Jörn Pachl

Railway Timetabling and Capacity

Edition 1.1

Author:

Prof. Dr.-Ing. Jörn Pachl, FIRSE

Professor of railway systems engineering at Technische Universität Braunschweig

Braunschweig, July 2025

Edition 1.1

PREFACE

The timetable is an essential element of rail traffic control. By inserting train paths into a timetable, the infrastructure capacity is assigned to the train moves. To establish a conflict-free timetable, models are needed to evaluate the capacity consumption both of individual train paths and of traffic patterns with a mix of different train types. In this E-book, after a short introduction to some basic terms on railway operations, the modelling of train paths is explained. Based on this, the following chapters cover the capacity evaluation of the infrastructure and the methods to establish feasible timetables.

This E-book was originally written as a tutorial for students at TU Braunschweig to support lectures on railway planning and operations.

Jörn Pachl

CONTENTS

1 BASIC OPERATING TERMS

To describe operating characteristics relevant for timetabling and capacity requires an understanding of basic terms of rail traffic control. Besides an introduction to the key operating tools, this chapter covers definitions of specific track areas and the classification of movements with railway vehicles. This is kept rather short and concentrates on the terms as used in this tutorial. Readers of this tutorial are required to already have basic knowledge on signalling and train control principles. That knowledge is not provided here but crucial for the understanding of timetabling and capacity research.

1.1 Key Operating Tools

In contrast to most other transportation systems, railway traffic is controlled from the infrastructure side. Trains cannot freely move through the network and make their own decisions which route to take. Instead, trains are guided on a path predefined in the timetable on which the safety is guaranteed by infrastructure control systems. This requires an interaction between three subsystems that may be seen as the key operating tools for rail traffic control:

- Interlocking and operation control systems
- Signals
- The timetable

The task of the interlocking system is to set a safe route for a train movement. A safe route means that the track section is clear, all movable track elements are locked in the proper position, and the train is protected against conflicting movements. Route setting for an approaching train is initiated from an operation control system either by manual action on a user interface or by an Automatic Route Setting system (ARS) that uses the timetable data for automatic route setting.

When a route has been set, a movement authority, i.e., the right to occupy the track section that is protected by the route, has to be transmitted to the train. This is the purpose of signals. Signals are a one-way communication system by which the trains are guided from the infrastructure side. There are two basic types of signals:

- Lineside signals
- Cab signals

While lineside signals are the traditional form of signalling, they are still the dominating form of train control and even used in many new installations. A lineside signal indicates to the driver whether or not the train has authority to proceed into the next section. So, by lineside signals, movement authorities can only be transmitted at discrete points. Cab signals indicate the guiding information directly on the driver's desk. For this, a continuous data transmission system from track to train is needed. Cab signalling is generally required on high speed lines, on most railways on all lines where the speed exceeds 160 km/h. When running at a high speed, the time window to safely get the aspect of a lineside signal would be too short. Besides enabling the control of high speed trains, cab signals will also improve capacity since the movement authority can be continuously upgraded. That is why with the availability of digital radio systems that can be used for continuous data transmission, lineside signals on conventional lines will also be more and more replaced by cab signals.

The timetable is not just a planning tool but contains a predefined path for each train from which the control data for guiding the train is derived. While the timetables used in control centres contain the relevant data for all trains running through the control district, the driver's timetable contains the data needed to safely drive the train on the entire train run. The driver's timetable does not only contain times but also speed limits, braking conditions, and other data needed to safely guide the train.

The so-called train path by which a train's run is described in the timetable has two aspects. It is both a route through the network, i.e., the sequence of track sections on which the train is guided, but also a time-distance line on this route including a sufficient time slot around the time-distance line to avoid conflicts with other train path's. This is the basis for establishing a conflict-free timetable.

On railways that are operated on an open access basis, i.e., shared use of the infrastructure by several train operating companies, the timetable is also an important tool for open access management. To run a train, the train operating company has to request a train path from the infrastructure manager. From the train paths requested by the different train operating companies, the infrastructure manager establishes a conflict-free timetable and assigns the final train paths to the train operating companies.

1.2 Lineside Signals

While being gradually replaced by advanced radio-based train control systems, in which trains are guided by cab signal indications, lineside signals are still the most common technology for controlling train movements. On railways where train movements are strongly separated from shunting movements (see Section 1.5), which is the case on most railways outside North America, there are usually also two basic kinds of lineside signals:

Main signals

Main signals authorise a regular train movement to enter a line section. In this tutorial, the term main signal is used in a generic way for all signals controlling regular train movements, no matter, whether or not an individual railway would separate these signals from shunting signals. The movement authority provided by a main signal is limited by the next main signal or a point specified in the operating rules.

Apart from lines with a low speed, a signal that authorises a train movement requires an approach aspect at the braking distance in approach to the signal because the stopping distance is generally greater than the range of vision. The approach aspect is necessary for a safe braking when approaching a stop signal. On lines where the distance between signals does not significantly exceed the braking distance, the approach aspect is usually provided by the signal in rear. On lines with very long distances between main signals, distant signals are placed at the braking distance in approach to a main signal. A distant signal can only provide an approach aspect for the signal ahead but it cannot show a stop aspect. Another common term for a distant signal used in the rulebooks of some railways is warner signal (*Chandra & Agarwal, 2008*).

Lineside signals are always used in conjunction with a fixed block system. A line with a fixed block system is divided into block sections for the purpose of safe train separation. A train must generally not enter a block section until it has been cleared by the train ahead. On lines with lineside signals, block sections are limited by main signals. However, a fixed block system may also be used in conjunction with cab signalling. Then, lineside signals are no longer needed. For degraded mode operations, marker boards may be placed at the block limits.

Shunting signals

Shunting signals are used to authorise shunting movements and to protect trains against shunting movements. On most railways, the stop aspect of a main signal also applies for shunting movements. On tracks where shunting movements may pass main signals, a shunt aspect is incorporated in the main signal, so that shunting movements may be authorised to pass main signals in stop position. For shunting signals, an approach aspect is not provided because shunting movements run at a very low speed that allows the driver to stop short of any vehicle or obstruction.

On some railways, the stop aspect of shunting signals is absolute, i.e., it indicates stop for all kinds of movements. Such shunting signals must also be cleared for train movements authorised by a main signal. Movements running under authority of a main signal may pass cleared shunting signals at the speed authorised by the main signal or the timetable. Some railways use shunting signals with a shunting stop aspect that does not apply to train movements. This is typical for several East European, Russian, and Chinese railways where a blue light is used for this purpose. On those railways, movements running under authority of a main signal may ignore the stop aspects of shunting signals.

On railways that do not use the distinction between main and shunting signals, there is a signal aspect that authorises a movement to pass a signal cautiously on sight prepared to stop short of any vehicle or obstruction. That aspect is used both for shunting purposes but also to authorise train movements to enter a section that may be occupied.

Control Principle of Lineside Signals

Concerning the control principle, there is a distinction between controlled and automatic signals. Controlled signals are all signals that protect track sections that contain movable track elements or points where conflicts with movements on conflicting routes may occur. The normal position of controlled signals is stop. To clear a controlled signal, a route is set in the control system that will lock all movable track elements in the proper position and lock out all conflicting moves. The route setting may be initiated by a human operator or an automatic route setting system. If route setting is automated, the signals are still referred to as controlled signals.

Automatic signals work automatically by the passage of the train through track sections. They are forced to stop position by track clear detection devices if the section beyond the signal is occupied. Automatic signals can only be used to protect track sections that do neither contain any movable track elements nor points where conflicts with movements on conflicting routes may occur. Depending on the type of control system, the normal position of automatic signals may be stop or clear.

Signal Arrangement for Double Track Operations

For double track operations, there is usually a specified direction of traffic for each track. While right-track operation dominates slightly worldwide, there is a significant number of countries where left-track operation is the standard form. On lines not equipped with a bidirectional signalling system for two-way working, all regular train movements have to be made with the normal direction of traffic. On such lines, movements against the normal direction (also called 'wrong line moves' or 'reverse movements') have to be authorised by special instructions under staff responsibility. On lines that are equipped with a signalling system for two-way working, movements against the normal direction can be authorised by clearing a main signal.

Many railways do not install intermediate block signals for reverse movements because on most lines, reverse movements are not carried out frequently. For temporary single track working in case of a track closure, the direction on the remaining track will change after almost every train. So, intermediate block signals would have no effect on capacity. Intermediate block signals for movements against the normal direction do only make sense on sections, where parallel moves on both lines are carried out on a regular basis.

Figure 1.14 shows typical examples of signal arrangements for double track operation. On many railways, a normal direction of traffic is only in effect outside of station areas.

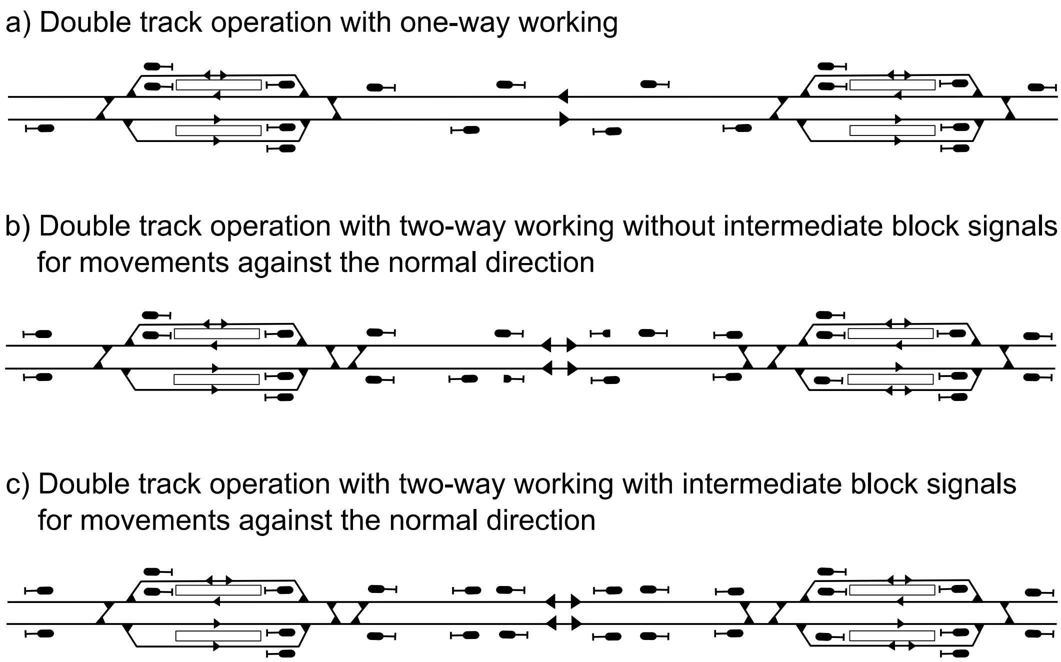


Figure 1.14 Signal arrangements for double track operations

1.3 Classification of Tracks

In railway operations, a track is often also referred to as a line. A route consisting of just one track is called a single line, while a route with double track operation, i.e., two parallel tracks and a specified direction for normal moves on both tracks is called a double line. For operational purposes, tracks are divided into two main classes. While they are called differently in the rulebooks of individual railways, the basic idea is always the same:

- Tracks that may be used for regular train movements
- Tracks that must only be used for shunting movements

The distinction between train and shunting moves is explained later in this chapter. The tracks used for regular train movements are called main tracks or running lines. The term main track is mainly used in North America, but was also adopted by some railways outside North America. It is also used in some international textbooks (*Theeg & Vlasenko, 2020*). Many other railways, in particular railways with roots in the British systems, prefer the term running line. The lines between stations and their continuation through stations and interlocking areas belong into this category. It also includes tracks for passing and overtaking trains which are called loops on most railways (Figure 1.10). On signalled lines, tracks used for train movements are equipped with signalling appliances for the safe passage of trains. Along the line a train passes through, points are usually interlocked with signals that provide the movement authority. Sidings are all tracks that must only be used for shunting movements. The points of sidings are often not interlocked.

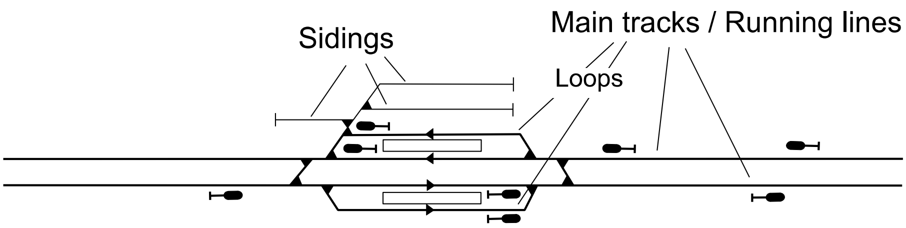


Figure 1.10 Classification of tracks

Note: In the North American terms, a line is a route that may consist of several parallel tracks. Furthermore, loops are called sidings; tracks other than main tracks are called yard tracks, secondary tracks, or industrial tracks.

1.4 Interlocking Areas

An interlocking area is a track area, where controlled signals are interlocked with points and other signals in a way that a signal can only be cleared when all points are locked in the proper position and all conflicting moves are locked out. Signals that govern routes in an interlocking area are called interlocking signals. The points and signals are controlled either by a local control station or from a remote control centre. Local control stations are called interlocking towers in North America, and signal boxes or signal cabins on most other railways. A locally staffed control station contains both the interlocking system and the user interface for the operator. Modern interlocking systems are usually remote controlled from a control centre.

There are two basic signal arrangements in interlocking areas (Figure 1.11). First, there are interlocking areas without consecutive interlocking signals. An interlocking signal provides authority to run through the entire interlocking area into the next block section. Second, there are interlocking areas with consecutive interlocking signals.

a) Interlocking area without consecutive interlocking signals

b) Interlocking area with consecutive interlocking signals

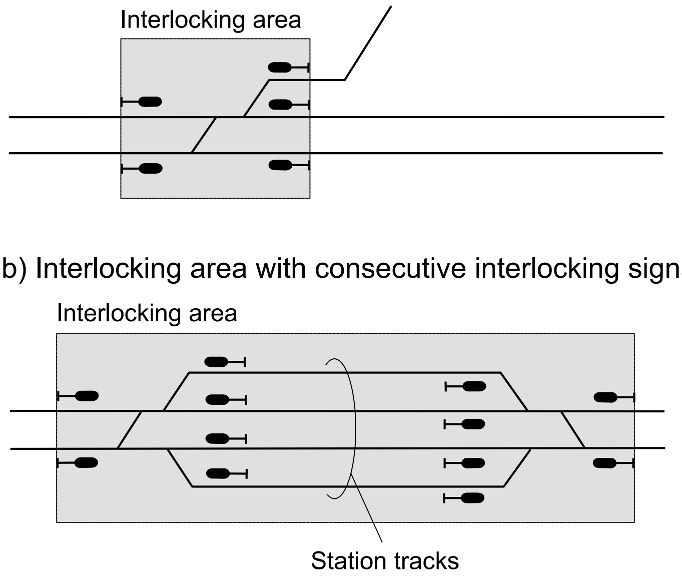


Figure 1.11 Types of interlocking areas

Such an interlocking area may contain tracks protected by controlled signals on which trains may originate, terminate, pass, and turn. In this tutorial, these tracks are called station tracks. Consequently, an arrangement of station tracks is called a station area. Also, in this tutorial, the tracks outside the opposing interlocking signals limiting the station area are called the open line. The terms station area and open line are used here in a generic way not referring to the practice of an individual railway. Due to the great variety of operating principles used worldwide, terms used in the rulebooks of individual railways may differ.

A platform station for scheduled stops of passenger trains is not necessarily associated with a station track in the sense as defined above, which needs not have a platform. There are also platform stations on the open line, which do not necessarily imply any provision of pointwork or the ability to reverse trains within the signalling system.

On most railways, the interlocking signals protecting a station area from both sides are called home signals. The interlocking signals that govern train movements to leave a station track into a section of the open line are often called station exit signals or just exit signals. These terms are also used in this tutorial. Other commonly used terms for these signals are section signals, starter signals, or leaving signals. Interlocking signals within the station area that are neither home nor exit signals are called intermediate interlocking signals (Figure 1.12). On some railways, they are also called inner home signals (when passed by arriving trains) and inner starter signals (when passed by departing trains).

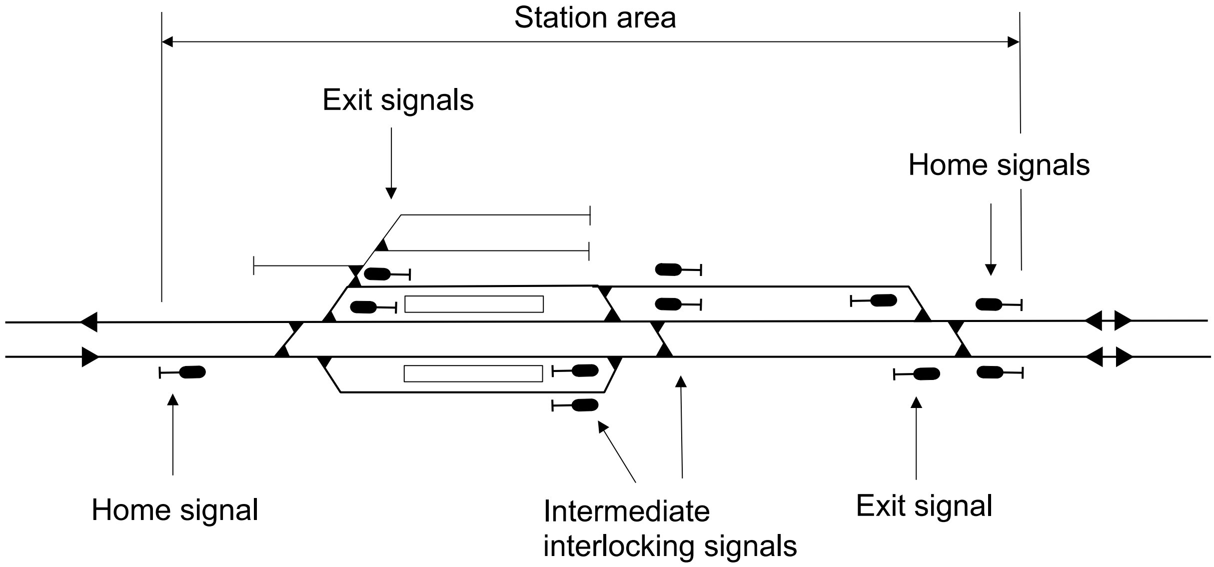


Figure 1.12 Station area with intermediate interlocking signals

In North America, interlocking areas with consecutive interlocking signals within the same interlocking limits are not common. The reason is that the distinction between station tracks and tracks of the open line does not exist in the North American rules (*Bisset et. al., 2008*). In Europe and most other railways outside of North America, station areas with consecutive interlocking signals are very common. In modern British signalling centres, there is also no longer any formal distinction between station tracks and sections of the open line. However, interlocking areas may contain consecutive interlocking signals. So, in a generic way, the terms used here will also fit to the British system.

1.5 Movements with Railway Vehicles

On most railways, the normal train movements are separated from the so-called shunting movements.

Train movements, also known as 'running movements', are movements of locomotives or self-propelled vehicles, alone or coupled to one or more vehicles, with authority to occupy a section of line under operating conditions specified in the timetable. Every train displays rear end markers (tail lights or marker boards) to enable the lineside staff to check the train completeness. All regular movements running along the line from station to station are train movements.

The authorisation of a train movement has two elements:

- A valid timetable as the authority to run through the network along a pre-defined route by specified operating conditions (timetable authority)
- A movement authority for every single section of track in the path of the train

The movement authority to enter a section of track is issued by the operator who is in charge of controlling train movements on that section of track. This way, a train is always under external guidance of a train control operator. The authority for train movements is given by:

- A proceed indication of a main signal

- A proceed indication of a cab signal display
- A call-on signal permitting a train to pass a signal displaying a stop aspect under special conditions
- A written or verbal instruction permitting a train to pass a signal displaying a stop aspect under special conditions
- A written or verbal authority on non signal-controlled lines

Shunting movements are movements for making up trains, moving vehicles from one track to another, and similar purposes. Shunting movements are accomplished without a timetable under simplified conditions at a very low speed that allows the driver to stop short of any vehicle or obstruction. On main tracks and on sidings controlled by an interlocking station or control centre, a verbal agreement between the shunting crew and the operator is needed before the operator may authorise the shunting move. That verbal agreement performs a similar function as the timetable authority for train movements. Block rules are not in effect. Shunting units may enter occupied tracks. Movements in industrial sidings are also carried out as shunting movements.

The authority of shunting movements is given by:

- A proceed indication of a shunting signal, which may be combined with a main signal to authorise a shunting move to pass the main signal in stop position
- Verbal permission

Concerning shunting movements, the railways have designated different limits in accordance with their individual operating practice.

On European railways, shunting units must not enter line sections outside the home signal limits of a station area. The same rule also applies on many railways outside Europe. The area between the home signals that may be used for shunting is usually limited by limit of shunt or shortly LOS boards (Figure 1.13).

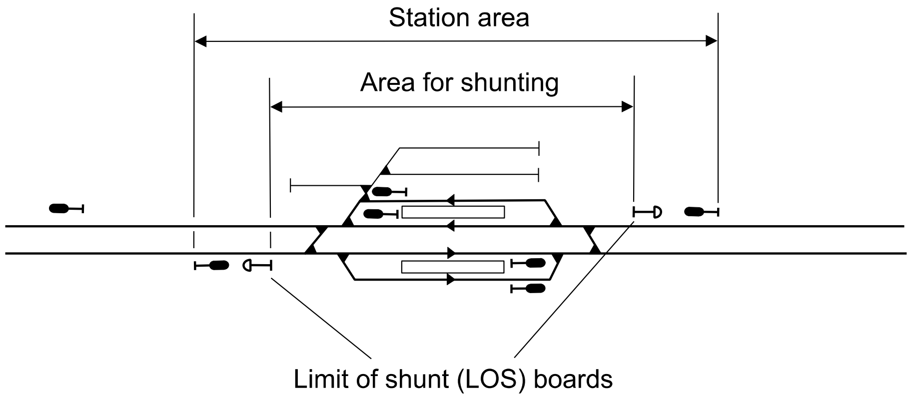


Figure 1.13 Shunting limits of a station area

The distance between the home signal and the LOS board equals the overlap of the home signal. Shunting units may pass the shunting limit boards only with a written permission from the operator. Before issuing authority to a shunting unit to pass the shunting limit board, the operator has to make sure that there is no train approaching the home signal.

While in modern British railway operations, the formal distinction between station areas and the open line doesn’t exist, shunting units must be prevented from entering a section on which they are not protected against opposing movements. If no main or shunting signal exists to limit the shunting move, limit of shunt signals are placed at the shunting limit. These LOS signals are inoperative shunting signals permanently displaying a stop aspect.

2 MODELLING TRAIN MOVEMENTS

For timetabling and capacity research, train movements must be modelled in a way that allows establishing a conflict-free timetable and to evaluate the capacity consumption. The key element of that model is the running time estimation that delivers the time-distance line of a train movement. Based on the time-distance line, a train path is modelled that describes the time slot around the time-distance line that is consumed by that train movement and must be kept clear to avoid conflicts with other trains. This chapter starts with some knowledge on train movement dynamics needed for running time estimation. After having explained the running time estimation, different principles are presented of how the train path can be modelled.

2.1 Relevant Forces and Resistances

2.1.1 Tractive Effort

The maximum tractive effort a locomotive could exert is limited by two factors:

- The maximum force that can be transmitted by adhesion between the wheels and the rail
- The maximum force that can be produced by the drive power of the propulsion engine

The force that can be transmitted between the wheels and the rail equals the weight force on the driven wheels multiplied by the coefficient of adhesion (also known as the coefficient of friction or friction factor). The coefficient of adhesion varies from 0.1 to 0.4 depending mainly on the rail conditions but also a little on the speed. The actual force exerted by the propulsion engine equals the installed power divided by the speed.

$$F_{adhes} = m \cdot g \cdot \mu_{adhes}$$

$$F_{power} = P / v$$

F_{adhes} tractive effort limited by adhesion
F_{power} tractive effort limited by drive power
μ_{adhes} coefficient of adhesion
g acceleration of gravity
P drive power
v speed

When both forces are displayed in a diagram as a function of speed, the result is the tractive effort curve (Figure 2.1). Both curves have an intersection at the so-called critical speed (v_C). Below that speed, the tractive effort is limited by adhesion, so that the installed power of the propulsion engine cannot be fully used. Above that speed, that limit is no longer relevant so that the tractive effort is limited by the installed power.

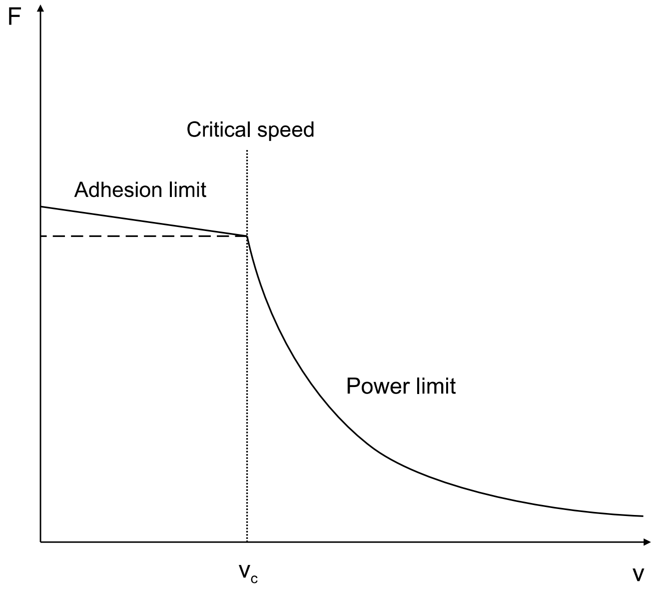


Figure 2.1 Tractive effort curve

2.1.2 Braking Performance

For running time estimation, it is in most cases sufficient to model the braking procedure by an average constant braking deceleration. When approaching a scheduled train stop, the train will not use the full braking performance but a lower braking deceleration. The full braking performance to get the shortest possible braking distance is mainly used in emergency situations or when a train unexpectedly encounters a stop signal. While that emergency braking deceleration may differ significantly between different train types, the braking deceleration when approaching a scheduled stop will more or less be the same for all trains of a specific train class.

According to *(Hansen & Pachl, 2014)*, the following values are recommended for European railways:

a_b = 0.525 m/s^2 for suburban trains (service braking)

a_b = 0.375 m/s^2 for passenger trains (service or comfort braking)

a_b = 0.225 m/s^2 for freight trains (service braking)

a_b = 0.7 m/s^2 for suburban trains (sharp braking)

a_b = 0.5 m/s^2 for passenger trains (sharp braking)

a_b = 0.3 m/s^2 for freight trains (sharp braking)

For timetabling, the service braking deceleration is used. The 'sharp braking' may be applied in the investigation of incidents and for the planning of optimum signal locations. For trains guided by cab signalling, the permitted speed curve of the cab signalling system must be considered when selecting a suitable value for the average braking deceleration.

There are also some timetabling systems that allow a more detailed modelling of the braking performance by entering parameters of the braking performance evaluation system. In software solutions developed in Europe, the user can often enter the brake parameters of the UIC braking performance evaluation system. In that system, the braking performance is described by two parameters:

- The brake position
- The braking percentage (also known as brake weight percentage).

The brake position, which has to be set by a control lever at each vehicle, describes the control behaviour of the brake system. In the UIC system, there are three main brake positions: G, P, and R. Position G is a slow action brake to be used on freight trains (G for ‘goods’). Position P provides a faster brake action and was originally developed for passenger trains (P for ‘passenger’). Today, it is also used for many freight trains. Position G is still needed for heavy freight trains. Position R provides a very fast brake action and was originally used for fast passenger trains (R for ‘rapid’). Today, also many other passenger trains run in R position.

The braking percentage is the so-called brake weight of the train divided by the total weight of the train multiplied by 100%. Both the brake weight and the total weight of the train are stated in tons (t). The brake weight describes the braking force of a vehicle but cannot directly be transformed into a braking force. Here is the official definition of the brake weight:

The brake weight multiplied by 1.25 equals the weight the vehicle may get to a stop from a speed of 120 km/h over a distance of 1000 m on an even track with no wind.

For new vehicles, the brake weight is determined by braking and coasting experiments. On vehicles where different brake positions can be set, the brake weight depends on the selected brake position.

Advanced timetabling tools that model the train consist by a list of all vehicles with their characteristics will automatically calculate the braking percentage for a given brake position.

2.1.3 Resistances

There are two essential resistances:

- Line resistance
- Train resistance

2.1.3.1 Line Resistance

The line resistance consists of two components: grade resistance and curve resistance. The grade resistance in its absolute form equals the weight force of the train multiplied by the sine of the angle of gradient.

$$R_{grade} = m \cdot g \cdot \sin \alpha$$

R_{grade} grade resistance
m mass
g acceleration of gravity
α angle of gradient

Since the grade resistance in its absolute form depends on the weight of the train, it is not suitable to put down the grade resistance in timetables or similar documents. Usually, the grade resistance is determined by a specific form, which is independent of the weight of a train.

The specific grade resistance equals the absolute grade resistance divided by the weight force. Therefore, the specific grade resistance equals the sine of the angle of gradient. Because of the relatively small angle of gradient of a railway line, the sine can be replaced by the tangent without a significant loss of accuracy. In doing so, the specific grade resistance equals the grade quotient.

$$r_{grade} = R_{grade} / (m \cdot g) = \sin\alpha \approx \tan\alpha = i$$

r_{grade} specific grade resistance
R_{grade} absolute grade resistance
m mass
g acceleration of gravity
i grade quotient

The grade quotient is either measured in ‰ or directly as the quotient 1 : n (e.g., a grade of 25 ‰ equals a quotient of 1 : 40).

The same way, any other resistance can be transformed into its specific form by dividing the resisting force by the weight ($m \cdot g$). Thus, all resistances can be easily expressed by an equivalent rising grade. The tractive effort can also be transformed into its specific form, thus being represented by an equivalent falling grade. In metric measures, this is extremely simple because the amount of a specific resistance or effort directly equals the amount of the grade without requiring any conversion. The specific resistance or effort is measured in N/kN (Newton per Kilonewton) with 1 N/kN = 1/1000 = 1 ‰. In non-metric measures as used in US operations, specific resistances are stated in lb/ton (pounds per ton). To transform this measure into an equivalent grade, it must be multiplied with a factor to eliminate the influence of the non-metric system.

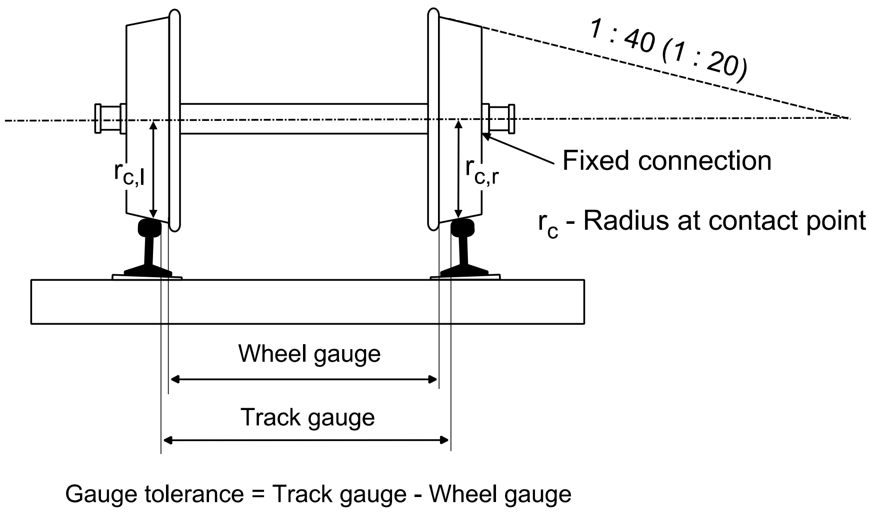


Figure 2.2 Wheelset on track

The curve resistance is caused by lengthwise gliding and crosswise gliding between the wheels and the rail. There is a fixed connection between the wheels of railway vehicles and the axle forming a wheelset (Figure 2.2). Thus, both wheels of a wheelset always have the same revolution speed. There is no differential like there is in highway vehicles. The running surface of the wheels forms a cone with the axle at an angle of 1 : 20 (USA and Russia) or 1 :

40 (Europe). So, if a wheelset is not exactly centred on the middle of the track, the radius at the contact points of both wheels will differ.

