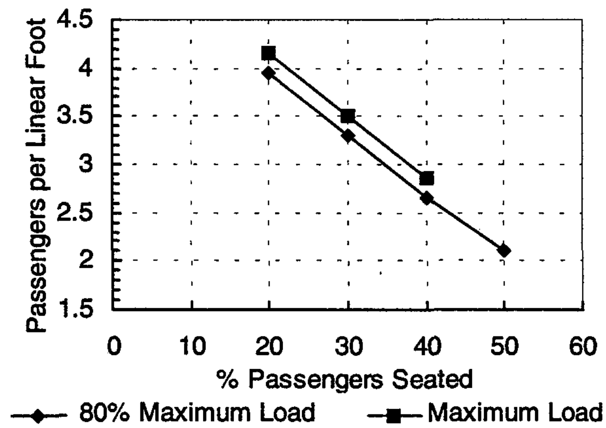


Figure 5.5 Passengers per length of car versus % seated

mended and actual values for the stated conditions. The report is one of the few to discuss the diversity of standing densities within a car—higher in doorways/ vestibules, lower in aisles and at car ends (unless the car has end doors). Table 5.5 is particularly interesting in that the design space allocation for light rail is slightly lower than for heavy rail.

Klopotov[R32] cites typical average peak-hour space requirements from an international survey (Table 5). Lang and Soberman[R39] discuss seating provisions relative to compromises between capacity and comfort. They suggest that all rapid transit cars are substantially similar in width. The report compares passengers per square foot with the percentage seated. This ranges from 0.3 passengers per square foot with 50% seated to 0.6 passengers per square foot with 15% seated. This is then translated into passengers per *linear foot of train,* as shown below in Figure 5.5. The maximum vehicle capacity is 4 passengers per linear foot—approximately 2.5 square feet per passenger. Lang and Soberman also discuss the importance of ease of ingress and egress, recommending minimum distances between seats and doorways and discouraging three abreast seating. Comfort levels are discussed relative to smoothness of operation and the issue of supply and demand. Where systems are oversubscribed and few attractive alternate forms of transportation are available, high levels of crowding will be tolerated. Where systems wish to attract passengers, higher comfort levels, i.e., less crowding, are desirable.

Levinson et al.[R43] and also the Transportation Research Board's Highway Capacity Manual[R67] introduce the concept of loading standards A through F (crush) similar to the alphabetized *level of service* for road traffic. The suggested *schedule design capacity* is 2.8 to 3.3 passengers per m^2, 25% below the "crush" capacity. The peak-hour factor is discussed for 15-min peak-within-the-peak. A range of 0.70 to 0.95 is suggested, approaching 1.0 in large metropolitan areas.

Vuchic(R71) suggests passenger space requirements of 0.30 to 0.55 m^2 per seat and 0.15 to 0.25 m^2 per standee. Vehicle capacity in passenger spaces per vehicle is shown as:

$$C_v = m + \frac{\xi A_g - A_l - m\rho}{\sigma} \qquad \text{Equation 5-1}$$

Where:
- ξ = vehicle floor area loss factor for walls
- A_g = gross vehicle floor area
- A_l = vehicle floor area used for cabs, stairwells and equipment
- m = number of seats
- ρ = floor area per seat
- σ = floor area per standing passenger

Young(R76) discusses a wide range of topics dealing with passenger comfort. He cites the "typical" transit vehicle as allowing 0.40 m^2 (4.3 sq ft) per seated passenger and 0.22 m^2 (2.4 sq ft) per standing passenger. The seating ratio is tabulated for a range of North American and European heavy rail and light rail systems. Heavy rail ranges from 25% to 100% seated and light rail from 40 to 50% in North America to 20 to 44% in Europe. Minimum seating pitch is recommended as 0.69 m (27 in.), 0.81 m (32 in.) to a bulkhead.

Several reports suggest vehicle passenger capacity can be stated as a multiple or percentage of the number of seats. Chapter 12 of the Highway Capacity Manual (R67) develops a measure of seated and total passengers per linear foot of car length, introduced in section 5.5 of this chapter.

Recommendations for a range of loading standards are developed in later sections of this chapter and applied in Chapter Seven, "Grade Separated Rail Capacity Determination," and the report's spreadsheet.

Wheelchairs There was no reference to wheelchair space requirements in the literature—much of which predates the 1991 Americans with Disabilities Act. Although wheelchairs come in several sizes, a common space allowance is 0.55 m^2 (6 sq ft), more for electric chairs and those whose occupants have a greater leg inclination, less for compact and sports chairs.

However, it is not the size of the chair that is a concern as much as the maneuvering and stowage space. Typically a chair occupies the space of a double seat whose seat squab folds up. Restraints and seat belts may be provided but the smoothness of the ride allows most rail transit systems to omit these. In certain vehicle layouts additional seats have to be removed to allow access to the designated wheelchair location.

In optimum designs wheelchair space occupancy should be assigned as the space of a double seat—0.8 m^2 (8.6 sq ft) with a 50% increase considered as an upper limit—1.2 m^2 (13 sq ft) No further allowance is necessary for maneuvering space as this will be occupied by standing passengers when circumstances dictate.

In several rail transit vehicle designs, capacity has actually increased with the removal of seats to provide a designated space for wheelchairs, or, selectively, bicycles. Where the designated space does not involve a fold-up seat the empty space is frequently used by standing passengers or to store baggage, baby strollers etc. Providing locations to store such potential obstacles away from doorways and circulation areas can assist in reducing dwell times.

Wheelchair effects on dwell times are discussed in Chapter Four, *Station Dwells*, and Chapter Eight, *Light Rail Capacity Determination.*

5.4 VEHICLE CAPACITY

In estimating the capacity of a rail transit vehicle one of the following approaches should be selected.

5.4.1 COMMUTER RAIL

Commuter rail capacity is based on the number of seats. Table A 3.5 in Appendix Three lists the dimensions and seating of all rail transit vehicles in North America. A summary extracted from this table is shown in Table 5.7. Commuter rail seating per car ranges from a maximum of 185 to below 60 on certain club cars and combination cars.[5] Seats will be reduced where staff, toilet, wheelchair, baggage or bicycle space is provided. The highest seating densities use 3+2 seating. Although suitable for shorter runs, 3+2 seating is not popular with passengers. The middle of the three-seats is often under utilized and capacity should be factored down accordingly by a suggested further 5%.

