

## Designing the Layout

---

### Peter Burns

MBA, BAppSci(Elect), CPEng, FIRSE, FIEAust  
Director, PYB Consulting Pty Ltd

### SUMMARY

Railways exist to meet operational needs.

The classic Signal Education starts with determining how to position some signals on a layout. The laying out of the signals determines the capacity, we are told. But there are quite a few process steps and quite a bit of design which must occur before there are signals to lay out.

The Signal Engineer cannot create a facility that the layout does not support.

Operational requirements start with what the passengers need in order for them to decide to travel on the railway.

Important to passengers will be (a) how many passengers can my layout transport per period of time (capacity), (b) what end-to-end travel times (looked at from the point of view of the passenger, not the train) are offered, and (c) what interchange facility is provided at start, midway (where needed), and at the end of the rail portion of the journey.

“Build a Metro” provides one route to answer such requirements, but there are many other more appropriate types of railways which can meet customer needs better than metros in particular situations.

Which situations?

A Metro is an example of a railway whose strategy is “saturating the roads”. But there is another strategy available called “saturating the nodes”. Anyone who has travelled by rail in Switzerland will have seen this strategy in action. It is also possible to adopt a strategy which is a hybrid of these two approaches.

A practical case example will be provided of what a “saturating nodes” solution would look like for one Australian Capital city.

Returning to the “Metro” concept, the first necessary step beyond the classic configuration is to support a junction. The headway aspects of junctions and the timetable constraints they impose are key to ensuring a metro (one with roads which are actually saturated) can provide its required capacity.

Two simplified illustrative examples will be considered.

The first will be the loop with diamond junction of the type originally present for the Clifton Hill Loop in Melbourne. Headway requirements for the diamond junction will be looked at together with the timetable transit time constraints imposed.

The second example will be the “split centre loop” type of the sort found on the Northern Line in London. Transit time constraints for the centre section will be the main focus here.

These two examples introduce the concept of “resonance” as a required component in timetable planning if capacity is to be achieved without needing to endlessly build new parallel tracks. The question as to why getting the trains to run on time is key to achieving capacity will be answered along the way.

### 1 Introduction

A railway can be modelled as a network comprising “roads” (the tracks available for operation) and “nodes” (stations and other passenger interchange points). In modelling the network in this way the focus is clearly put onto the passenger’s trip rather than the train’s trip.

It is, after all, passengers who have journeys. Trains are just there to facilitate those journeys.

Passenger journeys and train movements are different. This might seem obvious, but is worth re-stating here for the purposes of this paper. A passenger journey generally starts from home. It involves the passenger travelling to the station and boarding the train. The train runs (the train trip). The passenger may then interchange to another train (followed by another train

trip) or to another mode of transport (including walking) to complete their journey to destination. Later on, the passenger returns home using the same or similar mix of railway and other modes, including walking.

The journey starts and ends with the passenger walking. The journey in general involves interchange from one train service to another, or interchange between train and some other transport mode. Walking is always involved when the passenger moves from one segment of the journey to the next.

From the passenger’s perspective, they will use your train based on the attractiveness of the total offering, typically compared with the alternative of using their own car.

I have heard it said that passengers do not like to interchange. And this may be true. Based on my own

experience we can qualify that statement a bit to say "... unless they live in Paris, London, Tokyo, New York ...", a list which includes just about any international city with a train network. If passengers do not interchange on your network, then perhaps what you have is not a "network". All high performing networks manage interchange well.

Whether or not passengers enjoy interchange on our network depends on the interchange experience we provide them with. This is an experience which we as railway engineer have some control over.

Paris, London, Tokyo, New York are international cities where people rely on Metro systems to get around. We can contrast those with, for instance, Switzerland, which is not a city but a whole country. It also provides a high performing train network, but it does not achieve this outcome by providing a Metro. Yet it is still a network and interchange is an essential component which enables it to facilitate passenger journeys.

What is this type of network which is not a Metro?

## 2 Roads and nodes

Railway planning literature from the 1960s points out that there are two distinct pathways available to achieving network optimisation.

The first path, called "saturating the roads" is one familiar to anyone who has ever heard the term "Metro". It involves putting as many trains onto your train track as the signalled headway will support. We have had whole conferences around the considerations needed for providing a metro.