Because of the same number of revolutions, the wheel with the greater radius at the contact point moves at a greater speed. This results in a slant position of the wheelset on the track. Consequently, the wheelset will be ‘steering’ to the centre of the track. The same process takes place on the other side. As a result, the wheelset is running in the shape of a sine curve (Figure 2.3).

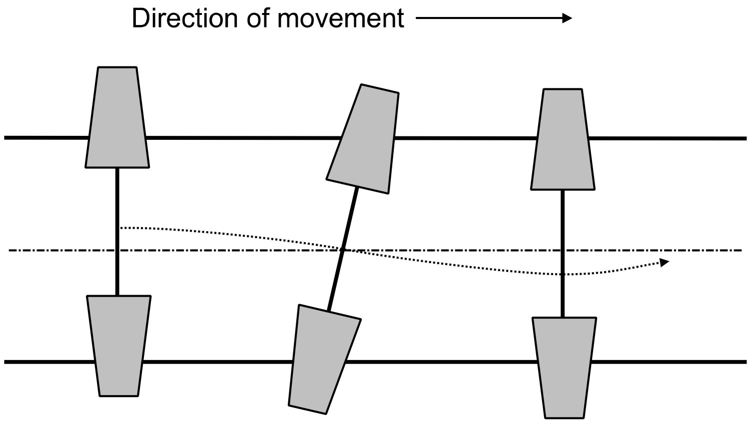


Figure 2.3 Sine running on a straight track

On a straight track, this ‘sine running’ leads to a self-centring of the wheelset without the guidance of the flanges. Although this is the great advantage of a fixed wheelset, a problem occurs when the train is running through curves. In curves, the running difference between the wheels of a wheelset results in lengthwise gliding producing a gliding resistance.

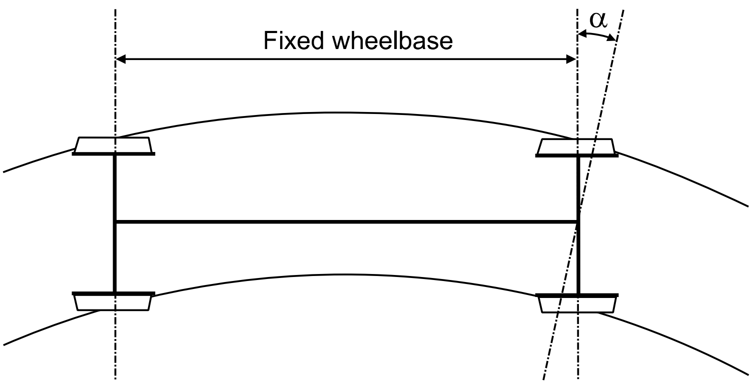


Figure 2.4 Crosswise gliding

When going through a curve, the wheels are guided by the flanges. Because of the parallel bearing of at least two wheelsets, the axles are not exactly aligned to the centre of the curve. The difference in angle is compensated by crosswise gliding creating an additional gliding resistance (Figure 2.4). The effect of crosswise gliding depends on the fixed wheelbase. It can be reduced by using 4-axle vehicles where the axles are grouped in radial steering bogies. On 2-axle vehicles, which are still very common on European railways, the same can be effected by steering single axles.

For standard gauge railways, the curve resistance can be approximated by the following tailored equation:

$$r_{curve} = 700 / R$$

r_{curve} specific curve resistance in ‰
R radius in m

Compared with the grade resistance, the curve resistance does not have a significant influence. But at places where curves are located on a steep grade so that the resulting line resistance is not acceptable for train operation, the curve resistance has to be compensated by reducing the grade. Therefore, the grade is reduced by the amount of the equivalent grade that equals the specific curve resistance.

2.1.3.2 Train Resistance

The train resistance is a result of the following components:

- Rolling resistance
- Bearing resistance
- Resistance of the traction system
- Dynamic resistance
- Air resistance

The rolling resistance is effected by elastic deformation at the contact points between the wheels and the rail. The bearing resistance and the resistance of the traction system are caused by gliding effects. The dynamic resistance results from the loss of kinetic energy by dynamic processes (vibrations) within the train consist. It is extremely complicated to find a mathematical model to exactly describe all these partial resistances. Therefore, the train resistance is not calculated from its components but approximated by empirical equations. These equations have the following form:

$$r_{train} = \alpha + \beta \cdot v + \gamma \cdot v2$$

r_{train} specific train resistance
v speed
α, β, γ empirically determined coefficients

On the basis of this general formula, railways have established lists of approximation formulas to calculate the train resistances of their own various types of trains. Examples are the formulas by Strahl and Sauthoff (*Hansen & Pachl, 2014*) which are used by German railways and that are very suitable for European train consists, and the Davis formula developed for North American operating conditions (*White & Krug, 2005*). The resistance of the locomotive is usually calculated separately from the rest of the train consist (*Wende, 2004; Pachl, 2022*). This makes it easy to calculate resistances of trains hauled by different types of locomotives.

In IT tools for timetabling and running time estimation, the equation is either automatically selected depending on the train type, or the user can select the desired equation manually. In tools where the train consist is modelled very detailed, the coefficients for the train resistance equation are usually calculated automatically. In tools that do not provide a detailed model of the train consist, standard values based on the train data are suggested that can be manually modified by the user.

2.1.4 Grade-Speed Diagram

The grade-speed diagram is a summarised description characterising the movement dynamics of a train. It shows the maximum grade at which a train can run, as a function of speed. There are different curves in the diagram for different train weights (Figure 2.5). The curves are calculated by the following equation:

$$i = (F - R_{loco} - r_{wagons} \cdot m_{wagons} \cdot g) / (m_{loco} \cdot g + m_{wagons} \cdot g)$$

i grade quotient
F tractive effort
R_{loco} absolute resistance of locomotive
r_{wagons} specific resistance of wagons
m_{loco} mass of locomotive
m_{wagons} mass of wagons
g acceleration of gravity

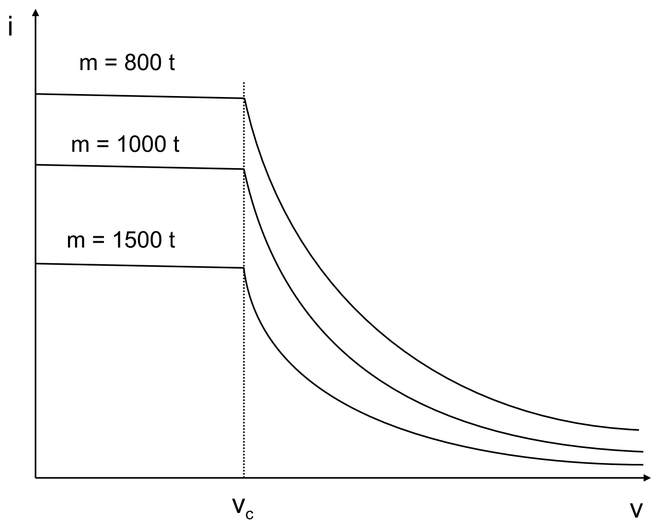


Figure 2.5 Grade-speed diagram

The number of i, which can be seen in the diagram, equals the specific tractive effort that is available either to keep moving at an equivalent grade or to accelerate the train.

$$i_{diagram} = i_{exist} + f_{accel}$$

$i_{diagram}$ number of i obtained from the diagram
i_{exist} existing grade quotient
f_{accel} specific force for acceleration

So, when the train is accelerating on a given grade, it is possible to calculate the tractive effort that will be available for further acceleration. This is a very important piece of information for running time calculations.

2.2 Running Time Estimation

The determination of the running time is done in two steps:

- Calculation of the speed curve
- Integration of the speed curve to determine the running time

When running between two stops, the movement of the train is a sequence of the following modes (Figure 2.6):

- Acceleration
- Cruising
- Coasting
- Braking

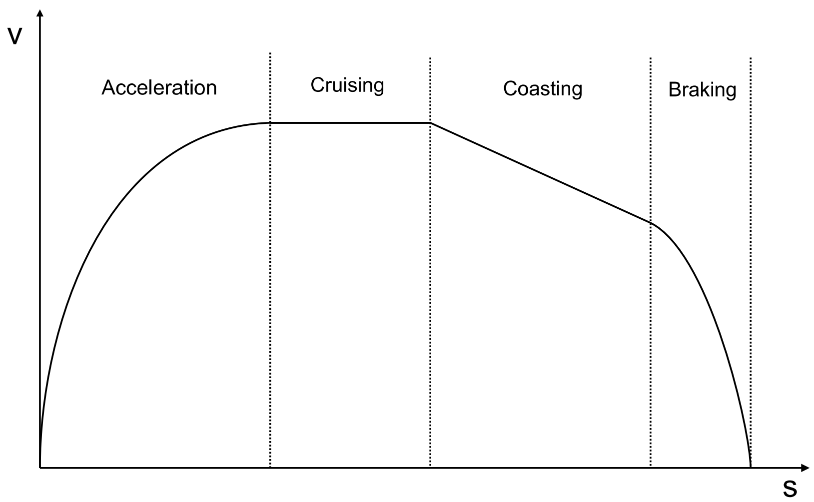


Figure 2.6 Elements of a train movement between two stops

During acceleration, the tractive effort equals the sum of all resistance forces plus the accelerating force.

$$F = \sum R + F_a$$

F tractive effort
R resistance forces
F_a accelerating force

When cruising, the tractive effort equals the sum of all resistance forces without any surplus for acceleration.

$$F = \sum R$$

In coasting, no tractive effort is present. The inertial force of the train is used up by the sum of the resistance forces.

$$R_{intertia} = \sum R$$

$R_{inertia}$ inertial force

While braking, the inertial force is used up by the sum of the resistance forces plus the braking force.

$$R_{inertia} = \sum R + F_{brake}$$

F braking force

Because of the locomotive’s tractive characteristics, an accelerating train has a non-constant acceleration. The line resistance may change at short intervals due to the track alignment characteristics. As a consequence, analytical calculation of the speed curve is impossible. The speed curve can only be approximated step by step in the form of a sequence of straight line portions. The accuracy depends directly on the point density of the polygon.

Here, the calculation in between two neighbouring points of the polygon is demonstrated exemplarily for the acceleration mode (Figure 2.7). In other modes of train movement the calculation follows the same principle.

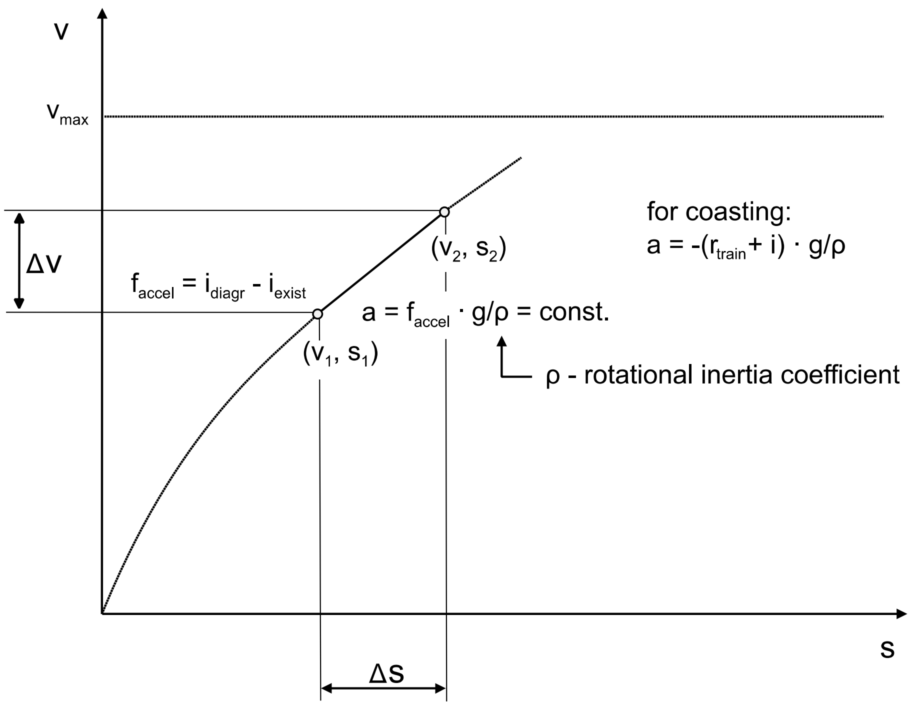


Figure 2.7 Calculation of a single polygon step in acceleration mode

At first, the current acceleration is calculated. The specific acceleration force equals the difference between the grade that can be read from the grade-speed diagram, and the actual grade at this point. From the specific accelation force, the accelaration can directly calculated without transforming the force into the absolute form. For this, the so-called ‘rotational inertia coefficient’ (also known as the 'rotational mass factor' or the 'factor of mass increase') has to be considered. When accelerating a train, some of the energy is lost by the inertia of rotating part of the vehicles. The rotational inertia coefficient describes that influence of the inertia of rotating masses. So, the kinetic energy of a train consists of two parts: translatory energy and rotational energy. The rotational inertia coefficient is calculated as follows:

$$\rho = 1 + E_{rot}/E_{trans}$$

ρ	rotational inertia coefficient
E_{rot}	rotational energy
E_{trans}	translatory energy

On average the rotational inertia coefficient amounts to 1.10. Locomotives have a greater rotational inertia coefficient due to the rotating parts of the propulsion system. For heavily loaded vehicles, the coefficient will decrease because of the lower share of the rotational masses.

With the calculated acceleration, the polygon is to be continued in the form of a straight line up to the next point where the described procedure starts over again. The position of the next points depends on the step width. There are two different ways to find the next point of the polygon:

- Speed interval method
- Distance interval method

The advantage of the speed interval method is that the point density increases with the sensitivity of the speed curve. From the speed curve, the running time can be calculated by numerical integration using the following equation:

$$t = \int (1/v(s))ds$$

The numerical integration is usually done with the same step width as the speed curve polygon.

On railways that work with manual scheduling, the calculated running times are used to establish running time lists. The running time list for a line consists of the running time of each train that may run on this line. In one column the running time is specified for through running without any stops. The extra times necessary for braking and accelerating (Figure 2.8) at stations with stops are specified in additional columns. So, when scheduling a train, the travel time is calculated by adding the extra times for braking and accelerating and the dwell times for all scheduled stops to the running time of a train without stops.

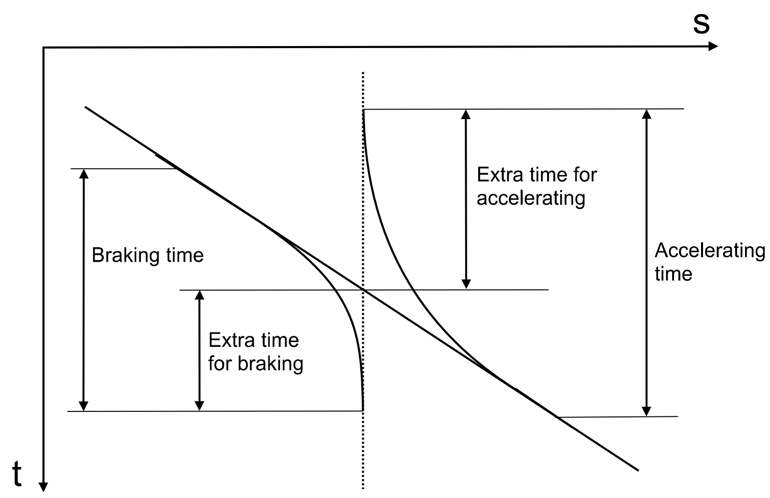


Figure 2.8 Extra time for braking and accelerating

2.3 Train Path Modelling

To model a train path by establishing the time slot needed around the time-distance line for the non-delayed passage of a train, two approaches exist. The blocking time model provides an accurate calculation of all times consumed by a train path. That model is used in advanced computer-based timetabling systems. It guarantees conflict-free train paths even in complex track layouts. A simplified train path modelling is a suitable solution for railway networks of lower complexity, where most conflicts can be detected without a very detailed calculation. Some railways use simplified methods also for complex layouts and apply additional feasibility checks by simulation to eliminate conflicts not detected by the timetabling software. In the past, simplified methods were also the standard procedure for manual timetabling, where conflicts detection mainly depended on the user.

2.3.1 Train Path Modelling by the Blocking Time model

Besides the data needed for running time estimation, the blocking time model also requires a very detailed infrastructure database that contains all relevant operational and signalling data, in particular:

- Signal locations
- Clearing points for the release of routes and block sections
- Route setting times

2.3.1.1 Blocking Time and Headway

For applying the blocking time model, it is crucial to understand the difference between blocking time and headway.

The blocking time is the time window that must be kept clear for the non-delayed passage of a train through a track section (*Hansen & Pachl, 2014; Pachl, 2018*). It is called blocking time (from the German term 'Sperrzeit') because it describes the capacity consumed by this train on this section by blocking the section for other trains.

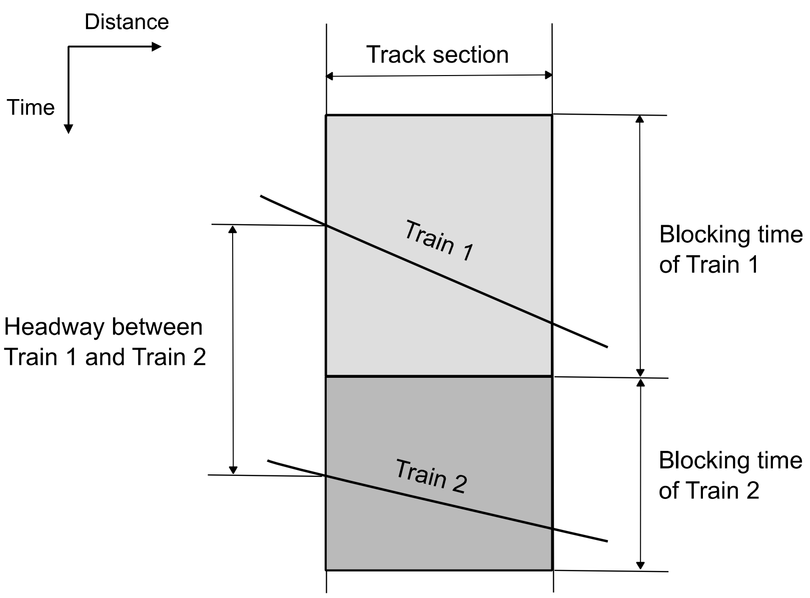


Figure 2.9 Blocking time and headway

The headway is the time interval between two successive trains. It is calculated from the head end of the first train to head end of the second train. To avoid a delay for the second train, the blocking times of the two trains must not overlap. At the minimum headway, the blocking times of the trains would touch each other without any tolerance. Figure 2.9 shows the relationship between blocking time and headway in a time-distance diagram

The two trains follow each other at the minimum headway. If the two trains had the same characteristics, which is not the case here, then the minimum headway would equal the blocking time.

2.3.1.2 Blocking Time in a Fixed Block System

For the non-delayed passage of trains, movement authority to enter block section must be issued, either by clearing a lineside signal or by cab signaling, before an approaching train is forced into a brake application.

The blocking time of a track section is usually much longer than the time the train occupies that section. In a territory with lineside signals, for a train without a scheduled stop, the blocking time of a block section consists of the following time intervals (Figure 2.10):

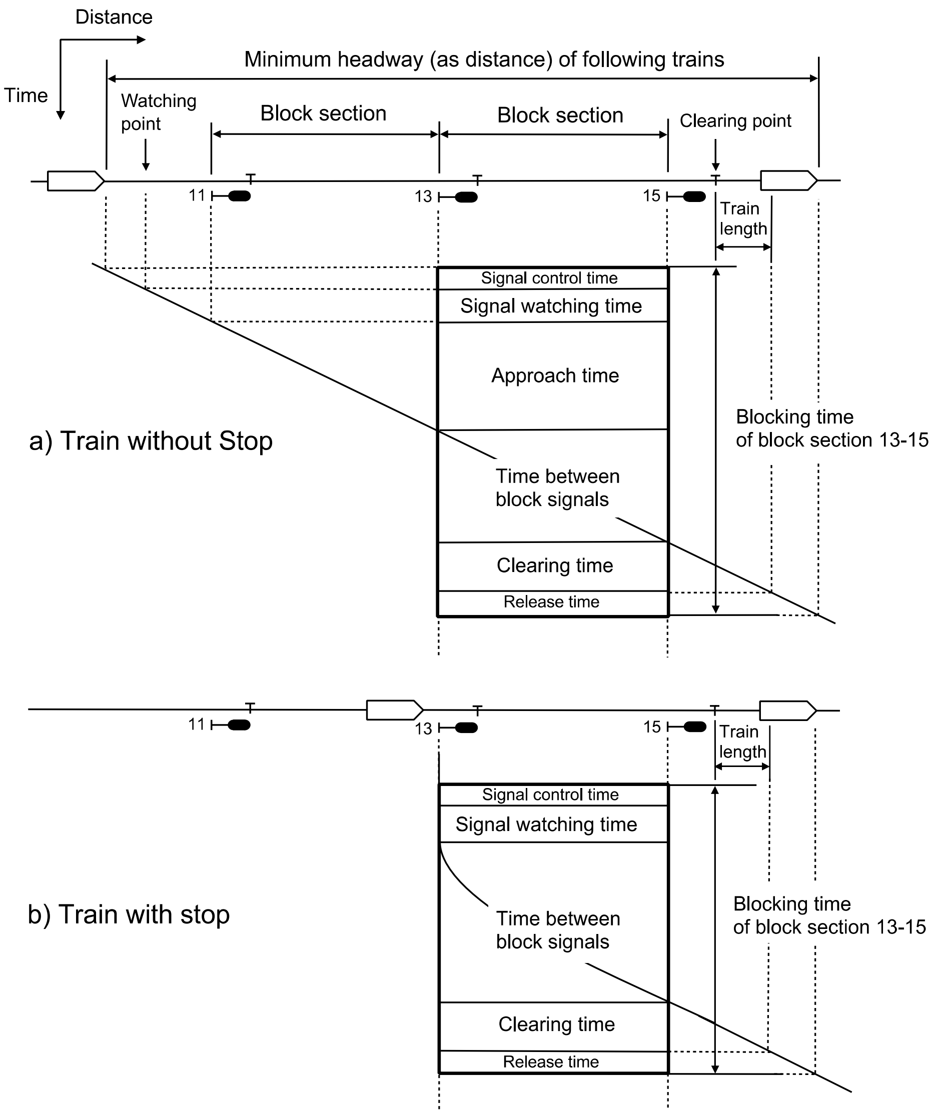


Figure 2.10 Blocking time of a block section with lineside signals

- The signal control time
- The signal watching time, i.e., a certain time the signal that provides the approach indication to the signal at the entrance of the block section must clear ahead of the train to prevent the driver from a brake application
- The approach time between the signal that provides the approach indication and the signal at the entrance of the block section
- The time between the block signals
- The clearing time to clear the block section and – if required – the overlap with the full length of the train
- The release time to 'unlock' the block system

The signal control time is the time from issuing the control command until the signal in the field has cleared. For controlled signals, this equals the route setting time. The approach time equals the time the signal has to be cleared ahead of a train to prevent this train from passing an aspect at the signal in rear that will force the train into a brake application. The signal watching time is needed as an additional reaction time for the driver to not miss the aspect change at that signal.

The approach time does not apply if the train has a scheduled stop at the signal at the entrance of the block section. In such a case, the signal watching time applies at that signal. Then, it is the reaction time from clearing the signal until the train starts moving. Drawing the blocking times of all block sections a train passes into a time-over-distance diagram leads to the so-called 'blocking time stairway' (Figure 2.11). This represents perfectly the operational use of a line by a train.

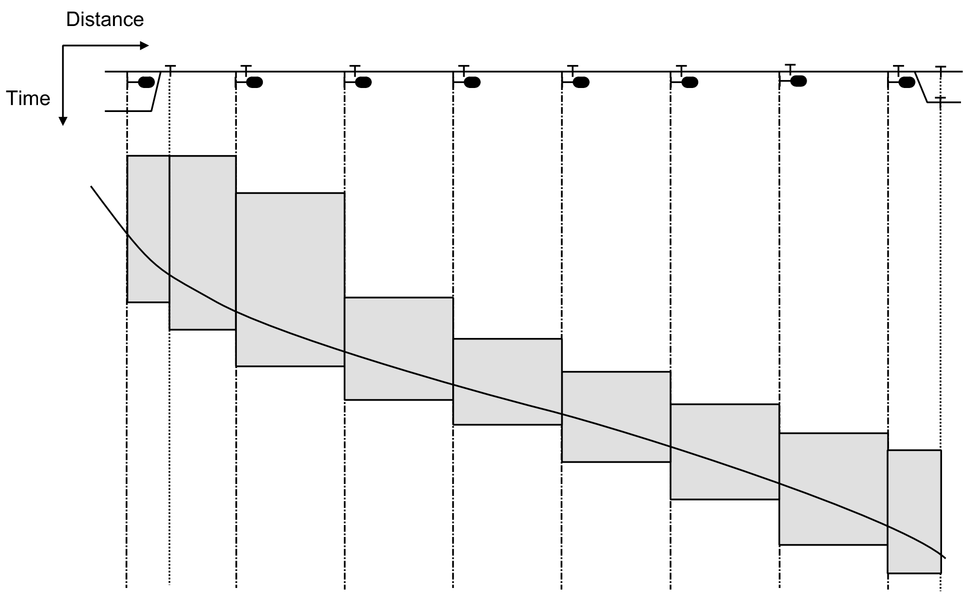


Figure 2.11 Blocking time stairway

On a fixed block line with cab signalling, the approach distance is independent from the location of lineside signals. The approach time is no longer the running time between two signals but the running time within the actual braking distance based on the supervision curves of the cab signal system. Also, due to the absence of lineside signals, a signal watching time to spot a signal aspect at a specific location is no longer needed. The other elements of the blocking time do not differ from a system with lineside signals (Figure 2.12).

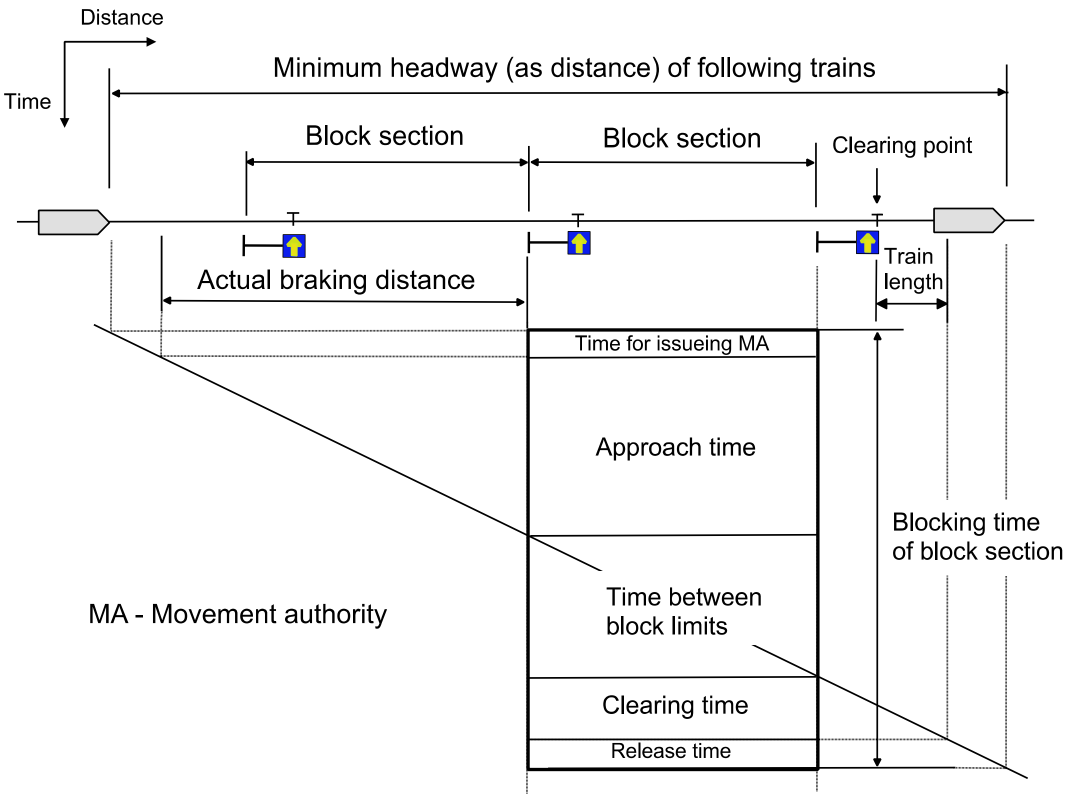


Figure 2.12 Blocking time of a block section on a cab signalling line

Compared with lineside signalling, cab signalling leads leads to some capacity improvement, since the approach time is automatically adjusted to the actual braking distance. In lineside signalling, the approach time is always the running time between the signal that provides the approach indication and the signal at the entrance of the block section even if the actual braking distance is shorter, e.g., for a train running at a lower speed. The elimination of the signal watching time has also a positive effect on capacity. On a line with driverless automatic train operation, the blocking time would actually look the same as in Figure 2.12.

Computer-generated blocking time stairways are a typical feature of advanced timetabling systems to establish conflict-free train paths. By means of the blocking time stairways, it is possible to determine the minimum headway between two trains. The blocking times directly establish the signal headway as the minimum time interval between two following trains in each block section. The line headway is the minimum headway between two trains not only considering one block section but the whole blocking time stairways on a stretch of line where the train sequence cannot be changed (Figure 2.13). In this case, the blocking time stairways of two following trains touch each other without any tolerance in at least one block section (the 'critical block section').

On lines with mixed traffic, the minimum line headway depends significantly on the speed differences between trains. On lines where all trains run at quite the same speed (typical on mass transit railways), the critical block sections are usually the block sections in which the blocking time includes the dwell time of platform stops (station sections, Figure 2.14). On such lines, signals should be placed in a way that keeps the blocking time of the station sections as short as possible.