Table 5.7 Commuter rail vehicle summary data

System — Car Type[4] Date Built	Length (m)	Width (m)	Seats
ConnDOT Comet II 1991	25.91	3.2	118
GO Transit Bi-Level 77-91	25.91	3.0	162
LACMTA Bi-Level V 92-3	25.91	3.0	148
LIRR M-1 1968-71	25.91	3.28	122
LIRR P-72 1955-56	25.2	3.18	123
MARC Coach 1985-87	25.91	3.2	114
MBTA BTC 1991	25.91	3.05	185
MBTA BTC-1 1979	25.91	3.2	099
Metra CA2E 1978	25.91	3.38	147
Metra Gallery 1995	25.91	3.33	148
Metro-North M-1A B 71	25.91	3.2	122
Metro-North M-6 B 1993	25.91	3.2	106
NICTD EMU-1 1982	25.91	3.2	093
NJT Arrow III 1977-78	25.91	3.2	119
NJT Comet I 1971	25.91	3.2	131
SEPTA SL IV 1973-77	25.91	3.2	127
CalTrain California 1993	25.91	3.05	135
STCUM Gall. Trailer 1970	25.91	3.03	168
Tri-Rail Bi-Level 1988-91	25.91	3.0	162
VRE Trailer 1992	26.01	3.05	120

[4] Bi-level cars are sometimes designated as tri-levelas there is an intermediate level at each end over the trucks.

[5] Not tabulated. Cars with baggage space, crew space or head-end (hotel) power.

Table 5.8 Light Rail Equipment Summary

System Car Designation Date Built	No.	Length (m)	Width (m)	Seats	Total Pass[5]	Door No. Width (m)
Bi-State U2A 1992-93	31	27.28	2.67	72		4—
Calgary U2, U2AC 1980-84, 86	85	24.28	2.66	64	158	4—1.3 m
Edmonton U2 1978-83	37	24.28	2.66	64	140	4—1.3 m
GCRTA Cleveland 800 1981	48	24.38	2.82	84	126	3—
LACMTA LRV 1989-94	69	27.13	2.67	76	137	4—
MBTA LRV Green 1986-88	100	21.95	2.69	50	112	3—
Metrorrey LRV 1990	25	29.56	2.65	58		
Sacramento MTA LRV 1991-93	35	28.96	2.9	85	201	4—
MUNI LRV 1995	40	22.86	2.74	60		
MUNI SLRV 1978	100	21.64	2.69	68		3—
NFTA Buffalo LRV 1983-84	27	20.37	2.62	51	180	2—
NJT PCC 1946-49	24	14.15	2.74	55	125	2—
PAT Pittsburgh U3 1986	55	25.73	2.54	63	125	4—
SCCTA SCLRV 1987	50	26.82	2.74	76	167	4—
San Diego U2 1980-89	71	24.26	2.64	64	96	4—1.3 m
San Diego U2A 1993	52	24.49	2.64	64	96	4—
SDTEO Guadalajara LRV 1989	16	29.56	2.65	52		
SEPTA LRV (S-S) 1980-82	112	15.24	2.59	51		2—
SEPTA N-5 1993	26	19.99	3	60	90	
SRTD U2A 1986-91	36	24.38	2.64	60	144	4—
STE Mexico LRV 1990-91	12	29.56	2.65	46		
Tri-Met LRV 1983-86	26	26.49	2.64	76	166	4—
TTC A-15 (PCC) 1951	22	14.15	2.54	45	103	2—
TTC L-1/2 (CLRV) 1977-81	196	15.44	2.59	46	102	2—
TTC L-3 (ALRV) 1987-89	52	23.16	2.59	61	155	3—

Table 5.9 Heavy rail equipment summary

System, Car Designation Date Built	# of Units	Length (m)	Width (m)	Seats	Total Pass[6]	Door #	Door Width
CTA 2600 B 1981-87	299	14.63	2.84	49	150	2	1.27 m
CTA 3200 (A&B) 1992	256	14.63	2.84	39	150	2	1.27 m
GCRTA Cleveland RT 84-85	60	23.01	3.15	80	128	3	1.27 m
LACMTA HRV 1991-93	30	22.86	3.2	59			
MARTA CQ 310 1979	100	22.86	3.2	68	136	3	1.27 m
MBTA 00600 Blue 1979	70	14.78	2.82	42	94		
MBTA 01200 Orange 1980	120	19.81	2.82	58	132		
MBTA 01400 Red 1962	86	21.18	3.18	54	160		
Metro-Dade Heavy Rail 1984	136	22.76	3.11	76	166	3	1.2 m
MTA Married Pair 1984-86	100	22.76	3.11	76	166	3	1.27 m
NYCT R46 1975-77	752	22.77	3.05	74		4	1.27 m
NYCT R62 1984-85	325	15.56	2.68	44		3	1.27 m
PATCO, PATCO II 1980-81	46	20.68	3.09	80	96	2	1.27 m
PATH PA-4 1986-88	95	15.54	2.81	31	130	3	1.37 m
SEPTA Single End: B-IV 1982	76	20.57	3.09	65	180	3	1.32 m
STCUM MR-73 1976	423	16.96	2.51	40			
TTC H6 1986-89	126	22.86	3.15	76	226	3	1.14 m
WMATA B3000 Chopper 1984	290	23.09	3.09	68	170	3	1.25 m

[6] Total passengers based on the agency's or manufacturer's nominal crush load.

Commuter capacity should be calculated as 90 to 95% of the total seats on a train, after allowing for cars with fewer seats due to other facilities. Where there are high incremental passenger loads for relatively short distances—for example the last few kilometers into the CBD—a standing allowance of 20% of the seats may be considered. However, this is unusual and standing passengers should not normally be taken into account on commuter rail.

5.4.2 EXISTING SYSTEMS

The vehicle capacity on existing systems should be based on actual loading levels of a comparable service. Actual levels on a specific system or line should be adjusted for any difference in car size and interior layout—particularly the number of seats—as outlined in section. If the average occupancy over the peak hour is used then the loading diversity factor should be omitted. If the higher peak-within-the-peak loading is used, then the loading diversity factor should be applied to reach an hourly achievable capacity.

