Back to basics: signals

Francis How

See page 36 for references to all the Back to basics articles published so far.

In this article in our "Back to basics" series we look at lineside railway signals. Signals are referred to frequently in other "Back to basics" articles, but without exploring the subject in detail.

Movement authorities

Trains and train drivers need to be provided with information telling them when to move, and where and when to stop. The generic term for information that permits a train to move onto a section of line is 'movement authority'. Although railways vary enormously in how they operate, almost all rely on one of three means by which movement authority information is given to the train:

Verbal instructions given to the train driver to proceed to a designated location on the line ahead. The instruction is usually issued via a train radio system by a person in the control (dispatching) centre or similar location, such as a signal box. These types of systems are often used on low-capacity lines (sometimes referred to as 'dark territory') where a costly signalling system cannot be justified. For more information about these systems and their more modern equivalents, you should read Robert Baird's Back to basics article on Low-Cost Electronic Authority Systems in issue 282 of IRSE News (November 2021) **irse.info/irsenews282**

Railways also use verbal instructions to drivers when the normal lineside or cab signalling system has failed.

Signals such as these in Germany are an essential link in the chain of safety for most railways.

Lineside signals (also called wayside signals), which comprise arrangements of visual symbols (coloured lights, mechanical devices, signs) positioned at pre-determined locations at the trackside for the driver to see and react to in the required manner. In the early days of railways lineside signals were mainly mechanically operated semaphores, and although many examples of such signalling can still be found around the world, all modern systems use colour light signals, which superficially look like road traffic lights/signals. This article focuses predominantly on lineside signals.

Cab signalling, where the movement authority information is displayed in the driving cab, rather than at the lineside. We will briefly consider cab signalling to understand how the provision of in-cab information compares with lineside signals. However, a full exploration of the topic, which includes systems such as Communications-Based Train Control (CBTC), the European Train Control System (ETCS) and Positive Train Control (PTC), is well beyond the scope of this article.

So, we are talking mostly about lineside colour light signals, the information they convey, and how to ensure that train drivers can see the signals and respond to them appropriately. However, we are not going to explore all the types of signals used in different parts of the world – a whole book could be written about that! Nor are we discussing the decision-making technology which enables a signal to display movement authority information to a driver. That topic is covered in two Back to basics articles about Interlocking – see issues 265 and 266 of IRSE News (April and May 2020) **irse.info/irsenews265** and **irse.info/irsenews266**

Figure 1 – Examples of lineside signals in Switzerland (left), the UK (centre) and Germany (right).

Why do railways need signals?

When you drive a road vehicle, you are ‘driving on sight’. You are responsible for observing the road ahead and other nearby traffic and making decisions accordingly. You choose which route to take at junctions, maintain a safe distance from any vehicle you can see ahead of you, and you slow down when you see the need to do so. If you can see that a collision may be about to occur, you can try to take avoiding action by emergency braking and by steering the vehicle. Although traffic lights (road signals) are provided at some road junctions, most have no signals and safety is dependent solely on all drivers complying with a set of rules.

Railways are quite unlike roads. It is not possible to steer a train; it will go wherever the rails and points (switches) take it. A train cannot be stopped quickly; a passenger train travelling at 200km/h will have a typical braking distance of over 2.5km. So, except for very low speed railways such as tramways, driving a train is not a ‘drive on sight’ operation.

To maintain safety in this very different operational environment, railway tracks are divided into ‘block’ sections, and normally only one train is allowed to be in each block section at any one time. A block section may be quite long (several kilometres), or very short (a couple of hundred metres or less, particularly on metros). Each block section may be plain line, or it may have points in the route; and in some cases trains may be permitted to travel through it in either direction.

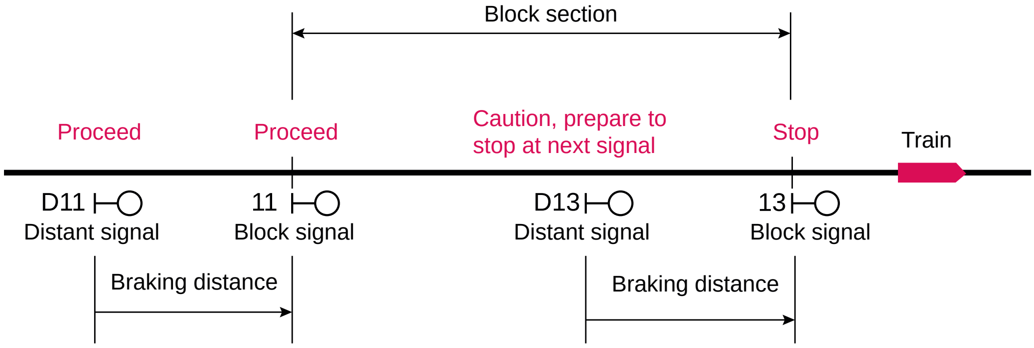
The driver of a train at the start of a block section may not be able to see all the track that comprises the section ahead. They cannot be sure whether points in the block section are set and locked correctly, nor know with certainty the movements of other trains already in (or approaching) the same block section. Because the driver cannot know if it is safe to proceed, a 'block' signal is positioned at the start of each block section to provide that information. An approaching train must stop at the signal unless it displays an indication to proceed. The signal is operated from the control centre via an interlocking, and it will display a proceed indication only when the block section is set and locked for the train, with no other conflicting train movements. The driver can then move the train through the block section to the signal that marks the start of the next section. If everything is working to plan (the timetable), train drivers should only infrequently encounter signals displaying ‘stop’, thereby minimising journey times and making maximum use of the line capacity.

