One of a series of papers on technical issues published by the RTW from time to time.

Basic Railway Signalling

by

Piers Connor[1]

Introduction

As any train driver will attest, driving a train is easy. The difficult bit is stopping it. It's easy to get the train going but much more difficult to stop in the right place and, to do this consistently and keep time, requires skill and concentration. The reason for this is simple – the adhesion available for a train with a steel wheel on a steel rail is such that the braking distance is considerably more than that obtained in a car with rubber tyres on the average road. The adhesion between a tyre and the road surface can be measured at over 85%. The UK main line railways calculate their braking distances on the basis of 8% adhesion, an order of magnitude less. This sort of calculation is standard across the world.

Figure 1: Train of Chiltern Railways approaching Princes Risborough station, showing various signals providing route control indications. Note also the fixed speed limit sign for movements over the crossover. Photo by Roger Marks.

To use a comparison between road and rail braking, we can look at someone who wants to stop a car from 70mph. The driver thinks about it (20m) and then brakes (80m), requiring a total of 100m stopping distance. Now transfer this process to the cab of a train running at 70 mph. The driver reaction distance is still 20m but you have to add tome for applying the brake and allowing for it to feed up to the required pressure on the whole train, which needs another 60m. Then the actual distance for the train to stop will be another 950m approx., giving a total of 1030m, or more than a kilometre. This would represent a full service brake application with no allowances for error or a safety margin. An emergency brake application at this speed might get the train to a stand in 700m if it's not raining. This means that a train's braking distance is likely to be an order of magnitude greater than that of a road vehicle and, for a driver in an emergency situation, it seems like a million times greater when you are at the front of a train with the emergency brake fully applied and you know there is nothing you can do to avoid hitting something in front of you.

---

[1] PRC Rail Consulting Ltd.

It is a sad fact of life that very few members of the general public understand that there is any difference between train braking and car braking, which is why a distraught mother once called a train driver in the UK a "murderer" when his train killed her 16-year old son who was trespassing on the railway with his friends. She thought he had not bothered to try and stop. Of course, this wasn't the case but the local press made a big issue of it. In another instance, there were calls by some individuals for the railways to be closed down altogether after the Ufton Nervet level crossing accident on 6 November 2004, "because trains aren't safe if they can't stop for a car at a level crossing". Such uninformed nonsense is widespread in the public mind and the media love it. Railway companies ignore it at their peril. Urban railway administrations in particular, must understand that a strong and carefully targeted public awareness campaign is essential for new railways and tramways if a low accident rate is to be achieved.

Signals

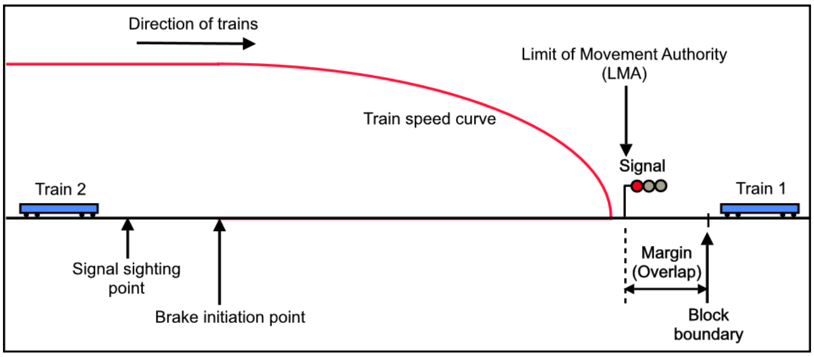

Figure 2: Schematic of train and signal on a manually operated railway. Note that it is necessary to allow a reaction time between the point where the driver sees the signal and his selection of a brake application. Diagram: Author

The long braking distances required by trains present a problem for the driver of a manually controlled train. He is unlikely to be able to see an obstruction or diversion requiring him to reduce the train speed or to stop. In effect, he is effectively driving "blind". To overcome this, he (or she) requires advice in advance of the point where it is necessary to apply the brake. Railways have developed such a system, which we call signalling.

Traditionally, signals are provided at fixed lineside locations to indicate to the driver of an approaching train how he should proceed or, possibly, the limit of his movement authority (LMA). To achieve this, the line is divided into sections, often called blocks, and a signal is placed at the entrance to each block to act as a sort of "gate keeper". The rule is that (normally) only one train is allowed into a block at any one time. Figure 2 above shows the general arrangement for a signal on a manually operated railway.