The second available path is called "saturating the nodes". This involves putting as many trains into the interchange station at once as you can. This is not a metro; it is what they do in Switzerland. It is a concept perhaps forgotten in Australia since the 1960s. Particularly for intercity, but even in suburban networks, understanding the operation of nodes and how to optimise them can open up some understanding as to how to improve the operation of our own networks.

I will introduce this topic with a simple example where the "saturating the nodes" concept can be applied at low cost, but with significant benefit.

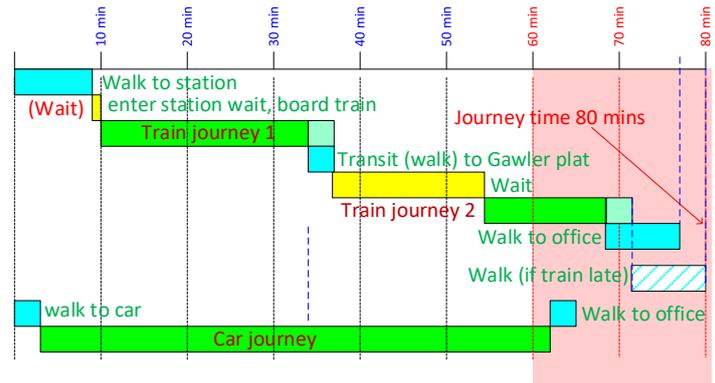
## 3 Dry Creek example

I was speaking some time ago to a manager at the Rail Operation Centre at Dry Creek in Adelaide about why he does not catch the train to work and back home again. I will fictionalise my example a tiny bit.

Dry Creek is on the Gawler line, 10km north of Adelaide. The manager lives at Seacliff, 10km south of Adelaide on the Seaford line. That is a total journey of 20km, something rail networks usually do well.

We can consider his journey to work for a night shift. Both going to work and returning from work are outside the peaks, but still at times when the trains are running.

We can draw up a chart showing the components of the journey to work:



[Figure 1: Chart showing indicative journey components]

Our controller starts his journey by walking to the station. This is Seacliff on the Seaford line. Our controller lives around 500m from the station as the crow flies. This translates to around 700m by the time the actual roads between home and station are negotiated. Ergonomics texts say average comfortable walking speed is around 1.42m/s, but we will assume the 1.33m/s recommended by Transport for London. This allows that we may be amongst other pedestrian traffic and there may be roads to cross. The walk will take around 8.8 minutes at that rate. We will allow 10 minutes for planning purposes.

The trains run on a clockface schedule at this time of evening at the rate of 2 trains per hour (10pm-ish). Our controller can plan to arrive at the station just on time for the train. The train trip to Adelaide then takes 24 minutes.

Then there is the need to change trains to a Gawler line train to travel to Dry Creek, the station which services the ROC.

This involves walking from the arriving train (Platform 2/3) to the departing train (Platform 6/7) then waiting for that train to depart. Altogether, this walk could be around 230m. At our assumed 1.33 m/s this will take 2.7 minutes.

For my reference train, the interval between the Seaford line train arriving in Adelaide and the Gawler line train departing (this is consistent across trains this time of night since both lines are on clockface timetables) is 5 minutes. Thus, he will catch the train with 1 minute to spare.

The journey to Dry Creek takes 14 minutes. The final walk to the office on the far side of the depot is around 700m, taking another 8.8 minutes.

Adding up all those components gives 62 minutes from door to desk. Note the time: 9:30pm to 10:30pm

Our controller can choose to drive instead. According to Google that will take around 55 minutes between garage and depot carpark. By the time he has walked from carpark to desk he has not really saved any time compared with just using the train.

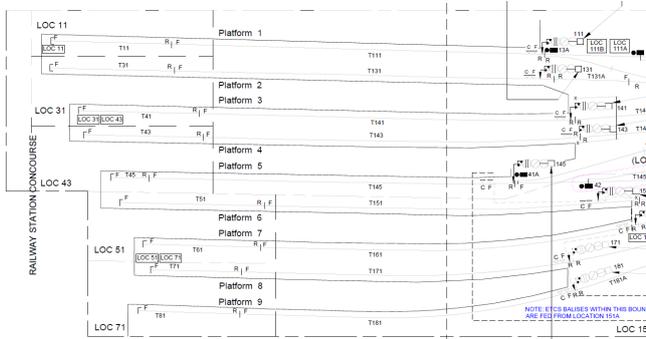
I need to point out that my discussion with the controller was 4 years ago now, but the train operation pattern I have quoted is taken off the published timetable today.

