Will ERTMS give us more capacity?

William Barter is sceptical about the capacity benefits of the standard European signalling system set to be adopted in this country

(This article was published in Modern Railways, September 2008)

Many technical texts and discussions focus on 'line capacity', and talk confidently about 'trains per hour'. But is that what we really mean by 'capacity' of a railway system?

What is capacity?

For each line in a network, the signalling system sets the 'headway' – the minimum possible interval between trains that avoids restrictive signal aspects. So long as the signals are located exactly as required to provide the braking distance for the intended maximum speed, and trains actually run at that speed, theoretical headways are remarkably low – around 90 seconds on four-aspect signalling for 100mph trains, and little over a minute at half that speed. But that is to overlook the realities of lineside signals, which must be clearly visible to drivers of approaching trains, not just the wrong side of bridges or tunnels, or out of sight round curves in cuttings. We tend not to place signals in the middle of station platforms so as not to stop trains frustratingly half in and half out of stations. Access for maintenance may militate against placing them in tunnels or on viaducts, which also avoids the risk of trains being stopped at locations that passengers might find unnerving. As signal sections cannot be shorter than is necessary to give braking distance, all these problems can only lead to longer sections and thus longer headways, and the worst group of sections sets the headway for the route.

What happens in practice

All in all, once the signalled headway has been rounded off for the convenience of timetable planners, and some allowance made for robustness in practice, a 125mph line will probably end up with a planning headway of three minutes, and a line on a suburban route, two minutes.

That is fine for one line in isolation, and for a continuous flow of trains running at the full permitted speed, but hardly describes any real railway system. In practice, trains stop at stations, so that their dwell time, which is completely independent of the signalling, adds to the separation. And some trains stop at stations while others don't, so that a wedge of unusable capacity builds up between a through train and a following stopping train.

Flighting

This loss of capacity can be mitigated by 'flighting' – running trains of the same speed in pairs or batches. However, intermediate stations may then find their stops concentrated into short periods, and a more passenger-friendly pattern may be laid down in franchise specifications.

Other factors combine to reduce the calculated capacity further. Flat junctions destroy opportunities to run trains simultaneously on conflicting routes. At each end of the line, trains need to turn back at terminal stations – the rate at which this can be done, determined largely by the turnround time and the number of platforms, is normally much less than the rate at which each approach line might feed trains in or out.

Finally, reality suggests it is unwise to work continuously to the limits of capacity.

In fact, terminal capacity is probably the binding constraint on usage of much of the national network. At Charing Cross, that little miracle of a London terminus, 30 trains arrive in the busiest 60-minute period of the morning peak, even though, based on headway alone, the Fast and Slow lines immediately outside the station could feed in 48 between them (the situation is of course complicated by the short stretch of double-track between London Bridge and Metropolitan Junction, offering just one line for Up trains).

The three Slow line platforms work continuously through the height of the peak at the minimum turnround of seven minutes. With three minutes between occupations of each platform, this comes to 18 trains, just 75% of the capacity of the Up Slow line, which in turn is set by the station stop at Waterloo East. Meanwhile, the three Fast line platforms handle only 12 trains, largely as many trains work back in service according to a clockface timetable rather than just at planned arrival plus seven minutes. Even so, four trains per platform per hour is fully comparable with terminals such as Victoria and Waterloo, and free time in the busiest hour equates to fewer than five minutes per platform.

Definition

So line capacity measured in 'trains per hour' is really a technical abstraction, useful for comparing some details of signalling schemes, but for little else. The Institution of Railway Operators' definition of network capacity, adopted in the DfT Rail Technical Strategy is: 'The number of trains that can be incorporated into a timetable that is conflict-free, commercially attractive, compliant with regulatory requirements, and can be operated in the face of anticipated levels of primary delay whilst meeting agreed performance targets'.

Even this definition refers only to the capability of the network to carry trains; true system capacity also takes account of the ability of the trains to carry passengers.

'Ten-aspect' signalling

So how will the European Rail Traffic Management System (ERTMS) affect all this? A rolling programme of ERTMS installation has been formulated by Network Rail ('Informed Sources', November 2007 issue) and it is expected to be in widespread use in this country by the middle of the next decade.

The features of ERTMS relevant to capacity derive essentially from cab signalling and Automatic Train Protection. If lineside signals are done away with (ERTMS Level 2 System D), the message to drivers becomes simply a safe speed at which to drive, calculated by the on-board computer and displayed in-cab. So a practical system could have shorter blocks and more of them between trains – without the need to display different aspects and expect a driver to comprehend them, perhaps the equivalent of ten-aspect signalling. Some things follow immediately from this:

* Any fixed block system puts one more section between trains than is actually required for braking distance (take three-aspect signalling, for example – braking distance is offered by any one signal section, but for trains to run on greens, they must be separated by sections from Green to Yellow, and Yellow to Red). With a given separation provided by a larger number of shorter blocks, this extra section adds less to the total separation – in effect, the benefit of four-aspect signalling taken to extremes.
* By decoupling block boundaries from the constraints of sighting lineside signals, block lengths can be closer to the theoretical minimum, minimising excess separation. We might, however, still be reluctant to split tunnels into more than one section, but is this really valid in these days of central door locking, good lighting, open stock and public address systems?
* With lineside signalling, trains running on greens are separated by the full braking distance for the maximum permitted speed, even if their own permitted speed is lower and their required braking distance shorter. ERTMS can give an unrestricted 'movement authority' to a slow train on the basis only of the braking distance it actually needs, rather than the worst case (probably the fastest) train, so a flight of slow trains can run with less time-separation than fast trains.
* With Automatic Train Protection a given with ERTMS, the risk of misjudged braking is virtually eliminated, so signal overlaps might be reduced significantly or even abandoned, further reducing separation.