It is the practice on many railways to provide overrun space beyond the signal in the form of a margin, often referred to as an overlap. This overlap can be a set distance (183m on UK main line railways), calculated on a site-by-site basis, (London Underground) or by separating trains by a whole block length. The choice depends on the administration, traffic requirements and the type of signalling installed. Metro systems normally require signalling to include calculated overlaps or fixed block safety margins. Similar principles are being applied to main line railways as they are modernised.

There is a wide variety of fixed signals used by railways around the world. They may be categorised into two broad groups, "speed signalling" or "route signalling". Speed signalling is the most common, where the indications for the driver represent an instruction to proceed at a speed depending on the type of train in his charge. Most administrations in continental Europe and North America have adopted this system. A driver is shown one or more lights in a specified pattern, which gives him the authority to proceed (or not) at the speed designated to his train. Thus it is common for the same indication to allow a passenger train to proceed at, say, 160km/h, while a freight train is restricted to 60km/h on the same route. Some systems use variable number displays to indicate maximum speed.

Figure 3: German main line railway signal with speed indication number showing maximum speed of 80km/h.

The route signalling system differs in that the speed limits are separate from the signalling. They are usually indicated by discrete and fixed lineside signs. Variations for different types of trains are assumed to be part of the driver's knowledge. Drivers are shown proceed indications for a block or a number of blocks ahead. If a diverging route is set, additional indications are provided and the driver is expected to adjust the train speed to comply with the separate lineside signs. This system is normal UK practice and has been adopted in many overseas countries where UK railway systems have been installed.

Some administrations use a combination of the two systems, the French railway SNCF, for example. Regardless of the system adopted, with manual driving, signal sighting will dictate the braking distances and therefore the block lengths. As train speeds increase, they require longer braking distances and therefore longer sighting distances or earlier warnings of conditions ahead. Very early on in the development of railways, routes were equipped with intermediate updates of signal information.

Signalling Developments

The first development was to provide advance warning of a signal's position or aspect[2] by the use of a "distant" signal. It provided an indication showing the condition of the next signal ahead, i.e. whether this signal showed "stop" or "proceed" (Figure 4 below).

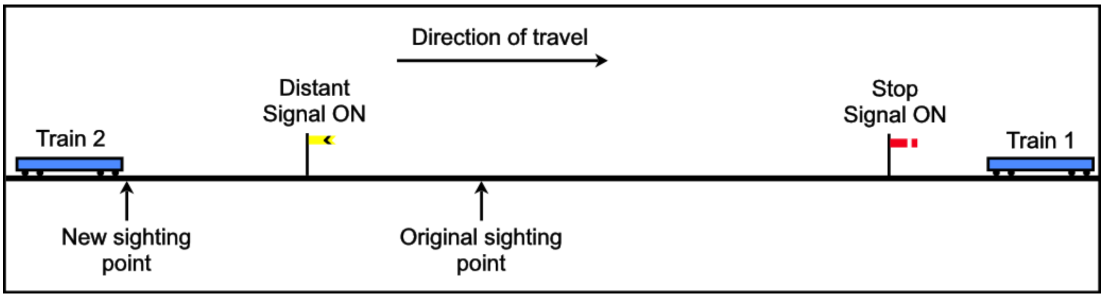

Figure 4: Schematic showing the relationship between a distant signal and its associated stop signal. This provides an increase in the warning time for a stop signal and therefore in the braking distance. Diagram: Author.

The introduction of the distant signal provided an earlier sighting point and thus increased the available braking distance, allowing higher train speeds. With the introduction of warnings, it

---

[2] Signals are said to show a "position" where a semaphore arm is used and an "aspect" where coloured lights are used.

became necessary to differentiate between the indications offered to drivers and yellow became the distant signal colour while red and green became "stop" and "proceed".

Another development was the "track circuit". Originally signals were operated manually from "signal boxes" provided for each block, with trains being passed from block to block but, towards the end of the 19th Century, automatically operated signals appeared. Their operation depends on the replacement of the visual train detection system by an automatic system using an electrically operated track circuit. In its most basic form, each block has a low voltage circuit passing through the running rails. This is connected into a relay, which registers the status of the circuit. The relay is used to change the aspects showing on the signal protecting the section. It is designed as a "vital" or "fail safe system".

Track circuits of various types are widely used, including variable frequency circuits that remove the need for mechanically insulated rail joints to separate block circuits. In some installations, axle counters are used to perform the same function.