Because a train takes a considerable distance to stop, a signal displaying ‘stop’ may not be visible to the driver at the location where they need to start slowing down. To ensure the driver is given adequate advance information and can brake appropriately, one or more ‘distant’ signals (also sometimes called 'approach' signals) are provided on the approach to each signal that can display a stop indication.

The distance from the first of these signals to the stop signal will be the longest braking distance (under normal adhesion conditions) of all the trains that are authorised to use the line, plus a

>>>

Figure 2 – Block signals displaying 'proceed' (11) and 'stop' (13), with associated distant signals displaying 'proceed' (D11) and 'prepare to stop' (D13) indications.
With acknowledgements to Jörn Pachl.



margin for safety. If a signal is displaying stop, the preceding distant signal(s) will display appropriate 'caution' information to warn the driver of the need to slow down and be prepared to stop at a signal further ahead. If a signal is displaying 'proceed' (rather than stop) then the preceding distant signal(s) will display information to tell the driver that there is no need to slow down. These two scenarios are shown in Figure 2.

On routes where the traffic density is higher, every signal may be capable of displaying 'proceed', 'caution, prepare to stop at next signal ahead' and 'stop', and a block section is simply the portion of track between one signal and the next, equal in length to the braking distance (plus a margin). This is shown in Figure 3.

On very busy railways, the block sections may be made shorter than full braking distance to further improve capacity (by reducing the headway). In such cases, more than one preceding signal will display a 'caution' indication on the approach to a signal displaying stop, with full braking distance from the first of these signals to the stop signal – see Figure 4 for an example. The aspects of the signals at caution (9 and 11) are usually different from each other, to avoid misleading the driver as to which signal ahead is displaying a stop indication.

Signal engineers, railway operators and train drivers use the word 'aspect' in association with signals, by which we mean all the information that is presented by a signal to the driver at a particular time, such as stop, proceed, and 'caution, prepare to stop at a signal further ahead'.

The aspects displayed by successive signals that a driver encounters is called an 'aspect sequence'.

How driving information is conveyed by lineside signals

In accordance with international technical standard UIC 732, most railways use the colour red for signals displaying a stop aspect (sometimes called a 'danger' aspect); green for 'proceed' (often called a 'clear' aspect); and yellow for 'caution' (sometimes called an 'approach' aspect). See Figure 5 for an example of an aspect sequence using these colours.

A signal may also provide additional information for the driver, most importantly for diverging junctions where the train needs to slow down to traverse the points safely. There are basically two different ways in which this is done.

Most of the world's lineside signalled railways use 'speed signalling systems', whereby the signals tell a driver how fast to proceed on the approach to diverging junctions and, in some systems, when approaching a signal showing a stop aspect. However, there is generally little or no information provided about which route is set for the train at a junction.

By way of an example, the Dutch speed signalling system, which is based on USA practices, uses three basic aspects (red, yellow and green), supplemented where required by illuminated numeric indicators which tell the driver of a speed limit ahead; for instance, '8' to indicate a maximum speed of 80km/h – see Figure 6.

Some other speed systems use combinations of colour lights to indicate speed limits, rather than displaying a number. For instance, the Italian national railway uses two lights displayed vertically (yellow over green) to warn the driver of the speed limit over a junction ahead that is set for the diverging route. In the example shown in Figure 7, the yellow/green aspect tells the driver that the junction ahead is set and must be traversed at a speed not exceeding 30km/h. For higher speed turnouts (60km/h and 100km/h), flashing yellow over green lights are used.

Figure 3 – Signal sequence with all signals capable of displaying 'proceed', 'prepare to stop' and 'stop' indications.
With acknowledgements to Jörn Pachl.