Particular care should be taken in applying any passenger loading level based on car specifications. The often cited *total*, *maximum*, *full* or *crush load* does not necessarily represent a realistic average peak hour or peak-within-the-peak occupancy level. Rather it reflects the specifier or manufacturer applying a set criteria—such as 5 or occasionally 6 passengers per square meter—to the floor space remaining after seating space is deducted. Alternately it can represent the theoretical, and often unattainable, loading used to calculate vehicle structural strength or the minimum traction equipment performance.

Tables 5.8 and 5.9 provide dimensions and capacity information of selected, newer, heavy rail and light rail equipment in North America.

Table A 3.5 in Appendix Three lists the dimensions and seating of all rail transit vehicles in North America.

5.4.3 VEHICLE SPECIFIC CALCULATIONS

Detailed calculations of vehicle passenger capacity are possible, however, given the wide range of peak hour occupancy that is dependent on policy decisions, elaborate determination of interior space usage is generally overkill. Reasonably accurate estimation of vehicle capacity is all that is needed. The following procedures offer a straight forward method.

Converting Exterior to Interior Dimensions

Rail transit vehicle exterior dimensions are the most commonly cited. Where interior dimensions are not available, or cannot be scaled from a floor plan, approximate interior dimensions can be estimated.

Typically the interior width is the exterior width less the thickness of two walls—0.2 m (8 in.). Heavy rail configurations are most commonly married pairs with one driving cab per car. The typical exterior length is quoted over the car anticlimbers. Although cab sizes vary considerably, the interior length can be taken to be 2.0 m (6.7 ft) less than the exterior length. This reduction should be adjusted up to 2.5 m if the exterior dimension are over the couplers and down to 1.5 m if only half width cabs are used, or 0.5 m if there is no cab.

Beware of rare pointed or sloping car ends which require this deduction to be increased. Curved side cars are measured from the widest point—waist level—allowing seats to fit into the curve and so increasing the aisle width. This maximum “waist” width should be used, not the width at floor level.

The first step after obtaining the interior car dimensions is to determine the length of the car side that is free from doorways. Deducting the sum of the door widths, plus a set-back allowance of 0.4 m (16 in.)[7] per double door, from the interior length gives the interior free wall length.

[7] A lower set-back dimension of 0.3 m (12 in.) may be used if this permits an additional seat/row of seats between doorways.

Seating can then be allocated to this length by dividing by the seat pitch:

- 0.69 m (27 in.)[8] for transverse seating
- 10.43 m (17 in.) for longitudinal seating

The result, in lowest whole numbers[9], should then be multiplied by two for longitudinal seating or by 3, 4, or 5, respectively, for 2+1, 2+2 or 2+3 transverse seating. The result is the total number of seats. A more exact method would be to use the specific length between door set-backs. Articulated light rail vehicles should have the articulation width deducted. Four seats can be assigned to the articulation, if desired.

The floor space occupied by seats can then be calculated by multiplying transverse seats by 0.5 m^2 (5.4 sq ft) and longitudinal seats by 0.4 m^2 (4.3 sq ft). These areas make a small allowance for a proportion of bulkhead seats but otherwise represent relatively tight and narrow urban transit seating. Add 10 to 20% for a higher quality, larger seat such as used on BART.

The residual floor area can now be assigned to standing passengers. Light rail vehicles with step wells should have half the step well area deducted. Although prohibited in many systems, passengers will routinely stand on the middle step, squeezing into the car at stops if the doors are treadle operated.

Articulated light rail vehicles should have half the space within the articulation deducted as unavailable for standing passengers, even if the articulation is wider. Many passengers choose not to stand in this space.

Standing passengers can be assigned as follows:

- 5 per square meter (0.2 m^2, 2.15 sq ft per passenger)—an uncomfortable near crush load for North Americans[10] with frequent body contact and inconvenience with packages and brief cases; moving to and from doorways extremely difficult.
- 3.3 per square meter (0.3 m^2, 3.2 sq ft per passenger)—a reasonable service load with occasional body contact; moving to and from doorways requires some effort

[8] Increase to 0.8 m (32 in.) for seats behind a bulkhead

[9] For more accurate results the sidewall should be divided into the lengths between each set of doors (and, when appropriate, between the door and any articulation) and checked, or adjusted, to ensure that an integer of the seat pitch is used. The computer spreadsheet carried this out by dividing the interior free wall length by the number of doorways plus one. The number of integer seat pitches in each space is then determined and used to calculate the total vehicle seating. The appropriate seat pitch is used automatically, 0.43 m for N=2, 0.69 m for N>2.

However, this approach can result in the seating changing radically with a small change in vehicle length, articulation length or door width, any of which are sufficient to add or remove a row of seats between each set of doors. On a four door car with 2+2 seating this results in the seating adjusting up or down by 20 seats at a time—five rows of four seats. Neither Equation 1.3 nor the computer spreadsheet can substitute for a professional interior design, which can optimize seating with a combination of transverse and longitudinal seats. Other design criteria can also be accommodated, including the provision of wheelchair spaces and maximizing circulation space around doorways.

[10] Loading levels of over 6 passengers per square meter are reported on Mexico City’s metro, lines 1 and 3. These are a unique exception in North America.

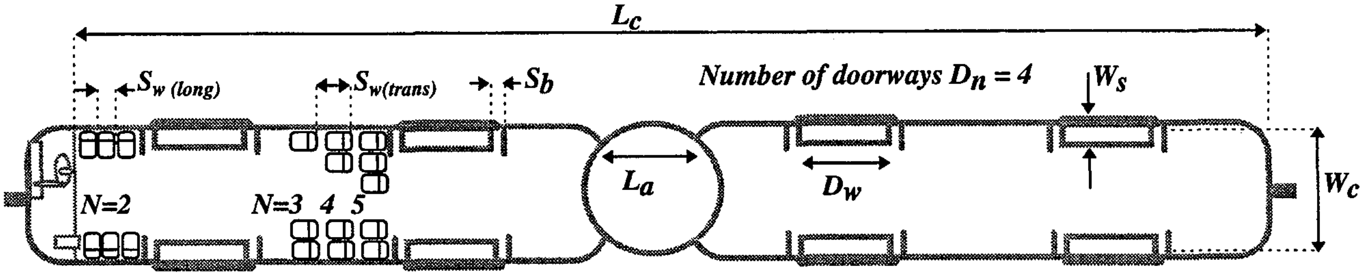


Figure 5.6 Schematic of rail car showing the dimensions of Equation 5.2

- 2.5 per square meter (0.4 m^2, 4.3 sq. ft. per passenger)[11]—a comfortable level without body contact; reasonably easy circulation, similar space allocation as seated passengers.