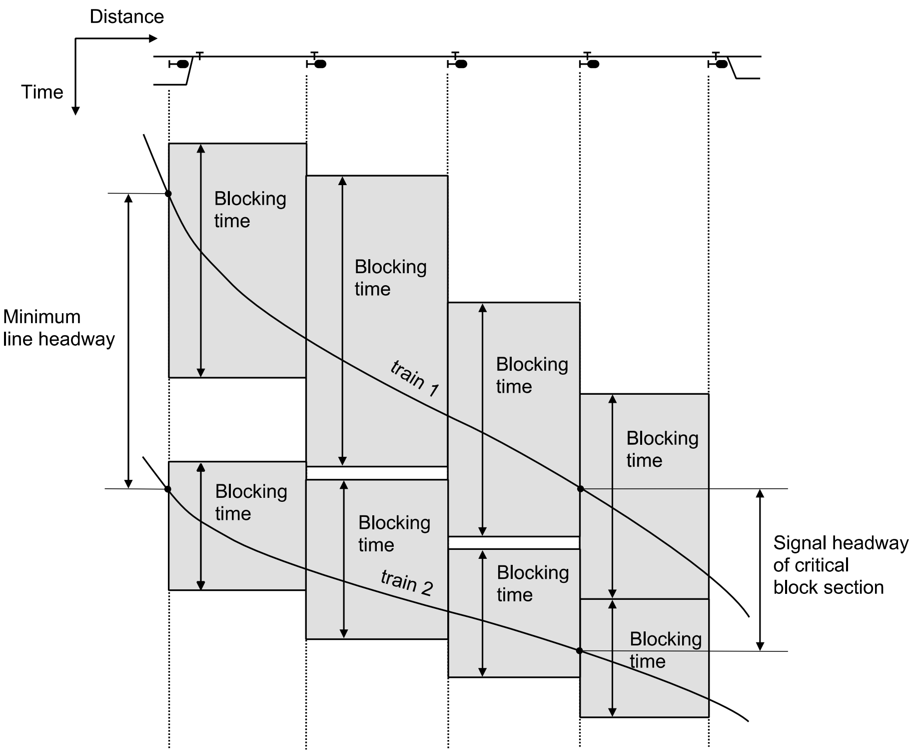


Figure 2.13 Signal headway and line headway

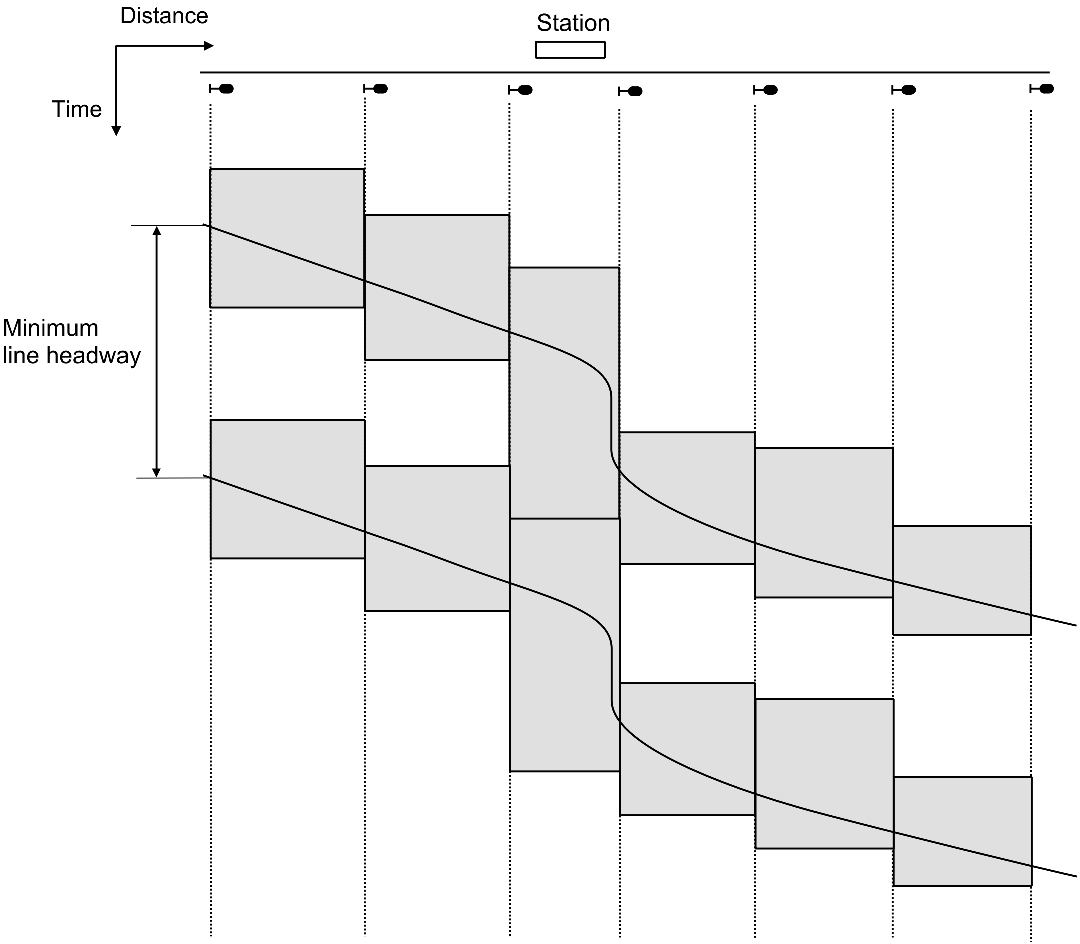


Figure 2.14 Blocking time stairways on a mass transit railway

2.3.1.3 Blocking Time in a Moving Block System

On a moving block line, the length of the block sections is reduced to zero. That means that the running time between the block signals will be eliminated. All other components of the blocking time can also be found in moving block, however. On most lines, the total of these other components is even much greater than the part of the blocking time that can be eliminated by moving block. The main effect of moving block is the elimination of the 'steps' of the blocking time stairways, so that the blocking time stairway will be transformed into a continuous time channel (Figure 2.15).

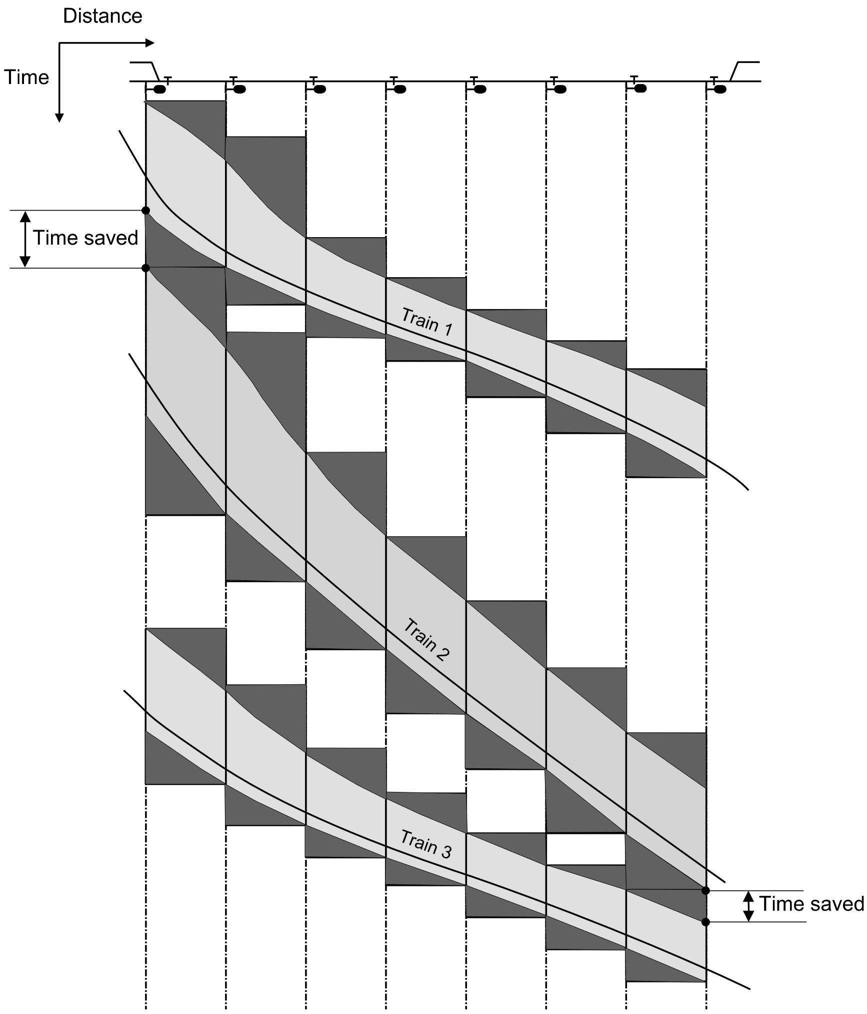


Figure 2.15 Blocking time of moving block compared with fixed block

When comparing moving block with a fixed block with lineside signals, the approach time may be reduced for trains running at a lower speed, since moving block will always use cab signalling or even automatic train operation. So, the approach time is always automatically adjusted to the actual braking distance. The increase in capacity by the introduction of moving block depends significantly on the speed differences between following trains. The best use of moving block is on mass transit lines and on similar railways where all trains run at the same speed profile. On lines with mixed traffic of fast and slow trains, the capacity improvement by moving block is quite limited.

2.3.1.4 Blocking Time in Interlocking Areas

When running through a route in an interlocking area, locked points will release after having been cleared by the rear end of the train. While the train has not yet passed through the entire section, the points that have released behind the train can already be moved to set a route for another train. In the blocking time model, a separate blocking time is estimated for each track section that is separately released. When a route is set, all track sections up to the next signal are blocked at once but released sectionally behind the train. Figure 2.16 shows the blocking time for a route leading through a junction. When the train has passed the route clearing point, the 1st section will release. Then, it will already be possible to start route setting for another route through this point zone. To let another train follow on the same path, all sections must have released, i.e., the train must have passed the signal clearing point beyond the next signal.

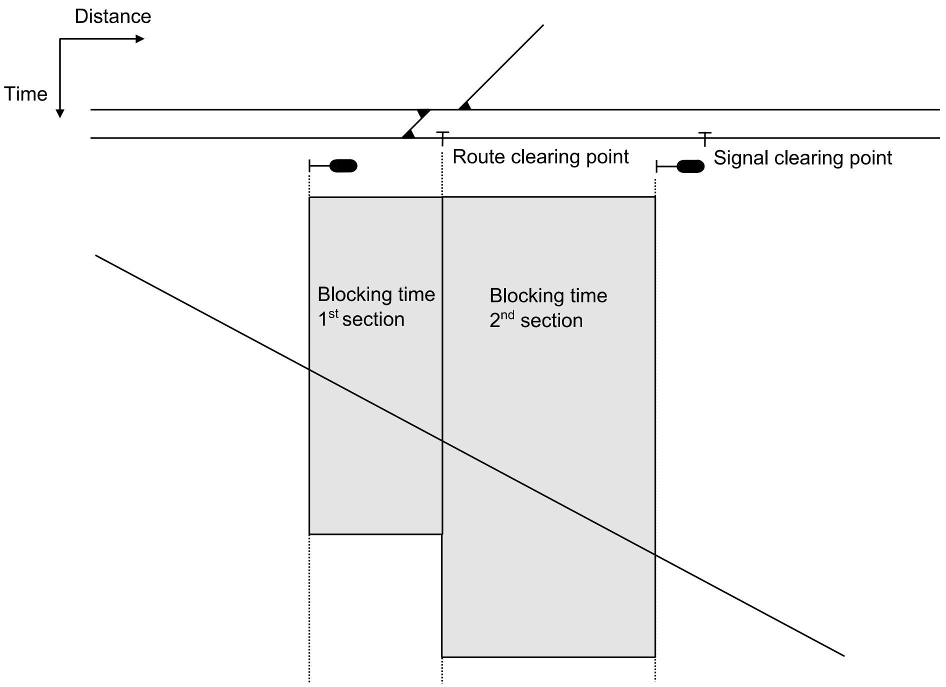


Figure 2.16 Blocking time diagram for a train movement running through a junction

For two trains following each other on the same route, the blocking time of the last track section on the approach to the next signal is relevant for the minimum headway. For two trains running on conflicting routes, i.e., diverging, converging, or crossing routes, the blocking times of the shared track sections are relevant for the minimum headway.

On railways where overlaps are used, the handling of overlaps in the blocking time model depends on the kind of the overlap. A pure block overlap that does not contain any points or crossings will never cause any conflicts with other routes. So, the effect of the overlap is already completely covered by the clearing time as part of the blocking time of the block section. While it would not be a fault, there is no need to establish a separate blocking time for the overlap. In Figure 2.17 this applies to the overlaps beyond signals 11, 13, and 17. In contrast to a pure block overlap, the overlap beyond the destination signal of a locked route may contain points and crossings locked by the route. This way, such an overlap may cause conflicts with other routes. To clearly identify such conflicts, a separate blocking time must be established for the overlap. In Figure 2.17, this applies overlap beyond signal 15. If a train has a scheduled stop and the overlap releases before the route to leave the station is set, the blocking time diagram will show that typical 'nose shape' as to be seen in case b).

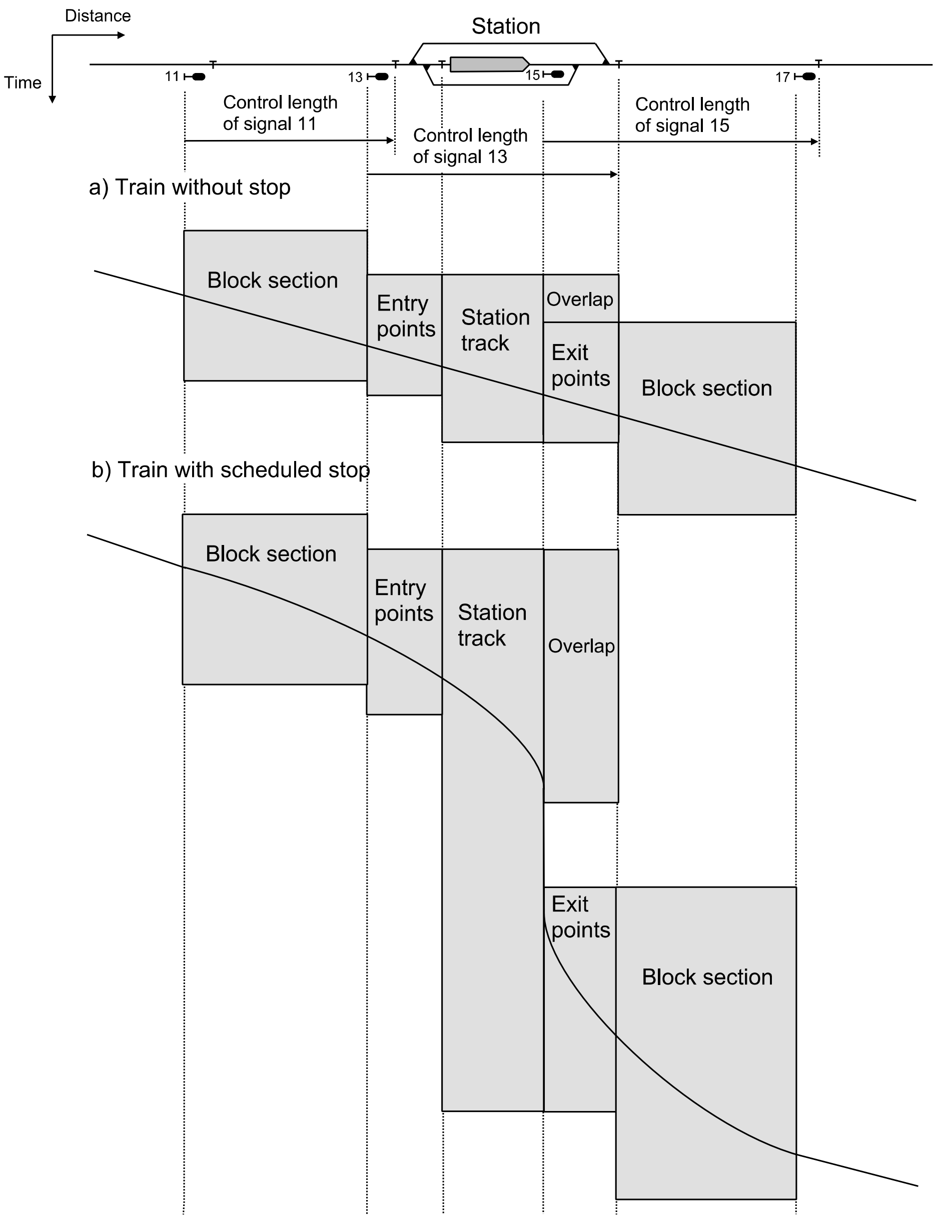


Figure 2.17 Blocking time in an interlocking area with overlaps

2.3.2 Alternative Approach: The Protected Zone Model

In some computer-based systems for timetabling and capacity research, train paths are modelled by a principle different from the blocking time model as explained above. It is here called the protected zone model. The basic idea is that behind every train, the signalling system or operating rules always provide a certain zone to protect that train against following trains. The protected zone can be considered as a time interval or as a track length behind a train. Figure 2.18 shows the protected zone as a track length. The protected zone can be divided into two parts (Figure 2.18 a). First, there is a part of absolute protection, which is marked by red colour. This part consists of all track sections the protected train has exclusive authority to occupy. Second, there is a part in which other trains are forced to slow down to prevent them from running into the part of absolute protection. That zone starts at the first restrictive signal encountered by a following train. In the diagram, this part is marked by yellow colour. The protected zone reaches its greatest length when a train is going to clear a block section (Figure 2.18 b). After the train has cleared the block section, the protected zone has its shortest length and starts to grow again (Figure 2.18 c). The change from the situation of Figure 2.18 b to the situation of Figure 2.18 c represents one 'step' in the blocking time stairway. Diagramming the protected zones with additional times for operating and watching the signals along the path of a train movement lead to picture similar to a blocking time stairway. The main difference is that the approach time is not added at the beginning but at the end of the occupation of the block section.

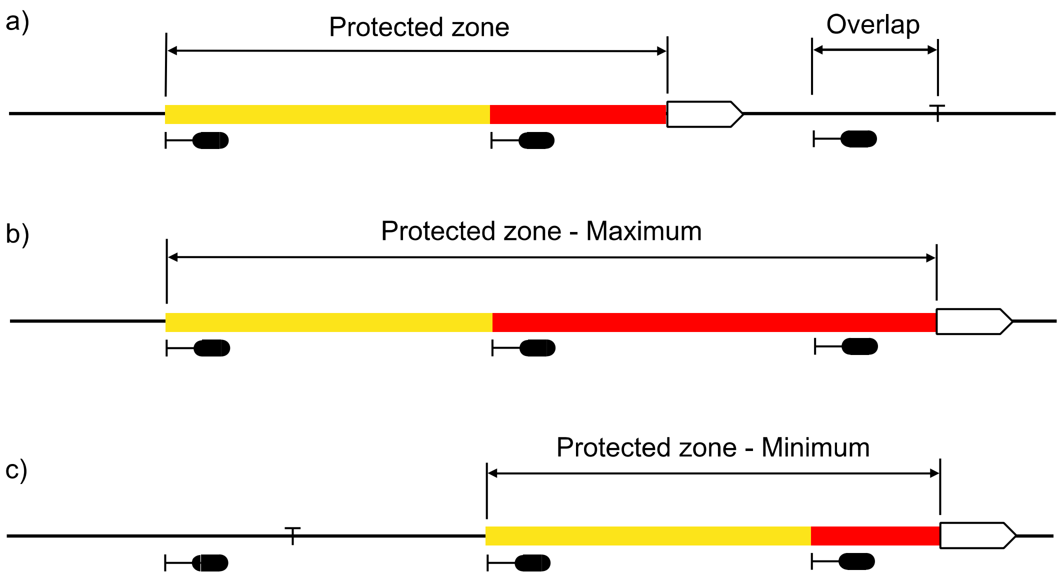


Figure 2.18 The protected zone model

While the protected zone model is in some way similar to the blocking time model, there is one essential shortcoming. The protected zone model cannot handle cab signalling systems in which the signalled braking distance depends on the actual speed of a train. This is the case in all moving block systems but also in most cab signalling systems based on fixed block sections, e.g. ETCS level 2. In such systems, a train can only calculate its own braking distance but not the braking distance for a following train. That is why a model that calculates the braking time ahead of a train, as the blocking time model does, is more appropriate than a model that considers the braking distance as part of a protected zone behind the rear of the train. In traditional signalling systems, in which all trains are governed by fixed signals, the protected zone model works quite well, however.

2.3.3 Simplified Approaches for Train Path Modelling

There are two simplified approaches to be found for train path modelling in timetabling systems:

- To model the occupation of 'timetable sections'
- Using pre-defined allowances at timing points

To model the occupation of timetable sections is based on a principle that has been used for more than a century in manual timetabling. The timetable section are the sections between the timetable locations, i.e., the timing points listed in the timetable. So, each station is just represented by a timing point (Figure 2.19). The track layout of station areas and the location of the signals are not modelled.

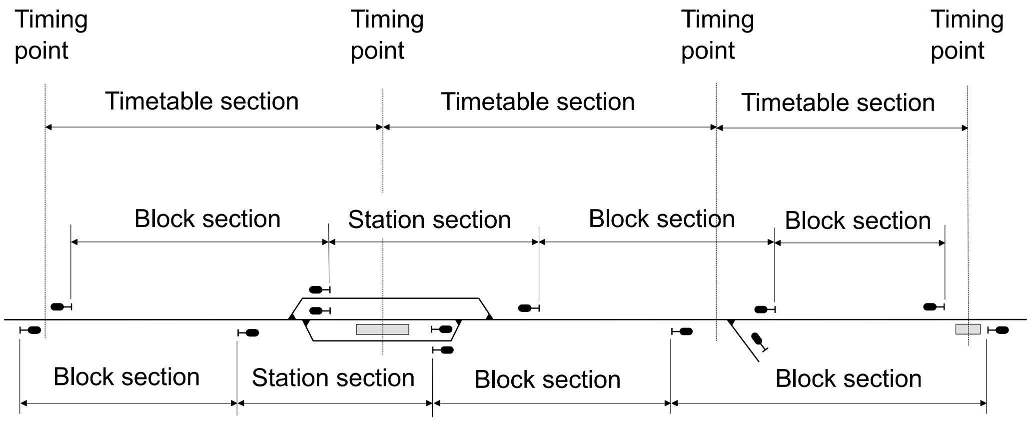


Figure 2.19 Timetable sections for simplified train path modelling

While the limits of the timetable sections do not exactly meet the block section limits, the number of timetable sections must meet the number of the block sections. Except for large station areas with different parts separated by intermediate interlocking signals, there is usually just one timing point for a station area. As a consequence, the station section does not get its own timetable section. However, advanced computer-based timetabling systems will consider the occupation of station tracks in the background. So, while the track layout is not modelled in detail, occupation conflicts on station tracks will be detected.

The occupation of a timetable section is calculated by taking the running time for that section and adding a supplementary time before and after that running time (Figure 2.20).

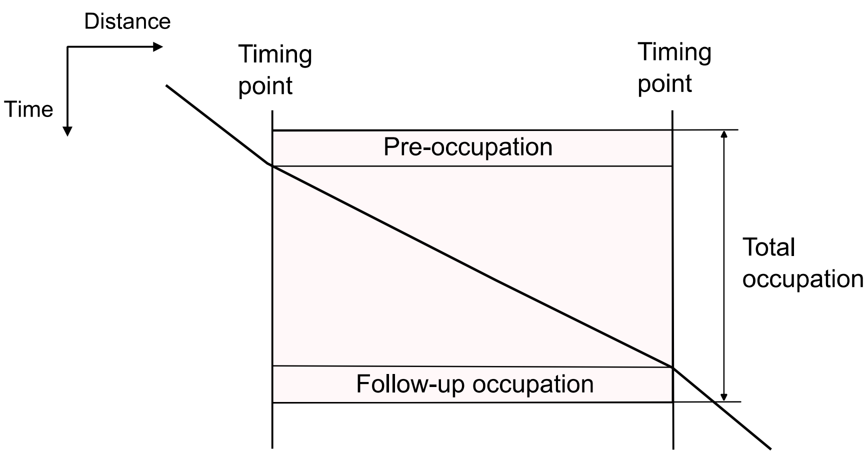


Figure 2.20 Occupation of a timetable section

The standard value of the total supplementary time is usually 1 min, but can be modified individually for each station. In particular, in large station areas where the block section on the approach to the home signal will release before the train has arrived at the timetable location, it might be recommended to adjust the follow-up occupation supplement for more realistic results.

At scheduled stops, the correct allocation of the dwell time to the timetable sections depends on the signal arrangement of the station. At a station with a signal both at the entrance and the exit of the platform, the dwell time neither belongs to the section in approach to the station nor to the section beyond the station (Figure 2.21 a). After the train has arrived, the section in approach to the station is released. The section beyond the station will be occupied when the train departs. At a platform track with a signal only at the exit side, the dwell time belongs to the section in approach to the station (Figure 2.21 b). The section in approach to the station is only released after the train has departed.

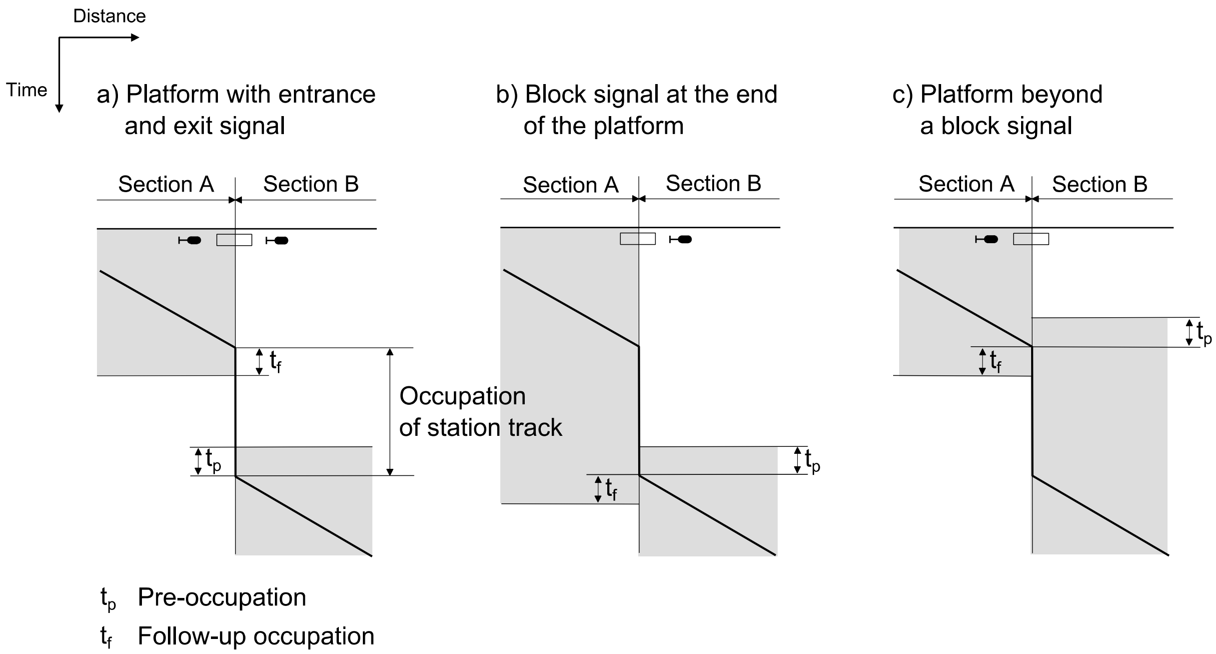


Figure 2.21 Allocation of the dwell time to timetable sections

At a platform track with a signal only at the entrance side, the dwell time belongs to the section beyond the station (Figure 2.21 c). The section in approach to the station is released after the train has arrived. It is not good practice to plan platforms with a signal only at the entrance side. When a train has to wait at the signal because of an occupied block section ahead, the train cannot enter the platform track and use the waiting time for the alighting and boarding of passengers.

The alternative solution is to establish pre-calculated time intervals, so-called allowances, that must be kept clear between trains at the timing points. Depending on the rules of an individual railway, these allowances may also contain buffer times and additional times for engineering works. By these allowances, conflict-free time slots are established around the time-distance lines. That simplifies timetabling significantly but causes some effort by requiring the railway to establish a database with the allowances to be applied at the individual timing points. Since the allowances already consider the minimum signal headways, there is no need to have a timing point at all block limits. That is an essential difference from the previously described method where the occupation was modelled by timetable sections.

3 CAPACITY RESEARCH

The methods of capacity research deliver quality measures for a timetable or traffic pattern on a given infrastructure. This can be used both to evaluate the feasibility and quality of timetables and to estimate capacity limits for infrastructure elements. These methods are also important tools for infrastructure planning to find the optimum track and signal layout design to handle traffic proposed for the future. When talking about capacity, it cannot be described by just one single value. Instead, capacity has to be seen as a relationship between the traffic flow (trains per time), the average speed, the heterogeneity of the traffic pattern, and the recoverability of the timetable.

3.1 Relevant Network Elements

A railway network consists of elements that have a different impact on capacity. While these elements are often investigated separately and evaluated by different quality measures, they are interconnected in many ways. So, the final picture has to look at the interaction of all the elements the network consists of.

The relevant network elements are (Figure 3.1):

- Line sections
- Route nodes
- Track groups

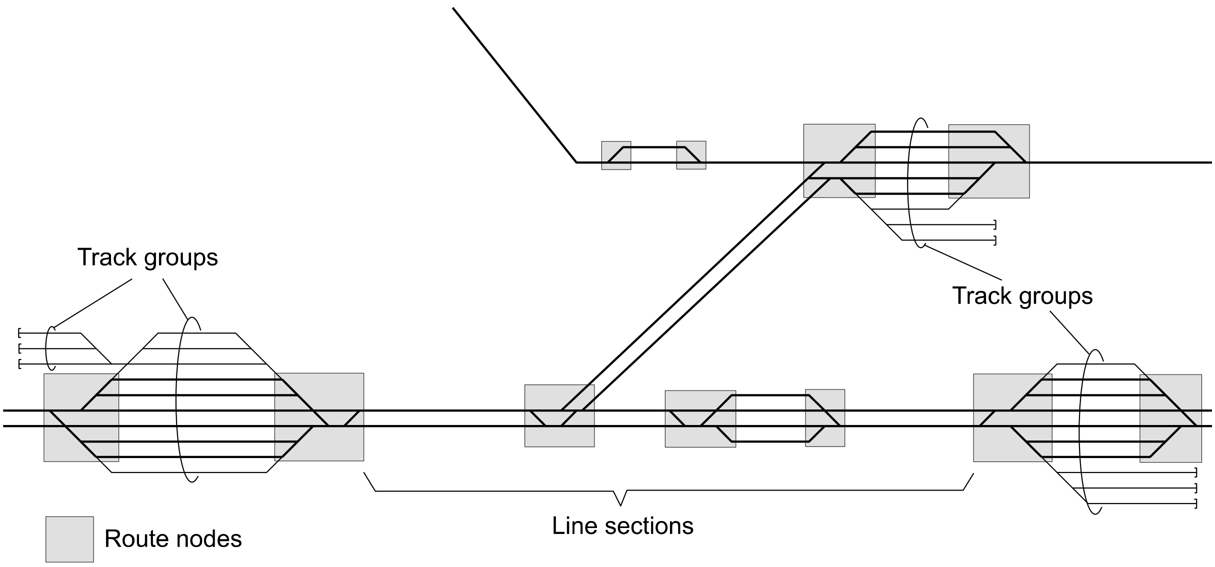


Figure 3.1 Elements of a railway network

The capacity of line sections is described by the traffic flow the line can handle. The research question is either to estimate the maximum traffic flow for an accepted level of operational quality or to evaluate the operational quality of a given timetable or traffic pattern. A capacity limit for a line section can only be stated if the traffic pattern does not change on that section. That means, all trains run through the entire section. If, at an intermediate station, a significant share of trains would leave the line (terminate or diverge from the line) or appear on the line (originate or join the line), the line will have to be divided at that station into sections, for which the capacity limit has to be estimated separately.

Route nodes are the point zones of interlocking areas where trains may take different routes. Typical examples are the route nodes limiting the track group of a station area and junctions on the open line. Conflicting routes in these areas may harm the capacity of the connected lines. So, for route nodes, the research question is not to state a capacity limit in the sense of a maximum traffic flow, but to estimate an acceptable level of route conflicts.

Track groups are parallel tracks in station areas trains could alternatively use. Here, the research question is to estimate the number of tracks needed to handle the traffic volume from the lines leading into and out of the station area.

Due to different research questions, the methods to investigate these elements do also differ. That is why after a general introduction into the methodology of capacity research, the specific methods used for these elements are presented in separate sections.

3.2 Traffic Flow Theory

Before the explanation of the research methods, a basic introduction to the traffic flow theory is provided. It is based on the research of *Hertel* in the late 1980s (*Hertel, 1992*). This can be seen as the rail traffic equivalent of the fundamental diagram of the traffic flow as used in road traffic engineering. While this pure form, it can hardly be found in any commercial research tool, a basic understanding is very valuable for the working with different tools for capacity research.

3.2.1 Waiting Time Diagram

The waiting time diagram shows the average waiting time per train as a function of the traffic flow (Figure 3.2). The waiting time is a measure of the quality of operation. The waiting time curve approaches a vertical tangent, which is the maximum possible traffic flow of the line. This is the theoretical capacity as the maximum output of the line regardless of quality. At this point, the average speed on the line would be zero because the waiting time would asymptotically go to infinity. That the average speed is zero at the maximum theoretical output is not a contradiction. When the input of the line exceeds that maximum capacity, an increasing queue of waiting trains will appear, the line can no longer handle. So, while the line runs at the maximum possible traffic flow, the waiting times will go to infinity. This will consequently bring the average speed down to zero.

There are two kinds of waiting times that may be considered separately:

- Scheduled waiting times
- Delays

Scheduled waiting times are waiting times that are already included in the timetable. These waiting times have to be added for passing and overtaking trains and to move train paths to conflict-free positions. Delays are non-scheduled waiting times that appear in a running operation caused by perturbations forcing trains out of their scheduled train paths so that conflicts may occur that have to be resolved by dispatching decisions.