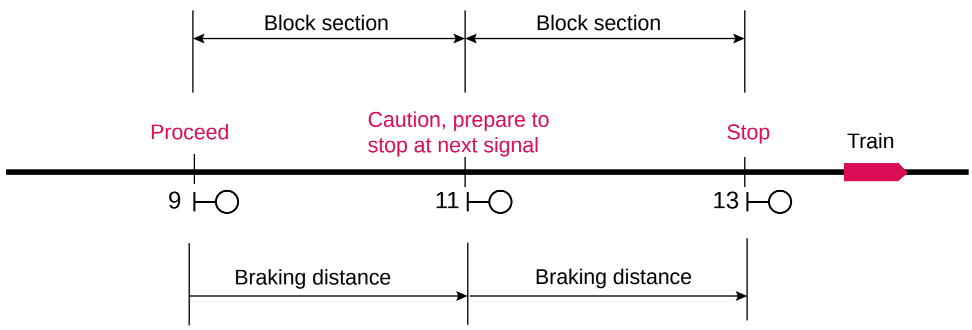


Figure 4 – Signal sequence with 'prepare to stop two signals ahead'.
With acknowledgements to Jörn Pachl.

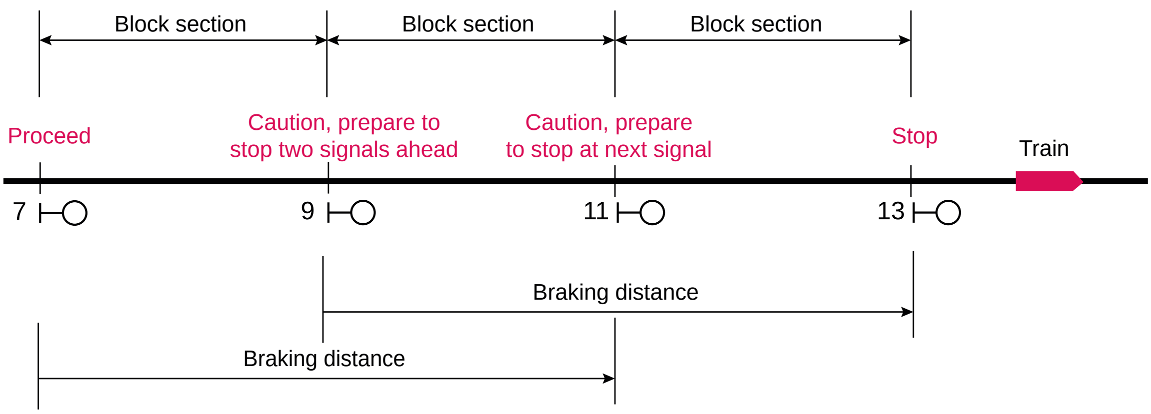


Figure 5 – A typical sequence of successive signals corresponding to Figure 3, displaying 'proceed', 'caution, prepare to stop at next signal ahead' and 'stop' (green, yellow and red respectively).

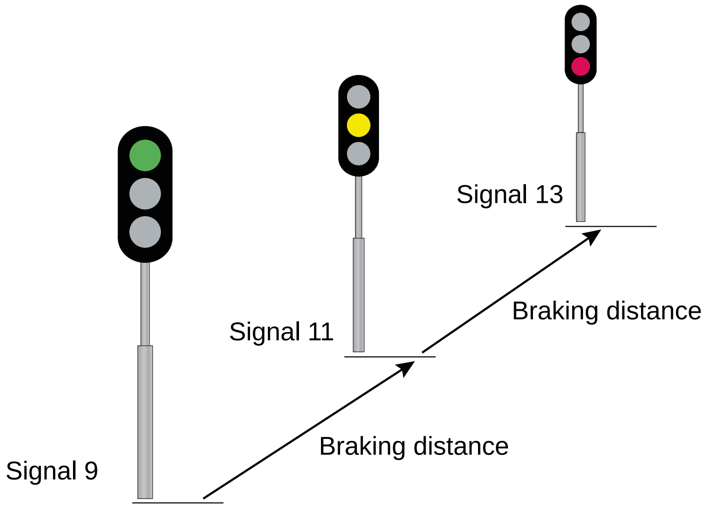


Figure 6 – An example of a Netherlands railway signal, instructing the driver to reduce speed to 80km/h before reaching the next signal ahead.

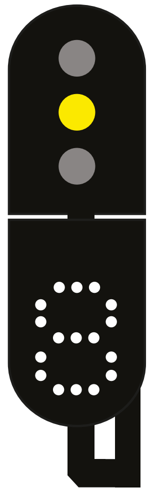
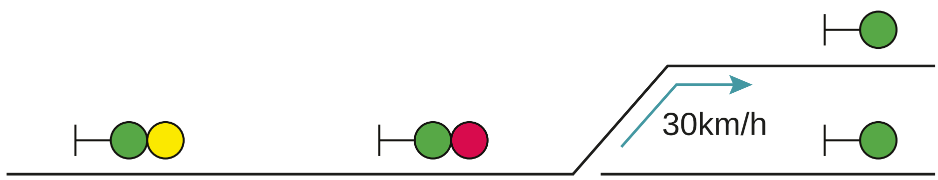


Figure 7 – An example of Italian speed signalling, where the diverging route is set and must be traversed at no more than 30km/h. The red over green (proceed) aspect displayed by the junction signal confirms to the driver that the diverging route is set and that the next signal ahead is also displaying a proceed aspect.