Multi-Aspect Signalling

The division a line into blocks or sections presents a restriction on the capacity of the railway. The degree of the restriction becomes dependent on the length of the blocks. Indeed, the ultimate restriction will be the length of the longest block. On routes where traffic developed to a level where more and more trains were required, block length became critical. Capacity could only be increased by reducing block length. However, shorter blocks will not allow the operation of higher speed trains unless the signalling is arranged to provide warning of the state of several blocks ahead. From this requirement, multi-aspect signalling was born.

The purpose of multi-aspect signalling is to allow a mix of trains to operate at high frequency and varying speeds of the same track. There are various forms of this type of signalling. One is shown in Fig 5 below.

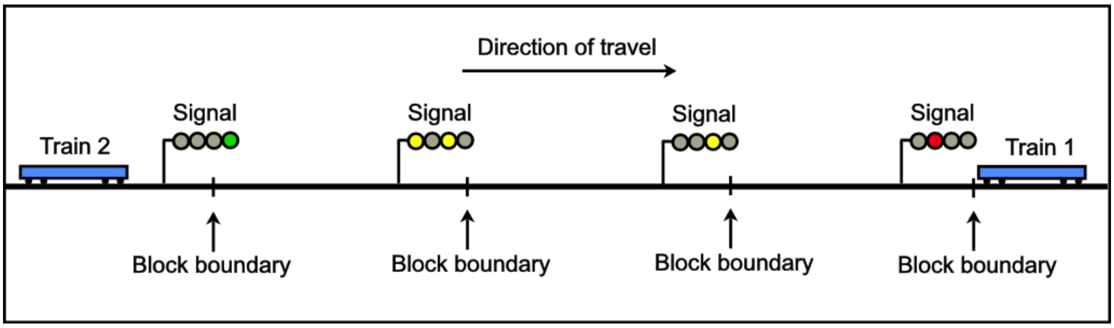

Figure 5: Schematic of UK 4-aspect signalling system. The occupied block is protected by a red signal. The signal protecting the block to the rear, shows a single yellow, the next a double yellow and the one three blocks to the rear, a green. Other railways have equivalent systems. Diagram: Author.

In the UK, a 4-aspect signalling system is common and allows trains to operate at speeds of up to 200km/h. Blocks are normally about 1100m in length. A green aspect allows full speed running, indicating at least three blocks ahead are clear. A double yellow aspect shows two clear blocks ahead, while one yellow indicates one clear block ahead. A similar 3-aspect system, with slower speeds or longer blocks, is also common, using red, yellow and green aspects. Other railways use similar systems, including sequences of flashing lights and multiple lights to indicate the number of blocks free or the required speed limit.

Routing Systems

Turnouts[3] are provided where tracks are divided or joined. In order to prevent the possibility of derailments, it is a requirement that, when selected, turnouts are mechanically secured and

---

[3] Also known as "points" in the UK and "switches" in the US.

electrically monitored whilst trains are approaching and traversing the route. This process, known as interlocking, is also used to prevent any conflicting routes or movements being set up. Routes are protected by signals which cannot show a proceed aspect unless the route over which they can admit trains is set, locked and proved locked. An interlocking can be controlled by a manned, local signal cabin, a remote control room or through an automated computer control programme supervised from the control room.

It is normal to provide the interlocking equipment and its safety systems local to the turnouts it controls. This was originally necessary because of the physical limitations of the mechanical locking systems and control rodding but now it is done to reduce the hard wiring usually mandated for such systems. Remote control need not be vital but the local interlocking system must be.

Signals in route-based systems provide drivers with visual indications showing the route set up. This assumes that information about the speed limits on diverging routes (if any) is provided separately, usually by a fixed sign at the track side. Speed-based systems do not necessarily provide route indications.

Junction Operation

The operation of frequent services through a junction (Figure 6) requires careful timetabling and accurate timekeeping if the planned throughput is not to be compromised. Each route has to be set and locked before the train approaches the sighting distance of the signal protecting the route, otherwise the driver will have to apply the brakes in case he is required to stop at the junction signal. For a train approaching at 160km/h, the route must be set and locked 82s before arrival. All conflicting moves will be prevented during this time. To this, the time for the train to clear the route must be added – about 10s. Assuming an opposing train movement is required through the junction after the first train in our example has passed, a second train on the same route can only pass through the junction 4.5 minutes later.