Change for the better has occurred in the interim. Back then I recall the same journey would have taken closer to 90 minutes.

Today’s timetable (for this trip) shows the “saturating the nodes” approach in action, at least on a small scale.

**3.1 How does it work?**

Adelaide suburban rail network comprises four operational groups all of which operate into a terminal station at Adelaide.



[Figure 2: Adelaide Station Platform]

There are a few elements of complexity about the operation of these groups, but these are not important for the purposes of this current discussion.

What is important is that all 4 groups are able to operate substantially independently of each other, each with its own set of tracks and platforms.

According to the classic “saturating the nodes” concept, the trains from each of the 4 groups arrive at Adelaide station simultaneously. They then wait in Adelaide station for sufficient time to allow any interchanging passenger to walk between the furthest trains to change services. All four departing trains then depart simultaneously. This occurs every half hour late at night.

In practice only an approximation to the ideal is achieved. Not all interchange times are just 5 minutes. Minimum change-ends time is around 6 minutes and 10 minute interchange times can also be seen in the timetable at this time of night for other service combinations.

Having each connecting service waiting to depart as the interchanging passengers complete their boarding is at the core of the “saturating the nodes” approach.

**3.2 How did the trains operate before?**

Prior to the possibility of “interchange at Adelaide” being considered important, the timetable still featured two trains per hour, but the wait time at Adelaide station for the connecting service was much more random.

It also varied randomly between versions of the timetable. For longer term planning of travel mode, assuming a full 30 minute wait time was safest.

The resulting 90 minute journey each way did not look nearly so attractive as the 60 minute journey which is actually achievable when the need for interchange is taken into account.

Our controller would likely decide to drive his car to work instead.

**3.3 Room for improvement**

Quoted outcomes above are based on trains performing to timetabled times. On time running in Adelaide is up quite close to 100%, but trains are treated as “on time” when they are running 3 minutes behind timetable.

Similar allowances are made in other suburban rail networks around Australia.

For my sample case, if the up Seaford train arrives in Adelaide 3 minutes late (as it is entitled to do and still be recorded as “on time”), our controller will still take 2.7 minutes to walk to the Gawler line train, but will just see its tail lights as it departs Adelaide station. He will need to wait 30 minutes for the next train.

A couple of approaches are available to control against this level of late running.

The first is to design the train operations to minimise the passenger walk time. This is done by introducing an off peak station use pattern which involves reducing the platform faces utilised to just four (say platforms 4 - 7). Maximum walk times are then maintained to less than 2 minutes and the interchange of trains can still be achieved.

I experienced this approach in actual use in Switzerland on my last visit. In that case passenger interchange from a train running 3 minutes late was achieved by having the arriving and connecting departing train located on either side of an island platform as the required interchange occurred.

**3.4 Measuring times taken**

To get interchange to work effectively requires that the time for passengers to walk from one place to another, and the time for a passenger to get on board a service or disembark from one be known. Signal engineers are perhaps not used to paying attention to passenger walking times when planning operations. It is something which would have been familiar when planning infrastructure even in Australia at various times in the past. The contemporary literature refers to it. As seen in this paper, some design can be done using reasonable assumptions combined with calculation. Many railways send out people with stopwatches to measure those times on the functioning railway in real life. A smaller number of railways feed the results from the people with the stopwatches back to the designers and timetable planners to inform their infrastructure designs and timetables.

That is how the Swiss know that they can tolerate a train running 3 minutes late whilst still achieving a planned passenger interchange.

One alternative (or additional) approach is to control the late running of trains to within better than 3 minutes of timetable. In Japan, where a mix of line saturation and node saturation (often on the same lines) is observed, late running on many lines is constrained to within 1 minute of timetable.

**3.5 Running on time**

This is challenging, but the benefits are not just improved reliability of passenger interchange (ultimately causing increased patronage of the railway), but also increased line capacity.

A network which can constrain train running to within 1 minute of schedule can run significantly more trains per hour and provide significantly better passenger journey time than a network which accepts that trains may acceptably run 3 minutes behind timetable. The benefits of (otherwise) billion dollar projects can be observed being achieved for mere millions. That is the why.

But there are a number of steps across a number of disciplines which must be achieved before the better on time running constraint becomes feasible. Many of these steps do not involve railway signal engineers, however there is one essential element where the signal engineer is right at the centre, and a couple of others where the signal engineer is a key contributor.