Although speed signalling predominates, some countries (including Britain and other countries that have systems based on British practice) use an alternative called 'route signalling'. Information is provided at signals that protect diverging junctions to tell the driver which route is set for the train. The driver is expected to know the prevailing speed for the route set, and when to start braking to traverse the junction safely. The signals provide no information about speed limits over junctions.

The route information may take the form of a junction indicator (a 'bar' of white lights set at an angle corresponding to the route set, usually mounted above the main signal, see the centre photo in Figure 1); or an illuminated alphanumeric character may be displayed above or alongside the main signal to indicate the route set; or, less commonly, a pair of signals may be provided alongside each other, one for the straight route and one for the diverging route. The one for the route set displays the appropriate aspect to allow the train to proceed over the junction, while the other remains at red.

To reduce the risk of a driver inadvertently traversing a diverging junction too fast on a route-signalled railway, which could result in derailment, 'approach release' (also known as 'approach control') of junction signals is employed. This typically involves holding the junction signal at 'stop' with the preceding signal(s) displaying the corresponding caution aspect(s) until the train has slowed and the driver can clearly see the junction signal ahead, at which point the interlocking allows the signal to display the appropriate aspect and route indication for the train to proceed over the junction. At some high-speed junctions in Britain the junction signal may initially be held at 'caution' (single yellow) with the junction indicator displayed, and the preceding signal(s) display flashing yellow aspects to warn of the need to slow down for the diverging route.

Although signalling systems can be classified broadly as 'route' or 'speed' in type, some speed signalling systems include some features of route signalling, and to a limited extent vice-versa. For example, the Italian speed system referred to earlier includes both speed and (limited) route information in the signal aspects on the approach to a diverging junction. Furthermore, not only do some countries have more than one type of speed or route signalling, a few (notably the USA and Australia) use route signalling on some railways and speed signalling on others. Wikipedia provides a useful summary of the signal aspects used in various Australian states. irse.info/idkn9

Fixed lineside speed reminder signs may be provided with both route and speed systems. This is arguably more important with route signalling than speed signalling, although it in no way replaces or lessens the need for route knowledge.

The IRSE's textbook "European Railway Signalling" provides information about the many different systems of signal aspects used by European railway administrations. It illustrates well just how much they vary from one country to another.

The driver's perspective

Having considered the information provided by lineside signals to drive a train safely, we will now think about how to ensure that each signal displays the information in a manner that drivers can assimilate and use.

Good practice for signal visibility is based on the needs of the driver. On approaching a signal, a driver has

>>>

The knowledge and information required for train driving

Train drivers need a wide range of information to do their job, of which signals are only one part. The information and knowledge they require includes:

- **Route and infrastructure information** (commonly known as 'route knowledge'), including the locations (and layouts) of junctions, stations, lineside signals, permitted speeds and gradients. This information is relatively static, with occasional changes when the infrastructure manager alters track layouts and signalling.
- **Rolling stock information**, including the train's maximum speed, braking capability and length. This information is particular to each type of train, and some of it (such as length and braking capability), may be particular to each individual train trip – particularly in the case of freight trains.
- **Timetable information**, telling the driver about planned departure, passing and arrival times at specific locations.
- **Driving rules for driving the train safely under normal operational conditions**, in degraded modes and in emergency situations; and for activities such as those associated with arrival/departure from stations (passenger trains) and loading/unloading operations (freight trains).
- **Movement authorities**. The driver must know the meaning of the information displayed by lineside signals (or the in-cab display), and how to respond to that information.

In the case of a metro, the operation of all the trains using any one line is more or less identical and driving is a relatively straightforward task, particularly if the trains are fitted with automatic train protection and operation (ATP, ATO). At the other extreme, the manual driving of long and heavy freight trains, where each one may have different characteristics, is much more complex.

firstly to read the signal (observe the information displayed), then interpret the aspect (understand what it means), and then drive the train accordingly.

Reading a signal requires that:

1. It is clearly applicable to the driver of an approaching train, and readily distinguishable from any other signals in the vicinity (for instance, other adjacent signals, and signals further ahead on the line).
2. All the illuminated lights that comprise the aspect displayed are clearly visible for a sufficient length of time for the driver to observe the aspect, taking into account the highest attainable or permissible approach speed.
3. It is visible under all conditions and at all times of day and night, including those where sunlight or nearby artificial lighting might make readability difficult.