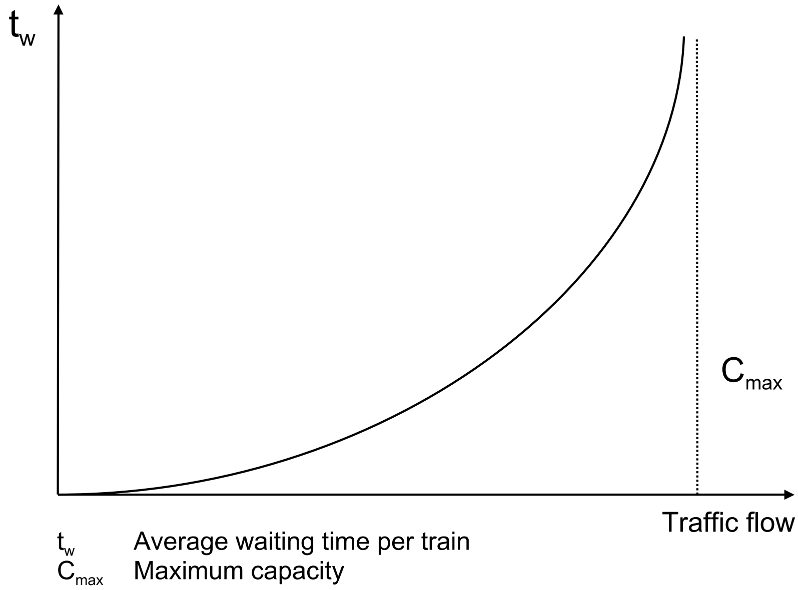


Figure 3.2 Waiting time diagram

The scheduled waiting times occur by solving train path conflicts in the timetabling procedure. Since with the number of trains to be added to a timetable, both the number of conflicts and the average waiting time needed to resolve a conflict by moving a train path to a conflict-free position will increase, the scheduled waiting times increase with the traffic flow and asymptotically go to infinity at a specific limit. One could argue that the scheduled waiting time cannot really go to infinity because the number of train paths that can be inserted in a timetable is limited. However, when the timetable has completely filled with trains without any leftover capacity to add more trains, the next train path to be added will wait forever.

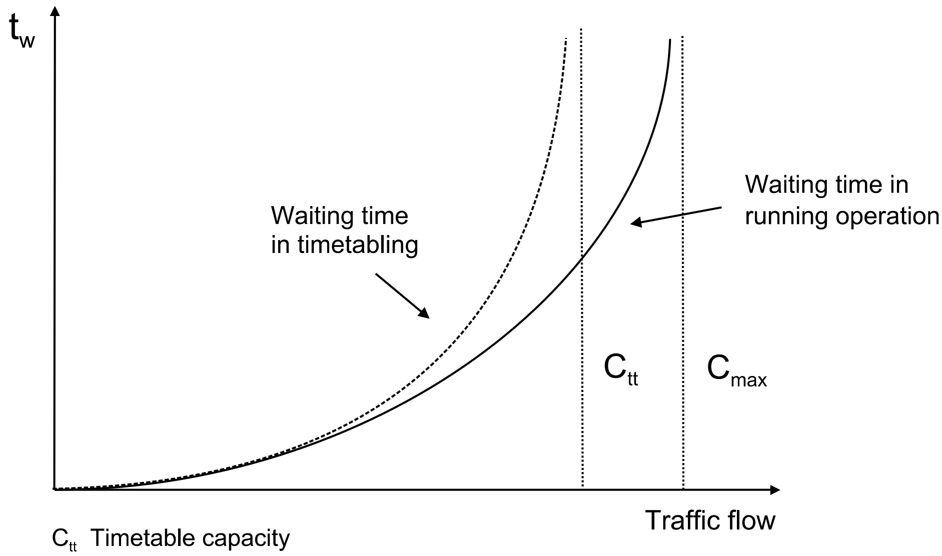


Figure 3.3 Timetable capacity

In Figure 3.3, the scheduled waiting time curve has been added as a separate curve to the waiting time diagram. The capacity limit of this curve is the timetable capacity. The timetable capacity may be interpreted as the maximum number of train paths that could be scheduled without considering buffer times. The waiting times in the running operation are not the arithmetical sum of scheduled waiting times and delays. In fact, the waiting times in the timetabling process are always higher than the waiting times in the running operation, and the timetable capacity is always less than the maximum capacity. The reason is that in the running operation, in case of a delay, slow trains may be caught up by fast trains. This would partially equalise the speed differences by slowing down the fast trains.

The consequences are reduced minimum headways and thus a higher capacity than it would be possible in the timetabling process. In timetabling, conflicts between train paths are usually solved by moving train paths to clear positions or by changing the train sequence. This procedure will produce a significant amount of scheduled waiting time. In case of delays in a running operation, a share of the scheduled waiting times may disappear by abolishing the timetable constraints that have resulted from speed differences. Slowing down a fast train to enable it to follow at a shorter distance behind a delayed slow train will of course delay the fast train by increasing the running time. However, that delay equals the delay the fast train would get by moving its path into a new conflict-free position behind the delayed slow train. So, there is actually no additional time lost, but by slowing down the fast train, transmission of secondary delay to other trains is minimised. Also, the reduced minimum headway will increase the maximum possible traffic flow. Figure 3.4 shows an example.

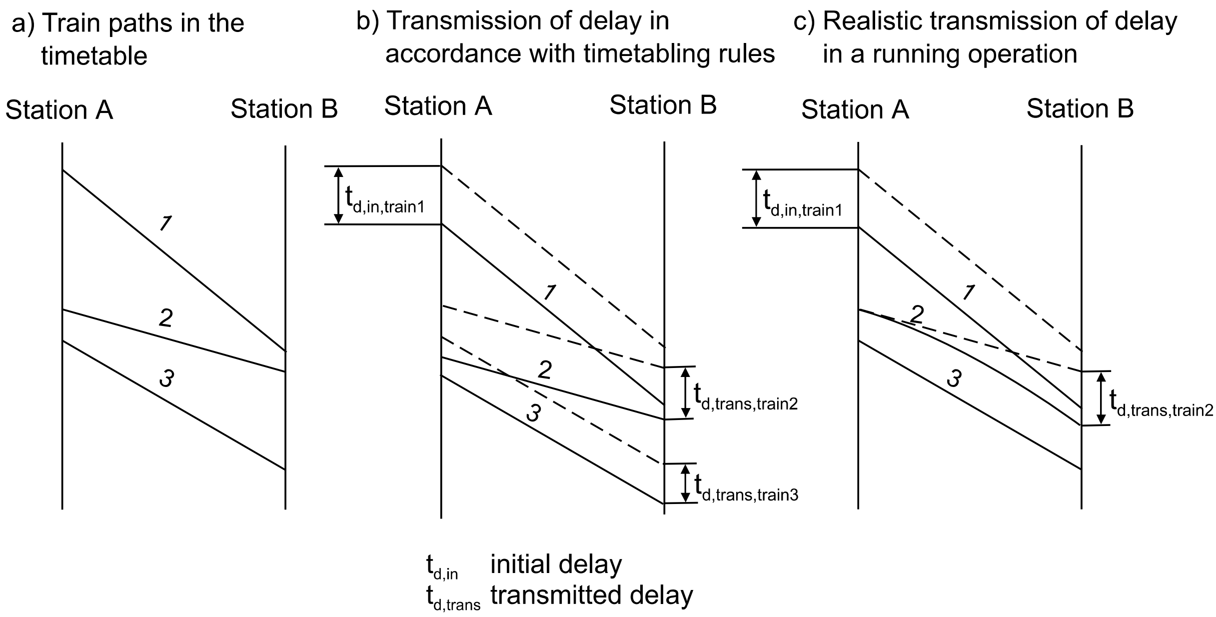


Figure 3.4 Transmission of delay

The left diagram shows a traffic diagram with three following trains. The other diagrams show how the trains 2 and 3 would react to a delay of train 1. The middle diagram shows the transmission of delay when complying with timetabling rules. In real operations control, this is not realistic. The right diagram shows how the transmission of delay is reduced by slowing down a fast train. The fast train 2 gets the same delay as in the middle diagram, but no delay is transmitted to the following train 3.

The fact that the timetable capacity, i.e., the maximum number of trains that can be put into a timetable, is lower than the maximum capacity in a running operation is actually very important. By this, the timetable becomes a valuable tool of capacity control by preventing a line from getting too much traffic. Operating a line near the timetable capacity limit by scheduling trains without sufficient recovery and buffer times may lead to severe delays and a bad quality of operation. However, it is always guaranteed that the line can at least handle the traffic flow. The difference between the timetable capacity and the maximum capacity is also a reserve for quickly reducing of queues in trouble situations by temporarily equalizing the speed differences

The difference between the timetable capacity and the maximum capacity of a line depends on the heterogeneity of the traffic pattern, i.e., the variation of the minimum line headways. The smaller the variation of the minimum line headways of the scheduled trains is, the smaller is the difference between the timetable capacity and the maximum capacity of the line (Figure 3.5).

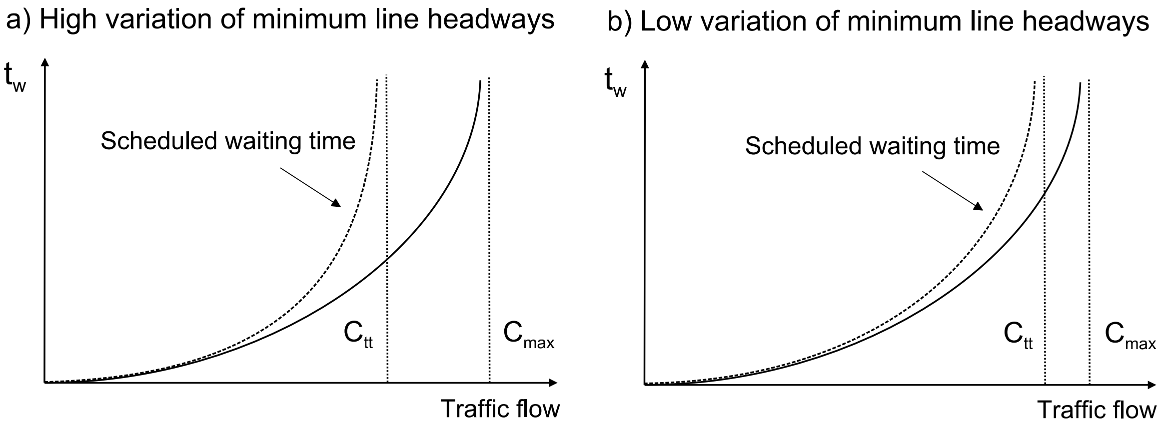


Figure 3.5 Influence of the Heterogeneity on the Difference between Timetable Capacity and Maximum Capacity

The best exploitation of the infrastructure could be achieved by a timetable in which all trains run on harmonised train paths without great speed differences. Such timetables are typical for mass transit railways. However, since the timetable capacity is already close to maximum capacity, there is not much reserve available to increase the traffic flow above the timetable capacity limit. Such a reserve exists on lines with a higher heterogeneity. While the heterogeneity leads to a lower timetable capacity, the traffic flow could exceed that limit in case of trouble situations by equalising train speeds.

On lines with a non-scheduled operation (typical for North American freight operations and also on a number of European industrial railway networks), the protective effect of the timetable capacity does not exist. In North American freight operations, a limit is defined as an acceptable delay ratio that equals the delay time divided by the total time. The accepted delay ratio amounts to 10–15%. The delay ratio can only be measured in the running operation. Thus, if the operation is not very carefully controlled, there is always some risk to overload the line with too much traffic that would lead to a heavy congestion the line cannot handle. On critical lines, the reintroduction of a scheduled operation can be a recommendable option against trying to solve the problem by installation of more complex and expensive dispatching systems.

3.2.2 Recommended Area of Traffic Flow

From the waiting time curve, a recommended area of traffic flow can be derived. If the traffic flow was below that area, the operational quality would be excellent, but there would also be a lot of unused capacity. Above that area, the waiting times would slow down the traffic in an unacceptable way. That area has two limits:

- The minimum of the relative sensitivity of the waiting time
- The maximum of the traffic energy

The sensitivity of the waiting time equals the gradient of the waiting time curve (i.e., the 1st derivative of the waiting time curve). The relative sensitivity of the waiting time results from the sensitivity divided by the absolute amount of the waiting time:

$$SEN_{rel} = t_w'/t_w$$

SEN_{rel} Relative sensitivity of the waiting time
t_w Average waiting time per train

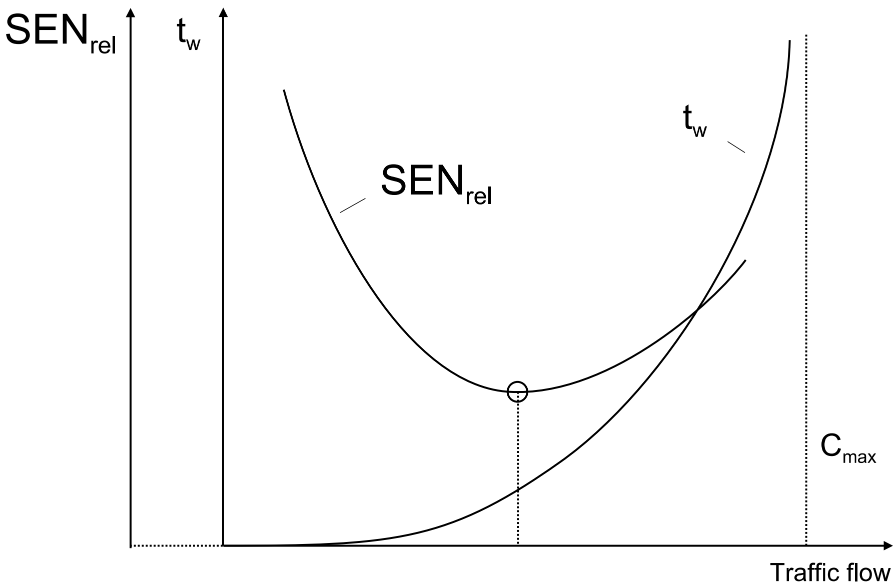


Figure 3.6 Relative sensitivity of the waiting time

The relative sensitivity of the waiting time has a global minimum (Figure 3.6). Up to that point, an increasing flow of traffic only causes a very moderate increase in waiting time. If the line is operated at a traffic flow below this point the capacity of the infrastructure will not be well exploited.

The traffic energy is calculated from the traffic flow multiplied by the average speed. If the traffic flow increases, the average speed will decrease. At the point of maximum line capacity, the average speed is zero. There is an analogy to kinetic energy in physics. The traffic energy may be regarded as the moving ‘mass’ in the form of the trains that are in motion on the line at the same time (the traffic density) multiplied by the square of the speed. This can be transformed into the traffic flow multiplied by the speed:

$$E_{traffic} = n/s \cdot v^2 = n/t \cdot t/s \cdot v^2 = n/t \cdot v$$

$E_{traffic}$ Traffic energy
n Number of trains
s Length of the line
t Time
v Average speed

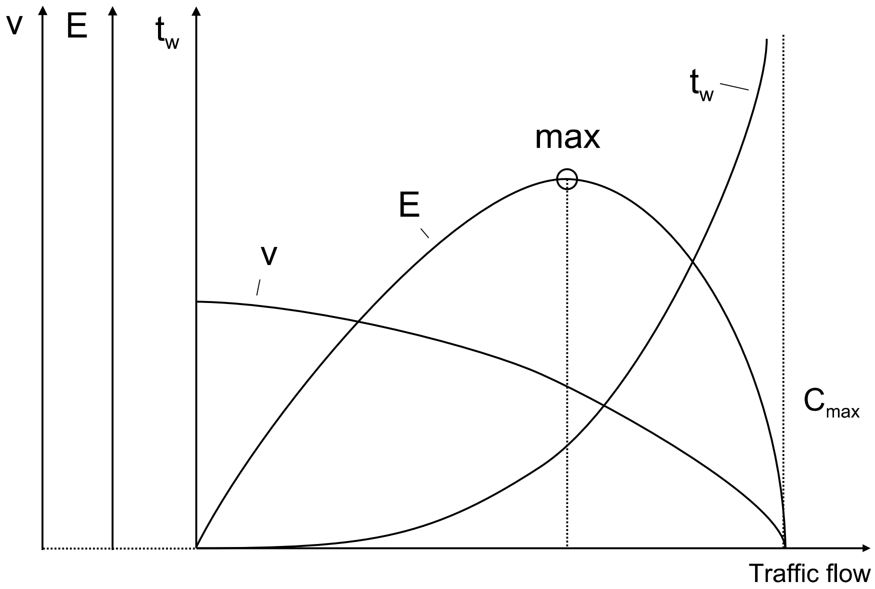


Figure 3.7 Traffic energy

The traffic energy curve has a global maximum (Figure 3.7). Above this point, an increasing traffic flow will cause an inadequate reduction of the average speed. Figure 3.8 shows both limits of the recommended area of operation in a combined diagram. To judge by practical experience, in passenger operations, the traffic flow should be close to the lower limit.

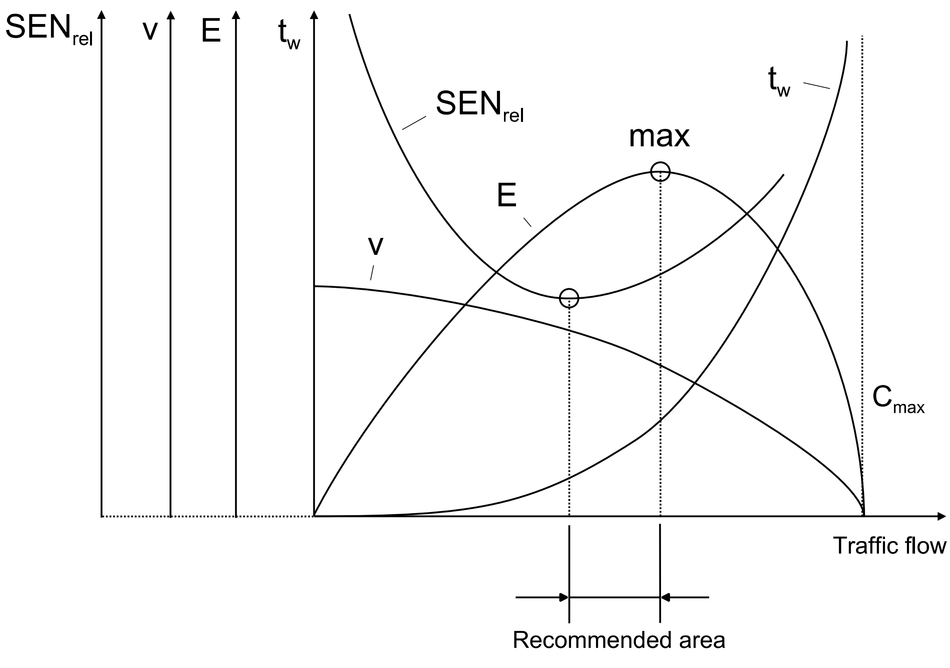


Figure 3.8 Recommended area of traffic flow

In contrast to earlier approaches for capacity evaluation, the capacity limit is not derived from an accepted amount of waiting time but from an analysis of the shape of the waiting time curve. Crucial for the capacity limit is not the amount of waiting time but the stability of the traffic flow. The accepted amount of waiting time at that limit may vary and depends on the traffic pattern. By this principle, this approach is very universal, since the waiting time curve may be derived for infrastructure areas of any size.

The simulation of a lot of European railway lines brought the result that the minimum of the relative sensitivity of the waiting time will be obtained at a traffic flow between 50 % and 60 %, and the maximum of the traffic energy at a traffic flow between 60 % and 80 % of the maximum capacity. In both cases, the higher values correspond to timetables with a lower heterogeneity. In a totally stochastic operation, the bandwidth of the recommended area of traffic flow would theoretically be zero. Then, both limits of the recommended area would meet at a traffic flow of 50 % of the maximum capacity.

3.3 Research Methods

As in many areas of industrial engineering, methods for capacity research of railway systems can be divided into two classes (Figure 3.9):

- Analytical methods
- Simulations

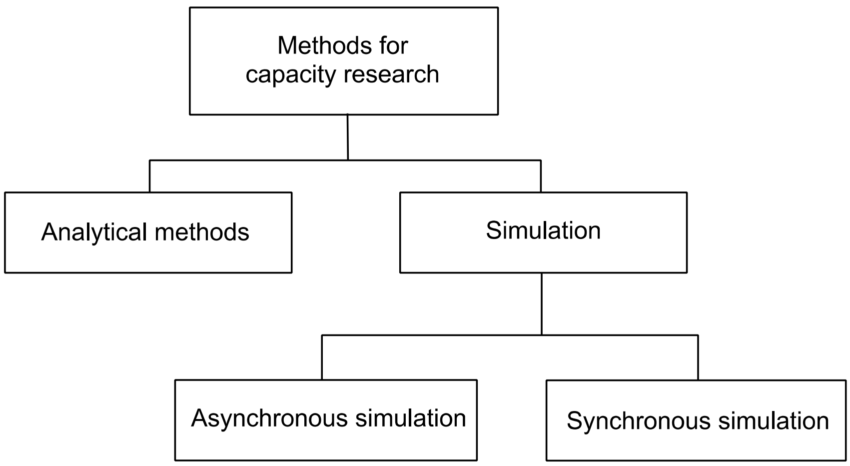


Figure 3.9 Methods for capacity research

The idea behind analytical research is to calculate data from the infrastructure and timetable characteristics to determine and describe the capacity of a line or other parts of a railway network. In contrast to this, a simulation follows the experimental method of natural science. But the experiment is usually done in a computer model instead of the real system. In a synchronous simulation, all processes are simulated simultaneously at a synchronised timeline. In an asynchronous simulation, the processes are simulated successively. Since parallel processes are synchronised on a timeline, each process will consider constraints that have resulted from a previously simulated process.

Analytical methods and simulations have their specific strengths. Analytical methods deliver capacity measures by which the capacity consumption of a given timetable or traffic pattern can be evaluated. However, analytical methods cannot directly check the operational quality to be expected for that timetable. This can be done with simulations that will deliver the delays by simulating the traffic of a given timetable. On the other hand, simulations cannot directly deliver capacity measures. So, whether analytical methods or simulation are the better option, depends on the specific research question.

3.3.1 Analytical Methods

There are two basic approaches in analytical capacity research:

- Estimation of the waiting times for a given traffic flow by queueing theory
- Estimation of the consumed capacity of a timetable by the ‘compression method’

By using queueing theory, the capacity is evaluated by comparing the estimated waiting times with limits at which an acceptable operational quality is to be expected. By using the compression method, the capacity is evaluated by comparing the percentage of capacity consumption with a limit recommended for an acceptable operational quality. By this, the two approaches deliver a similar result. The capacity limit is just stated in a different way.

When applying queueing theory, the calculation of the waiting times requires the infrastructure to be divided into occupation elements that can only be occupied successively keeping a minimum headway between successive trains. This way, such an occupation element may be treated as a single-channel system from the viewpoint of queuing theory. Typical occupation elements are (Figure 3.10):

- A section of line consisting of just one track that may contain several consecutive block sections
- A part of a route node that may contain several points and/or crossings but that must not allow any parallel routes. That means that all routes leading through this element must lock out each other.

While several trains may be present on an occupation element at a time, which is typical for elements consisting of several block sections, they can only enter that element one after another. The key criterion is that it must always be possible to calculate a minimum headway between successive trains for the passage through this element. If a part of the infrastructure contains parallel routes so that trains may enter that part simultaneously, it must be divided into separate occupation elements.

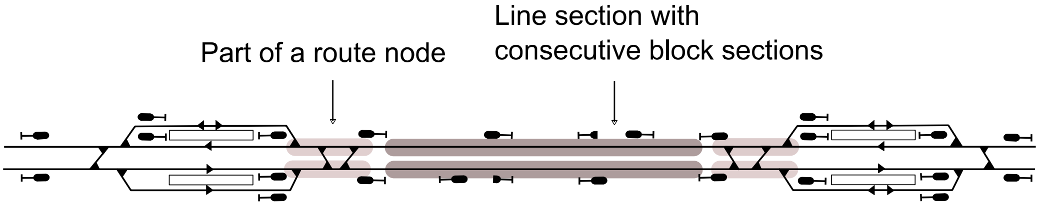


Figure 3.10 Examples for occupation elements

The minimum headway between two successive trains on an occupation element is calculated by the rule that the blocking times must not overlap on that element. Waiting times appear if the time interval between two trains arriving at this element is shorter than the minimum headway. From the distribution of time intervals between trains in the traffic flow, the expected waiting times at occupation elements are estimated by the queueing model. Due to the complexity of the queueing models, this is only possible by the support of computer tools. By this principle, waiting times are always assigned to an occupation element. All waiting appears at the entrance of an occupation element, usually at a signal that protects that element. The interdependencies between the occupation elements cannot be completely modelled.

With queueing models, the feasibility of a timetable cannot be checked. The proof that the waiting times will not exceed a given quality limit is not yet a proof that a particular timetable will work under given constraints, e.g., a cyclic timetable with fixed intervals between trains. So, for traffic patterns with specific timetable constraints, a timetable study could be added to the waiting time estimation.

When applying the compression method, a given traffic pattern is virtually compressed by moving all train paths to the closest positions without any buffer times left. From this, the consumed capacity can be calculated and compared with limits that should not be exceeded for an acceptable operational quality. The compression can be performed either by searching the critical buffer path in the timetable structure, or it can be calculated by the average minimum line headway. For the calculation approach, a finished timetable is not yet needed. It is sufficient to estimate the minimum line headways for all train combinations and the frequency of these train combinations. This will be demonstrated in an example in chapter 3.4. The compression method is mainly recommended for the investigation of longer stretches of line. In complex interlocking areas, it will reach its limits.

3.3.2 Simulations

In contrast to analytical methods, simulations can record waiting times at any location without assigning them to a specific occupation element. Since trains may run without keeping conflict-free minimum headways, conflicts may occur in the following line section. So, waiting times can not only occur at the entrance of an occupation element but also at any location within an occupation element. As a consequence, capacity limits developed for analytical methods cannot directly be adopted for simulations. Simulations allow capacity research of complex infrastructures up to large networks. However, a high level of skill and experience is required from the user to get reasonable results from the great amount of data that may be delivered by simulations.

By simulating an increasing traffic flow, it is theoretically possible to derive the waiting time curve. A challenge for applying this method is to find a suitable procedure to gradually increase the traffic flow in a way that the traffic pattern does not significantly change. There has been some research on that approach, but it has not become common practice.

The maximum capacity of the line can also be determined without generating the whole waiting time curve. For this, an unlimited input of trains is simulated at the entrance of the line. Then, the traffic flow that could be measured at the exit of the line is the maximum possible traffic flow. An unlimited input means that the first block section will always be immediately occupied by a new train after it has been cleared by a train ahead. At this point, the line is overloaded with too much traffic and no longer able to handle the traffic flow. While a line should never be operated near that limit, it might be useful to determine that maximum possible traffic flow. Just by that limit, the recommended area of traffic flow can be roughly estimated using the percentages given in section 3.2.1.

The main use of simulations is to check the operational quality of a timetable or traffic pattern on a given infrastructure. For this, there are three strategies:

- Single simulation run of a given timetable without any stochastic delays
- Multiple simulation run of a given timetable with stochastic delays
- Multiple simulation runs with a stochastically generated traffic pattern

The only use to perform a single simulation run without any stochastic delays is to check the feasibility of a given timetable, i.e., to check that the timetable does not contain any undetected conflicts. If the timetable was established by an advanced timetabling tool based on the blocking time model, it is already guaranteed that there are no train path conflicts. However, there might be undetected conflicts with shunting moves that do not appear in the timetable. A timetable created by a timetabling tool with simplified train path modelling may also contain undetected train path conflicts.

When a conflict-free timetable already exists, the recoverability of the timetable can be checked by multiple simulation runs with stochastic delays. Multiple simulation runs without a timetable but a stochastically generated traffic pattern may be used to test how the infrastructure will react to different operating patterns. This is rather used by infrastructure planners to optimise the infrastructure design.

There are different strategies to evaluate the operational quality of a timetable or traffic pattern:

- Recording delays at specified locations to evaluate infrastructure elements by comparing entrance and exit delays for these elements
- Tracking the average delay over a longer stretch of line to detect critical locations at which the delay would sharply increase
- Calculating the travel time quotient, which is the actual travel time divided by the shortest possible travel time

The travel time quotient is a very valuable measure since it states the waiting time not as an absolute value but in a relation with the travel time. Depending on the question to be investigated, the shortest possible travel time is either the scheduled travel time without any delays, or the travel time that would be possible by eliminating all scheduled waiting times that resulted from solving train path conflicts in the scheduling procedure. The waiting probability can be directly calculated from the travel time quotient.

$$q_{tr} = t_{tr,act} / t_{tr,min} = (t_{tr,min} + t_w) / t_{tr,min}$$

$$P_w = (q_{tr} - 1) / q_{tr}$$

q_{tr} Travel time quotient
$t_{tr,act}$ Actual travel time
$t_{tr,min}$ Minimum travel time
t_w Waiting time
P_w Waiting probability

Most simulation tools provide a lot of statistics. But it is not an easy task to get reasonable answers from this mass of statistical data. They can also be used to develop entirely incorrect answers. Interpreting simulation statistics requires a great experience in railway operations. So, a sophisticated simulation is a great tool for capacity research but would never replace the knowledge and experience of the user.

As already stated, there are two simulation principles concerning the internal working of the software tools:

- Asynchronous simulations
- Synchronous simulations

Asynchronous simulations are based on the control principle of rescheduling. The timetabling process is usually simulated separately from the process of a running operation. When simulating the timetabling process, the simulation system puts stochastically generated train paths into a timetable. In case of a conflict, a train path to be added has to be moved to a conflict-free position. Existing train paths are not changed. Since the train paths are inserted in the timetable by the order of their priority, conflicts are always solved by adding waiting time to the inferior train.

The result is a conflict-free timetable. For a given number of trains, the total of scheduled waiting time is derived from the amount of time for which the train paths have to be postponed in order to get a timetable without any overlapping blocking times.

The simulation of the running operation follows the same procedure based on the timetable that was generated previously by simulating the timetabling process. But now, stochastic delays are generated, which will disturb the timetable by causing additional conflicts. These conflicts are again solved by rescheduling rules with respect to the train priorities. In case of overlapping blocking times, train paths are postponed or the train sequence is changed depending on the priority of the trains.

Thus, the simulation of a running operation is done like a ‘disturbed scheduling’. Conflicts are always solved by moving the train paths and their blocking times but not by ‘bending’ them. The effect that a fast train may be slowed down by a slow train ahead is not simulated. As in analytical methods, a train will always wait in front of an occupation element. Consequently, the simulated operation will differ from the real railway operation. Delays found out by asynchronous simulation in a running operation should always be looked upon with some doubt.

In synchronous simulations, all partial processes of railway operation are simulated in real time sequences. Because of that, the results of synchronous simulations are quite close to the data to be expected in a real operation. In contrast to asynchronous simulations, the process of scheduling cannot be simulated directly. But with the help of a computer-based scheduling tool, it is possible to create timetables which could be evaluated by synchronous simulations. Some software packages for synchronous simulations already contain a scheduling module that allows the user to create a timetable for the traffic to be simulated. Today, synchronous simulation is the dominating approach for capacity research.

3.4 Evaluation of Line Capacity by the Compression Method

3.4.1 Calculation Approach

Before applying the compression method, the line has to be divided into suitable compression sections. The selection of the compression sections has a significant impact on the result and should be done with care.

For this, there are the following recommendations:

- On single lines, compression sections should be limited by all stations with scheduled meets and all other stations that can be used for train meets in case of delay.
- On double lines, compression sections should be made as long as possible to cover the running time difference in a sufficient way.
- Compression sections on double lines should be limited by stations at which more than 5% of the trains leave the line or appear on the line.

The compression method can either be performed by calculation of the consumed capacity or by analysing the structure of a given timetable. Figure 3.11 shows the calculation approach. The infrastructure data must contain all data needed for running time estimation and for determining the blocking time stairways for the train paths. For train characteristics, trains with similar movement characteristics may be put into groups as reference trains. This will significantly reduce the calculation effort by limiting the number of train combinations for which minimum headways have to be calculated. The timetable is only needed to get the frequencies of the train combinations. If a completed timetable does not yet exist, the frequencies of the train combinations can also be calculated as a probability from the number of relevant reference trains.