These will be introduced in the next section.

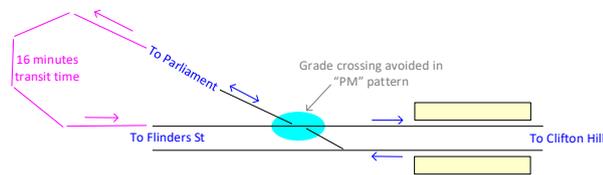
As an initial observation, note that for our Adelaide example, achievement of full classic “node saturation” at Adelaide station is prevented by a “hidden” infrastructure constraint. No Operations Department can implement any high performing operations strategy (eg asking train drivers to maintain his/her train running to within 1 minute of posted timetable) in the face of a blocking infrastructure constraint.

If we require our operations to be high performing we require that the infrastructure constraints blocking the strategy need to be removed first.

Step 1 in this process is to be able to recognise the existence of the “hidden” infrastructure constraints. Step 2 is to know how to remove them. This paper only seeks to introduce those topics.

#### 4 Train running at junctions

To understand the impacts of conflict points on network capacity and on time running, we will consider the case of the simple junction with loop. This is the equivalent to the hydrogen case in chemistry. In real life we consider an approximation of the Clifton Hill loop in Melbourne as it was operated in the early 2000s.



[Figure 3: Junction at Clifton Hill; Loop]

The operating pattern here is that each train from Clifton Hill operates over the junction on its way into the city. It then travels around the city loop, then returns to cross the junction again on its return path towards Clifton Hill.

Melbourne network planners at the time had carried out simulations and identified that the practical impact of trains needing to cross at the junction each way reduced the practical capacity of this loop from 20 trains per hour (when operated such that the junction cross was not required) to 15 trains per hour. The simple solution implemented at the time was to change the operation such that the junction cross was never used. But simply looking at a simulation does not provide a true design perspective of what factors are driving that outcome and what design parameters can be adjusted to reduce or eliminate the negative impacts observed at the junction.

At the other end of the discussion lies the economic basis used by all high capacity railways in the world to justify replacement of flat junctions with flyovers.

Applying a common rule of thumb, to timetable for 20 trains per hour we may design the scheme with 2 minute signal headways. But what do we mean by “headway” for the tracks passing through this junction.

I will provide a definition:

“Headway” at a signal is the minimum time between a train passing that signal unrestricted by the signal and the time when, in the normal course of operations, a second train may pass that signal similarly unrestricted.

I know there are design contractors out there who will look at the 2 minute headway requirement and provide back a design they say complies just for following trains and ignoring the impact of the junction.

For my simple case AS 7711 will be applied but not any particular set of local standards. The parameters used in calculation are found in Appendix A. A scheme can be proposed with 2 minute headways assuming no returning trains across junction.

In practice we know that each up train must return as a down train and use the junction. Practical operational sequence at junction is: Up train, then Down train, then Up train etc.

For this “2 minute headway” layout, the minimum time interval between the two up trains is calculated to be 282s (4.7 minutes). This is rather more than our 120s required design headway. In addition, the arrival window for the return down train is heavily constrained for this interval to be met. Not even 15 trains per hour can be achieved with such a layout.

To see what can be achieved, a layout with minimum signal spacing (refer again to appendix A for this) can be considered. For such a layout the minimum time interval between the two up trains is calculated to be 127s (2.1 minutes), just shy of the 120s specified requirement. Improvements in technology (eg intelligent application of axle counters) can allow the 120s to be achieved. The specifics are beyond the scope of this paper.

The next step in translating from signal headway to practical timetable headway is to consider the possibility of trains running late.

If the return train is timetabled to achieve the available window and no train is more than 1 minute late, a 3 minute timetable interval is practical (20 trains per hour). If, on the other hand, late trains can be up to 2 minutes late, a 4 minute timetable interval (15 trains per hour) only is achievable.

Thus it can be seen that by improving on time running such that trains run within 1 minute of timetable, rather than within 2 minutes, significant improvements in practical timetabled capacity are available.

Practical examples (though not in Australia) can be found where railways in other countries use that method to improve timetabled capacity as a method to avoid the need for costly new infrastructure. But that is not the main focus of this paper.

There is another affect to be observed on our section of railway also.