Interpreting a signal requires that:

1. The aspect displayed is one of an established set of aspects applicable to the railway, with a clearly defined meaning for drivers.
2. The information conveyed by the aspect is unambiguous, so that the driver is not placed in a position of uncertainty as to what action to take; nor be 'conditioned' by previous driving experience to expect to take a particular course of action, only to find on rare occasions that a different action is required even though the aspect displayed is the same.
3. The aspect is not corrupted (all relevant lights illuminated, and no others), to avoid the driver either not understanding the information presented or interpreting it as having a different meaning to that intended by the signalling system.

When the driver has read and interpreted a signal, they then decide on what action to take, making use of their knowledge of the rolling stock, the route ahead and the applicable operational rules.

Driveability of the signal requires that:

1. The aspect forms part of a logical aspect sequence encountered by the driver, except of course in emergency or failure conditions, for instance if a signal reverts to red to stop a train because of an unexpected unsafe condition ahead.
2. Where the aspect requires the driver to reduce speed, it must be possible to meet the speed reduction requirement by means of normal (service) braking, taking into account factors such as the maximum attainable/permissible speed on the approach to the signal, rolling stock performance and braking capability, gradients and distance to next signal.

Figure 8 – A lineside signal in Japan, indicating a route set to the right at a diverging junction. Japan predominantly uses speed signalling, but with route signalling at junctions.

Figure 9 – Signal structures in the USA, with signals on a cantilever structure above the main running lines. To the right is a signal on a post for trains exiting the loop line.

3. The information displayed is of fairly immediate relevance to the driver. Displaying information well before the driver needs to act on it could risk them forgetting it by the time it becomes relevant.

Two important corollaries to these requirements are firstly that drivers must meet the relevant eyesight (colour vision and visual acuity) and competence requirements; and secondly that driving cabs should be designed to provide the driver with an optimal view of lineside signals, not only from a distance but also when close to, or stationary at, a signal.

Positioning signals

It should be clear by now that designing a lineside signalling system requires careful consideration of the train drivers' needs. In particular, the locations of signals for optimum viewing and correct driver response are of critical importance. The locations are initially determined by headway and signal spacing calculations, taking account of the topography of the railway infrastructure (including track layout, gradients and speed limits), operational requirements, the rolling stock that will use the railway, and capacity requirements. Other factors may also need to be considered, such as:

- Line curvature and structures (such as bridges, station infrastructure and overhead line equipment) that may impede signal visibility.
- Viaducts and tunnels, where it is inadvisable to place signals unless unavoidable because of the risks associated with evacuating a train in an emergency.
- Nearby artificial lighting (on stations, buildings and roads) which could impair the clarity with which signals can be seen.
- The effects of sunlight (shining either into the signal or into the driver's eyes).
- The possible consequences of a train passing the signal at danger (stop) – commonly known in Britain and some other countries as a SPAD (Signal Passed At Danger').

The proposed location and light beam alignment of each signal is usually subject to review to ensure that in practice a driver will be able to read the signal aspect reliably and accurately, and that there are no unforeseen hazards associated with the location.

This review process, which may involve site visits and/or the use of simulation systems, is known as 'signal sighting' (in Britain, at least), and is a joint activity involving signalling designers and train operators. Signal maintainers may also participate so that maintainability issues can be considered and addressed.

Where the time for which a signal is visible to a driver is unacceptably short (for example, because of tight curvature on the approach) and it is not possible to place the signal in a better location, repeater signals are sometimes used, located a short distance before the main signal. As their name implies, they repeat some or all of the information displayed by the main signal to give the driver an extended time to read and interpret the aspect. It should be noted that if the signal is displaying 'stop', the driver is expected to stop at the main signal, not the repeater.

Countdown marker boards are also sometimes provided on the approach to signals to aid the driver. On some railways this is done routinely; on others it is done only in special circumstances, for instance where the signal sighting time is poor.

Most railways have conventions about whether signals are located to the left or the right of the track, usually dependent on whether, on double track lines, trains are normally on the right-hand or left-hand track in the direction of running.