The middle level above is slightly relaxed from the often stated standard of four standing passengers per square meter. So-called crush loads are frequently based on 5 or 6 passengers per square meter, the latter being more common in Europe. Asian standards for both maximum and crush loads reach 7 or 8 standing passengers per square meter.

The resultant sum of seated and standing passengers provides a guide for the average peak-within-the-peak service loading level for the specific vehicle. Peak-hour loading should be adjusted by the vehicle loading diversity factor. No specific allowance has been made for wheelchair accommodation or for reduced standing densities away from doorways. The above range of standing densities makes such small adjustments unnecessary. Cars intended for higher density loading should have a greater number of doors. Space inefficiencies at the extremities of a car are unavoidable unless the London Underground arrangement of doors at the very end of each car is adopted.

The above process can be expressed mathematically as

$$V_c = \left\lfloor \frac{(L_c - 0.5L_a)W_c - 0.5D_nW_sD_w}{S_{sp}} \right\rfloor + N\left\lfloor \left(1 - \frac{S_a}{S_{sp}}\right)\left(\frac{L_c - L_a - D_n(D_w + 2S_b)}{S_w}\right)\right\rfloor$$

Equation 5-12[12]

where
V_c = vehicle capacity—peak-within-the-peak
L_c = vehicle interior length
L_a = articulation length for light rail
W_s = stepwell width (certain light rail only)
W_c = vehicle interior width
S_{sp} = space per standing passenger
0.2 m^2 (2.15 sq ft) maximum
0.3 m^2 (3.2 sq ft) reasonable
0.4 m^2 (4.3 sq ft) comfortable

N = seating arrangement
2 for longitudinal seating
3 for 2+1 transverse seating
4 for 2+2 transverse seating
5 for 2+3 transverse seating [13]
S_a = area of single seat
0.5 m^2 (5.4 sq ft) for transverse
0.4 m^2 (4.3 sq ft) for longitudinal
D_n = number of doorways
D_w = doorway width
S_b = single set-back allowance
0.2 m (0.67 ft)—or less
S_w = seat pitch
0.69 m (2.25 ft) for transverse
0.43 m (1.42 ft) for longitudinal

Figure 5.6 shows these car dimensions.

The equation can be worked in either meters or feet. An expanded version of this equation is included on the computer spreadsheet. The spreadsheet calculation automatically applies the S_w seat pitch dimension through an IF statement acting on N, the seating arrangement factor, using the longitudinal dimension *if N=2.*

Offset Doors A small number of rail vehicle designs utilize offset doors. These do not merit the complexity of a separate equation. Provided that each side of the car has the same number of doors Equation 5.2 will provide an approximate guide to vehicle capacity with a variety of seating arrangements and standing densities.

Fast Alternative A fast alternative method is to divide the gross floor area of a vehicle (exterior length *x* exterior width) by 0.5 m^2 (5.4 sq ft) and use the resultant number of passengers as the average over the peak hour—without applying a vehicle loading diversity factor. An average space over the peak hour of 0.5 m^2 (5.4 sq ft) per passenger is the U.S. comfortable loading level recommended in several reports and is close to the average loading on all trunk rail transit lines entering the CBD of U.S. cities.

5.4.4 RESULTS OF THE CALCULATION

Light Rail Applying the calculations of section produces passenger loading levels for typical light rail vehicles as shown in

[11] This upper level is a peak-within-the-peak occupancy level for standing passengers. Over the peak hour, it corresponds closely to Pushkarev(R50) and Jacobs(R30) estimates of a United States rush-hour loading average of 0.5 m^2 per passenger—both seated and standing. It also corresponds to Pushkarev and Batelle's(R12) recommendation for an *adequate* or *comfortable* loading level.

[12] $\lfloor\ \rfloor$ = expression rounded down to nearest integer (whole number).

[13] 2+3 seating is only possible on cars with width greater than 3 meters, not applicable to light rail or automated guideway transit.

Table 5.10 Calculated light rail vehicle capacity

RAIL MODE	EXTERIOR WIDTH	EXTERIOR LENGTH	STANDING SPACE	DOOR NUMBER	SEATING FACTOR	TOTAL SEATS	STAND PAX	TOTAL PAX
	Wc (m)	Lc (m)	Ssp (m^2)	Dn	N	Sc	Ps	Vc
Siemens	2.65	25	0.2	4	4	52	151	203
Siemens	2.65	25	0.3	4	4	52	101	153
Siemens	2.65	25	0.4	4	4	52	75	127
Baltimore	2.9	29	0.2	4	4	76	189	265
Baltimore	2.9	29	0.3	4	4	76	126	202
Baltimore	2.9	29	0.4	4	4	76	94	170

Table 5.11 Calculated heavy rail vehicle capacity

RAIL MODE	EXTERIOR WIDTH	EXTERIOR LENGTH	STANDING SPACE	DOOR NUMBER	SEATING FACTOR	TOTAL SEATS	STAND PASS	TOTAL PASS
	Wc (m)	Lc (m)	Ssp (m^2)	Dn	N	Sc	Ps	Vc
Generic	3.1	23	0.2	4	4	60	192	252
Generic	3.1	23	0.4	4	4	60	96	156
Generic	3.1	23	0.2	3	4	80	157	237
Generic	3.1	23	0.3	3	4	80	104	184
Generic	3.1	23	0.4	3	4	80	78	158
Generic	3.1	23	0.2	4	2	60	207	267
Generic	3.1	23	0.3	4	2	60	138	198
Generic	3.1	23	0.4	4	2	60	103	163
Vancouver	2.6	13	0.2	2	4	36	75	111
Vancouver	2.6	13	0.3	2	4	36	50	86
Vancouver	2.6	13	0.4	2	4	36	37	73
Chicago	2.84	14.7	0.2	2	3	36	98	134
Chicago	2.84	14.7	0.3	2	3	36	65	101
Chicago	2.84	14.7	0.4	2	3	36	49	85

Table 5.10. Two articulated light rail vehicles are shown, the common Siemens-Düwag car used in nine systems (with some dimensional changes) and the largest North American light rail vehicle used by the MTA in Baltimore. The resulting capacities are for a generic version of these cars. Reference to Table 5.9, *Light Rail Equipment Summary*, shows that the actual number of seats in the Siemens-Düwag car varies from 52 to 72 while rated total capacity varies from 96 to 201. This stresses the wide, policy related, car capacity issue.