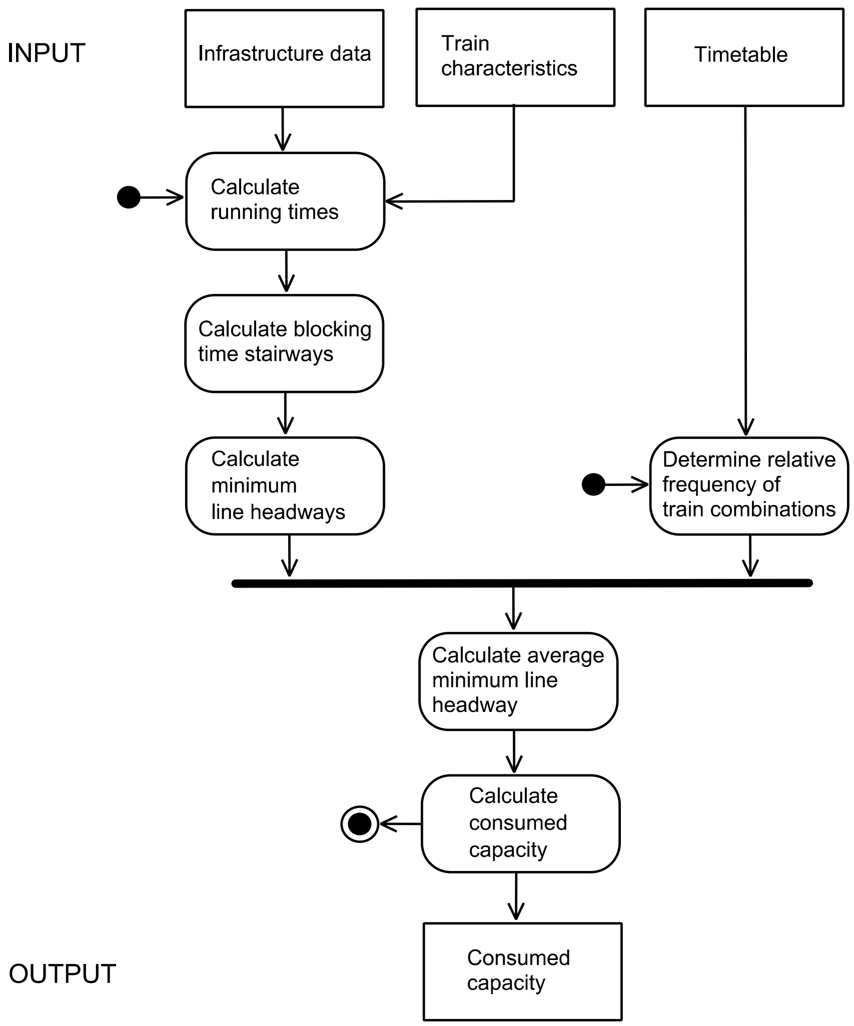


Figure 3.11 Procedure to calculate the consumed capacity

3.4.2 Calculation of the Minimum Line Headways

From the infrastructure data and the train characteristics, running times and blocking time stairways are calculated for all reference trains. So, for each reference train, there is now a completely modelled reference train path.

By these train paths, the minimum line headway is calculated for all combinations of reference trains. The minimum line headway is the shortest time interval at which two trains could pass through the section without any overlapping blocking times (Figure 3.12).

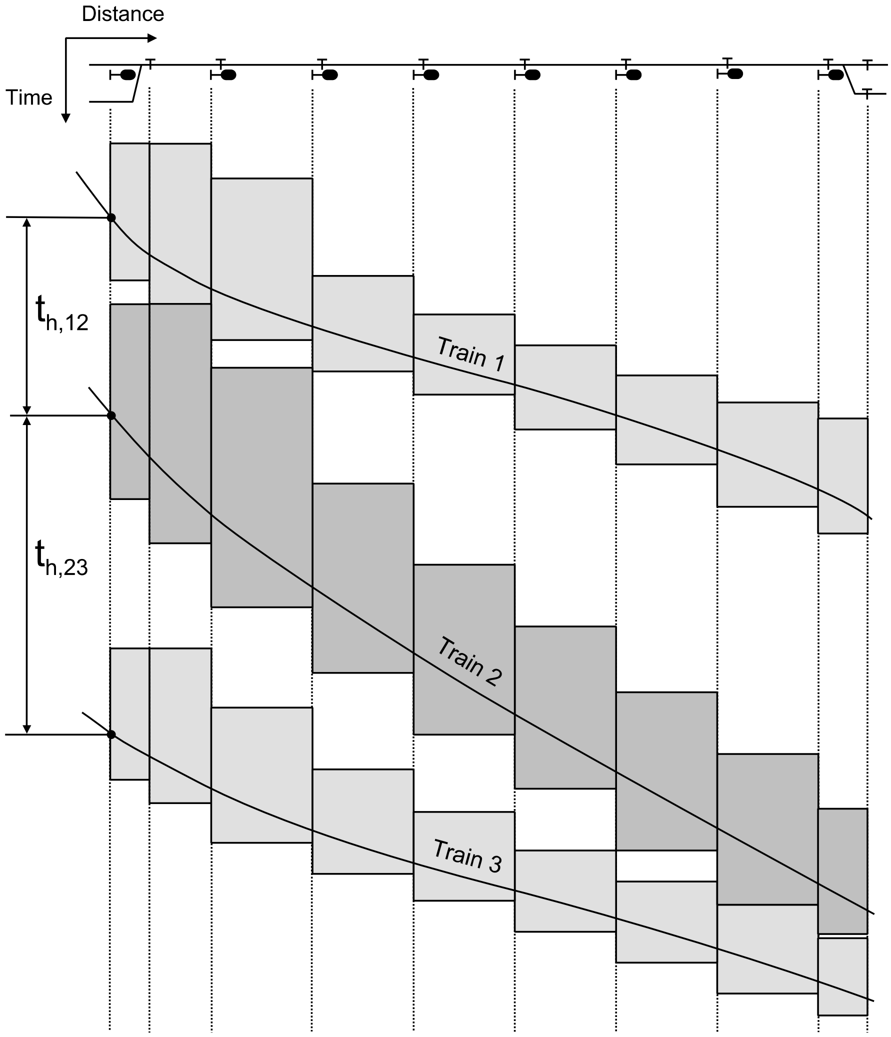


Figure 3.12 Determining the Minimum Line Headways

The section the minimum line headways are calculated for may contain several passing tracks where trains can meet and/or overtake each other. The minimum line headway for a specific train combination (a pair of trains entering the section one after the other) has to be calculated for every single subsection between locations where these two trains could be scheduled to pass each other. The highest value is relevant for the entire section. Instead of assigning headways to a section of line as described here, headways can also be assigned to the timing point of a station. That principle is still common in timetabling, see chapter 4. For capacity research, the section-related approach is more appropriate, however.

The following example demonstrates the calculation of the minimum line headways.

Example 3.1
For three reference trains, the minimum line headways should be calculated for a track of a double track section between two stations with 6 intermediate block signals (see Figure 3.12). Table 3.1 shows the results of running time and blocking time calculation. All times correspond to the theoretical departure time 0 at the left station.

Abbreviations: t_{bb} beginning of blocking time
t_{sig} time when passing the signal
t_{be} end of blocking time

Table 3.1 Running times and blocking times of the reference trains (in minutes)

	Signal	**11**	**13**	**15**	**17**	**19**	**21**	**23**	**25**
Train 1	t_{bb}	-0.9	-0.5	0.7	1.2	1.6	2.0	2.3	2.5
	t_{sig}	0.0	0.9	1.4	1.8	2.1	2.5	3.0	3.5
	t_{be}	1.1	1.5	1.9	2.2	2.7	3.1	3.7	4.0
Train 2	t_{bb}	-1.4	-0.6	0.8	1.9	2.6	3.5	4.3	5.0
	t_{sig}	0.0	1.3	2.2	3.0	3.8	4.5	5.3	5.9
	t_{be}	1.6	2.4	3.2	4.0	4.7	5.5	6.2	6.5
Train 3	t_{bb}	-1.1	-0.3	0.4	1.0	1.5	1.9	2.1	2.6
	t_{sig}	0.0	0.7	1.2	1.7	2.0	2.3	2.7	3.1
	t_{be}	0.9	1.4	1.8	2.1	2.5	2.9	3.3	3.5

When the blocking time stairways of two trains are put one on top of the other (i.e., headway = 0), a blocking time overlap can be calculated for every block section (blocking time overlap = end of blocking time of the 1st train - beginning of blocking time of the 2nd train, see Figure 3.13).

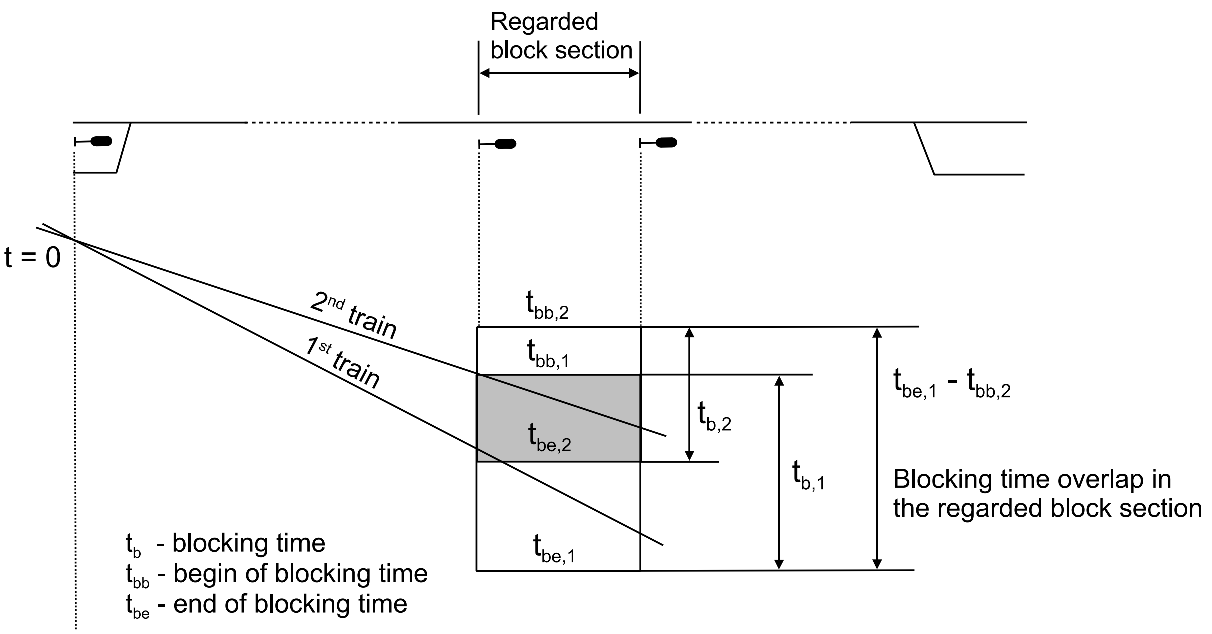


Figure 3.13 Principle of calculating the blocking time overlap of a block section

The blocking time overlap of a block section equals the amount of time the train path of the 2nd train has to be postponed to eliminate the blocking time conflict in this block section. When the blocking time overlaps for all block sections have been calculated, the maximum value equals the time the train path of the 2nd train has to be postponed to eliminate all conflicts between the blocking time stairways. Thus, this time also equals the minimum line headway between these two trains. Table 3.2 shows the calculation sheet with the blocking time overlaps of all train combinations calculated from the data in Table 3.1. The maximum values which equal the minimum line headways are marked.

Table 3.2 Calculation sheet with blocking time overlaps of all train combinations (in Minutes)

Combination		Signal							
1st train	2nd train	11	13	15	17	19	21	23	25
1	1	**2.0**	**2.0**	1.2	1.0	1.1	1.1	1.4	1.5
1	2	**2.5**	2.1	1.1	0.3	0.1	-0.4	-0.6	-1.0
1	3	**2.2**	1.8	1.5	1.2	1.2	1.2	1.6	1.4
2	1	2.5	2.9	2.5	2.8	3.1	3.5	3.9	**4.0**
2	2	**3.0**	**3.0**	2.4	2.1	2.1	2.0	1.9	1.5
2	3	2.7	2.7	2.8	3.0	3.2	3.6	**4.1**	3.9
3	1	1.8	**1.9**	1.1	0.9	0.9	0.9	1.0	1.0
3	2	**2.3**	2.0	1.0	0.2	-0.1	-0.6	-1.0	-1.5
3	3	**2.0**	1.7	1.4	1.1	1.0	1.0	1.2	0.9

In this example, the headways were calculated from the times the trains pass the signal at the entrance of the section. It is also possible to calculate headways from the beginning of the blocking time of the first block section between the two stations. That principle is used in several computer-based systems for analytical capacity research. Although the minimum line headways calculated by these two principles for a specific train combination may differ slightly, both methods will lead to the same average minimum line headway.

3.4.3 Calculation of the Average Minimum Line Headway

If the timetable is known, the frequency of every pair of reference trains can be read directly from the timetable. The average minimum line headway can be calculated by the following equation:

$$t_h = \Sigma\,(t_{h,ij} \cdot f_{ij})$$

t_h average minimum line headway
$t_{h,ij}$ minimum line headway for train j following train i
f_{ij} relative frequency of combination: train j following train i

If the timetable is unknown, the calculation can be made by using relative frequencies which would result from a randomly chosen train sequence. The frequencies can be calculated by the following equation:

$$f_{ij} = (n_i \cdot n_j) / n^2$$

f_{ij} relative frequency of combination: train j following train i
n_i number of train i
n_j number of train j
n number of all trains

A complete calculation example follows after the explanation of the consumed capacity.

3.4.4 Calculation of the Consumed Capacity

The consumed capacity is described by the percentage of use within a considered period of time. For a given timetable, this can be visualised by pushing the blocking time stairways together as close as possible without any buffer times left (Figures 3.14 and 3.15).

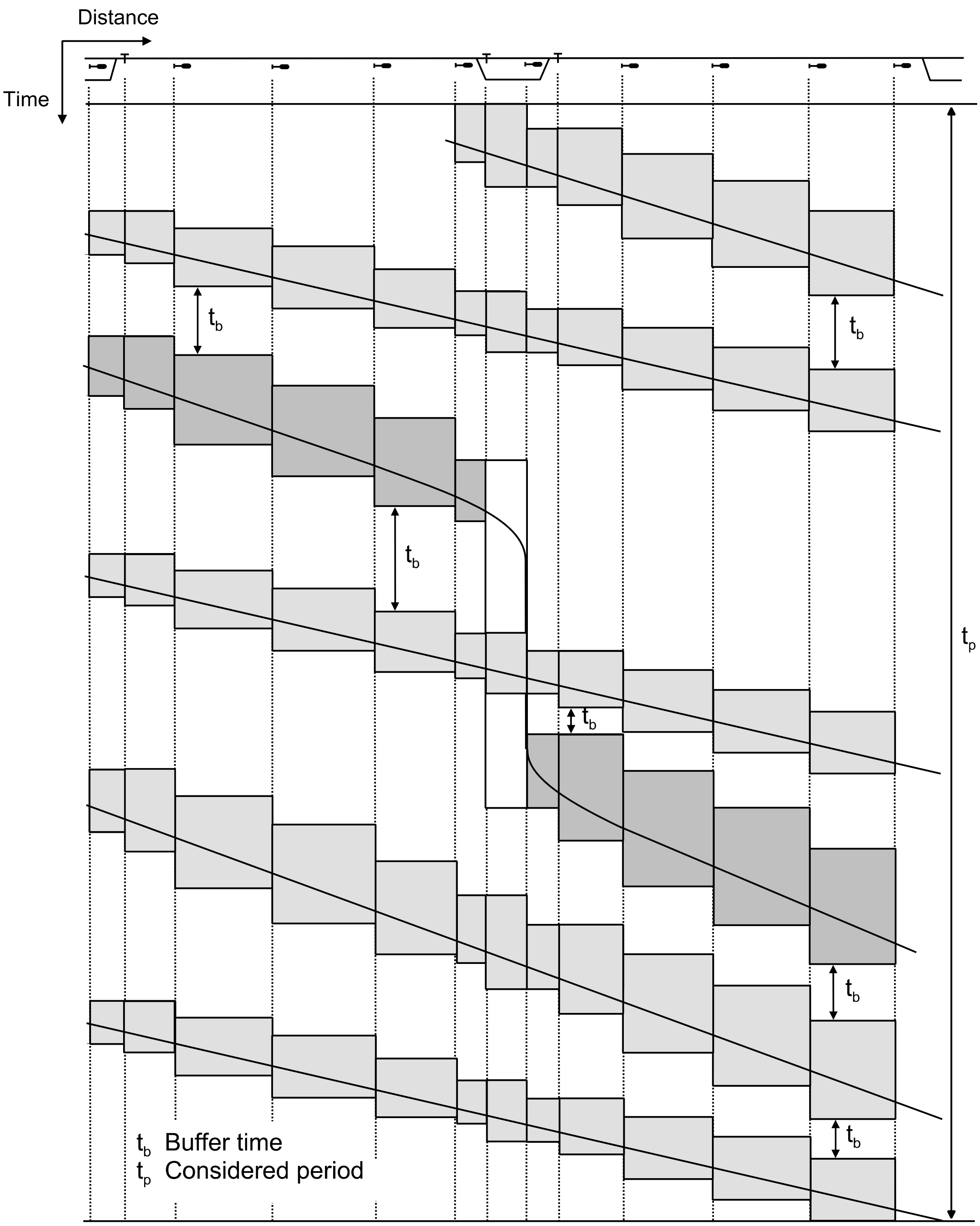


Figure 3.14 Timetable example (one direction of a double line)

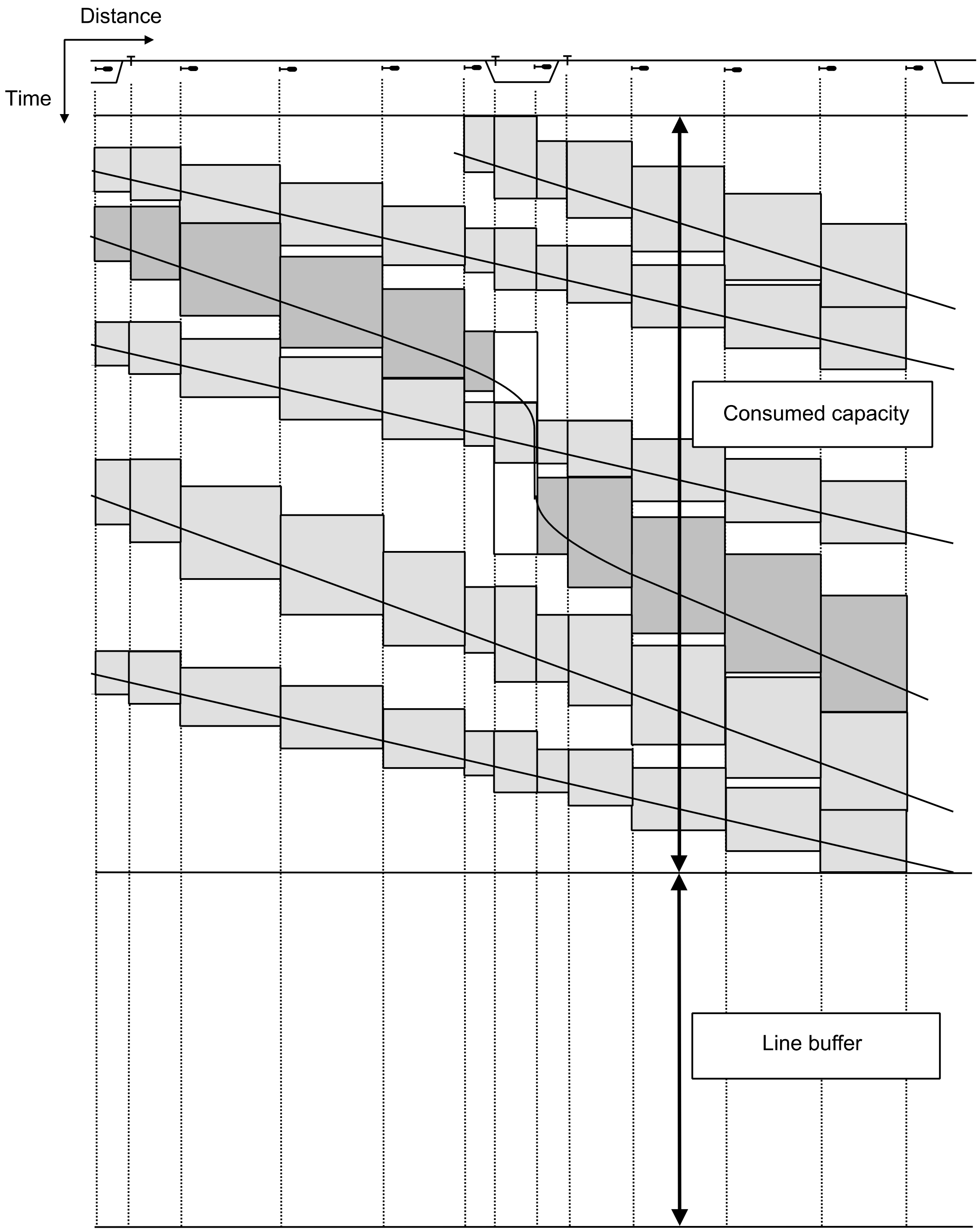


Figure 3.15 Visualisation of the Consumed Capacity for the Timetable of Figure 3.14

This can be directly calculated from the number of trains and the average minimum line headway:

$$\eta = n \cdot t_h / t_p$$

η ratio of consumed capacity
n number of trains
t_h average minimum line headway
t_p total elapsed time of the considered period

Example 3.2
On a double track section between two terminals, it is planned to run the following trains within a period of 4 hours:

Table 3.3 Planned number of trains

Train class	**Number of trains per direction**
High speed	8 trains in 4 hours
Regional passenger	4 trains in 4 hours
Local passenger	8 trains in 4 hours
Freight	9 trains in 4 hours

For these trains, the following minimum line headways $t_{h,ij}$ were calculated:

Table 3.4 Minimum line headways

	2nd train	High speed	Regional	Local	Freight
1st train					
High speed		3.2 min	2.4 min	2.4 min	2.6 min
Regional		6.1 min	3.5 min	3.1 min	4.1 min
Local		8.2 min	6.5 min	3.4 min	6.2 min
Freight		7.9 min	6.3 min	2.8 min	3.3 min

A detailed timetable has not yet been established. Therefore, the frequencies of the train combinations are calculated by a randomly chosen train sequence. As a result, Table 3.5 contains the relative frequencies f_{ij} of all train combinations.

Table 3.5 Relative frequencies

	2nd train	High speed	Regional	Local	Freight
1st train					
High speed		7.6 %	3.8 %	7.6 %	8.6 %
Regional		3.8 %	1.9 %	3.8 %	4.3 %
Local		7.6 %	3.8 %	7.6 %	8.6 %
Freight		8.6 %	4.3 %	8.6 %	9.6 %

From these data, the average minimum line headway can be calculated:

$t_h = 3.2\text{ min} \cdot 0.076 + 2.4\text{ min} \cdot 0.038 + ... + 3.3\text{ min} \cdot 0.096 = \mathbf{4.5\text{ min}}$

This leads directly to the ratio of consumed capacity:

$$\eta = \frac{29 \cdot 4.5\text{ min}}{240\text{ min}} = \mathbf{0.54}$$

According to the quality limits presented later in section 3.4.6, the line is well exploited and problems should not be expected. As shown here for one track of a double track section, the same calculation is required for the other track in the opposing direction.

3.4.5 Estimation of the Consumed Capacity by the Critical Buffer Path

If there is already a given timetable, the consumed capacity can be estimated by searching for the critical buffer path. As can be seen in Figure 3.14, all buffer times are in a critical path that represents the minimum sum of buffer time. From this total line buffer, the consumed capacity can be directly estimated. In complex timetable structures, e.g., on lines with many single track sections, performing the critical path search might become a rather complicated job.

The consumed capacity estimated by that approach may slightly differ from the consumed capacity calculated from the average minimum line headway. Since the average minimum line headway cannot represent all constraints of the timetable structure, the calculated capacity consumption is sometimes a bit lower than the value estimated by analysing the critical buffer path in a given timetable structure. A typical example of such a constraint is the so-called 'three-train inequality problem'. This is a situation in which a shorter train path fits with some tolerance into the gap between two other train paths that already touch each other without any buffer. Figure 3.16 shows an example.

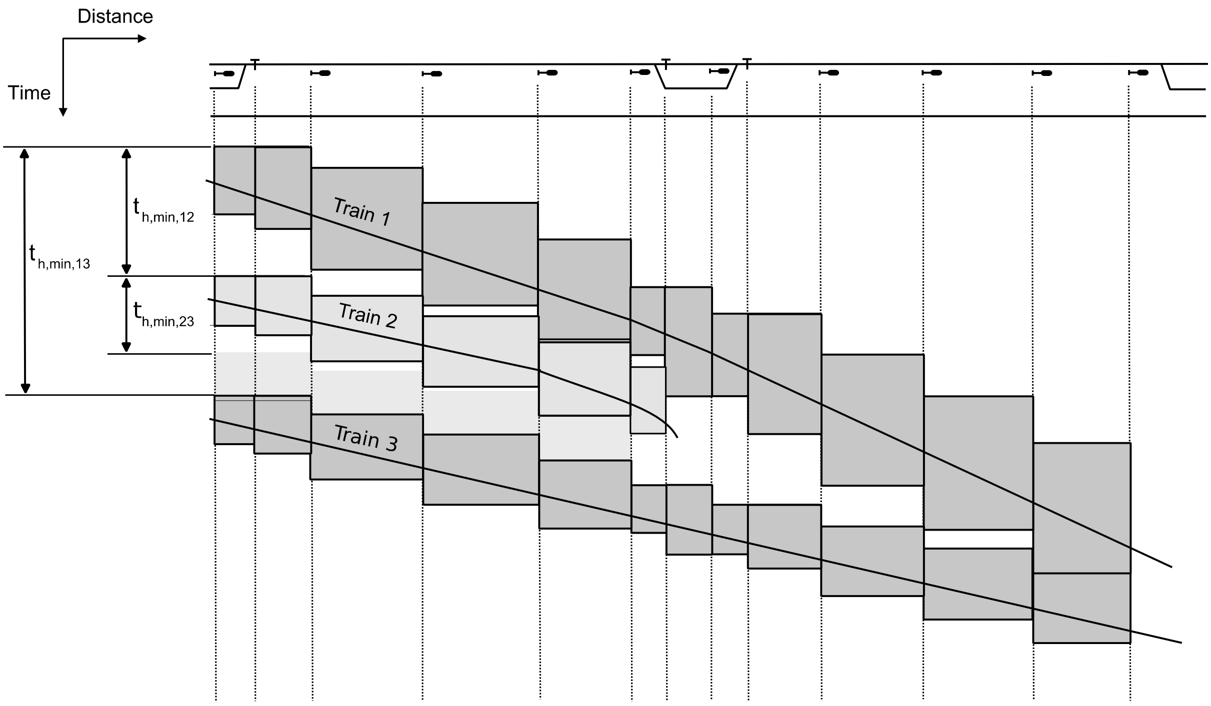


Figure 3.16 The 'three-train inequality problem'

In this example, there are two pairs of trains, train 1 followed by train 2, and train 2 followed by train 3. By calculated compression, the consumed time for these three trains would be the sum of these two minimum headways. However, because the minimum between train 1 and train 3 is greater than the sum of the other two minimum headways, the three train paths cannot be compressed up to this limit.

$$t_{h,min,13} > t_{h,min,12} + t_{h,min,23}$$

$t_{h,min,ij}$ Minimum headway between trains i and j

In a three-train inequality situation, the relevant headway is between two trains that have a third train between them. Since that third train is not in the critical buffer path, the situation will be correctly analysed when searching for the critical path. In contrast to this, when calculating minimum headways just between combinations of two trains, such situations may remain undetected. That is why a calculated compression may sometimes deliver results that are a little bit too optimistic.

3.4.6 Quality Limits

To achieve a reasonable quality of traffic in a double track operation with mixed traffic (as is typical on European railways), the ratio of capacity consumption over 24 hours should not significantly exceed 50 %. That means, the average buffer time between two trains should not be much less than the average minimum line headway. During peak hours, the consumed capacity of the line should not significantly exceed a ratio of 80 %. In single track operations with long distances between the meeting stations, and in double track operations with low speed differences (e.g., mass transit railways), a higher ratio of consumed capacity rate can be accepted.

In single track operations, the required buffer depends significantly on the distance between passing tracks in comparison with the distance between scheduled meeting points. On lines with many passing tracks, where, in case of a delayed train, the transmission of delay to opposing trains can be avoided by switching the scheduled meeting point to another passing track, a lower buffer is required than on lines where suitable passing tracks to change the meeting point do not exist. That means, a line with low reserves in infrastructure requires high reserves in time (buffer times) and vice versa.

In 2004, UIC (International Railway Association) issued a code with more detailed recommendations regarding the limits of the consumed capacity, see Table 3.6 (*UIC, 2004*). However, these limits are mainly based on European operating experience. On railways with very different operating conditions, e.g., in North America, the limits may differ.

Table 3.6 Recommended Limits of the Consumed Capacity in UIC Code 406

Type of line	Peak hour	Daily period	Comment
Dedicated suburban passenger traffic	85%	70%	The possibility to cancel some services leads to a high capacity exploitation
Dedicated high speed line	75%	60%	
Mixed traffic lines	75%	60%	Can be higher when number of trains is low (fewer than 5 per hour) with strong heterogeneity

While the percentage of the consumed capacity allows a good evaluation of the exploitation of a line, the available remaining capacity cannot directly be estimated. There is no fixed relation between the number of trains to be added and the increase of the consumed capacity caused by these trains. This is because the consumed capacity very much depends on the traffic pattern. By adding trains, the heterogeneity of the traffic pattern may change. Through this effect, when adding more trains, the increase of the consumed capacity may be greater or lower than the increase of the traffic flow. Also, the percentage of the consumed capacity cannot be used to compare the capacity of single and double lines. On single lines, a rather high percentage of consumed capacity may be acceptable if, in case of delay, secondary delays are minimised by moving train meets to other stations. On such lines, the infrastructure design has a kind of buffer effect that improves the capacity of a single line but that is not to be seen in the compression procedure.

3.5 Capacity of Route Nodes

Like line capacity, the capacity of route nodes, i.e., the point zones of interlocking areas, can also be investigated either by analytical methods or by simulation. Simulation is more common, but in either case it is not possible to determine one single capacity for a whole route node. For analytical capacity research, the layout of a route node has to be divided into occupation elements, which may be regarded as single channel systems from the viewpoint of queueing theory. Sometimes, these elements are also called sub route nodes. Such an occupation element may consist of several points and crossings but must not contain any parallel routes (Figure 3.17).

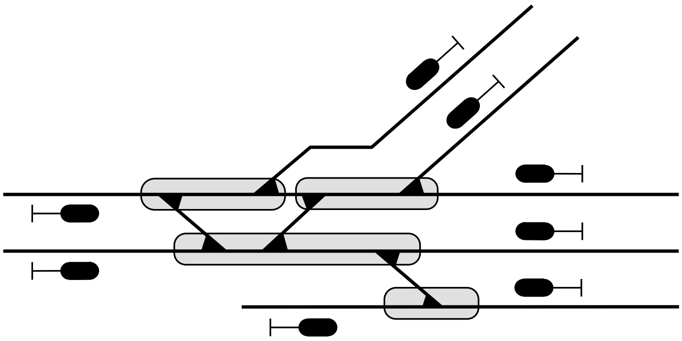


Figure 3.17 Dividing a route node into occupation elements

This means all routes leading through such an element must lock out each other. While it is not possible to set two parallel routes through an occupation element, a route may already be set while parts of a previous route through the same occupation element have not yet been released. As a consequence, the occupation times of points and crossings belonging to the same occupation element may differ. This is not a problem, however. The basic requirement for an occupation element to be treated as a single-channel element in queueing theory is that it must always be possible to calculate a minimum time interval between two trains. This is always the case even with sectional route release within an occupation element.

Since a minimum headway can be calculated between all trains passing through an occupation element, it is also possible to determine an average minimum headway and the consumed capacity for every occupation element. The advantage of this form of research is to get information about the most critical elements in a complex layouts regarding the capacity. But the problem is that the interdependencies between the occupation elements have not been considered yet. When two routes have a route conflict on an occupation element, it could happen that both of the two routes have also a route conflict with a third route that does not touch this element. Figure 3.17 shows an example.

