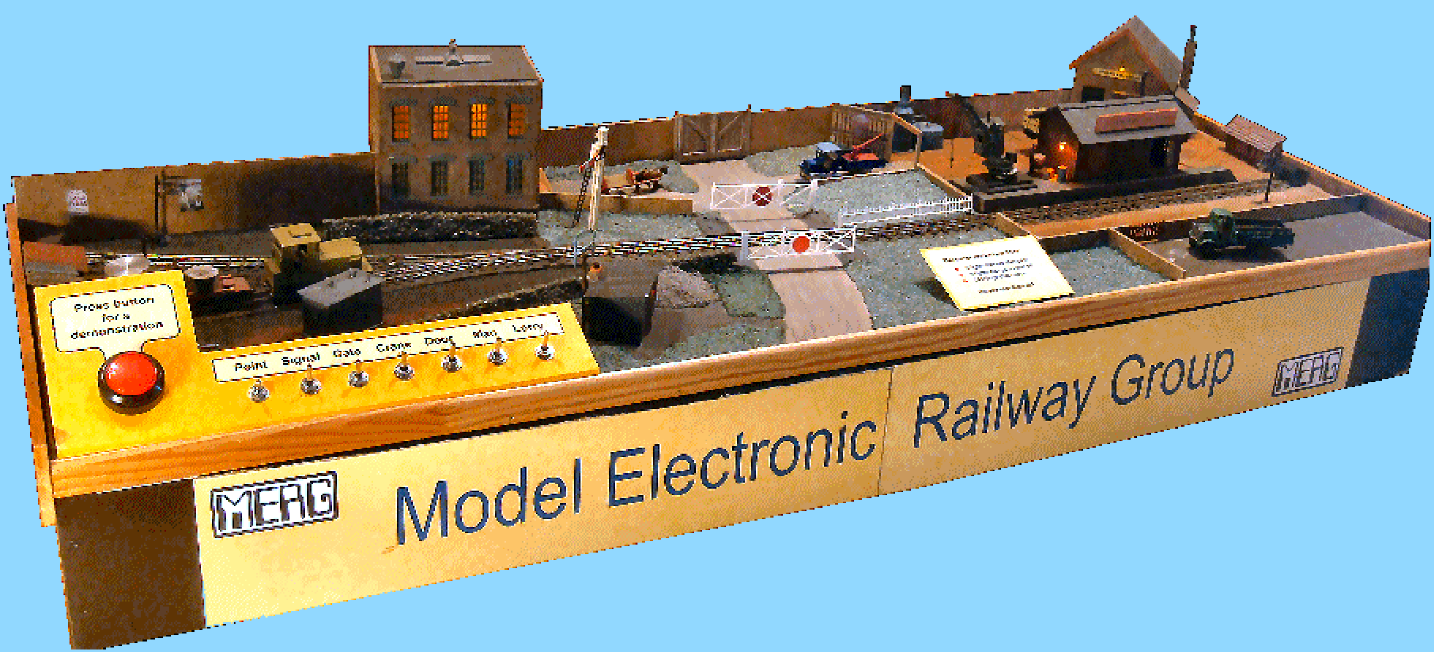
Electronics for Model Railways

Part A



by Davy Dick

Electronics for Model Railways

Part A

By Davy Dick

In memory of Margaret

Introduction

Unless you are running an antique clockwork train set or have a large outdoor layout running steam trains, you are already using electrical and electronic equipment.

For most railway modellers, power is supplied to the loco's motor through the track. With the help of electronics, our layouts can be further improved. How about control of points, signals, lights, turntables, crossing gates and barriers, uncouplers, sound effects – and much more.
We can even detect where trains are on our layouts or automate activities, with or without computers.

At first sight, this may seem complicated – but layout wiring and controls often consist of lots of little individual circuits, each carrying out its own function (e.g. switching a point or lighting an LED).

The chapters try to look at each issue and show how they fit together.

If you like, you can read through from start to finish. However, it is not meant to be read as a book. If you are new to electronics, this book covers a lot of ground.
You are not meant to understand it all after just a quick read through.
You will also find that you already know some parts – and are not currently interested in other parts.

I would suggest that everyone have a read of the first chapter – even if you think you already know all the basics.

> Also, as you are working with a hot soldering iron, knives and cutters, drills and so on , be aware of safety at all times.
> In particular, please read the safety notes on page 50 of Part A and page 72 of Part B.

Hopefully, there is something of interest in here for all railway modellers.

Davy Dick

Model Electronic Railway Group member 1853

Contents

Click on the page numbers to jump to chapters

Part A

Part B

Chapter 1

Basic electronics

Basic electronic terms

Units of measurement

Before we start looking at volts, amps, etc., you should note that there are large variations in electrical and electronic units of measurement.
For example, the steel pylon that brings power to your home might be carrying up to 400,000 Volts, while an LED (light emitting diode) in your layout might be working at around 3 Volts.