The calculation cannot encompass all options. However, the calculation provides a policy surrogate in the form of the allocated standing space,—0.2, 0.3 or 0.4 m^2 per passenger. Seating should be adjusted accordingly. A need for high standing levels would suggest longitudinal seats, low standing levels, the 2+2 transverse seats.

Heavy Rail Applying the calculations of section produces passenger loading levels for typical heavy rail vehicles as shown in Table 5.11. Data is shown for a generic 23 meter heavy rail car with variations of seating arrangements and standing space allocations. Two data sets follow for the smaller cars used in Vancouver and Chicago.

5.5 LENGTH

In this section the above calculations are converted to the passengers per unit length method suggested by Lang and Soberman[R39] and others, stratified into classes, then compared with actual peak-within-the-peak loading levels of North American rail transit. Given the variation in loading levels that depend on policy—the standing density used and seat spacing—this simplified method is appropriate in most circumstances. It is the recommended method of estimating peak-within-the-peak car capacity except for circumstances and rolling stock that are out of the ordinary.

Light Rail Applying the calculations of section produces passenger loading levels for typical light rail vehicles as shown in

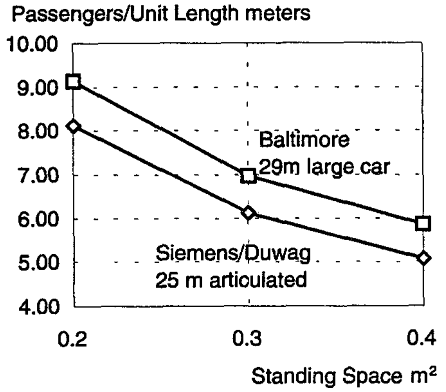


Figure 5.7 Linear passenger loading of articulated LRVs

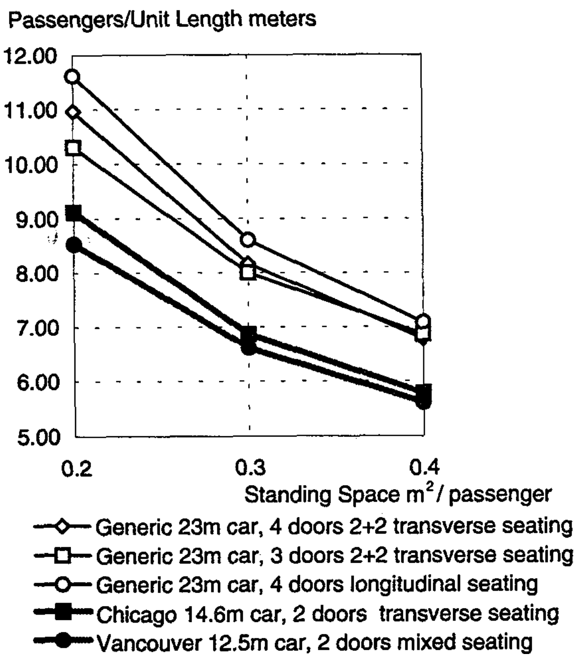


Figure 5.8 Linear passenger loading of heavy rail cars

Table 5.10 and as passengers per unit length in Figure 5.7. As would be expected, the wider and longer Baltimore car has proportionately higher loadings per meter of length. The typical Siemens-Düwag car used on nine systems (with some dimensional changes) has a range of 5.0 to 8.0 passengers per meter of car length. The lower level of five passengers per meter length—with a standing space per passenger of 0.4 m^2—corresponds closely with the recommended *quality* loading of a an average of 0.5 m^2 per passenger.

Heavy Rail Applying the calculations of section 5.4.3 produces passenger loading levels for typical heavy rail vehicles as shown in Table 5.11 and, as passengers per unit length, in Figure 5.8. As would be expected, the smaller and narrower cars in Vancouver and Chicago have lower loadings per meter length.

The more generic 23-m-long cars used in over 12 North American cities have a remarkably close data set for each of the three variations, 4 and 3 door versions, and transverse or longitudinal seating—with a range of 7.0 to 11.5 passengers per meter of car length. The lower end of the range of seven to eight passengers per meter length—with a standing space per passenger of 0.4 to 0.3 m^2—is an appropriate range for higher use systems. A lower figure of six corresponds closely with the recommended *quality* loading of an average of 0.5 square meters per passenger and is appropriate for a higher level of service on new systems. In both cases a reduction by one should be used for smaller, narrower cars.

These calculated linear loading levels can be compared with actual levels on major North American rail transit lines shown in Table 5.12 and summarized in Table 5.13.

Heavy Rail outside New York shows a level comparable with the recommended *comfortable* level of 6 passengers per meter of train length. New York is higher by some 25%, averaged over 11 trunk routes. Commuter rail, with most passengers seated, has an average only 13% lower than the average of heavy rail outside New York. Only two light rail lines are running close to capacity and peak-within-the-peak ridership is not available for these.

5.6 LOADING DIVERSITY

Passengers do not load evenly into cars and trains over the peak hour. This unevenness is the diversity of passenger loading. There are three different types of loading diversity: unevenness of passenger loading within a car; unevenness of passenger loading within cars of a train; unevenness of passenger loading within peak-hour trains. The loading diversity factor developed in this section essentially encompasses all three.

In individual cars, the highest standing densities occur around doorways, the lowest at the ends of the cars. Several European urban rail systems add doors, sometimes only single stream, at the car ends to reduce this unevenness. London Transport's underground system is the most notable with this feature on most rolling stock,[14] except at car ends with a driving cab. The end door on the low-profile cars are 0.75 m (2.5 ft) wide compared to the main doors of 1.56 m (5.1 ft). These exceptionally wide doors, with their 0.17 m (6.8 in) set-backs often accommodate three streams of passengers.

No data exist to determine such loading diversity within a car and the variations are accommodated in the average loadings of the previous sections. It is important in cars designed for high occupancies to minimize this effect by using wide aisles, uncluttered vestibules and suitable hand holds that encourage passengers to move into the extremities of a car. Very little information was found on car interior design efficiency in the literature search with the exception of Young[(R76)] *Passenger Comfort in Urban Transit Vehicles.*

A second level of diversity occurs in uneven loading among cars of a train. This second level is also included in the average loading data of the previous sections and in the application chapters. Cars that are closer to station exits and entrances will be more heavily loaded than more remote cars. This inefficiency can be minimized by staggering platform entrances and exits

[14] London's Docklands Light Railway does not have end doors.