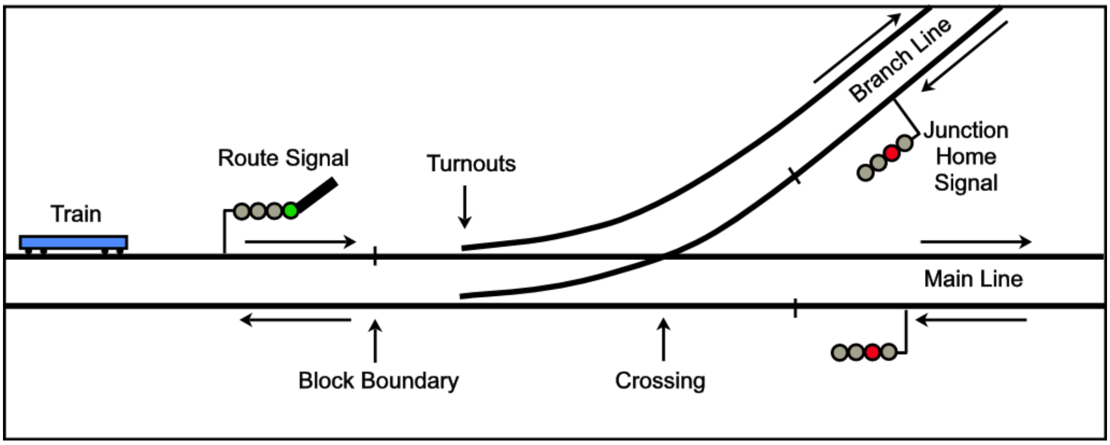

Figure 6: Schematic of double track flat junction. The route is set for a train to pass through the junction on the main line. The junction home signal must be held at danger in order to protect the move over the crossing. For trains approaching the junction at 160km/h not to see an adverse signal, the route needs to be set up 82s before arrival. A further 10s is required to set up the other route from the branch line in the opposite direction. The conflicts restrict the throughput of trains. Diagram: Author.

For urban railways, grade separation of junctions must be the preferred option. This allows unobstructed throughput for converging routes and prevents delays in one direction affecting service in the other direction. Grade separation requires additional construction and land take but it is essential for the efficient operation of a high frequency metro service. It is invariably money well spent.

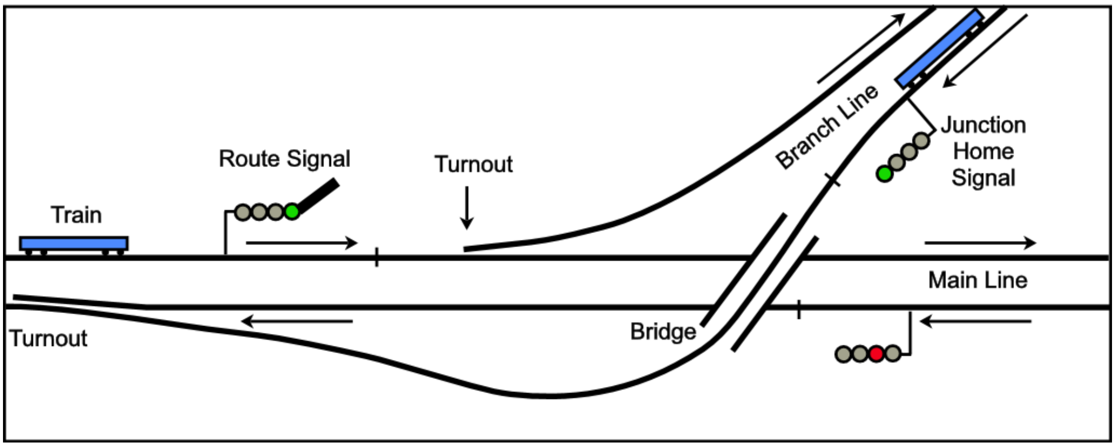

Figure 7: Schematic of grade separated junction, allowing removal of the conflict at the crossing point by lifting the branch line track over (or under) the main line. This is an expensive option to construct and requires considerable land take but it permanently removes conflicts and can increase throughput by 30%. Diagram: Author.

Enforcement

From the earliest days of railway operation, mistakes by drivers led to trains passing signals at danger, nowadays known in the UK as SPADs[4]. Some of these incidents resulted in serious accidents involving the death of staff and passengers. Huge publicity is always given to train accidents and there was always (and still is) constant pressure on railway administrations to provide improved safety. Various schemes to provide some form of enforcement of signal stops were devised, some more successful than others. Some main line railways use a warning system in the cab to advise drivers of an adverse signal aspect. This requires the driver to acknowledge the alarm to prevent a brake application.