So ERTMS reduces headways, if track circuit arrangements and block boundaries for ERTMS Level 2 are redesigned specifically, rather than simply being ported over from the previous MAS (multiple-aspect signalling) schemes. All things considered, a three-minute planning headway on four-aspect MAS might become two minutes under ERTMS. That sounds excellent – line capacity goes up from 20 trains per hour to 30.

Constraints

The problem is that very few lines with three-minute headways now actually carry anything like 20 trains per hour, for all the reasons of junctions, differing speeds, and terminal capabilities outlined above. What will ERTMS do for those problems? Not a lot!

With regard to the mix of train speeds, the underlying issue is one of differing running times, not of headway. True, at the point where trains enter a 'corridor', a slow train might follow a fast a bit more closely to start with, but the lost capacity en route will not change. Perhaps once the fractions of minutes mount up, another complete train might be run, but which sort of train – another fast, another stopper, or what? The benefit of improving headways is only felt when trains of the same speed and stopping pattern follow each other.

At junctions, some benefit might be found. Trains, especially slow ones, could approach the point of conflict more closely before the interlocking needs to 'deny' it to other trains. But using one route over a point of conflict still prevents trains running on all conflicting routes. And the ability of terminals to accept, turn and despatch trains will not change. In a suburban operation, the limiting factor is the time taken for drivers to change ends (crew changes at the terminus in the peak are not a good idea). For long distance trains, other factors come into play, such as servicing and stocking the buffet, as well as a robustness allowance.

How could it be done?

Let's play devil's advocate for a moment. Amongst the figures that have been bandied around for the potential capacity benefits of ERTMS is 10%, a nice round number. So for 30 trains-per-hour Charing Cross, that means three more, or 33 trains per hour on six platforms. Is that possible?

In terms of platform capacity, the answer is easy – the Fast platforms will have to work as hard as the Slow platforms. The clockface timetable will be at risk, and outer-suburban trains will have turnrounds as short as inner-suburban. The extra trains will have to be formed of rolling stock that can inter-work with the outer suburbans. In theory that gives us six more trains per hour, but let's not overdo things.

Work back from Charing Cross itself, and see what other constraints arise. Waterloo East comes next, where all trains stop. Again the Fast lines could do what the Slow lines already do.

The problem comes at Metropolitan Junction, where the two-track section from London Bridge changes to the four 'paired by use'. That means a diamond crossing, where the 18 trains up the Slow line have to cross the Down workings on the Fast line, which, on the basis that what goes Up must come Down, now total 15. 33 trains per hour over a diamond crossing is a lot. But the current Rules allow two minutes as standard, with $1^1/_2$ minutes 'not for successive moves'. Exploiting that to the full with 33 trains would lead to the diamond being locked out for 58 minutes out of 60 – too high for a reliable service. But if overlaps short of junctions can be eliminated, and speed of trains controlled to keep them moving, maybe ERTMS brings enough to make it realistic. But first the 33 trains have to use the one Up line from London Bridge, and we must assume the headway benefits of ERTMS will allow this.

Once the Thameslink Project has done its stuff, London Bridge can be a Yes. There will be two platforms for Up Charing Cross trains, with 16 or 17 each per hour. This is fewer than the current single Up Charing Cross platform does now off-peak when almost all trains call, albeit with off-peak dwell times. However, we can hope that peak dwell times will reduce, partly as there will be 10% more trains, and partly as without signal overlaps the stopping point of trains can be brought back towards the station access and alleviate the tendency for people to concentrate in the first coach. And as signal locations are currently heavily constrained on this complex layout, we can hope that ERTMS will reduce platform reoccupation times.

Below London Bridge, headways effectively set capacity, as the intermediate station platforms are on the Cannon Street lines. We can reasonably hope for success, at least until we get to Parks Bridge Junction (south of St Johns). Here trains are transferred between the Fast and Slow lines so as to sort them out for Cannon Street and Charing Cross. The numbers of trains making these crossing moves will increase, especially as the 10% increase we are aiming for should apply to Cannon Street as well, and much will depend upon how well the pattern of service exploits scope for parallel moves. This is a big unknown, especially as, to make the London terminal work, clockface patterns are jeopardised.

From Orpington to Parks Bridge headway is again the main constraint on the Fast lines, although usage is lower as the network fans out into branches. But trouble starts again below Orpington, where the line via Sevenoaks is only double track, and capacity is limited by the speed differential between fast and stopping trains. The fast headway is already 2 minutes, and there are long tunnels in which we may still want to limit the number of trains. Three extra trains over this section? Probably not.

And where does that leave us?

It all very well identifying a bit of capacity here and a bit there – to put a train in a timetable, these bits have to link up into a path, with platform slots at origin and destination. But with the Thameslink Project to help at London Bridge, some doubts at Parks Bridge Junction and Metropolitan Junction, and some heroic assumptions about signal overlaps, a 10% increase in trains can be made to sound plausible – in the inner suburban area.

But that is Catch 22! Remember, our starting point was that the extra trains at Charing Cross would have to be capable of working round with outer suburban trains. So what scuppers the potential of ERTMS in this thought-experiment is a commercial desire to have trains that suit the passengers they carry, just the sort of elephant trap in the realities of preparing a timetable that is overlooked by glib talk of 'trains per hour'.

The irony is that if we can depart from clockface timetables, depend upon benefits from Thameslink 2000, and standardise the rolling stock between inner and outer suburban services, running extra trains might be possible without praying ERTMS in aid at all. Only over the double track bottleneck between London Bridge and Metropolitan Junction does success depend upon ERTMS doing something that conventional signalling cannot.

ERTMS Level 2 has valuable attributes in that it can take away the need for lineside signals and thus render maintenance savings feasible. But the capacity argument is another kettle of fish!