Table 5.12 Passengers per unit train length, major North American trunks

System	Trunk Name	Length	Width	Seats	15 min. peak	Average passengers per car	Passengers per meter of car length
CalTrain	CalTrain	25.91	3.23	146	932	117	4.5
GO Transit	Lakeshore East	25.91	3	162	4094	152	5.9
LIRR	Jamaica - Penn Stn.	25.91	3.28	120	12380	117	4.5
Metra	Metra Electric	25.91	3.2	156	4765	113	4.4
CTS	Northeast Line	24.28	2.66	64	1495	125	5.1
CTS	South Line	24.28	2.66	64	1840	153	6.3
BCT	SkyTrain	12.4	2.49	36	2056	73	5.9
CTA	Dearborn Subway	14.63	2.84	46	2616	82	5.6
CTA	State Subway	14.63	2.84	46	3601	75	5.1
MARTA	East/West	22.86	3.2	68	926	77	3.4
MARTA	North/South	22.86	3.2	68	1796	82	3.6
NYCT	53rd Street Tunnel	18.35	3.05	50	15154	210	11.4
NYCT	60th Street Tunnel	22.77	3.05	74	7534	126	5.5
NYCT	Broadway Express	15.56	2.68	44	7962	119	7.6
NYCT	Broadway Local	15.56	2.68	44	5398	135	8.7
NYCT	Clark Street	15.56	2.68	44	4873	102	6.6
NYCT	Joralemon St. Tunnel	15.56	2.68	44	7305	122	7.8
NYCT	Lexington Ave. Express	15.56	2.68	44	9800	123	7.9
NYCT	Lexington Ave. Local	15.56	2.68	44	8648	144	9.3
NYCT	Manhattan Bridge	22.77	3.05	74	12306	162	7.1
NYCT	Rutgers St. Tunnel	22.77	3.05	74	3937	123	5.4
NYCT	Steinway Tunnel	15.56	2.68	44	6318	144	9.3
PATH	33rd St.	15.54	2.81	31	3080	88	5.7
PATH	World Trade Center	15.54	2.81	31	5595	92	5.9
TTC	Yonge Subway	22.7	3.15	80	8285	197	8.7

Table 5.13 Summary of linear passenger loading (per meter) Additional passenger loading per unit length data are compiled in Tables 7.4, 7.5 and 7.6 of Chapter Seven.

	Average	Median	Standard Deviation
All Systems	6.4	5.9	2.0
Commuter Rail	4.8	4.5	0.7
Heavy Rail	6.8	6.3	2.0
Heavy Rail less NY	5.5	5.6	1.5
New York City alone	7.9	7.8	1.8

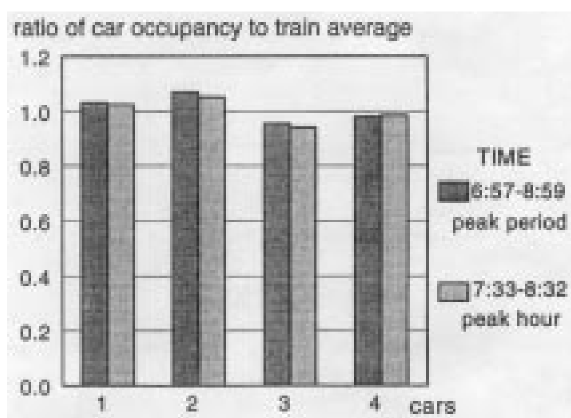


Figure 5.9 Vancouver, Broadway Station inbound peak-hour passenger distribution between cars of train. October 27 1994, 50 trains, 12,173 passengers

between ends, centers and third points of the platforms. This is not always possible or practiced. The busiest, most densely occupied rail lines in North America, lines 1, 2 and 3 of Mexico City's metro all have stations with center entrances/exits. Even so, relatively even loading occurs both here, and on rail transit lines at or near capacity elsewhere, due to the duress factor that encourages passengers to spread themselves along the platform during heavily traveled times—or risk being unable to get on the next arriving train.

Few systems count passengers by individual cars when these are *crush* loaded. This is difficult to do with any accuracy and the results differ little from assigning a set *full* load to each car of a fully loaded train. Data are available from two Canadian properties.

BC Transit operates four car trains on headways down to 90 sec. Pass-ups are routine at the busiest suburban station, Broadway with an end and two third-point entrances/exits. The relative loading of the four cars is shown in Figure 5.9. The main entrance/exit is provided with escalators and lies between the

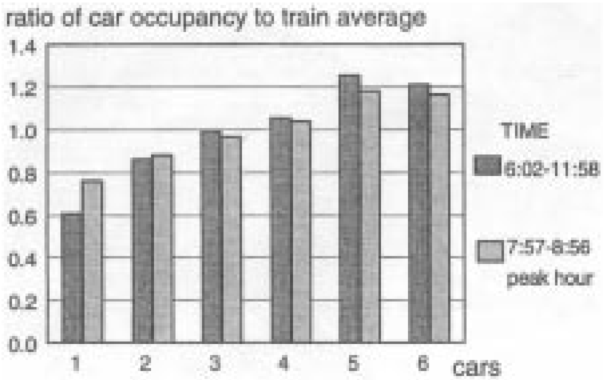


Figure 5.10 TTC Yonge Subway, Wellesley Station southbound, a.m. peak-period average passenger distribution between cars of train. Jan 11, 1995, 99 trains with 66,263 passengers

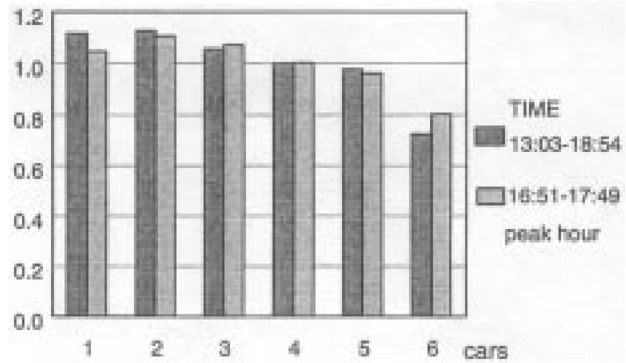


Figure 5.11 TTC Yonge Subway, Wellesley Station northbound, p.m. peak-period average passenger distribution between cars of train. Jan 11, 1995, total 69,696 passengers on 108 trains

second and third cars of the train. While the second car is the most heavily loaded, the third is the lightest loaded indicating the influence of entrance/exit locations at other major stations.