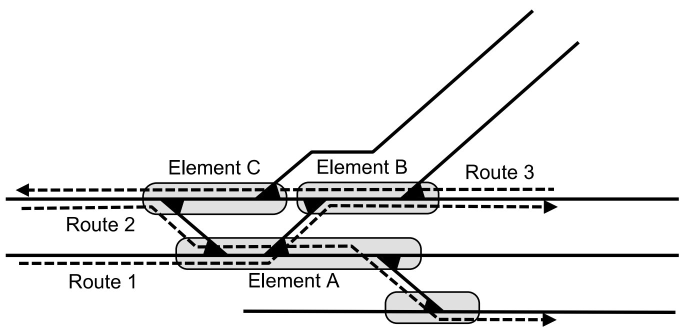


Figure 3.17 Example of interdependencies between three routes

Route 1 conflicts with route 2 on element A. Both routes have also a conflict with route 3 outside of element A. So, when a minimum headway between trains on route 1 and route 2 on element A has to be determined, the influence of a train that runs on route 3 cannot be ignored. There might occur time slots in which both route 1 and route 2 are blocked, although element A is not occupied. That means, a train on route 3 may produce some kind of an indirect occupation on element A. There are computer models that can handle such indirect occupations but only up to a specific degree of complexity (in large layouts, there may be involved much more than three trains).

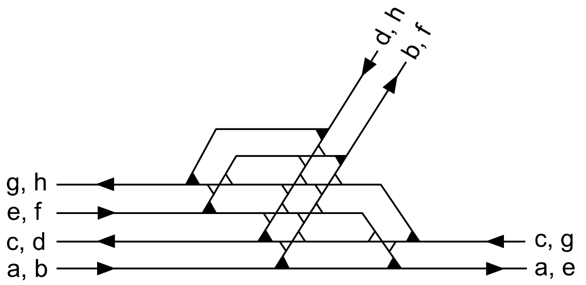
This problem can only be solved by simulation. However, in very complex layouts, choosing a suitable simulation strategy to identify the critical elements of the infrastructure requires a high degree of experience in railway operation and detailed knowledge about the possibilities and limits of the computer models used.

Before starting extensive investigations, it is often very helpful to compare different designs of a complex layout by simplified methods. A typical method is to use route conflict tables. They are also known as route locking tables but must not be confused with control tables as used for the design of interlocking systems. In such a conflict table, all routes are represented by both a row and a column (Figure 3.18). For simplification, different from normal interlocking nomenclature, each route is labelled by a single letter at the entrance and the exit. All table elements representing conflicting routes are marked with an abbreviation to characterise the kind of conflict (crossing, diverging, converging). Besides the conflict types in this example, conflicts may also be caused by opposing routes and overlaps. With help of the route conflict table, the conflict rate can be determined as the number of conflicting route combinations divided by the total number of route combinations.

$$\eta = \Sigma\,(c_{ij}) / n^2$$

η route conflict rate
c_{ij} state of route combination ij (1 = conflict, 0 = no conflict)
n total number of routes

This simple form of a route conflict rate does not consider the number of trains on the different routes. If the traffic flow on the different routes differs significantly, the route conflict rate will not represent the advantages and disadvantages of different layouts very realistically. To achieve better results, it is recommended to weight the route combinations by the number of trains.



	a	b	c	d	e	f	g	h
a	S	D	-	-	C	-	-	-
b	D	S	X	-	X	C	X	-
c	-	X	S	C	X	-	D	-
d	-	-	C	S	X	X	X	D
e	C	X	X	X	S	D	-	-
f	-	C	-	X	D	S	X	-
g	-	X	D	X	-	X	S	C
h	-	-	-	D	-	-	C	S

D - Diverging routes
C - Converging routes
X - Crossing routes
S - Self correlation

Figure 3.18 Route conflict table

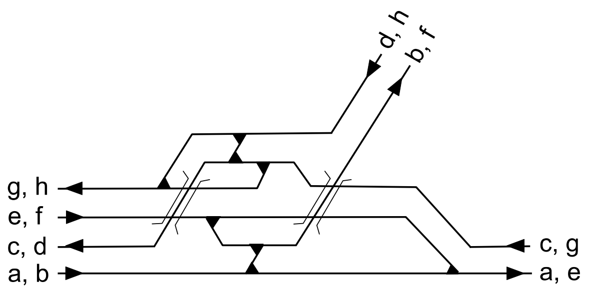
$$\eta_w = \Sigma\,(c_{ij}) \cdot f_{ij}$$

$$f_{ij} = n_i \cdot n_j / n^2$$

η_w weighted route conflict rate
c_{ij} state of route combination ij (1 = conflict, 0 = no conflict)
f_{ij} relative frequency of route combination ij
n_i number of trains on route i
n_j number of trains on route j
n total number of all trains

Of these route conflicts, only conflicts caused by crossing routes or by overlaps are relevant for capacity. On diverging routes, only one train can be approaching at a time. On converging routes, the conflict is not caused in the route node but by the blocking times in the following section. The same is true for opposing routes, where the conflict is caused by the blocking times in the adjacent section used by the opposing trains. So, for diverging, converging, and crossing routes, the minimum headway is always determined by the blocking times in an adjacent section commonly used by the relevant trains outside the route node. So, these route conflicts have actually no impact on capacity. They can also not be eliminated by changing the track layout since they result from the entrance-exit relations of the route node. For crossing routes and overlap conflicts, the involved trains have no common section outside the route node. Only these conflicts can be removed by changing the track layout. By marking the different conflict types in the conflict table, e.g., by letters as in Figure 3.18, the conflicts that can be removed can be easily identified.

By removing all the capacity-relevant route conflicts, one would get an ideal layout that cannot be further improved. In such an ideal layout, all conflicting routes are either diverging, converging, or opposing routes. To evaluate different designs of a track layout, it is very illustrative to compare their route conflict rates with the route conflict rate of an ideal layout with the same route connections. Figure 3.19 shows the ideal layout that corresponds to the layout of Figure 3.18. The following example illustrates the calculation of the route conflict rates for the inferior and the improved layout.



	a	b	c	d	e	f	g	h
a	S	D	-	-	C	-	-	-
b	D	S	-	-	-	C	-	-
c	-	-	S	C	-	-	D	-
d	-	-	C	S	-	-	-	D
e	C	-	-	-	S	D	-	-
f	-	C	-	-	D	S	-	-
g	-	-	D	-	-	-	S	C
h	-	-	-	D	-	-	C	S

D - Diverging routes X - Crossing routes
C - Converging routes S - Self correlation

Figure 3.19 Improved (ideal) layout for the route node of Figure 3.18

Example 3.3

For the track layouts of Figure 3.18 and 3.19, the conflict rates without considering the number of trains on the different routes can directly be calculated from the route conflict tables:

Inferior layout: η = 40/64 = 0.625

Improved layout: η = 24/64 = 0.375

Relative improvement: 1 - (0.375 / 0.625) = 0.4 = **40.0 %**

That means, the improved layout contains 40 % fewer route conflicts than the inferior layout. A further improvement is impossible because the improved layout is the ideal arrangement. To calculate the weighted up conflict rate, the number of trains on the different routes must be considered. In this example, the numbers of trains are given as follows:

Table 3.7 Trains on different routes

Route	Trains per day	Route	Trains per day
a	80	e	30
b	40	f	60
c	80	g	30
d	40	h	60

To calculate the weighted route conflict rates, a calculation sheet is used. The elements of the sheet contain the relative frequencies of the route combinations ($f_{ij} = n_i \cdot n_j / n^2$) that are excluded. The sum of all elements is the weighted route conflict rate.

Table 3.8 Calculation sheet of inferior layout

Route		a	b	c	d	e	f	g	h	Sum
	Trains	80	40	80	40	30	60	30	60	420
a	80	0.036	0.018			0.014				0,068
b	40	0.018	0.009	0.018		0.007	0.014	0.007		0.073
c	80		0.018	0.036	0.018	0.014		0.014		0.100
d	40			0.018	0.009	0.007	0.014	0.007	0.014	0.068
e	30	0.014	0.007	0.014	0.007	0.005	0.010			0.056
f	60		0.014		0.014	0.010	0.020	0.010		0.068
g	30		0.007	0.014	0.007		0.010	0.005	0.010	0.053
h	60				0.014			0.010	0.020	0.044
Sum	420	0.068	0.073	0.100	0.068	0.056	0.068	0.053	0.044	**0.529**

Table 3.9 Calculation sheet of improved layout

Route		a	b	c	d	e	f	g	h	Sum
	Trains	80	40	80	40	30	60	30	60	420
a	80	0.036	0.018			0.014				0,068
b	40	0.018	0.009				0.014			0.041
c	80			0.036	0.018			0.014		0.068
d	40			0.018	0.009				0.014	0.041
e	30	0.014				0.005	0.010			0.029
f	60		0.014			0.010	0.020			0.044
g	30			0.014				0.005	0.010	0.029
h	60				0.014			0.010	0.020	0.044
Sum	420	0.068	0.041	0.068	0.041	0.029	0.044	0.029	0.044	**0.364**

Relative improvement: 1 - (0.364 / 0.529) = 0.312 = 31.2 %

In this example, the improvement of the weighted route conflict rate is significantly less than the improvement calculated by the simplified method without considering the number of trains. The reason is that, in this simplified calculation, the influence of route conflicts that have only a low number of trains is over-estimated. In the inferior layout of this example, this is true for the crossing routes b–g and e–d. The calculation of the simple route conflict rate will only deliver reasonable results if the number of trains on the different routes does not differ much.

The conflict rate can be used to calculate the relative improvement of a changed layout design. Because the conflict rate is calculated from all conflicts including the conflicts that are not capacity-relevant, there is no direct relationship between the the conflict rate and the performance of a layout. This is the reason, why this method only allows comparing different designs of layouts that are based on the same route arrangement, i.e., the number of routes and the entrance-exit relations of the individual routes must be the same. The conflict rate is not of any value to compare layouts with a different route arrangement.

As a general rule, when increasing the size of a layout, i.e., increasing the number of routes, the conflict rate will tend to decrease. This is caused by the effect that with increasing the number of routes, the square of the number of routes in the denominator of the equation will grow faster than the number of route conflicts in the numerator. A practical measurement to compare layouts with a different route arrangement is the average number of routes locked out by a single route. This measurement can by calculated as the product of the conflict rate and the total number of routes.

$$n_{locked} = \eta \cdot n$$

n_{locked} average number of routes locked out by a single route
η route conflict rate
n total number of routes

As another alternative approach, a conflict table may be established that only contains capacity-relevant conflicts. Then, for an ideal layout, the conflict rate will be zero.

Another method of comparing different versions of a route node is the diagram of non-conflicting routes. It had been introduced by Potthoff in the 1960s (*Potthoff, 1979; Potthoff, 1980*). In such a diagram, every route is represented by a node. The nodes are usually arranged in the shape of a circle. The nodes of routes that can be set at the same time are connected by edges. By doing so, a geometric figure appears, which describes all route conflicts of a route node completely. Figure 3.20 shows such diagrams for the same layouts as in the Figures 3.18 and 3.19.

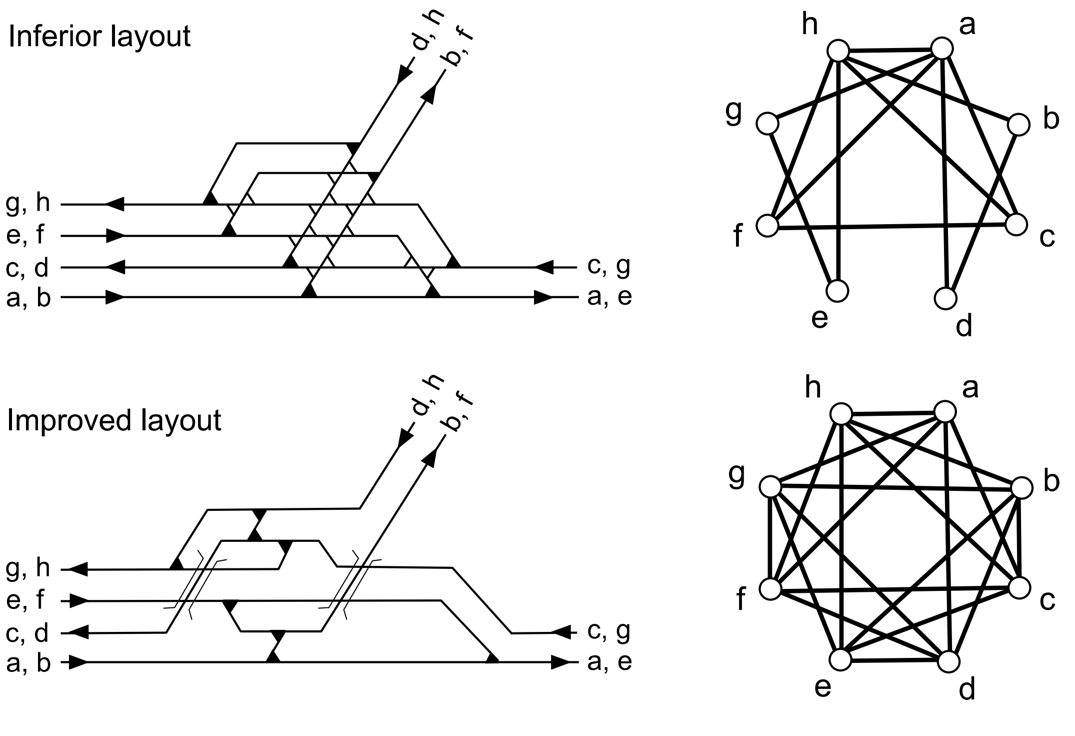


Figure 3.20 Diagram of non-conflicting routes

In contrast to the route conflict table, which only shows combinations of exactly two routes, the diagram of non-conflicting routes also shows all possible combinations of more than two routes. In larger layouts containing many routes, this kind of diagram tends to become very complex and might be a bit confusing. To get a full view of all possible states of the layout, another form of diagram is very helpful. In this diagram all possible combinations of non-conflicting routes are shown in the shape of a tree (Figure 3.21). In this tree, every element represents a possible combination of non-conflicting routes (*Corazza & Musso, 1990*).

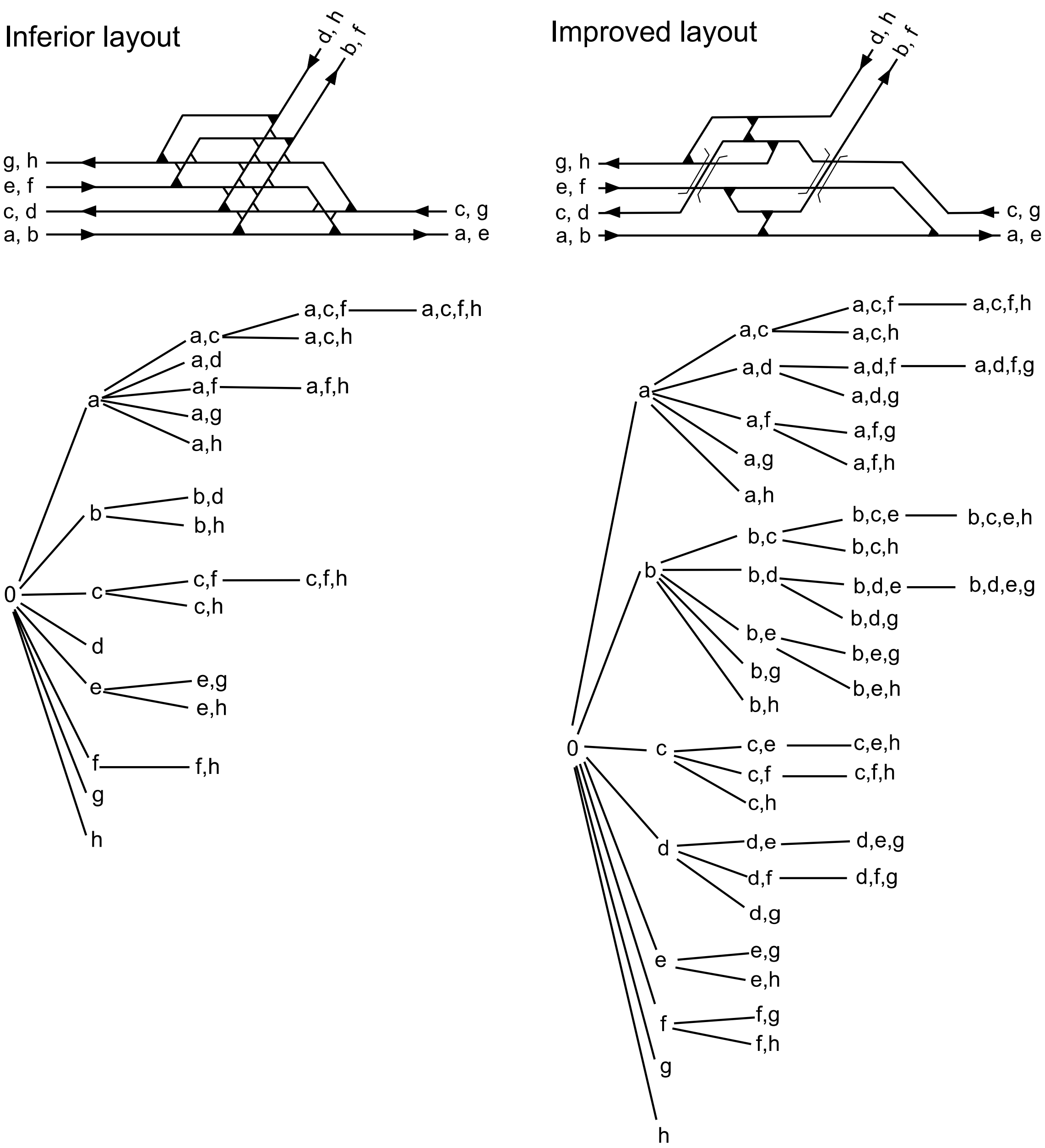


Figure 3.21 Tree diagram of possible states of a route node

3.6 Capacity of Track Groups

The capacity of a track group depends on the number of tracks and the dwell times of trains that are served in the group. On the other hand, the required number of tracks depends on the inbound and outbound traffic flow that the track group has to handle. On average, the inbound traffic flow should equal the outbound traffic flow. The minimum number of tracks to handle the traffic flow may be estimated by the following equation:

$$n_{tracks,min} = t_{dwell,average} \cdot (n_{trains}/t)_{average}$$

$n_{tracks,min}$ minimum number of tracks
$t_{dwell,average}$ average dwell time
$(n_{trains}/t)_{average}$ average traffic flow

To determine the number of tracks that are needed for real operations, not only the average traffic flow but also the variation of the inbound and outbound traffic over time has to be considered. There may be peak hours, in which, for a limited period, the inbound traffic flow exceeds the outbound traffic flow. To avoid congestion, the track group must be able to buffer such peak traffic. A diagram that shows the inbound and outbound traffic of a track group over time is very helpful for this purpose (*Potthoff, 1979; Potthoff 1980*). In such a diagram, not the traffic flow (trains per unit of time) but the absolute amount of inbound and outbound traffic over time is displayed with separate curves (Figure 3.22). The gradient of the curves equals the traffic flow.

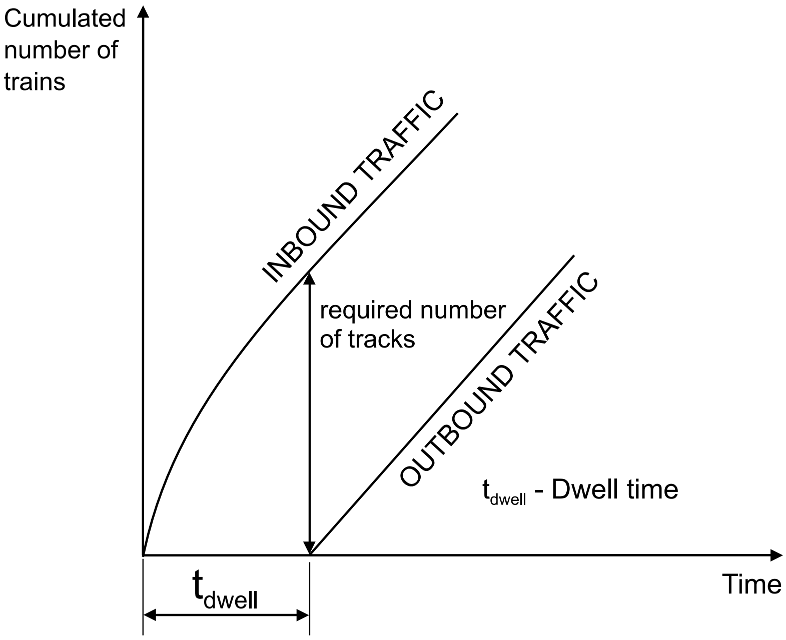


Figure 3.22 Inbound-outbound traffic diagram of a track group

The area limited by these two curves visualises the usage of the tracks. The difference between these curves on the vertical scale equals the number of tracks that are occupied at the same time. Thus, the number of tracks needed to handle the traffic of the peak hours can be determined. The difference between the curves on the time scale equals the dwell time.

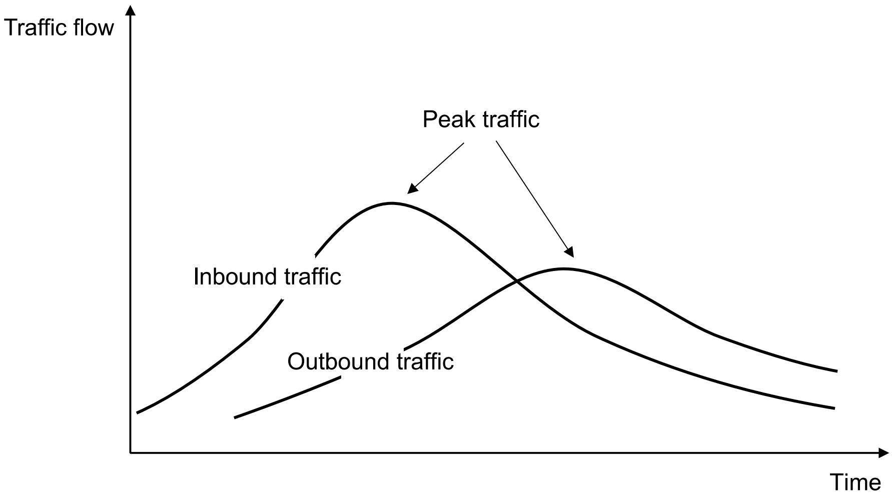


Figure 3.23 ‘Tidal Wave’ of peak traffic flow running through a large freight terminal

An alternative approach is to apply queueing theory to track groups, which is used in some computer models. For this, a track group is modelled as a multiple server system in which the tracks are treated as parallel service stations. By approximation formulas, the waiting probability and the average waiting time in front of the track group can be estimated from the distribution characteristics of the traffic flow and the service process. The key distribution characteristics are the average headway and its coefficient of variation for the traffic flow, and the average dwell time and its coefficient of variation for the service process on the individual tracks. For an introduction to the mathematical background, see the chapter on queueing in (*Hansen & Pachl, 2014*).

Typically, the inbound traffic has a much higher variation than the outbound traffic. Particularly in large terminals, the outbound traffic curve is often of an almost linear shape, which indicates a rather constant traffic flow. The reason is that a large terminal tends to equalise the traffic flow by buffering traffic peaks. The large terminals act as the ‘dampers’ of a railway network. That is why Potthoff compared the peak traffic flow that is running through a large freight terminal with a tidal wave that is partially subsided by the terminal (Figure 3.23).

3.7 Improving Capacity

In case of a lack of capacity of a line or a part of a network, the capacity could be improved in the following different ways:

- Modifying the timetable or the operating procedure
- Removing slow speed sections
- Modifying the signalling arrangement
- Modifying the track layout

At first, one should try to solve capacity problems by means of operation. This could be done by scheduling trains over other lines to reduce the traffic density on lines or parts of the network where sufficient capacity is not available. Instead of reducing the traffic density, the capacity of a section of line could be improved by bundling the train paths in the timetable (also known as 'fleeting'). The idea is to reduce the variation of the minimum line headways to achieve a higher timetable capacity. In double track operations, bundling means to bundle trains of the same speed, and in single track operations to bundle trains of the same direction. However, the direction bundling on single lines hardly works in reality. It only works on short single track sections. If, on a longer single line, two opposing bundles had to meet, a large station would be needed that could take all trains of a bundle to let the opposing bundle pass. In reality, such passing stations hardly exist. So, for a meet of opposing bundles, the trains have to be distributed to several smaller passing stations. This will finally destroy the bundle. On some railways, e.g., in Russia, passing stations on longer single lines often have two passing tracks. That allows the railways to run bundles of two trains. This will improve the capacity but requires a carefully scheduled operation.

Bundling may also lead to increasing waiting times in the adjacent sections or terminals. Assembling bundles of trains may force a train that is ready to depart to wait for the next bundle of its direction or speed class. The waiting time will increase with the length of the section where bundling is in effect. The effects of bundling should be carefully analysed before being implemented on a line. The adjacent terminals must also be able to handle the bundles. Bundling is rather recommended on short sections on the approach to large terminals, where the traffic of different converging lines is put together, and also on connecting lines within extended terminal areas.

Slow speed sections may seriously reduce the capacity by increasing the blocking time of block sections and by producing waiting time for trains that are slowed down.

Increased capacity by modifying the signal arrangement could be achieved by

- Replacing an operation with verbal or written authority by a signal-controlled operation
- Introduction of continuous cab signalling
- Adding block signals to reduce the length of block sections
- Introduction of moving block (i.e., to reduce the block length to zero)
- Optimise the location of interlocking signals and of approach and clearing points of interlocking routes

On a railway still controlled by verbal or written authority, implementation of a signal-controlled operation will not just improve safety but significantly increase capacity. While a signal-controlled operation is often associated with Centralised Traffic Control (CTC), this is not necessarily the case. Already in the late 19th century, European railways introduced signal-controlled operation with locally staffed mechanical control stations and semaphores in which trains were completely governed by signal indication without any written authority.

In publications on North American operations, it is often postulated that the capacity of a line can be significantly improved by the introduction of CTC. That is true, but it should be noted that in the North American philosophy, CTC has two elements: a signal-controlled operation plus centralisation of signal control. Both components of CTC have different effects on the capacity. The increase of capacity does not come from centralisation but from the introduction of signal-controlled operation.

On a line that already has a signal-controlled operation with local operators, the introduction of CTC would not necessarily increase the capacity. Sometimes, the capacity may even deteriorate. On a line with locally controlled interlockings, a lot of signal control can be done simultaneously by the local operators, so the dispatcher can concentrate on traffic regulation. On a CTC line, all the signal control has to be done by the dispatcher. That does not mean that centralisation is not useful. But the main positive effect of centralisation is improved efficiency but not increased capacity. That is the reason why, in a highly centralised operation, a lot of automation technology is needed to reduce the workload of the dispatchers that was caused by eliminating the local operators.

On a line with continuous cab signalling, the approach time is automatically adjusted to the actual braking distance. This will improve capacity if there is a significant share of slow trains with an actual braking distance that is shorter than the braking distance the lineside signalling system is based on. Also, the elimination of the signal watching time contributes to shorter minimum headways. The value of reducing the length of the block sections and of the introduction of moving block is often overrated. Of course, a reduced block length will reduce the running time inside this block section. However, the running time inside a block section is only one part of the blocking time of this section. So, a reduced block length would never reduce the blocking time (and thus the minimum signal headway) by the same ratio. The shorter the block sections, the lower the capacity improvement that could be achieved by a further reduction of the block lengths.

When having estimated the minimum headway for a moving block system (block length = 0), the minimum headway of a fixed block system depending on the number of block sections on a given stretch of line can be calculated by this formula:

$$t_{h,FB} = t_{h,MB} + l_{Str} / (v \cdot n_{Bl})$$

$t_{h,MB}$ minimum headway for moving block
$t_{h,FB}$ minimum headway for fixed block
l_{Str} length of the stretch of line
v train speed
n_{Bl} number of block sections

This equation corresponds to the diagram in Figure 3.24, which shows the situation on a stretch of line of 10 km and a speed of 140 km/h. At that speed, for a service braking deceleration of 0.5 m/s^2 and a train length of 400 m, the minimum headway for moving block is about 1 min. To support the statement of the decreasing influence of reduced block lengths on capacity, the diagram also shows the sensitivity of the fixed block headway depending on the number of block sections on a fixed stretch of line (1st derivative of the equation above) as a measure of the relative improvement that may be achieved by shorter block sections.

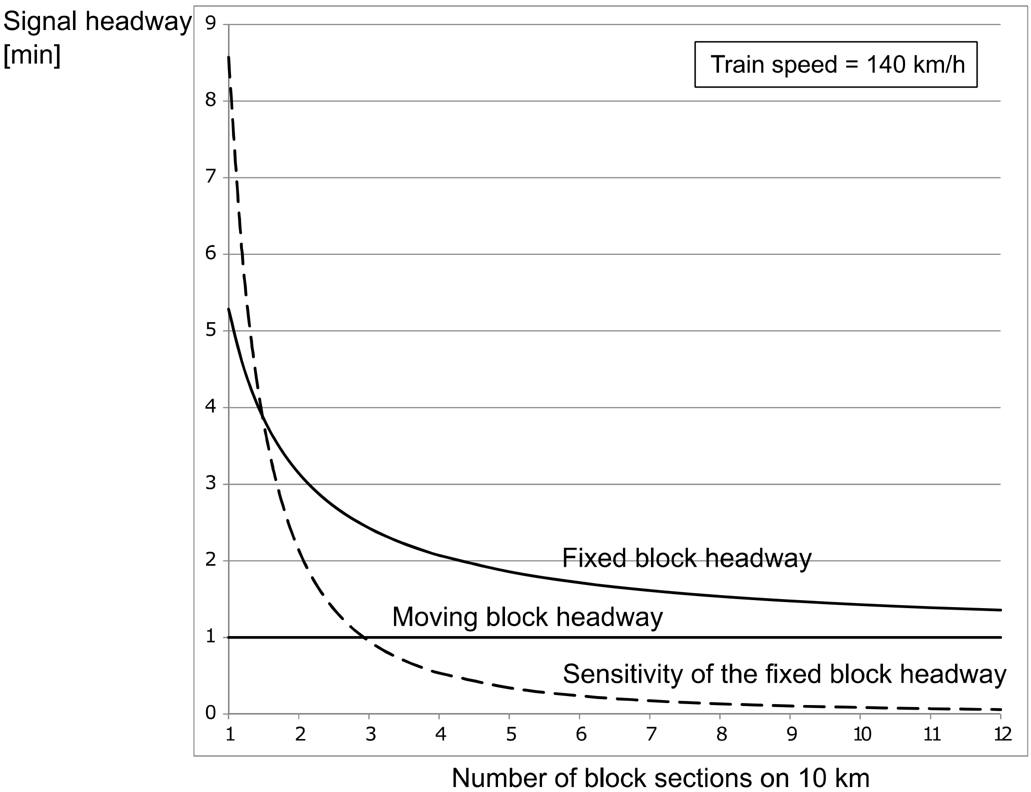


Figure 3.24 Impact of shorter block lengths on the signal headway

In double track operations with mixed traffic, the capacity depends significantly on the speed differences between the different train classes. On most of such lines, the average difference of the running time between stations where fast trains can overtake slow trains is a much greater part of the average minimum line headway than the running time inside the block sections. In typical European double track operations with mixed traffic, even moving block would increase the capacity by no more than about 10 %. In single track operations, as explained above, bundling of trains of the same direction is hardly possible. So, the direction changes after almost every train. For this reason, moving block makes hardly any sense in single track operations. So, reduced block lengths are useful to improve the capacity in double track operations with low speed differences between the trains (e.g., on mass transit railways). The same is true for moving block.

The optimisation of the location of interlocking signals and of approach and clearing points of interlocking routes would increase the capacity of an interlocking by reducing the blocking times of conflicting routes. A typical example is the introduction of a sectional route release to reduce the blocking times of locked points. The capacity of the interlocking would often not increase very much, but the improvement could be done with little modifications in the signalling layout.

Modifying the track layout is the most expensive way to improve the capacity. It is only recommended when other ways to solve capacity problems fail. The capacity could be improved by:

- Adding tracks to a line
- Adding tracks to a terminal
- Removing conflicting routes in an interlocking

Conflicting routes in route nodes could be removed by replacing track crossings at grade by bridges and by adding points for parallel movements (Figure 3.25).