TPWS

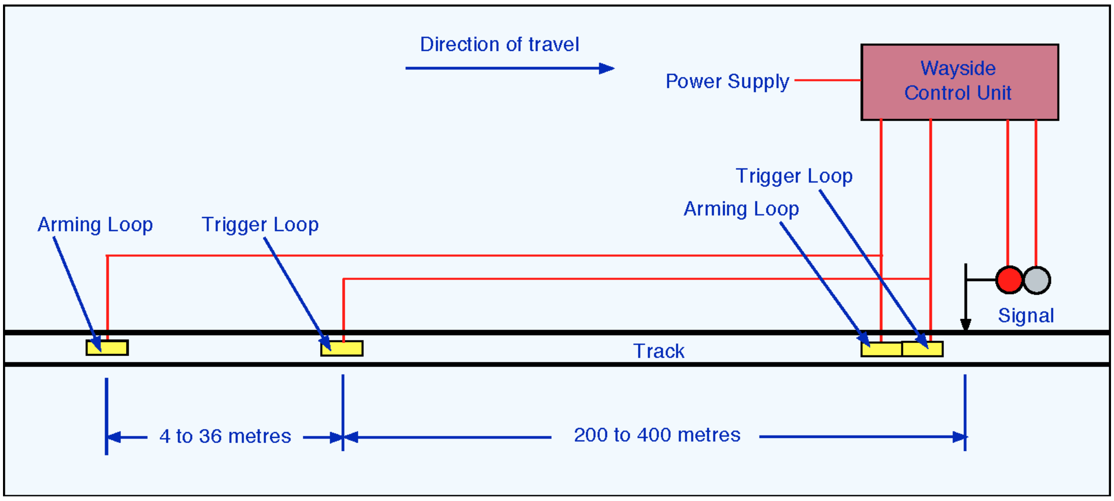

Figure 8: Schematic of TPWS setup on the approach to a stop signal. The Arming Loop switches on a timer and the Trigger Loop assesses the time elapsed to determine the speed of the train. If the time is too short, showing the speed is too high, the trigger will activate the train brakes. Diagram: Author.

The UK uses the Train Protection and Warning System or TPWS. The idea behind TPWS is that, if a train approaches a stop signal showing a danger aspect at too high a speed to enable it to stop at the signal, it will be forced to stop, regardless of any action (or inaction) by the driver. The equipment is arranged as shown in Figure 8.

---

[4] SPAD = Signal Passed At Danger.

For each signal equipped with TPWS, two pairs of electronic loops are placed between the rails, one pair at the signal itself, the other pair some 200 to 450 metres on the approach side of the signal. Each pair consists of, first an arming loop and secondly, a trigger loop. The loops are activated if the signal is showing a stop aspect.

Figure 9: A pair of TPWS loops fixed to the track close to a signal. The two loops act as a tripping activator if a train attempts to run past a signal showing a stop aspect. Most main line signals is the UK are fitted with this system. There will also be two further loops on the approach to the signal set to match the required train approach speed. Photo: Author's Collection.

The pair of approach loops first met by the train at 400 to 200 metres before the signal, are set between 4 and 36 metres apart. When the train passes over the arming loop, an on-board timer is switched on to detect the elapsed time while the train passes the distance between the arming loop and the trigger loop. This time period provides a speed test. If the test indicates the train is travelling too fast, a full brake application will be initiated. In case the train passes the speed test successfully at the first pair of loops but then fails to stop at the signal, the second set of loops at the signal will cause a brake application. In this case, both loops are together (Figure 9) so that, if a train passes over them, the time elapsed will be so short that the brake application will be initiated at any speed.

London Underground and some other metro systems use a mechanical trainstop device at all stop signals. These cause an emergency brake application on any train passing the signal at danger. This is a successful system but it represents a significant maintenance burden and it is not suitable for train speeds much over 100km/h.

In the last 60 years or so, beginning in the US, some main line railways provided cab signalling indications and these have been coupled to enforcement systems on the train. During the same period, metros have adopted various forms of automated enforcement of signal aspects coupled with automatic driving of trains. These systems have become known as Automatic Train Control (ATC) and they incorporate enforcement, known as Automatic Train Protection (ATP), and automated driving, known as Automatic Train Operation (ATO). They will be described in a future paper.