Here is a chart that lists the most likely values that you will come across in model railway electronics.

pico	One millionth of a millionth (10^{-12})	Tiny measurement, usually of capacitance
nano	One thousandth of a millionth (10^{-9})	Tiny measurement, usually of capacitance
micro	One millionth (10^{-6})	1μF = one millionth of one Farad
milli	one thousandth (10^{-3})	1 amp = 1,000 milliamps
kilo	1,000 times (10^{3})	1 kilohm = 1,000 ohms
mega	1 million times (10^{6})	1 megohm = 1,000,000 ohms (or 1,000 kilohms)

If you have not come across terms such as 10^3 or 10^{-12} before, its not that difficult.
6×10^3 means 6 multiplied by ten three times, in other words 6 x 10 x 10 x 10 = 6,000
Put another way 10^3 means add three zeros to the end of the number.
Similarly, 6×10^{-3} means 6 divided by ten three times, in other words 6 /10 /10 /10 = 0.006

Examples

An N gauge loco motor might draw a current of 250 milliamps (written as 250mA); this is the same as a quarter of an Amp.
An electric fire might be rated at 1 kilowatt (written as 1kW); this is the same as 1,000 Watts.

Note

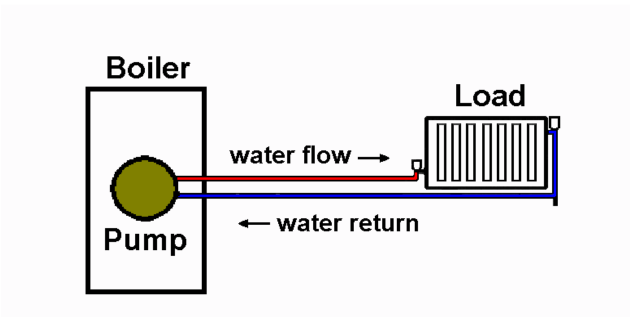
Computers use a different definition of some terms, as they operate in binary (i.e. off and on) states. So, everything is in multiples of 2.
Here, kilo is 1024 (2^{10}) and mega is 1,048,576 (2^{20}).
These multiples are used to described memory size and disc drive and memory stick capacity.

Current

You cannot see electricity. However, you can see, and often feel, its effects.
Electrical current drives motors, lamps, relays, electric heaters, etc.
It produces light, heat, radiation and induction.

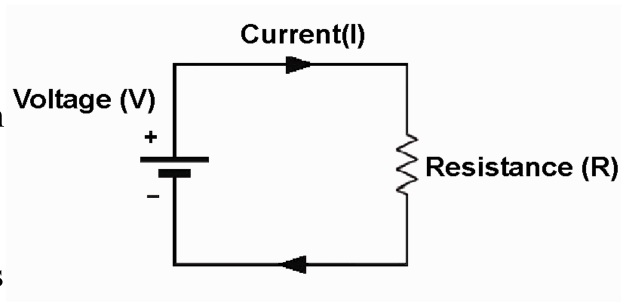
So, what exactly is current?
Current is simply the flow of electrons along a material that allows its passage. Electrical current always flow in a loop and a common analogy is to imagine marbles circulating inside a hula hoop.
You will already have noticed that a battery has two terminals and the mains supply has a live and a neutral. Both have an out and a return.
Of course, something has to push the electrons round the circuit.
An often used comparison is made between the flow of water and the flow of electrons.

This illustration shows a basic central heating system.
The water drops flow round the system, passing through the boiler's pump, the pipes and the radiator. The water is heated before being pumped into the outgoing pipe. The water dissipates much of its energy in the form of heat in the load (the radiator).
The cooler water returns to the boiler to be reheated and pumped back round again.



Compare that system with how a lamp on a bicycle works. As the wheel turns, it spins a dynamo (an electrical 'pump'); the electricity is taken by wires to the load (the lamp bulb) and returned again to the dynamo. The energy created by the dynamo is dissipated in the lamp as light.
In fact, the electricity in your home is also organised this way, except on a much larger scale, with huge power stations and the National Grid distribution network.

Now look at this basic electrical diagram. It shows a battery as the electron 'pump', the wire that carries the current to the load and the return wire that conducts the electrons back to the battery.
In this case, the load can be a lamp, a relay, a locomotive motor – or one of many other things that will convert the electrical energy for its own use.



While the water flow is measured in gallons that are pumped round, the flow of electrons is measured in Amps (short for Amperes), or in milliamps.

Just for the record (there is no need to remember this), one Amp is 6,250,000,000,000,000,000 electrons passing a particular point in one second!