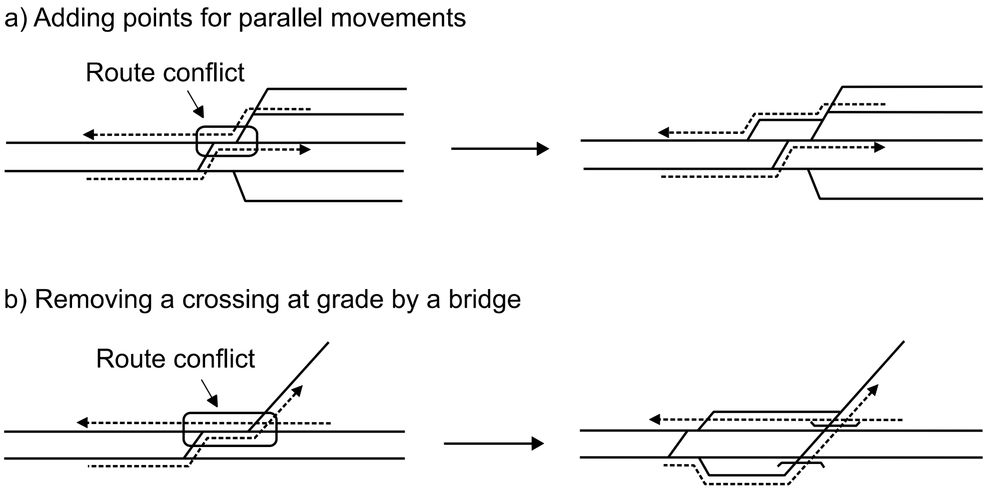


Figure 3.25 Improving the Capacity of a Route Node by Modifying the Track Layout

4 TIMETABLING

4.1 The Role of Timetabling in Traffic Control

Timetabling performs the following functions:

- It coordinates the train paths in the planning process for optimum use of the infrastructure.
- It ensures the predictability of train traffic.
- It produces timetable data for passenger information.
- It is essential for traffic control, locomotive and rolling stock usage and crew scheduling.

By the rule to establish conflict-free train paths, the timetable will always keep the number of trains below the maximum capacity. This way, the timetable is not just a planning tool but has a valuable function for capacity control. Even if the number of trains is above the recommended limit for an acceptable operational quality, it is always guaranteed that the line will never be overloaded with more traffic than it can handle.

In passenger operations, the timetable is essential to provide predictable travel times for passengers. In freight operations, the situation is different. Due to changing demands from the shippers, establishing long-term timetables is often not possible. Some freight operators abolished timetables completely and run their trains entirely on-demand with priority-based train dispatching. Train crews just get a 'timetable' that does not contain any times, but only the data required for safe train driving (e.g. speeds, braking conditions). This is typical in North American operations but also on a number of mining railways in other parts of the world. On such railways, the lack of conflict-free train paths will significantly increase the dispatching effort. Many conflicts that are normally resolved in the timetabling procedure, have to be resolved in a running operation. Since the capacity control function of the timetable does not exist, there is also the risk of putting too much traffic on a line. This requires advanced dispatching systems that can compensate for the lack of pre-established train paths.

European freight operators are also facing changing demands. In the past, this problem was solved by establishing on-demand paths in the yearly timetable. These paths could be used to run scheduled freight trains on a daily basis. Today, computer-based timetabling systems allow the principle of pre-constructed on-demand paths to be replaced by running an increasing number of freight trains as extra trains. However, running 'extra' does not mean running 'unscheduled'. An extra train has to be scheduled as accurately as a regular train, but on a more dynamic basis. On many railways, a freight train operating company can order a train path for an extra train just a few hours in advance. Compared with a completely unscheduled operation, this kind of flexible scheduling will lead to a much higher degree of predictability of traffic. On lines with a mixed traffic of freight and passenger trains, it will also ensure the required quality of service.

On railway networks that are operated on an open access basis, timetabling is not just a planning procedure but also the commercial interface between infrastructure managers and train operating companies. There, the scheduled train path is the product sold by the infrastructure manager to the train operating company. In open access systems, timetabling means coordinating the train paths ordered by competing train operating companies. By assigning a train path, the right is granted to a train operating company to run a train on that path under specified operating conditions.

4.2 Traffic Diagrams

On most railways, traffic diagrams are used both as the basis of all planning of railway traffic and also as essential documents for the control of the running operation. The only exception is North American operations where traffic diagrams (there called 'string graphs') are mainly only used for capacity analysis and in very early stages of operation planning while tabular sheets are preferred in running operations.

The amount of traffic on a line is described in the form of a time-distance diagram that consists of a time axis and a station axis. Every station is represented by a line at the timing point of the station. The train paths are represented by time-distance graphs with a train description inscribed on them. At the intersections of train paths and timing points, the current time is marked by minute numbers (Figure 4.1).

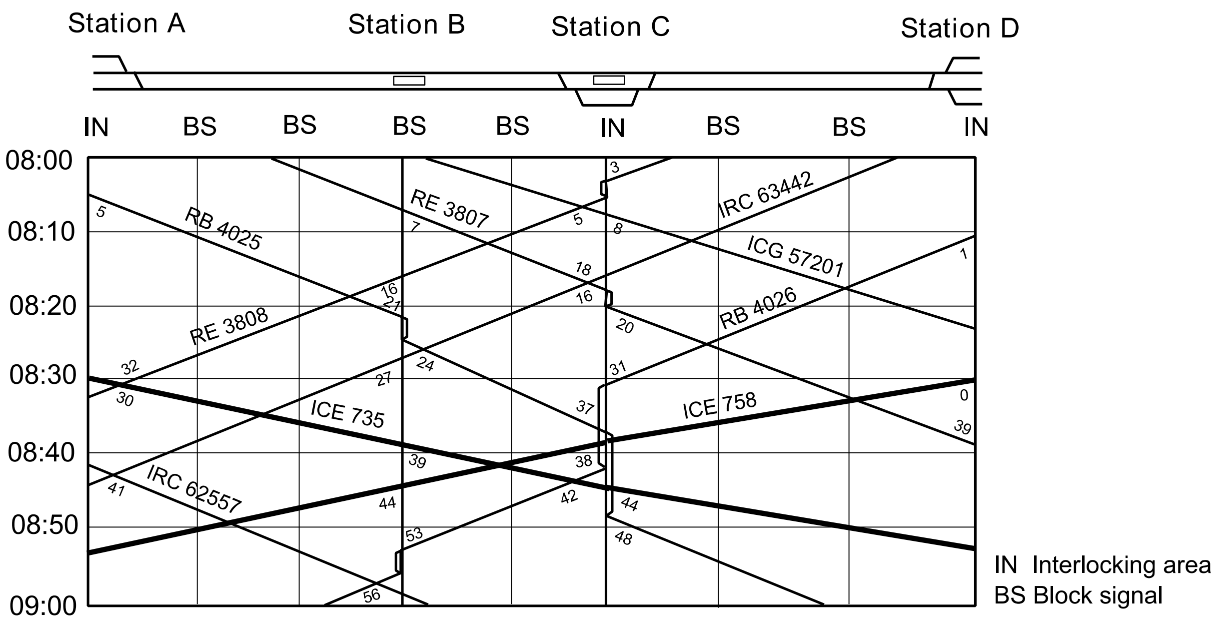


Figure 4.1 Principle of a traffic diagram (double line)

Traffic diagrams may be shown with a horizontal or with a vertical station axis (Figures 4.2). The information contained in both diagrams is of the same value. On individual railways, it is a matter of tradition which one of these two designs is used. Some railways even use both principles, e.g., the Swiss railways, the vertical station axis for timetable documents and the horizontal station axis for online time-distance diagrams in control centres. This is done because the horizontal station axis fits better to the 'mind model' of traffic controllers that work with control interfaces where the infrastructure is displayed in a horizontal way. For all traffic diagrams in this chapter, the horizontal station axis is used.

In addition to the traffic diagram of a line, additional diagrams are used to describe the traffic inside stations or interlocking areas where different tracks are available per direction. In such a station traffic diagram, each track is represented by a line parallel to the time axis.

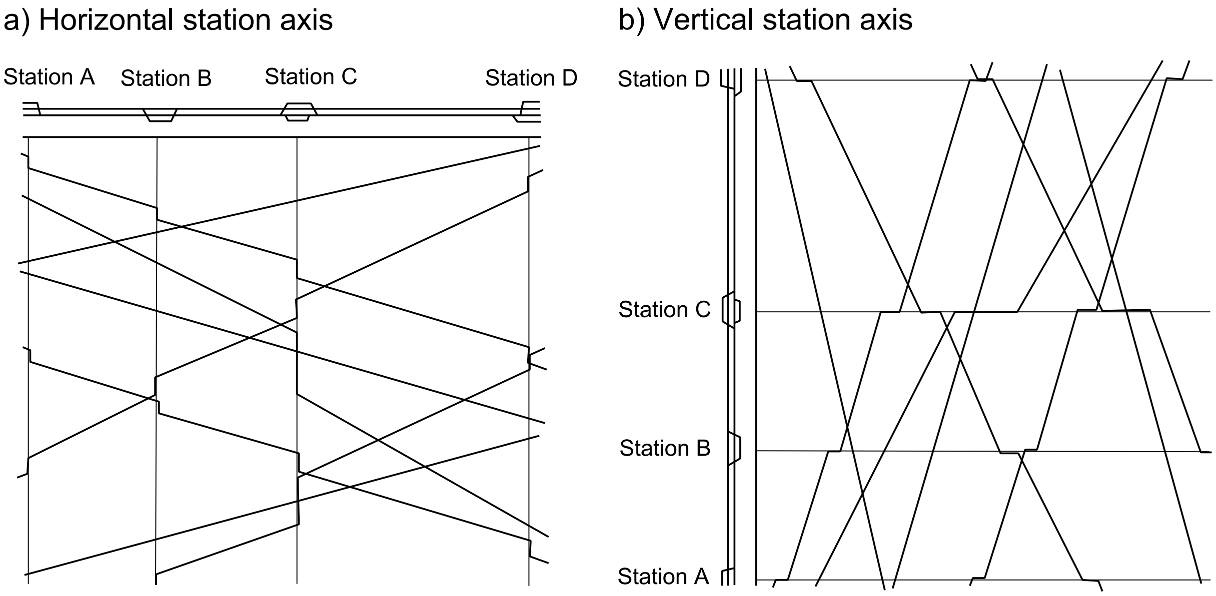


Figure 4.2 Traffic diagrams with horizontal and vertical station axis

The track occupancies are marked by stripes which have the description, arrival, and departure times inscribed on them (Figure 4.3). Small bent pieces at the end of the stripes are suggested time-distance graphs to show the direction of movement. On small layouts, station timetables are often used instead of those more complex station traffic diagrams. On some railways, the occupation of station tracks is integrated in the line traffic diagrams (Figure 4.4). For smaller intermediate stations, this is sufficient. For larger terminals, a separate station traffic diagram leads to a better overview.

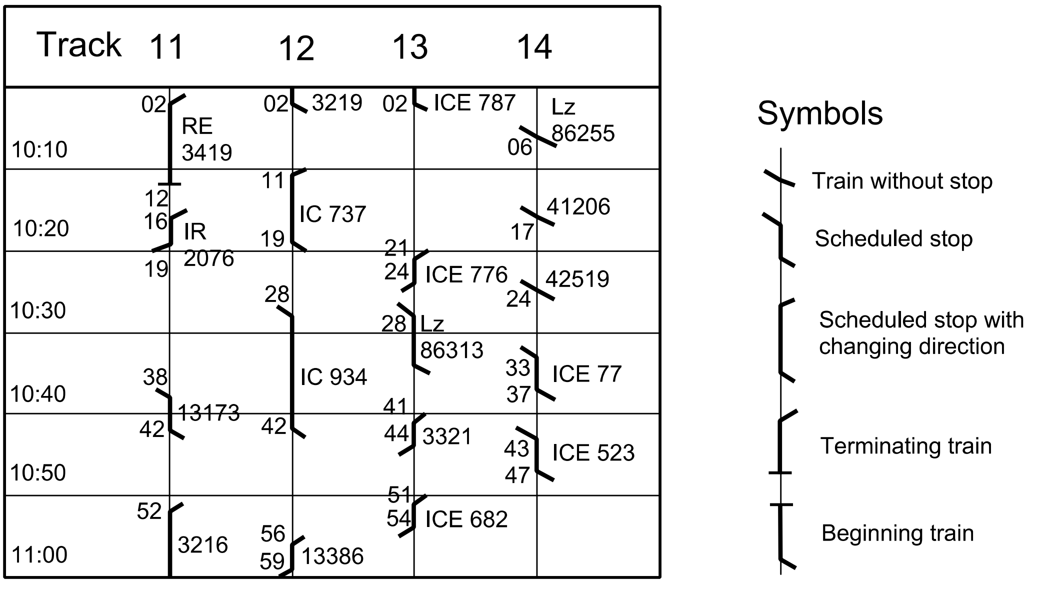


Figure 4.3 Station traffic diagram

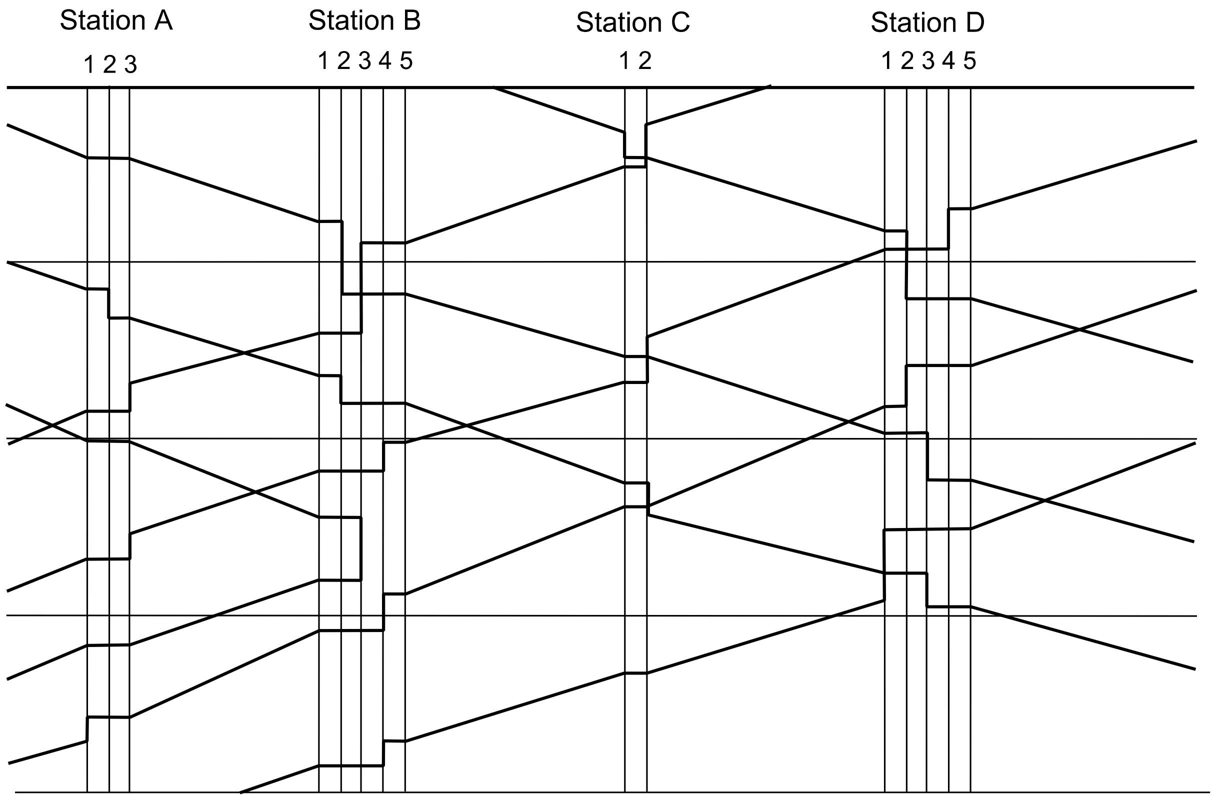


Figure 4.4 Combined traffic diagram that shows the occupation of line and station tracks

4.3 Scheduled Running Time

The scheduled running time of a train consists of the following components:

- The pure running time between scheduled stops
- The dwell time at scheduled stops
- Recovery time
- Scheduled waiting time

The pure running time between scheduled stops is the shortest possible running time as a result of a running time calculation (see chapter 2). To enable a train to make up small delays, recovery time must be added to the pure running time. There are two kinds of recovery time:

- Regular recovery time
- Special recovery time

The regular recovery time is added to every train path as a percentage of the pure running time. The typical time supplement is 3–7% on European railways and 6–8% for North American passenger trains. On some railways, the regular recovery time is evenly spread over the train path while other railways prefer to concentrate the recovery time at the end of the run and at large intermediate terminals (Figure 4.5). Sometimes, at large terminals, the recovery time is not added to the running time in the section in approach to the terminal but to the dwell time in the terminal.

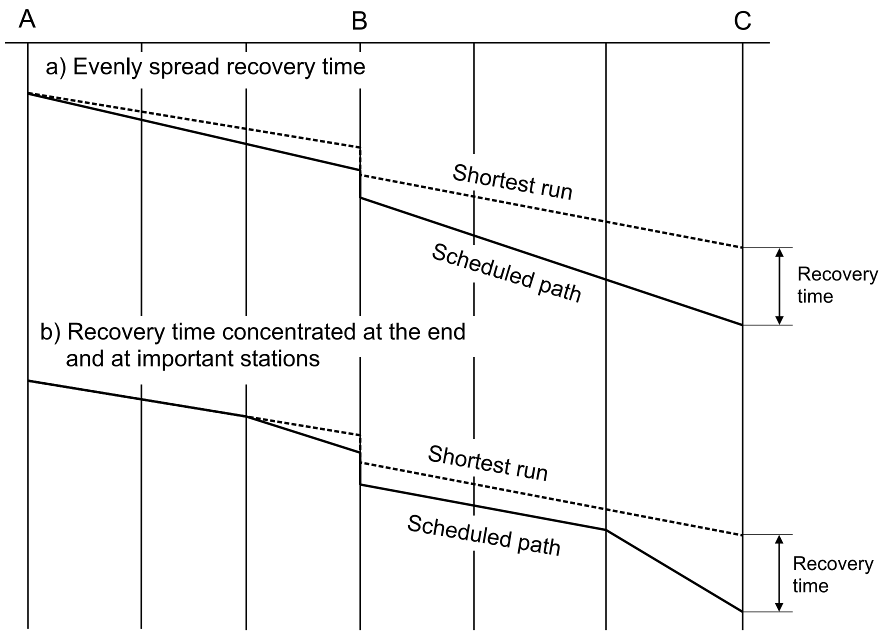


Figure 4.5 Principles of Adding Recovery Time to a Timetable

The special recovery time is used to compensate for the influence of maintenance and construction works and of sections with temporarily bad track conditions. But different from the regular recovery time, it is not added as a percentage of the running time but as a fixed supplement to the running time of the concerned section.

Scheduled waiting time is added for scheduling reasons, e.g., to synchronise timetables of different passenger lines at connecting stations, to synchronise departure times of a cyclic timetable, and to wait for a scheduled passing or overtaking. Scheduled waiting times are mostly added to the dwell time of scheduled stops but sometimes also to the running time.

4.4 Headways and Buffer Times

Headways can be considered in two different ways (Figure 4.6):

- Headways assigned to the stations that limit a section
- Headways assigned to the section between two stations

While assigning headway to a section is the standard practice in capacity research, assigning headways to stations is still more common in timetabling. By this principle, line headways can be divided into four types:

- The headway between two trains that depart onto the same line ('depart-depart' headway)
- The headway between two trains that arrive from the same line ('arrive-arrive' headway)
- The headway between the arrival of a train and the departure of an opposing train towards the same line ('arrive-depart' headway)

- The headway between the departure of a train and the arrival of an opposing train from the same line ('depart-arrive' headway)

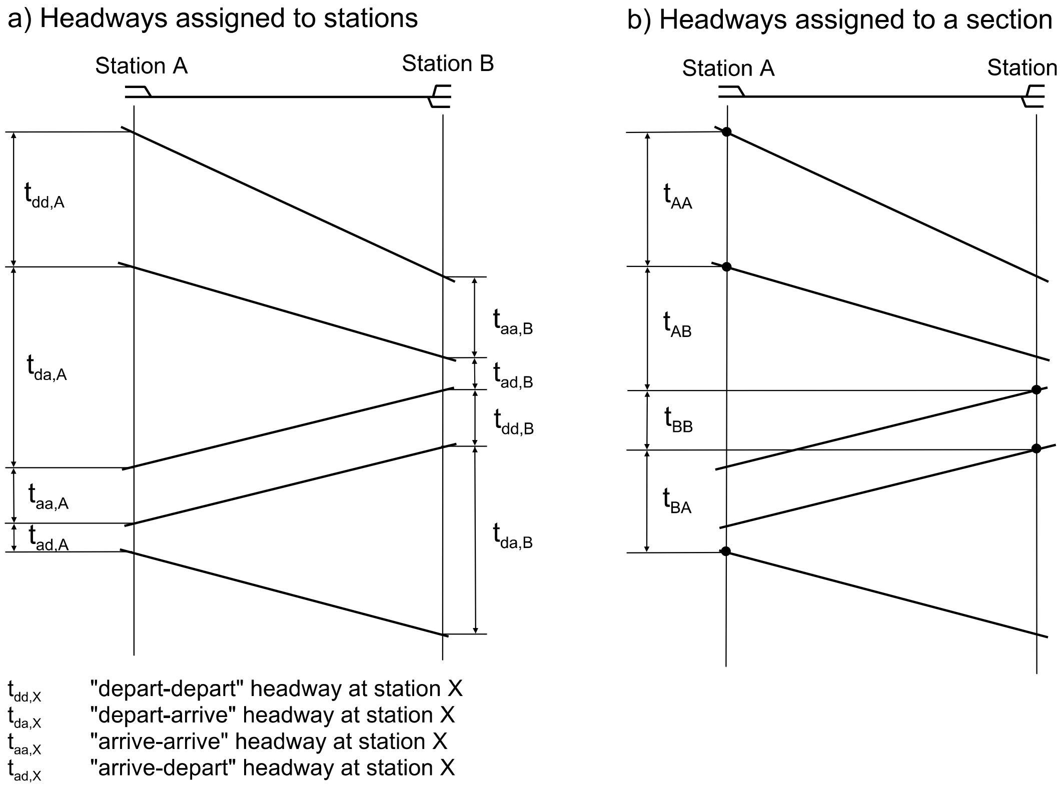


Figure 4.6 Principles of assigning headways

The departure and arrival times refer to the timing points of the stations. Assigning headways to stations leads to different headways at both sides of a section.

The scheduled headway between two trains must consist of the minimum line headway plus the required buffer time to compensate for small delays. The buffer time is the smallest slot between the blocking time stairways of two trains (Figure 4.7 and 4.8).

The buffer time must not be confused with the recovery time. The recovery time enables a train to make up a small delay while the buffer time prevents a small delay from being transmitted to other trains. The recovery time extends the running time of a train while the buffer times reduce the number of trains that can be scheduled. The amount of buffer times depends on the required quality of traffic. Usually, the buffer time is determined depending on the priority of trains. Most railways use the following basic rules:

- Long buffer time when the second train has a higher priority than the first train
- Short buffer time when the first train has a higher priority than the second train
- Moderate buffer time when both trains have the same priority

An exact allocation of a buffer time to each train path in accordance with these rules makes sense only in an accurately scheduled operation where the train sequence rarely changes. On lines where trains often run out of their schedules or with a lot of extra trains, the traffic may be scheduled without assigning buffer times to individual trains. This is typical for freight lines and for connecting lines within large terminal areas. On such lines, the buffer that is required for a good quality of operation could be better provided by the rule that a certain number of trains must be followed by a clear buffer path (Figure 4.9).

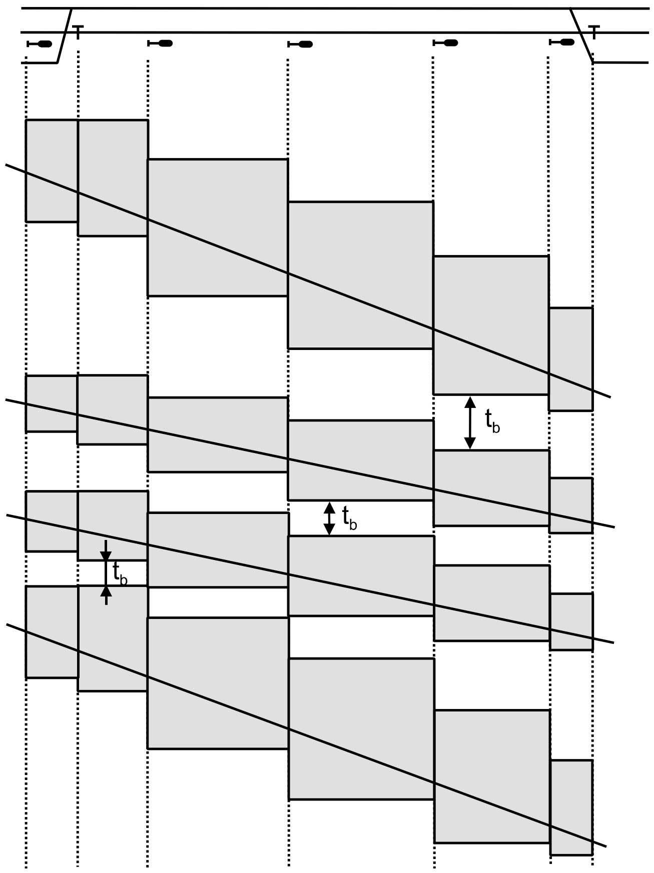


Figure 4.7 Buffer times on a double line

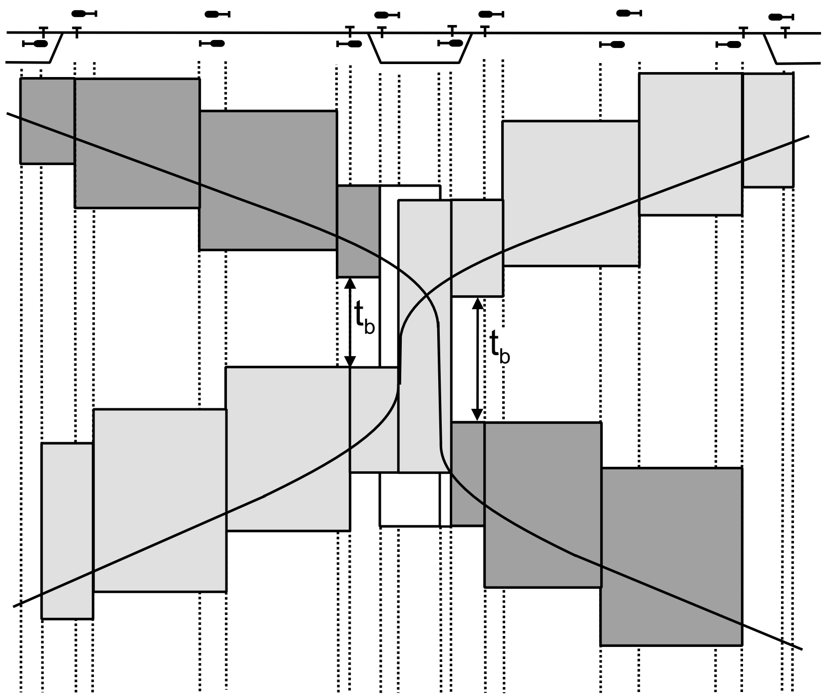


Figure 4.8 Buffer Times on a single line

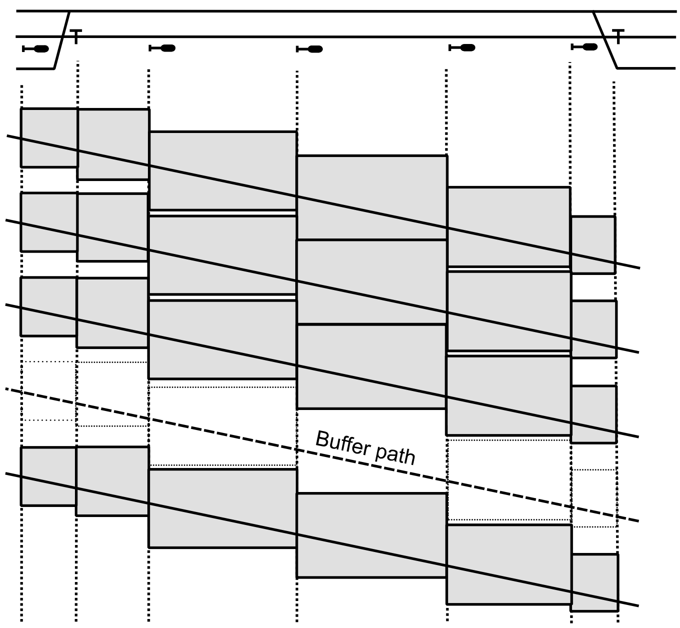


Figure 4.9 Buffer path

Besides the train path buffer times, buffer times are also required at connecting stations where passengers can change between trains and where crew or equipment change from one train to another. To avoid the transmission of delay at such points, a buffer time must be added to the time that is required for the changeover (Figure 4.10).

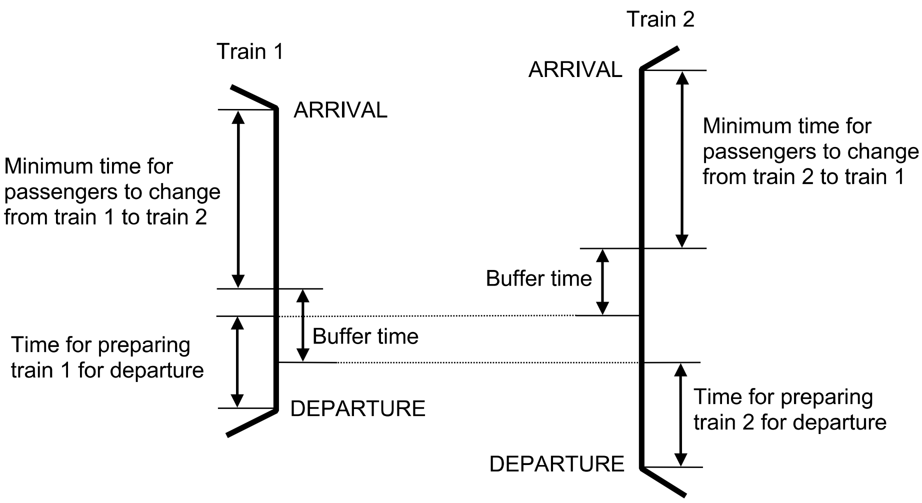


Figure 4.10 Buffer time for changing of passengers at a connecting station

4.5 Cyclic Timetables

A cyclic timetable (also known as a clock face or periodic timetable) is based on a traffic pattern that repeats every hour. Within this pattern, trains of the same route are scheduled at fixed intervals. Cyclic timetables are very common on European passenger lines. On a route, trains will always meet opposing trains at time intervals that equal half of the fixed intervals between trains. On single lines, the scheduled meeting points (passing tracks) must also be located at a distance that meets that time interval (Figure 4.11). When the distance between two meeting points is too short, the running time or the dwell time has to be extended up to this amount. This is a major constraint for the design of a cyclic timetable in single track operations. Due to the mathematics of the situation, the number of meeting points required for a timetable with time intervals can easily be determined by the following equation:

$$n_{MP} \geq 2 \cdot t_{tr} / t_{fix} - 1$$

n_{MP} number of meeting points
t_{tr} running time of the line
t_{fix} fixed scheduled time interval between trains

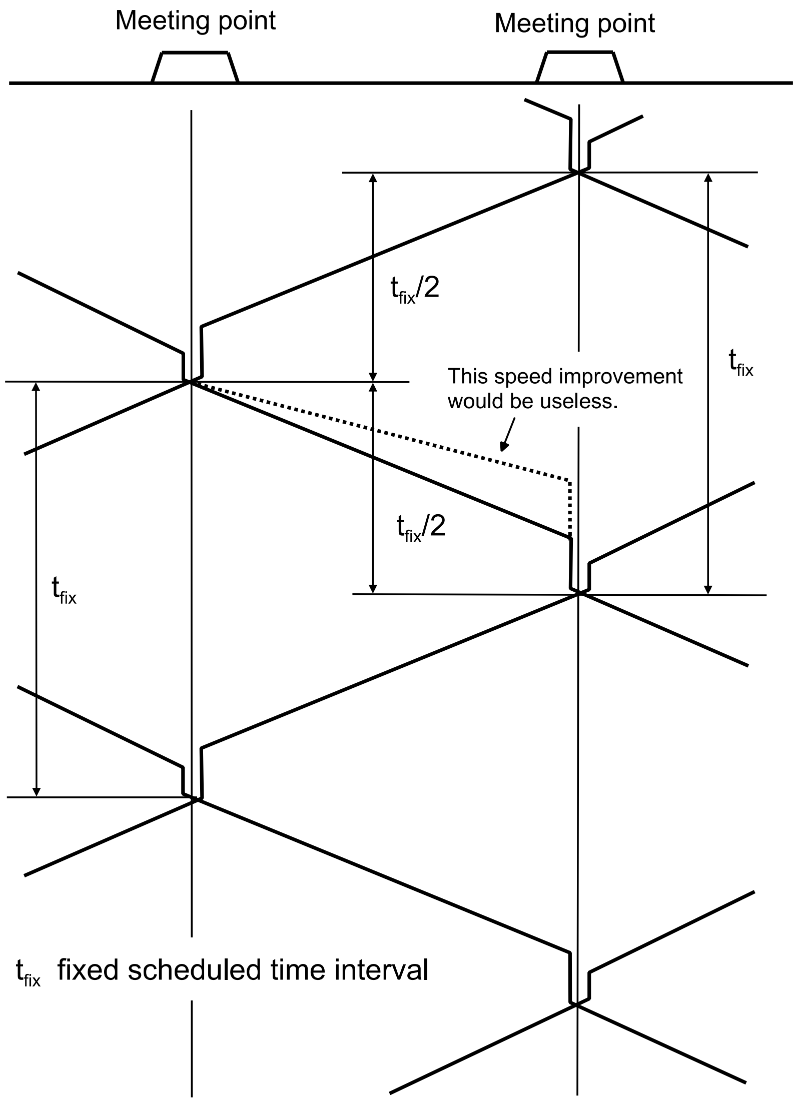


Figure 4.11 Constraint of a cyclic timetable in a single track operation

The number of train sets needed to operate a route on fixed time intervals is calculated as follows:

$$n_T = t_c / t_{fix}$$

n_T number of train sets
t_c cycle time
t_{fix} fixed time interval between trains

The cycle time is the total time between two successive departures of the same train set at the same station in the same direction (Figure 4.12).