There is no significant variation in the average loading diversity between the peak hour and the peak-period both of which remain within the range of +5% to -6%. The unbalance for cars on individual trains ranges from +61% to -33%. The uniformity of loading can be attributed to four factors—the short trains, wide platforms, close headways and dispersed entrance/exit locations between the stations of this automated, driverless system. The Toronto Transit Commission's Yonge Street subway shows a more uneven loading between cars in Figure 5.10. In the morning peak period the rear of the train is consistently more heavily loaded reflecting the dominance of the major transfer station at Bloor with the interchange at the northern end of the Yonge platform. As would be expected, there is little variation in the average car loading diversity between the peak hour and the peak period due to the pressures on passengers to spread along the platforms at busy times. The average diversity of individual car loading over the peak period has a range of +26%to -39%. The unbalance for cars on individual trains ranges from +156% to -89%.[15] In the afternoon peak period shown in Figure 5.11, the reverse occurs with the front of the train most heavily loaded—despite the principal entrances at the two major downtown station being toward the rear of the train. There is less variation in the average car loading diversity between the peak hour and the peak period than in the morning. The average diversity of individual car loading over the peak period has a range of +13% to -28%. The unbalance for cars on individual trains ranges from +113% to -72%. These ranges are lower than in the morning reflecting the less intense peak-within-the-peak in the pm rush hour.

It is this peak-within-the-peak that provides the third and most important diversity factor, termed the *peak-hour loading diversity factor* and defined by:

$$D_{ph} = \frac{R_{hour}}{4R_{15\text{min}}} \qquad \text{Equation 5-1}$$

where D_{ph} = Diversity factor—peak hour
R_{hour} = Ridership in peak hour
R_{15min} = Ridership in peak 15 min

Passengers do not arrive evenly and uniformly on any rail transit system as shown dramatically over the extended peak period in

Table 5.14 Diversity of peak hour and peak 15 min[16]

Type	System	Routes	Diversity factor
CR	CalTrain	1	0.64
CR	GO Transit	7	0.49
CR	LIRR	13	0.56
CR	MARC	3	0.60
CR	MBTA	9	0.53
CR	Metra	11	0.63
CR	Metro-North	4	0.75
CR	NICTD	1	0.46
CR	NJT	9	0.57
CR	SCRRA	5	0.44
CR	SEPTA	7	0.57
CR	STCUM	2	0.71
CR	Tri-Rail	1	0.25[17]
CR	VRE	2	0.35
CR	**Sum/Average**	**74**	**0.56**
LRT	CTS	2	0.62
LRT	Denv. RTD	1	0.75
LRT	SEPTA	8	0.75
LRT	Tri-Met	1	0.80
LRT	**Sum/Average**	**12**	**0.73**
RT	BCT	1	0.84
RT	CTA	7	0.81
RT	MARTA	2	0.76
RT	MDTA	1	0.63
RT	NYCT	23	0.81
RT	PATCO	1	0.97
RT	PATH	4	0.79
RT	STCUM	4	0.71
RT	TTC	3	0.79
RT	**Sum/Average**	**46**	**0.79**
All	**Sum/Average**	**133**	**0.67**

[15] One car of one train was completely empty (-100%), possibly due to an incident or defective doors. This outlier was excluded from the data set.

[16] This peak-hour diversity factor is the same as the peak-hour factor (phf) in the Highway Capacity Manual[(R47)].

[17] Service is only one train per hour and is not included in the average.

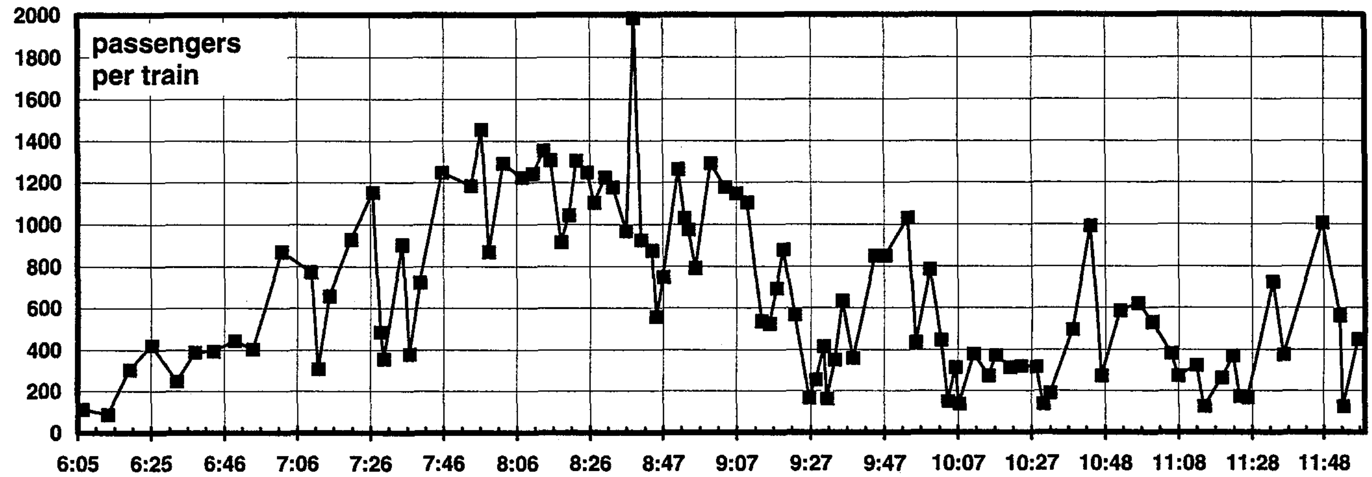


Figure 5.12 Individual train loads, TTC Yonge Subway, Wellesley Station southbound Jan. 11, 1995 (5-min tick marks)