Note

Early pioneers of electricity thought that current flowed from positive to negative potentials. In fact, electrons move from negatively charged to positively charged areas.
To 'save confusion', you will often find texts talking about *'conventional current'* (the long-standing but wrong description) and *'electron flow'* (which is what actually happens). So, all the arrow symbols that you see in diagrams (e.g. diodes, transistors) are showing conventional current flow, not actual electron flow. Just something we have to live with!

Voltage

Let's continue with the central heating analogy for a moment.
The pump is pushing out water at one end, thus creating a shortage of water at the other end.
This difference in water pressure results in the flow of water round the house.
As you would expect, increasing the pump pressure results in more water being pumped round.
With electrical circuits, the power source creates an excess of electrons at one end and a shortage at the other. This is known as the 'potential difference' and this results in the electrons flowing round the circuit. This is true of any two points in a circuit; if one point is at a more negative potential than another point, electron current will attempt to flow towards the less negatively charged point.

Water pressure is measured in 'psi' (pounds per square inch) or 'bars' (one bar is the force required to raise water to a height of 10 metres).
The difference in potential of an electrical power source is measured in Volts (or kV or mV).
The power source can be a battery (as used in your watch, mobile phone, laptop, car, etc.) or a mains supply (us used for your cooker, washing machine, house lights, etc.).
For model railways, the mains supply is too high a voltage and has to be brought down to a lower level. Many loco controllers either use an external power unit, or an internal unit, to reduce the voltage from the 240 volt mains supply to the 12 volts commonly used for model railways. (more later).

Resistance

The amount of water flowing through your central heating pipes depends on the water pressure, as already mentioned. But, it also depends on how easily the water is able to flow. So, for example, if we replaced old wide pipes with new narrow pipes, it would lower the flow. Similarly, if we screwed down the radiator valve, the flow would be lowered.

The flow of electricity will similarly be affected by the material it has to pass through.
Some materials, such as copper, offer only a very low resistance to the passage of electrons and such materials are known as *'conductors'*. Other materials, such as plastic and glass, offer a very high resistance and these are known as *'insulators'*. That is why wire is made from copper inside a PVC sheath. The copper conducts the electrons and the PVC prevents shorts to any other conductors.
Different materials have different resistances and properties, and this is used to good effect

in model railway electronics.   You can deliberately introduce extra resistance into a circuit to reduce the voltage on other components (e.g. to control loco speeds) or limit the current flowing through other components (e.g. when illuminating LEDs).
On other occasions, the resistance is inherent in the materials you use. Your layout wiring, your loco motors, your point actuators, etc. all add their own resistance.

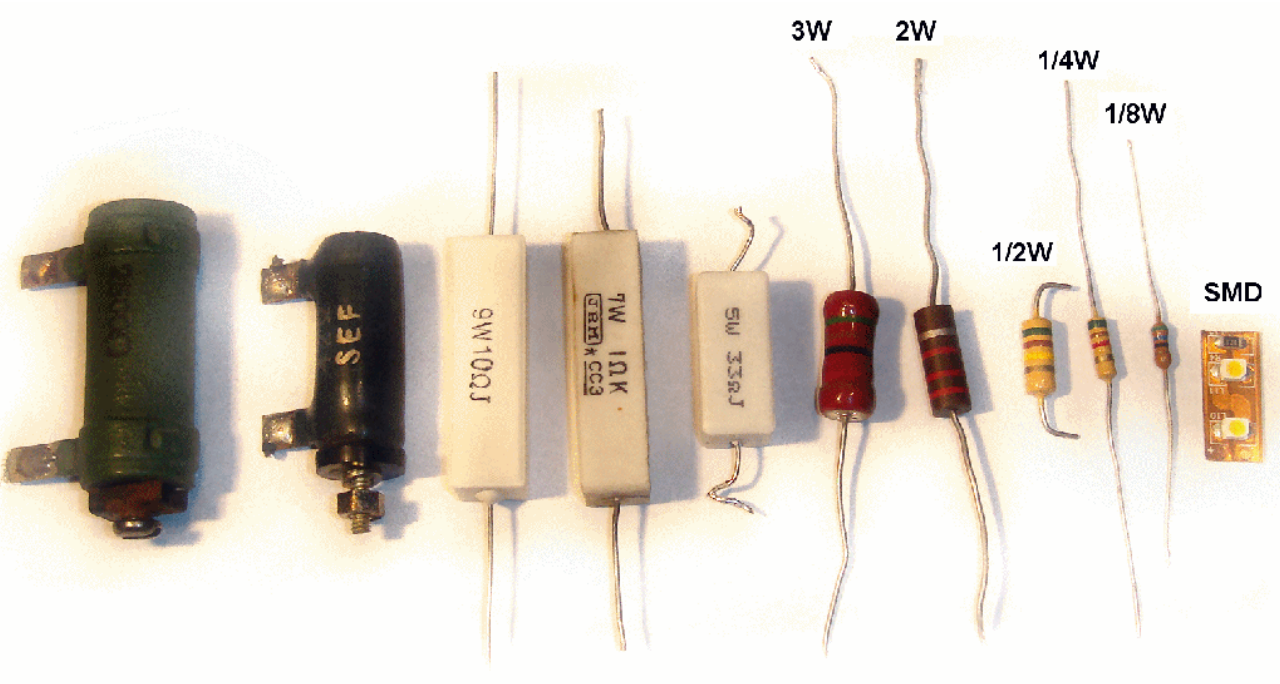
The measurement of resistance is in Ohms (named after Georg Ohm, a German physicist). The symbol for Ohms is Ω, (the Greek letter Omega) but this is seen less often in diagrams and textbooks nowadays.