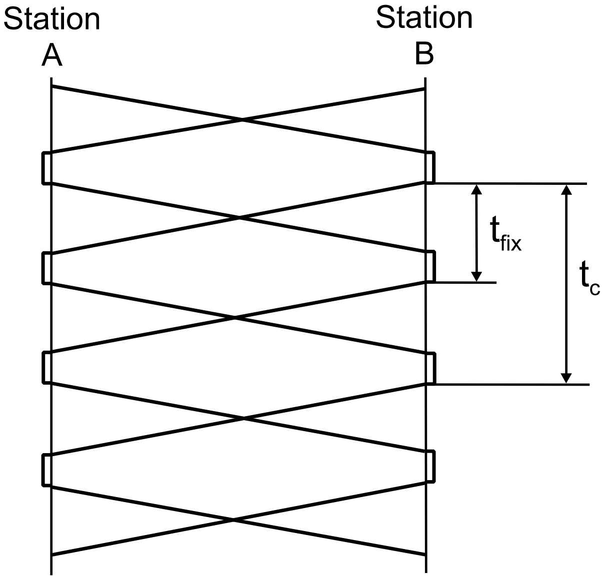


Figure 4.12 Cycle time

Concerning the combination of several train routes, there are three scheduling strategies used in cyclic timetables:

- Non-symmetrical cyclic timetables
- Symmetrical cyclic timetables
- Integrated cyclic timetables

As shown in Figure 4.12, trains from opposing directions meet twice within the fixed scheduled time interval. Thus, for the trains of one route, the timetable is always symmetrical at the meeting time of opposing trains. This time is also called the ‘symmetry time’ of the timetable. If different train routes run over the same line, these train routes may have different symmetry times. In such a case, the timetable is called a non-symmetrical timetable. If all train routes have the same symmetry time, it is a symmetrical timetable.

Some railways with extensive passenger operations connected the cyclic timetables of different routes of the network. In such an integrated cyclic timetable, the different routes are scheduled in a way that, in a connecting terminal, all trains meet always at the same time to enable the passengers to change between all routes (Figure 4.13). That means, in these terminals, the trains of connected routes must have the same symmetry time. For this, the travel time between such terminals must always be an integer multiple of the half fixed time interval between trains. Also, if several terminals are located in a loop, the travel time around the loop must be an integer multiple of the fixed time interval. Otherwise, it could happen that the symmetry times of routes meeting at a terminal are shifted by half of the fixed time interval. This makes the design of cyclic timetables in complex networks quite a challenge. Because of the many constraints on every route and between the different routes, the operation of a network with an integrated cyclic timetable also requires a very carefully scheduled and controlled traffic.

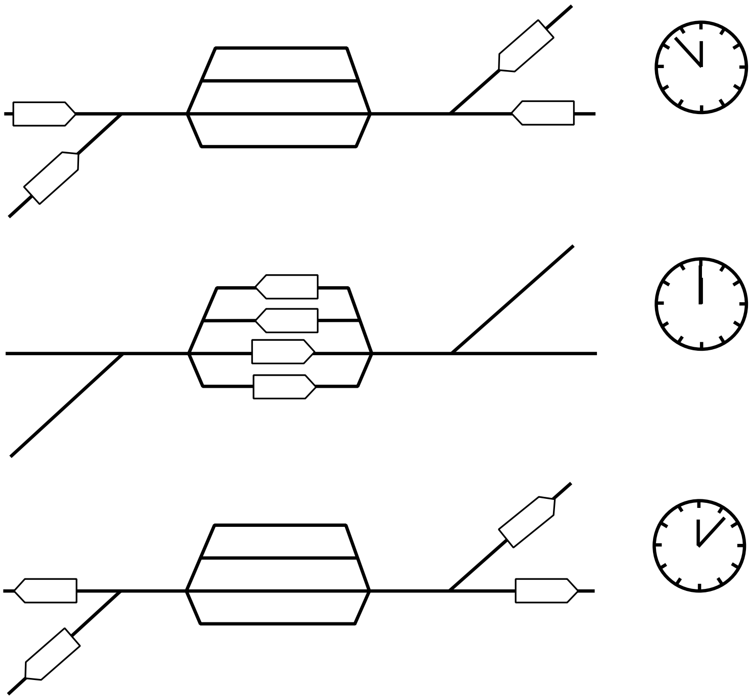


Figure 4.13 Principle of an integrated cyclic timetable

4.6 Timetabling Methods

While timetabling is usually supported by powerful computer tools, manually timetabling is also still in use on some railways. However, even for manual timetabling, computer tools are used for plotting the traffic diagrams.

In manual timetabling, the train path is constructed as a polygon from station to station. The running times are read from running time lists, which consist of the running times between all stations and the extra times required for acceleration and braking. These running times contain the regular recovery time. Special recovery times for maintenance must be added manually. Since the running time calculation is so complex, these tables are usually prepared by a computer. Thus, manual scheduling is not completely conducted without the support of computers (regardless of some railways with only a few different train classes, so that the running time can be determined by ride checks). Since the blocking time model is too complex to be applied manually, simplified models as described in chapter 2 are used, i.e., either by considering the occupation of timetable sections by the running time between timing points and an additional supplementary time, or by adding pre-calculated allowances at the timing points. Furthermore, a buffer time is added, which usually lies within the range of one to three minutes.

In computer-based timetabling, some tools provide a full blocking time modelling of the train paths, while other models use simplified approaches. When using simplified approaches, some scheduling conflicts may not be detected. That is the reason why these systems still require quite a lot of experience and control by the user. Some railways use the simulation systems that were originally developed for capacity research also for timetabling. But with pure simulation software, it is not easy to detect the exact location of train path conflicts where buffer times have to be added. To meet the demands of the railway companies, many suppliers of simulation software have developed timetabling modules as add-ons to their software. These modules can calculate and display blocking time stairways to the simulated train paths in a way very similar to pure timetabling systems. In all systems that use the blocking time model, train path conflicts are accurately detected by the overlapping of the blocking time stairways (Figure 4.14).

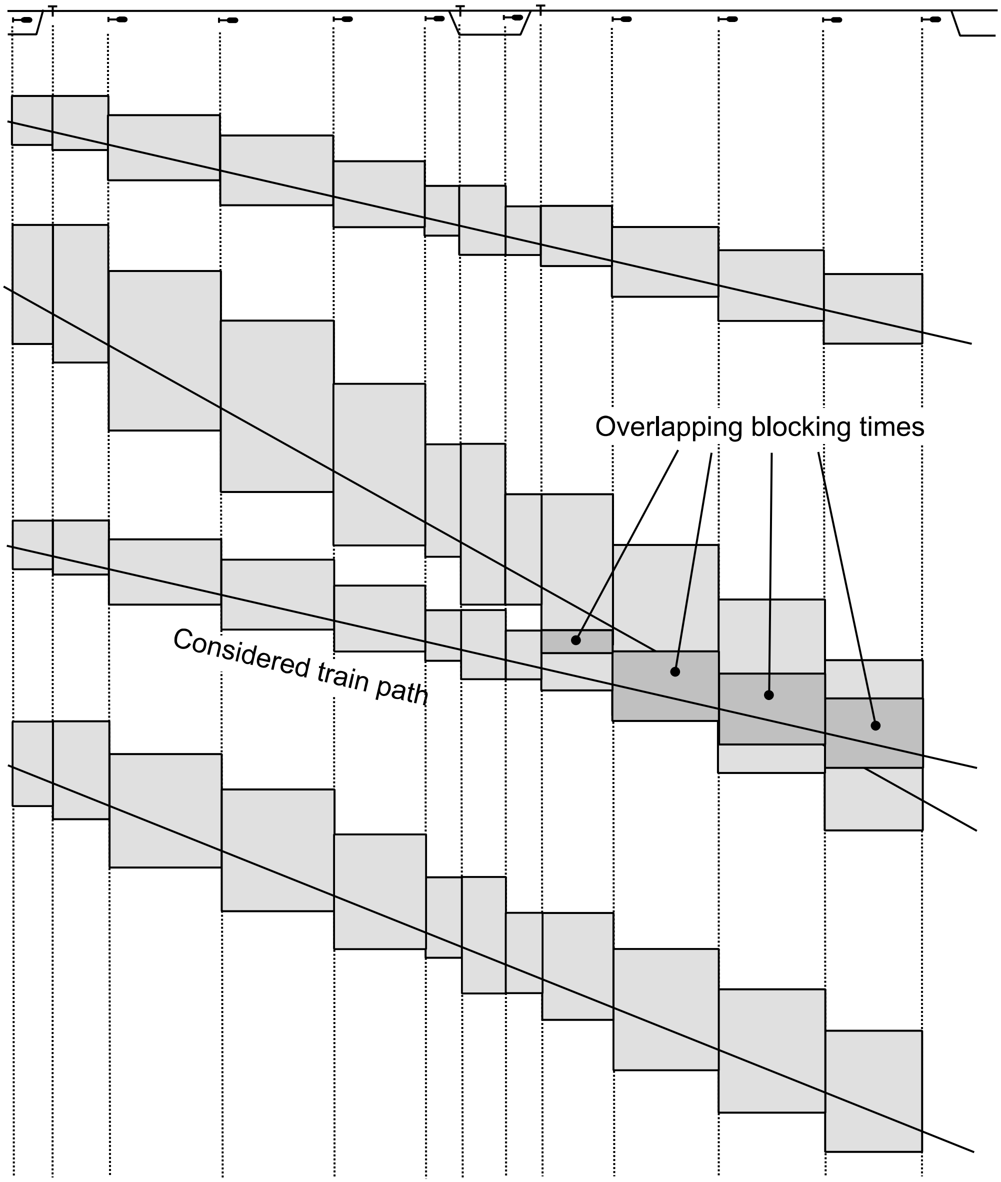


Figure 4.14 Visualisation of a train path conflict

This kind of computer-based tools requires a very sophisticated infrastructure database, which contains all data needed for running time estimation and the calculation of the blocking times. The user can solve train path conflicts by moving train paths to conflict-free positions, by changing the train sequence, or by slowing down fast trains on congested sections. In the near future, there will be advanced systems that assist the user not only by visualisation of conflicts but also by suggestions for solving conflicts based on background knowledge. Some systems can already automatically search a conflict-free path for a train to be inserted into the timetable. However, since they cannot yet consider all constraints, the user is still required to perform final adjustments.

4.7 Timetable Evaluation

Timetable evaluation has several aspects: capacity consumption, feasibility and timetable quality. The capacity consumption of a timetable can be evaluated by the compression method as explained in chapter 2. The remaining two aspects are covered here.

4.7.1 Feasibility

Checking the feasibility of a timetable means proving that the timetable does not contain any unsolved conflicts. So, the timetable will be feasible as long as all trains run with no delay. The strategy of how to detect conflicts in a timetable depends on how the train paths are modelled.

If train paths are modelled by an exact calculation of the blocking times, train path conflicts are always safely detected by overlapping blocking times and completely resolved in the timetabling procedure. So, no further checks are needed to guarantee conflict-free train paths. However, conflicts caused by shunting movements are not automatically detected. If there are just a few shunting moves in a track layout of moderate complexity, it might be sufficient to check that adequate time windows for the shunting moves are available on the relevant tracks. On more complex layouts, further checks might be required.

One method is to perform a single simulation run without initial delays. Besides the train moves, the simulation must also cover all shunting moves. If the output of the simulation does not deliver any delays, the simulation has proved that the timetable does not contain any conflicts with shunting moves.

As an alternative method, all shunting moves may be modelled in the time-distance diagram in the same way as train paths, i.e., by time-distance lines and blocking time stairways. Then, conflicts with shunting moves will be detected by overlapping blocking times. Depending on the timetable rules for an individual railway, the shunting moves may be later deleted from the final timetable documents.

In timetabling systems with simplified train path modelling, train path conflicts on the open line are detected at an acceptable level. The capability to detect train path conflicts in route nodes may be limited, however. The degree of conflict detection depends on the principles used in a particular timetabling software. If undetected conflicts are not obvious for the user, the timetable may be checked by a single simulation run with no initial delays.

4.7.2 Timetable Quality

The timetable quality has two aspects:

- Planning Quality
- Recoverability

A good measure of the timetable quality is the ratio of scheduled waiting times to the total running time. This can be expressed by the travel time quotient as described in chapter 3. For the timetable quality, the travel time quotient should be calculated as the scheduled running time divided by the shortest possible running time without any scheduled waiting times. The acceptable limits of the travel time quotient depend on the kind of traffic. In freight traffic, a much higher share of waiting times is accepted than in passenger traffic. As an example, the following limits are recommended on German railways:

Intercity trains:	1.05
Fast regional trains:	1.07
Normal regional trains:	1.15
Fast freight trains:	1.20
Normal freight trains:	1.40

The recoverability (also called stability or robustness) of a timetable results from the ability to reduce the consequences of delays. The basic requirement for the recoverability of the timetable is that the sum of delays at the output of the system (trains leaving the system and trains terminating in the system) is lower than the sum of delays at the input of the system (trains entering the system and trains beginning in the system). The criterion for the recoverability of the timetable may be described by the following equation (Figure 4.15):

$$t_{d,leave} + t_{d,terminate} \leq t_{d,enter} + t_{d,begin}$$

$t_{d,leave}$ delays of trains leaving the system
$t_{d,terminate}$ delays of trains terminating in the system
$t_{d,enter}$ delays of trains entering the system
$t_{d,begin}$ delays of trains beginning in the system

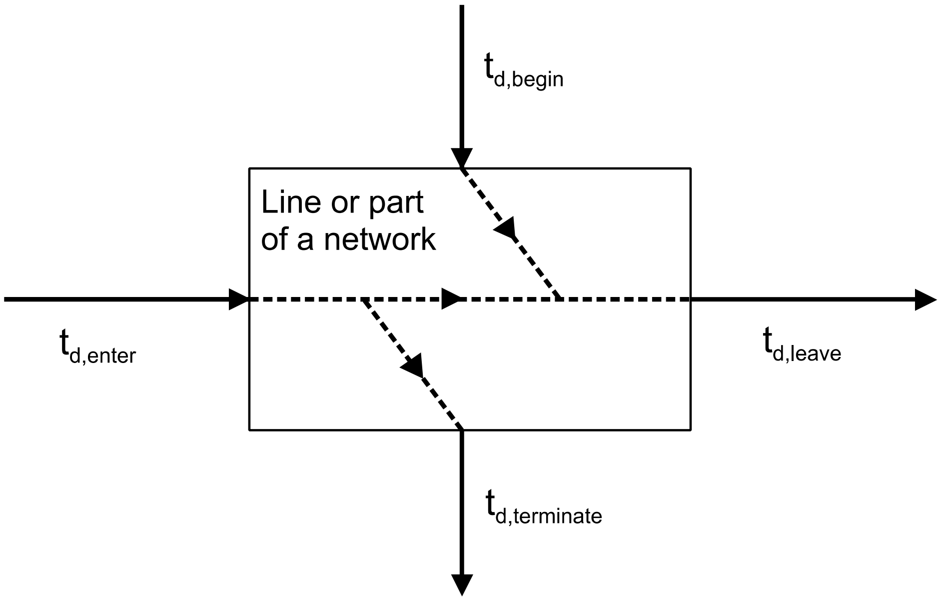


Figure 4.15 Criteria for the recoverability of the timetable

That inequality condition just proves the general ability of the timetable to recover. Besides, for a sufficient evaluation of recoverability, the time to recover from a specific initial delay has to be estimated. Figure 4.16 demonstrates the recovery curve by showing the total delay over time as a consequence of a serious initial delay.

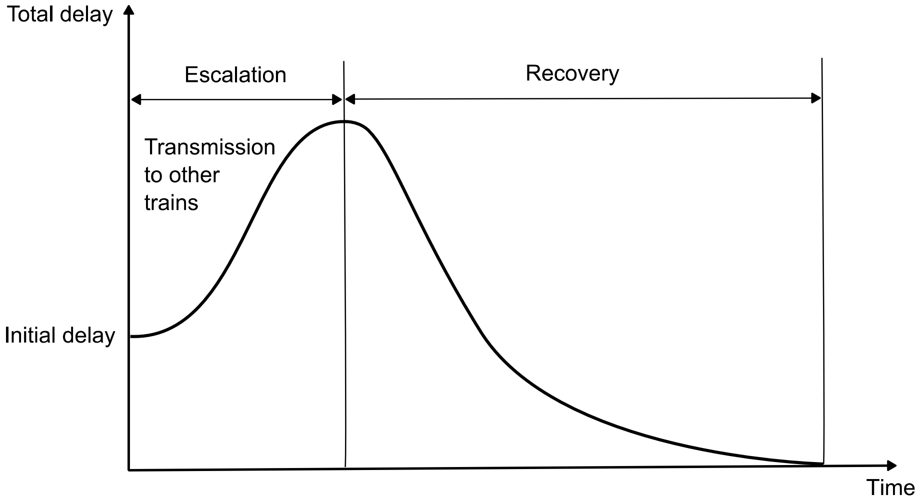


Figure 4.16 Recovery from an initial delay

A serious initial delay means that it significantly exceeds a delay that can be easily compensated by buffer and recovery times. The total delay in this diagram is the average delay per train multiplied by the number of delayed trains. After the initial delay has occurred, the number of trains that are affected by transmitted delays is increasing while the average delay per train is decreasing due to buffer and recovery times. After the initial delay, the total delay will at first sharply increase (escalation phase), then reach a peak, and then decline (recovery phase). The recovery phase takes often much more time than the rather short escalation phase. The recoverability may be evaluated by the total time needed to recover from a specific initial delay.

The recoverability is usually verified by simulation. Under simplified conditions (single track operation with low traffic), it can be verified manually by constructing the transmission of delays that would result from a defined input delay in the traffic diagram (Figure 4.17). As a practical measure to evaluate the recoverability of cyclic timetables, a recoverability quotient can be calculated:

$$q_{recov} = t_{d,enter} / (\Sigma\, t_{b,cycle} + \Sigma\, t_{res,cycle})$$

q_{recov}	recoverability quotient
$t_{d,enter}$	delays of a train entering the section
$\Sigma\, t_{b,cycle}$	sum of buffer time of one cycle between two meeting points
$\Sigma\, t_{res,cycle}$	sum of recovery and other reserve time to reduce delays of one cycle between two meeting points

The recoverability quotient equals the number of cycles that are needed to compensate for the delay completely. This measure has to be determined for each section that is limited by two meeting points. The section with the highest recoverability quotient is the critical section of the line. The German rules recommend a recoverability quotient of less than 2.00 for an initial delay of 10 minutes. That means a line should have completely recovered from an initial delay of 10 minutes within two timetable cycles.

Example 4.1
In Figure 4.17, the first train at station A enters the line with a delay of 10 minutes. At station B, it may make up 1 minute by reducing the dwell time. This has no effect on the opposing train. The delay that is transmitted to the opposing train equals the initial delay reduced by the buffer time of 1 minute. Now,

both trains are running with a delay of 9 minutes. At both ends of the line, these delays are again transmitted to opposing trains, reduced by the buffer times at stations A and C. The train from station C can make up 1 minute in station B by reducing the dwell time. The train from station B cannot make up any delay because it has to wait in station B for the opposing train, which has a greater delay. After the second cycle, all delays are eliminated.

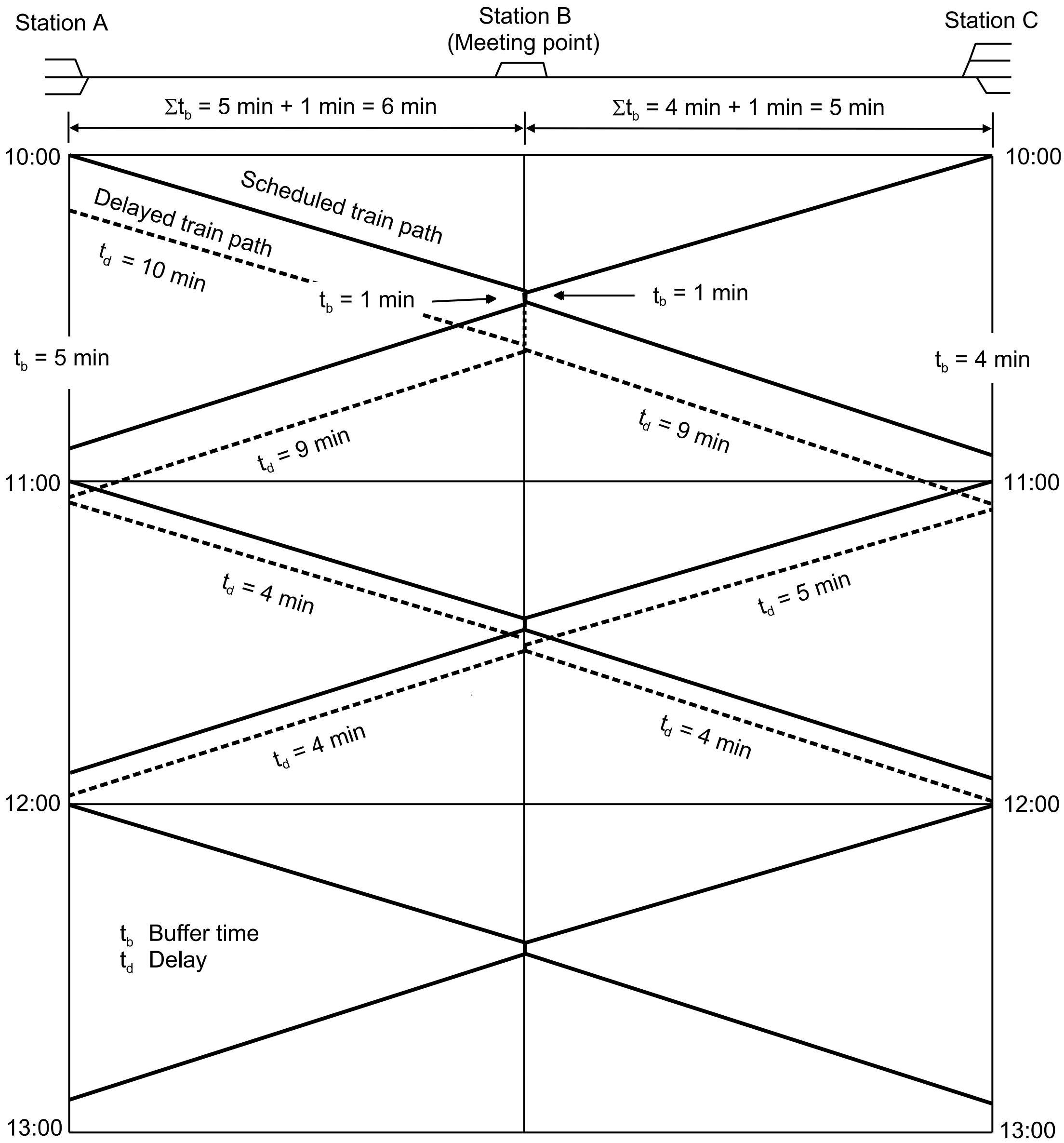


Figure 4.17 Transmission of delays in a single track operation with a cyclic timetable

The recoverability quotients are:

Section A–B: 10 min / (5 min + 1 min) = 1.67
Section B–C: 10 min / (4 min + 1 min) = 2.00

Thus, section B–C is the critical section that has the main influence on the recoverability of this line.

The recoverability of the timetable is not only a question of the train path scheduling. Delays can also be transmitted at connecting stations. For connections between passenger trains, there is often a rule that the connection will get lost when the delay exceeds a specified amount that would cause serious problems in the network. At connecting stations where crew or equipment has to change from one train to another, the situation is much more difficult. When, in case of a great delay no reserve crew or equipment is available, the connection cannot be cancelled. Thus, a carefully planned circulation of crews and equipment is a very essential condition for maintaining acceptable recoverability of the operation in a railway network.

REFERENCES

Chandra, S.; Agarwal, M.M.: Railway Engineering. Oxford University Press 2008

Corazza, G. R.; Musso, A.: Methodologie zur Planung der Gestaltung von Bahnhofsgleisanlagen. Schienen der Welt (1990) 1, p. 11–18

Hansen, I.; Pachl, J. (editors): Railway Timetabling & Operations – Analysis, Modelling, Optimisation, Simulation, Performance Evaluation. Eurailpress, Hamburg 2014

Hertel, G.: Die maximale Verkehrsleistung und die minimale Fahrplanempfindlichkeit auf Eisenbahnstrecken. Eisenbahntechnische Rundschau 41(1992)10, p. 665–671

Pachl, J.: Railway operation and control, 4th Edition, VTD Rail Publishing, Mountlake Terrace 2018

Pachl, J.: Systemtechnik des Schienenverkehrs – Bahnbetrieb planen, steuern und sichern. 11th Edition, Springer Vieweg, Wiesbaden 2022

Potthoff, G.: Die Eisenbahn – Fahren und Leiten. Transpress Verlag Berlin 1979

Potthoff, G.: Verkehrsströmungslehre, Band 1 – Die Zugfolge auf Strecken und in Bahnhöfen. 3rd Edition. Transpress Verlag Berlin 1980

Theeg, G.; Vlasenko, S. (editors): Railway Signalling and Interlocking – International Compendium. Eurailpress, 3rd Edition, Hamburg 2020

UIC Code 406 – Capacity. 1st Edition 2004

Wende, D.: Fahrdynamik des Schienenverkehrs. Teubner Verlag Stuttgart, Leipzig, Wiesbaden 2004

White, T.: Elements of Train Dispatching, Vol. 1. VTD Rail Publishing, Mountlake Terrace 2003

White, T., Krug, A.: Managing Railroad Transportation. VTD Rail Publishing, Mountlake Terrace 2005

GLOSSARY

Analytical capacity research - A method of capacity research that calculates data from infrastructure and timetable characteristics to determine and describe the capacity without simulation.

Approach time - The running time between a signal that provides an approach indication and the following signal.

Asynchronous simulation - A method to simulate railway operations by stochastically generated train paths. The different train classes are generated one after the other (i.e., asynchronous) in accordance with their priority. Train path conflicts are solved by scheduling rules.

Automatic signal - A signal that works automatically by the passage of the train through track sections.

Block section - A section of track in a fixed block system, which a train may only enter if the section is not occupied by other vehicles.

Block signal - A main signal that governs train movements into a block section.

Blocking time - The minimum time interval that must be kept clear for the non-delayed passage of a train through a track section.

Blocking time stairway - A graph displaying the blocking times of all block sections a train passes into a time-over-distance diagram.

Buffer time - An extra time that is added to the minimum line headway to avoid the transmission of small delays.

Cab signalling - A signalling system that displays the movement authorities on the driver's desk.

Capacity research - The usage of methods of railway operations research to determine the capacity of lines, route nodes, terminals, and yards.

Clearing point - A point a train must have cleared completely before a signal in rear may be cleared or a locked route may be released.

Compression method - A method to estimate the consumed capacity by pushing the blocking time stairways of a given timetable together until there are no buffer times left.

Consumed capacity - The sum of headway within a considered period of time.

Controlled signal - A signal that is controlled by an operator.

Control length of a signal - The length of track beyond a signal that must be clear and safe before the signal can be cleared for a train movement.

Cycle time - In a cyclic timetable, the total time between two successive departures of the same train set at the same station in the same direction.

Cyclic timetable - A timetable in which trains that belong to the same route are scheduled with fixed time intervals between their train paths.

Distant signal - A signal that provides an approach indication to a signal ahead but that cannot show a stop aspect. A distant signal does not limit a block section.

Dwell time - The total elapsed time from the time that a train stops in a station until the time it resumes moving.

Exit signal - 1) A controlled signal that governs train movements to leave a station track. It is also called a station exit signal. 2) A controlled signal at a route exit. It is also called a destination signal.

Fixed block system - A block system in which the track behind a train is sectionally cleared in accordance with fixed block sections.

Headway - The time or distance interval calculated from 'head to head' between two successive trains.

Home signal - 1) A signal governing entrance to an interlocking area. 2) A signal governing entrance to a station area.

Interlocking - An arrangement of points and signals interconnected in a way that each movement follows the other in a proper and safe sequence.

Interlocking signal - A controlled signal that governs a route within an interlocking.

Intermediate interlocking signal - An interlocking signal that is neither a home signal nor a signal that governs a route to leave the interlocking area.

Line headway - The headway that results from the blocking time stairways of two successive trains.

Main signal - A signal that governs regular train movements. This term is used by many railways to distinguish these signals from shunting signals.

Main track - A track that may be used for regular train movements.

Moving block system - A block system in which the track behind a train is cleared continuously.

Occupation element - A part of the infrastructure that can only be occupied successively keeping a minimum headway between successive trains.

Open line - Main tracks outside of station areas.

Operator - An employee who is in charge of authorising train and shunting movements.

Overlap - A certain length of track beyond a signal that must be kept clear as long as a train movement is approaching that signal.

Points - The movable parts of a turnout that are moved to set different routes.

Pure running time - The shortest possible running time between scheduled stops as a result from a running time calculation.

Queuing theory - A branch of operations research to model and analyse server systems so that queue lengths and waiting time can be predicted.

Recoverability of the timetable - The ability of a timetable to reduce the consequences of delays.

Recovery time - A time supplement that is added to the pure running time to enable a train to make up small delays.

Route node - The point zone of an interlocking area in which conflicts between different routes may occur.

Scheduled waiting time - The waiting time that is needed for a scheduled passing and overtaking and to synchronise the schedules of a cyclic timetable.

Shunting - All movements other than train movements.

Shunting signal - A signal that is used to authorise shunting movements.

Siding - A track that may not be used for regular train movements.

Signal aspect - The appearance of a lineside signal, as viewed from the direction of an approaching train, or the appearance of a cab signal.

Signal headway - The headway that results from the blocking times of two successive trains in a single block section

Station area - An arrangement of station tracks limited by opposing home signals.

Station track - A main track protected by controlled signals within an interlocking area on which trains may originate, terminate, pass, and turn.

Station traffic diagram - A diagram that displays the scheduled occupation of the tracks in large stations and interlockings.

Timetable capacity - The maximum number of train paths that could be inserted into a timetable.

Timing point - A location listed in the timetable at which the arrival, departure, and passing times apply.

Track clear detection - A device that detects the occupation and clearance of a track section.

Traffic density - The number of trains that are running at the same time on a certain portion of line.

Traffic diagram - A diagram that contains the train paths of all trains that run on a line.

Traffic energy - The traffic flow multiplied by the average transportation speed.

Traffic flow - The number of trains per unit of time.

Train movement - A locomotive or self-propelled vehicle, alone or coupled to one or more vehicles, with authority to occupy a section of line under operating conditions specified in the timetable**.**

Train path - A time-distance graph that represents the schedule of a train in a traffic diagram.

Waiting time diagram - A diagram displaying the waiting time as a function of the traffic flow.