The time at which the return train arrives as a down train at the junction is constrained by the normal running time around the loop. Inspection of the current Clifton Hill loop timetable puts this time at around 16 minutes, though this seems to allow for a longer-than-minimum stop time at Flinders St.

If we consider a timetable with trains running at 3 minute intervals, there will be a train expected at the up signal at 15 minutes after the up train we are following, then another at 18 minutes. Inspection of the train graph shows that our train returning at the 16 minute mark will find the gap between the up trains. If, on the other hand, the timetabled stop at Flinders St is either reduced or increased, the return train will miss its gap and a delay against specified timetable delay will occur.

We may say that this is just normal and we don't really care about it, but note the comment above that it is precisely by controlling this type of impact that the world's high capacity railways achieve their high capacity. We should care.

Another interesting impact is the observation that we could try to address any issues with operational reliability by reducing the trains in the timetable to 15 per hour, a 4 minute interval. This was where we started. The simulation run suggested we could only expect 15 trains per hour to be achieved.

But notice that if we make this change, we will find that there will be a timetabled up train at 16 minutes after our up train, exactly at the moment our initial train returns as a down train. This timetable would ensure that 100% of up trains are in conflict with 100% of down trains. Late running of services will be unavoidable.

This is an illustration of a resonance in the timetable. Successful timetables are created by understanding and designing to the operational resonances in your rail network. If you want to see a timetable which rings, you only need to look to Switzerland. You will find that management of resonances is a feature of high capacity railways as an enabler to achieving high capacity.

You may say it is not the signal engineer's responsibility to design the timetable. You would be right. It is an interdisciplinary area. The signal design is an essential component together with a number of other interdisciplinary components. Whilst the signal engineer may not alone be able to assure success, with an inappropriate design (see the one presented at the start of this discussion) the signal engineer can by themselves assure failure against the required outcomes.

#### 4.1 Beyond the simple case

Beyond the simple case and beyond the scope of this paper, practical railways can contain multiple at-grade junctions, each with its own resonance characteristics.

With knowledge of required timetable outcomes, it can be possible to design the signalling and operation (using local headway and appropriate regulation points) such that the railway can operate to timetable with trains approaching only green signals. Sometimes incompatible resonances are encountered such that no solution is available without unacceptable sacrifice of capacity (due to the need for too many timetable regulation points or the need to hold the train too long at each one).

When that stage of realisation has been reached, the way out is by removal of resonances. This is done by removing at-grade junctions and replacing them with flyovers. A quick inspection will show that such flyovers are found in all high capacity railways which achieve high capacity.

In the 1930s and 1940s when both Melbourne and Sydney were creating the bases for their high capacity rail networks, flyovers were being built as part of those networks. Burnley and East Camberwell in Melbourne; on the approaches to Central in Sydney.

These schemes involved the track engineer and the signal engineer working together to achieve successful outcomes. We did not get to where we are by operating in silos.

#### 4.2 Flyovers enable interchange

Looking at the flyover schemes referred to above, we can see another set of benefits achieved from the perspective of the passenger.

Passengers in both Sydney and Melbourne approaching the city face a choice as to which direction they wish to negotiate their underground loop. This choice involves a decision about interchange. For the passenger the question starts with "How long will it take to walk to the alternate service and then how long will I have to wait for the train?"

For Sydney Central with its system of flyovers and Richmond on the Burnley group (assisted by the Burnley flyover) the interchange between one type of service and the other occurs on one side or the other of an island platform. Walk time is in the vicinity of 10s and there are few passenger capacity constraints (half the passengers on each train wishing to interchange) to the interchange choice being made.

For Richmond on the Caulfield group (with no support from any flyover) the answer is quite different. In this case interchange requires changing platforms. This means walking along the platform to one of three mid capacity interchange passages, negotiating the passage to the alternate train (amidst the flow of passengers wishing to interchange the other way) then walking along the train to find a carriage with space to enter (carriages closest to the interchange passages fill first from interchange for the obvious reasons).

Rather than taking around 10s, the interchange takes closer to 2 minutes at low traffic times, 3-4 minutes at peak times due to insufficient capacity in the tunnels for free flow of interchanging passengers.

The flyover thus has an added benefit apart from helping the trains to run on time (thus boosting network capacity). It is also an enabler for passengers to effectively negotiate the interchange features which the network provides. In many situations the flyover can be part of a cost effective means of managing capacity constraints whilst helping passengers decide that the rail network is worthwhile to use.