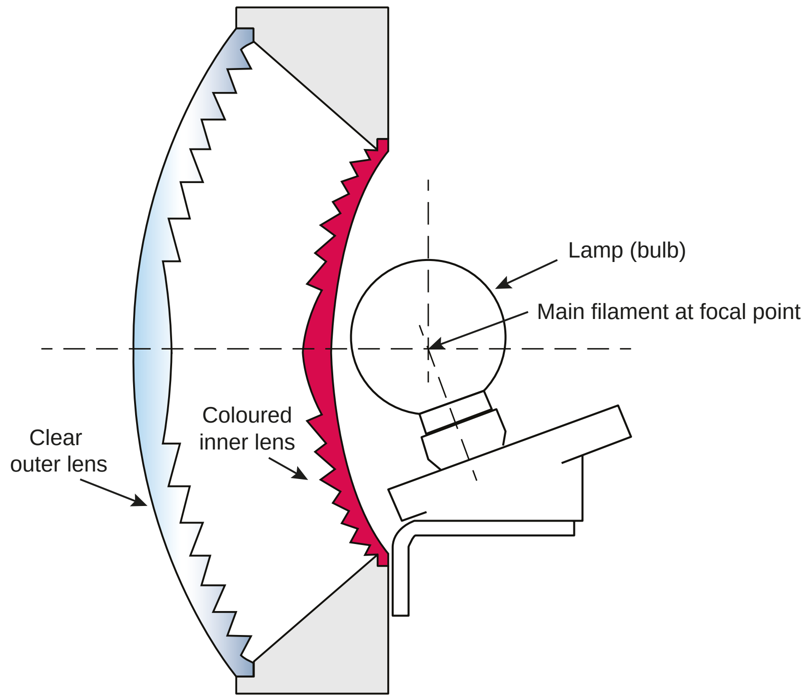
Main signals are usually positioned at or above driver's eye level, on either posts (masts), cantilever structures or gantries. The latter two are used mostly on busy routes with several parallel lines where there may not be room to place individual signal posts between the tracks, see Figure 9.

>>>

On some railways, notably underground metros, there may be insufficient room to place signals in locations that are clear of the swept envelope of the train. In such cases, smaller signal heads might be used and/or they might be positioned near ground level instead of being elevated, where there is more room for them.

Lastly on the subject of signal positions, each signal may be assigned an alphanumeric identity that is unique within the area of the control centre. Typically the identity is shown on a plate attached to the signal structure, and on the signalling display system in the control centre. If a driver and signaller need to speak with each other (by radio or phone) when a train is stationary at a signal, the identity is used to verify the train's location. This is vital for safety, particularly if the signal is displaying 'stop' because of a failure and the driver has to be given verbal instructions to proceed.

Figure 10 – A cutaway side view of a typical filament lamp lens system.



Signal head design

Signal heads have many different designs. The predominant form is a metal structure within which are separate lamp (bulb) and lens assemblies for each coloured light that the signal requires. Each assembly has an electrical connection to the signalling system which switches the lamp on or off.

Some railways use lamp and lens assemblies that allow multiple colours to be displayed through a single lens, rather than having a separate assembly for each colour. The so-called 'searchlight' signal (an electromechanical device) is one example of this. Other types use dichroic mirrors or colour LEDs.

The filament (incandescent) lamps traditionally used for signals are generally low power devices (typically 24W or less) specifically designed for the purpose, in front of which is a lens system, see Figure 10. This arrangement produces a very concentrated beam of light that is visible as much as a kilometre away in good conditions, with a narrow angle of divergence to help ensure that the signal can be seen predominantly by drivers on the line to which it applies but less so by drivers on other (parallel) lines. Lens systems can also be wide-angled when necessary to aid viewing in special circumstances, such as when approaching on a sharp curve.

LED technology has led to a revolution in railway signals, enabling heads to be much more compact, lighter and reliable, with a less complex lens system, whilst still meeting the requirement to produce a concentrated beam of light with beam intensities suitable for long, medium or short-range viewing. For these reasons many railways now use LED signals in preference to filament lamp types for signalling schemes and renewals.

Although the visibility of a signal from a distance is of prime importance, a driver also needs to be able to see the aspect displayed when stopped close to the signal. Various techniques are used to ensure this, including a specially designed part of the lens which shines light down towards the driving cab; using small repeater (or co-acting) signals at cab height for close-up viewing; and requiring drivers to stop a few metres before the signal, where it is clearly visible to them.

The visibility of signals can be problematical in locations where at some times of the day and year the sun is either shining directly into the signal or is shining directly behind it. Signals are usually equipped with hoods to reduce the problem, and lenses are designed so far as possible to minimise the reflection of sunlight, which could otherwise produce so called 'phantom aspects' where the lamp appears to be alight but is not.

Signals must be highly reliable, for both safety and train performance reasons. On many railways, failure of a signal light is considered sufficiently hazardous that the signalling system has built-in mitigations to minimise the risk of a driver encountering an un-illuminated ('black' or 'dark') signal. Typical arrangements include double filament lamps, one being the standby for the other; proving within the signalling system that the signal ahead is alight before a driver is given a movement authority towards it; and restricting aspects and their sequences so that drivers only encounter correctly illuminated signals. Railways' operating rules usually state that in the unlikely event of a driver encountering an un-illuminated signal, they should treat it as displaying 'stop'.

The use of LEDs in signals has made the sudden failure of a light very infrequent compared with filament lamps. Each light typically comprises a matrix of multiple LEDs and therefore a single LED failure only marginally degrades visibility.

Maintenance considerations

Signals need to be maintained, which may include cleaning the lenses, replacing failed lamps, and periodically checking beam alignment and lamp voltages.