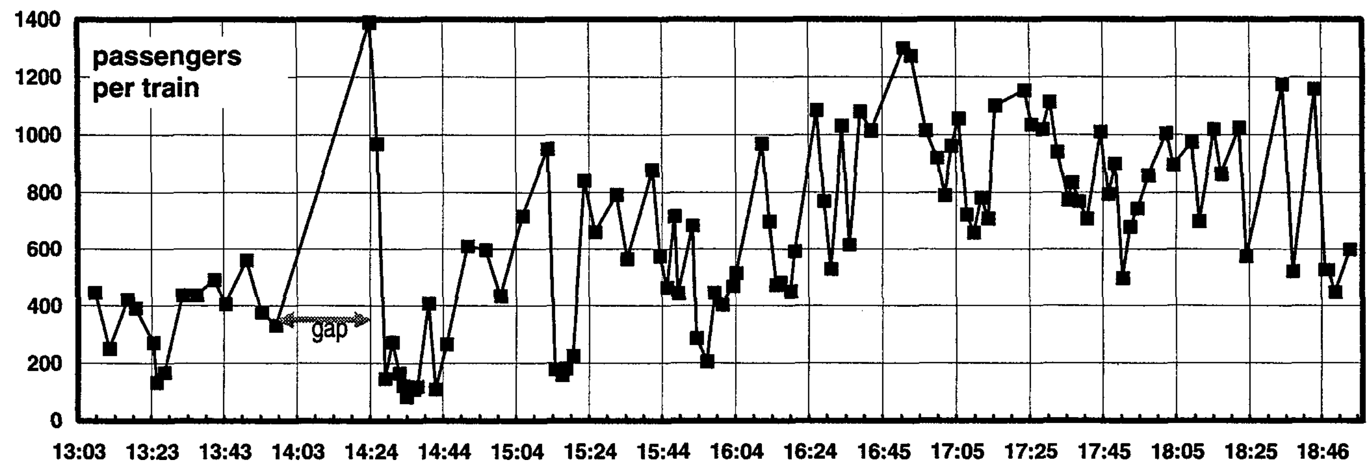


Figure 5.13 Individual train loads TTC Yonge Subway, Wellesley Station northbound Jan. 11, 1995 (5-min tick marks) Note cluster of low occupancy trains at 14:24 to 14:44h following a crush load train after a 29-min gap in service.

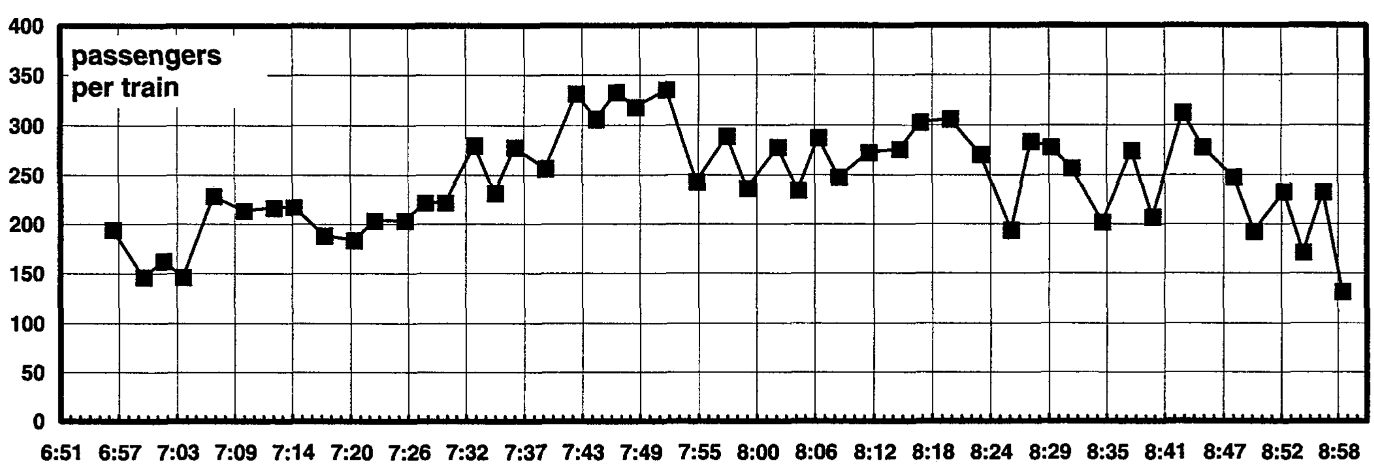


Figure 5.14 Individual train loads Vancouver, Broadway Station inbound October 27, 1994 a.m. peak (1-min tick marks)[18]

[18] The courtesy of the Toronto Transit Commission and British Columbia Rapid Transit Company in providing car by car and train by train checker data is acknowledged. The willingness of the Toronto Transit Commission to allow use of data with unusual erratic headway operation is particularly appreciated.

Figures 5.12 and 5.13 for the Toronto Transit Commission's Yonge subway.

These figures do not show the smooth peaks-within-the-peak often displayed in texts but rather the realities of day-to-day rail transit operation. The morning peak-within-the-peak has a pronounced abnormality at 8:35h following a short gap in service.

The afternoon peak actually occurs at 14:24h following a 26-min delay due to a suicide. Next are two abnormally low troughs as the delayed trains move through—and the commission's control center strives to normalize service prior to the start of the real peak hour.

In both charts the different loading, train by train, is striking and it is difficult to visually pick out the peak hour or the 15 min peak-within-the peak. This entire data set of car by car loadings and headways, representing 1,242 individual car counts of 135,000 passengers, is contained on the computer disk.

Figure 5.14 shows an a.m. peak-period for BC Transit that, although without major delays, shows the irregular loading from train to train due to the interlacing of short-turn trains with regular service from 07:30h onwards. The loading diversity factor was obtained for most systems. The principal data deficiency was for light rail where few systems count passengers by train.

The diversity of train loading over the peak hour is shown in Table 5.14. Note that the values can be strongly affected by the level of service provided. This is particularly true of infrequent commuter rail lines. (Infrequent service on two of GO Transit's lines contributes to GO's relatively low average.) Rail rapid transit (RT) is generally the most frequent mode and so has relatively low values for the diversity factor. Values for light rail transit are intermediate.

Diversity of loading within a car and among cars of a train are included in the recommended peak-within-the-peak loading levels. The peak-within-the-peak *loading diversity factor* is not so included and must be used to adjust passenger volumes from the estimated *design capacity* to a more practical *achievable capacity*. This important loading diversity factor is discussed further in Chapter Seven, *Grade Separated Rail Capacity Determination*, and subsequent mode specific chapters. Here suitable values are recommended for use in calculating the maximum achievable capacity.