### 5 London Underground and Line Splitting

The material in this paper is informed roughly by the configuration of the London Northern line 10 years ago. There seem to have been a couple of projects there since then and even one in progress as we speak. These projects are no doubt seeking to address some of the issues observed back then.

For the purposes of this paper, the Northern Line is just an example of a line which splits and then reforms further along. If I catch a train at Camden Town to Stockwell, the train can go via Bank, or via Leicester

Square. Either way the journey will take around 24 minutes.



[Figure 4: London Northern Line in 2014]

This type of configuration is an example of a line with another class of junction conflict. On the Northern Line, because the run times for each leg are the same, each train has a slot available to merge into at the other end. The existence of this slot is not affected by timetabled train frequency, but it does require that the run times on each leg are the same. If one leg is actually a bit faster (eg just taking 22 minutes) than the other, the train will be slowed (eg by a longer train stop at a regulating station) so that they end up being the same.

This family of configurations does not always feature run times which are the same on each leg. Often there will be a fast train leg and a slow train leg. However when different legs have different run times, the requirement remains that each train must find a time slot at the other end of the split to allow it to rejoin comfortably.

The existence of this slot depends on the timetabled train interval together with the difference in run times for each leg. These in turn are informed by the signalled headway at the far end junction and the late running constraint time for the railway, as was the case for the simple junction with loop.

For our Northern Line example, if we introduce the “fast underground” on the Bank leg such that run time is just 12 minutes instead of 24 minutes, the infrastructure will still provide smooth operation for timetabled train intervals of 3 minutes. However, if the operator decides to reduce frequency a bit (perhaps hoping to improve reliability) by going to a 4 minute timetable interval, the result will be a minor disaster. The trains merging at the far end of the section will be found constantly in conflict with each other. This will be the case regardless of signalled headway or tightness of on time running. For each layout there will be such forbidden combinations of run time and train interval as was found for the simple junctions.

For the simple junction case we were able to manage this type of issue with a flyover, but that solution is not available here.

For this case a solution can be found by providing a train regulating platform stop on a separate platform face at one end of the loop or the other. This introduces a whole new topic, but for the purposes of this paper I should just say that solving the “problem” by providing an extra platform face can convert a quite negative feature into a very desirable positive feature.

To see how this positive feature can emerge, we return to our “fast” Northern Line example.

The Northern Line is timetabled using some “saturate the roads” assumptions and methods. When mixing fast and slow lines those assumptions and methods can still be good, but the ability to interchange between a fast and a slow service can also be desirable.

People like to change from local trains to expresses and vice versa. The facility for efficient interchange of this type can be important to your railway.

In our simple example, we have found a way (by using transit time and timetable interval to inform our network design) to have a fast train and a slow train on each side of an island platform at the same time. This can be good. Some high performance railways find it to be a desirable feature enabling interchange.

On the other hand, some other high performance railways who do not provide for interchange at the junction find it undesirable.

When laying out the network for your high performance railway, there can be benefit in knowing what the operational requirement in this space is. There can be further benefit in knowing how to design the layout to provide for it.

## 6 Conclusion

It can be seen that the way a railway layout needs to be designed is inextricably linked to how it is intended to be operated.

In general designing the layout is an integrated task involving Operations, Track and Structure design and Signalling.

In collaborating in this task, the Signal Designer should be familiar with design concepts relevant to “saturation of nodes” and “interchange” as well as concepts associated with the perhaps more familiar “saturation on roads” (metro) approach.

For a high performance railway, active signal design with calculation is needed. Simple reliance on operational simulations (a testing tool) is unlikely to suffice by itself.

## 7 Appendix A

The following parameters were used to calculate headways and other aspects of performance for the sample junction conflict example:

Line speed = 65kph, including across points reverse.

Overlap and junction lengths = 200m.

Minimum Signal Spacing = 200m.

Sighting distance = 6s

Point run time = 10s

Track pickup time = 4s

Calculation time = 2s

**AUTHOR**



**Peter Burns**

Director of PYB Consulting.

A Professional Engineer with 40 Years experience providing Railway Signal Engineering to the industry, but currently sadly unemployed.

Commenced designing Level Crossings for the Victorian Railways in 1981. Most recently over 5 years advising DIT and providing Design Reviews for major Signalling Projects in Adelaide.

Have worked on projects in all mainland states in Australia in between.