Safe access for maintenance staff is therefore an important consideration in the design of signal structures and, to some extent, in their positioning as well (although the needs of the driver take precedence). Signals are usually equipped with ladders for access, although some modern lightweight single post structures have a fold-down arrangement so that the signal can be lowered to the ground for maintenance. Cantilever and gantry structures are sometimes used in preference to

Figure 11 – An example of a shunting signal. This one is in Great Britain; the appearance and aspects of shunting signals vary widely on different railways.

Figure 12 – Two tram and two bus signals on a single supporting structure in Germany.

single posts to facilitate access for maintenance purposes without requiring staff to cross the tracks, a walkway for maintainers being an integral part of the structure.

Some signals use optical fibres to transmit the light to the signal head, allowing the lamp unit to be placed closer to ground level to make maintenance easier and allowing the signal head to be mounted on a simpler structure, or even on the overhead electrification supports if space and visibility requirements permit.

Signals for low-speed movements

So far, we have only discussed the main signals used for running movements, where a train is travelling at speed from A to B. Railways must also make provision for a variety of other train movements, for which distinctly different signals are generally used. These movements include:

- Shunting into and out of sidings. Such movements may be in the opposite direction to the normal traffic flow.
- Detaching/attaching locomotives and wagons from/to trains.
- Enabling a train to enter a section of line already occupied by another train, for instance where a platform is to be shared by two trains or where two trains are to be joined together.
- Moving a train over a crossover from one line to another so that it can reverse direction and travel back on the other line.

The signals used for these purposes are generally known as shunting or position light signals, although when associated with a main signal (i.e., attached to the same structure) they are sometimes called subsidiary signals. The colour red is commonly used to denote stop, and authority to proceed is indicated by white or yellow lights (often in pairs), see Figure 11. This is not universal practice however, and railways use many different forms. The USA, for instance, has several types of position light signals (and, incidentally, main signal aspects as well), their existence being due to the varied historical development of the railroad companies.

Because shunting signals are used only for low-speed movements, they are not designed to be visible over great distances, nor is there any need for aspect sequences to warn drivers to slow down when approaching a shunt signal displaying stop. The signals are generally located close to ground level, except where they are associated with a main signal, when they are mounted alongside or immediately below the main head.

Tramways (streetcar systems) also generally operate at low speeds and braking distances are sufficiently short that a tram (streetcar) can be brought to a stand when the driver sees a signal displaying stop, without the need for 'distant' signals. Signals are provided to indicate when and where to stop at intersections with roads and other tram lines (including entry to single line sections), and they may also indicate which route is set.

Tramway signals take many different forms, and streets both intersect tram tracks and share the same space. It is therefore important to ensure that signs, road markings and signals for trams are distinct and different from those applicable to motorists and pedestrians. Both motorists and tram drivers need to be aware of signals applicable to them, and to some extent also understand the meaning of those which apply to the other mode of transport, see Figure 12.

How train warning and protection systems help drivers to obey lineside signals

A separate "Back to Basics" article deals with the subject of train protection systems, and it is not the intention to go into the details of such systems here. However, it is worth mentioning the role of both warning systems and protection systems in the context of lineside signalling.

Train-borne warning systems alert the driver to take note of information presented by lineside signals, and they may initiate braking if the driver fails to acknowledge the warning. After acknowledgement, the train is usually solely under driver control.

Train-borne protection systems monitor the speed of the train and cause a brake application if the train is being driven outside safe limits. Some, such as train-stops, intervene to apply the brakes if a train passes a signal at danger (stop) but without providing a visual or audible warning before doing so. At the opposite end of the spectrum, continuous >>>

Figure 13 – Typical cab display for ETCS (left) and a CBTC system (right).

Automatic Train Protection systems (ATP) provide comprehensive information about the speed profile that the driver should be following, warn the driver of over-speeding, and intervene to slow the train down if the train speed is not reduced appropriately.

Warning and protection systems that are provided in conjunction with lineside signals can assist the driver in several ways:

1. They can help maintain the driver's vigilance by providing in-cab visual/ audible indications or warnings on the approach to each signal, repeating some of the information provided by the signal.
2. They can provide an in-cab reminder of the aspect or information displayed by a signal after the driver has passed it.
3. They can aid the driver in observing signals in conditions of poor visibility, such as fog.
4. They can intervene to slow the train down if the system judges that the train speed is too high.

Cab signalling

Before we go further, note that trains which are driven using Automatic Train Operation (ATO) are not part of this discussion, nor are we considering Driver Advisory Systems (DAS), which assist the driver in driving optimally (better time-keeping and energy-efficient driving) but are not a substitute for lineside signals. See ITC Report 44 for more information about DAS. **irse.info/itc44**

Cab signalling may be provided for various reasons, including better operational performance, improved safety and reduced costs. Bob Barnard's article "What can cab signalling ever do for us?" in issue 239 of IRSE News (December 2017) **irse.info/irsenews239** provides an excellent broader perspective on the challenges and benefits of cab signalling.

Modern cab signalling systems provide comprehensive in-cab driving information for the driver to operate the train safely and optimally without the need for lineside signals (even if they are still provided for unfitted rolling stock). On lines without signals, lineside signs (marker boards) may be provided to tell the driver precisely where to stop the front of the train at the end of a movement authority, but apart from that almost all the information the driver requires is available via the cab display (see Figure 13).

The display is often called a driver-machine interface (DMI), but we will continue to use the term 'cab display' in this article.

Most cab signalling systems include ATP which, as we have already noted, monitors the driver's actions and intervenes if the train exceeds the maximum safe speed. Under normal circumstances the driver is in full control of the train. Even if the system warns them that the train is travelling too fast, the driver remains in control unless they fail to heed the warning and slow down. Like lineside signals, the cab display must be readable, interpretable and driveable.

Readability of the cab display means that:

1. It must be easily visible by the driver from their normal position in the cab, even under difficult lighting conditions such as bright sunlight.
2. Warnings and other important 'alert' information must be particularly prominent and supported by clear and distinguishable audible warnings.

Interpretability of the cab display means that:

1. The information displayed must be clear, unambiguous and consistent with any other information that may be available to the driver (whether in the cab or by lineside signals and signs).
2. The information must make sense to the driver without the need for mental computations or geo-spatial interpretation.

Driveability requires that:

1. Information must be provided to enable the train to be driven safely, including, for example, current and upcoming speed limits/profiles, the distance to the next stopping position, gradients and other relevant driving information.
2. Information must be provided on a timely basis, so that the driver does not need to take abnormal action (such as emergency braking) to comply with it.
3. Information that requires the driver to take action (either immediately or later) must remain displayed for as long as it is relevant.
4. The system must be designed so that the driver can control the train speed without the system intervening – provided of course that they comply with speed limits and do not pass the end of any movement authority. Where a system intervention may become necessary (for instance, due to over-speeding), it should normally be preceded by a warning so that the driver can take the appropriate corrective action.

Most cab signalling systems are designed on the 'head up' principle. In other words, the driver looks out of the cab window at the track ahead most of the

A rainy day in Glasgow, but good sighting and modern signals give drivers a clear indication of their movement authority.

time, glancing at the display only when necessary. This is particularly important on railways where it is possible that an obstruction could arise on the line ahead – although of course there is no guarantee that the driver could take action to prevent a collision. The 'head up' principle makes it doubly important that the in-cab system, and the display in particular, is designed to facilitate this driving style, for instance by providing audible alerts when new information is displayed that requires a response by the driver.

Cab displays are usually interactive, requiring the driver to input information (for example, at the start of a journey, or when changing the driving mode, or to acknowledge warnings). Other information, such as the speedometer, may also be integrated into the display. Ergonomic factors are therefore important in the design of the display and associated input keys.

Closing remarks

In this article we have attempted to present a broad picture of lineside colour light signals, their purposes, and their variations. We have explained the most commonly used aspects, the differences between speed and route signalling, and the importance of meeting the driver's need for signals to be readable, interpretable and driveable.

We have considered how train protection and warning systems support the driving task in lineside signalled territory. Finally, we have acknowledged that, although full cab signalling (with ATP) removes many of the issues associated with lineside signals, the need to make the information displayed in the cab readable, interpretable and driveable remains important.

If you are new to railway signalling, we recommend that you learn more about the lineside signals that are used on railways with which you are familiar. What aspects and aspect sequences are used? How are diverging junctions signalled? How well do the signals meet the needs of drivers, as set out in this article?

You might also investigate types of signals that have not been mentioned – for instance, special signals used to authorise a train to depart from a platform when it is ready to leave; or situations where fixed lineside signs are used instead of colour light signals; or signals used in conjunction with some types of level (grade) crossings.

The contributions of others to the production of this article are acknowledged, with thanks in particular to Robert Baird, Richard Barrow, Hugh Barton, Wim Coenraad, Piers Connor, David Fenner, Mark Glover, Ian Mitchell and Allan Neilson.

///

About the author

Francis has been a long time member of the IRSE. His career has seen him work for British Rail/Railtrack, Atkins, the Railway Industry Association and finally as chief executive of the IRSE before his retirement.

Francis was an IRSE Exam Thorrowgood Scholar and served on Council for many years, becoming president in 2012-2013.

He has previously written "Back to basics" articles on interlockings and points.

If you found this article useful we're sure you will enjoy other Back to basics articles, including:

Interlockings

Francis How

Part 1 April 2020 **irse.info/irsenews265**
Part 2 May 2020 **irse.info/irsenews266**

Points

Francis How

Part 1 February 2022 **irse.info/irsenews285**
Part 2 March 2022 **irse.info/irsenews286**

Train protection

David Fenner

July 2020 **irse.info/irsenews268**