Although resistances are measured in decimal amounts, the decimal point is not used in texts and diagrams as it is easy to miss.
The letters R, k and M are used, where R represents the decimal point, k represents one thousand an M represent one million.
Here are some example resistor values:

560R is 560 ohms
R47 is 0.47 ohms
47R is 47 ohms
3R9 is 3.9 ohms
2k7 is 2,700 ohms
1M2 is 1.2 megohms

Resistors

These are available in various shapes, sizes and performance.



The main characteristics of resistors are:

Value

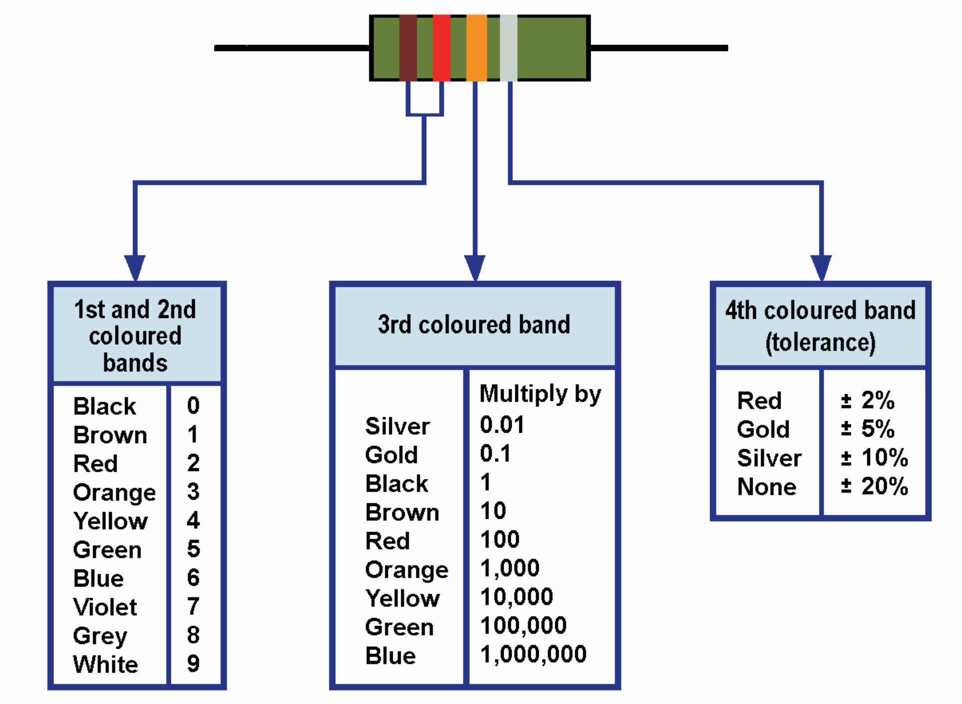
You can buy resistors with values as small as a fraction of an Ohm, up to resistors with millions of Ohms of resistance.
Clearly, it is not possible to manufacture and stock every possible value in between.

Instead, resistors are manufactured in limited fixed ranges of values. These were the standards set up by the now defunct Electronic Industries Alliance, commonly known as *'preferred values'* and these are shown in the table below.

Preferred series	Values available in that range
E6	1.0, 1.5, 2.2, 3.3, 4.7, 6.8
E12	1.0, 1.2, 1.5, 1.8, 2.2, 2.7, 3.3, 3.9, 4.7, 5.6, 6.8, 8.2
E24	1.0, 1.1, 1.2, 1.3, 1.5, 1.6, 1.8, 2.0, 2.2, 2.4, 2.7, 3.0, 3.3, 3.6, 3.9, 4.3, 4.7, 5.1, 5.6, 6.2, 6.8, 7.5, 8.2, 9.1

Colour coding

Larger-sized resistors have their values printed on them. In most cases, the values are deduced from the coloured marking round the resistor body.
Most resistors use a numbering code that uses four coloured bands.



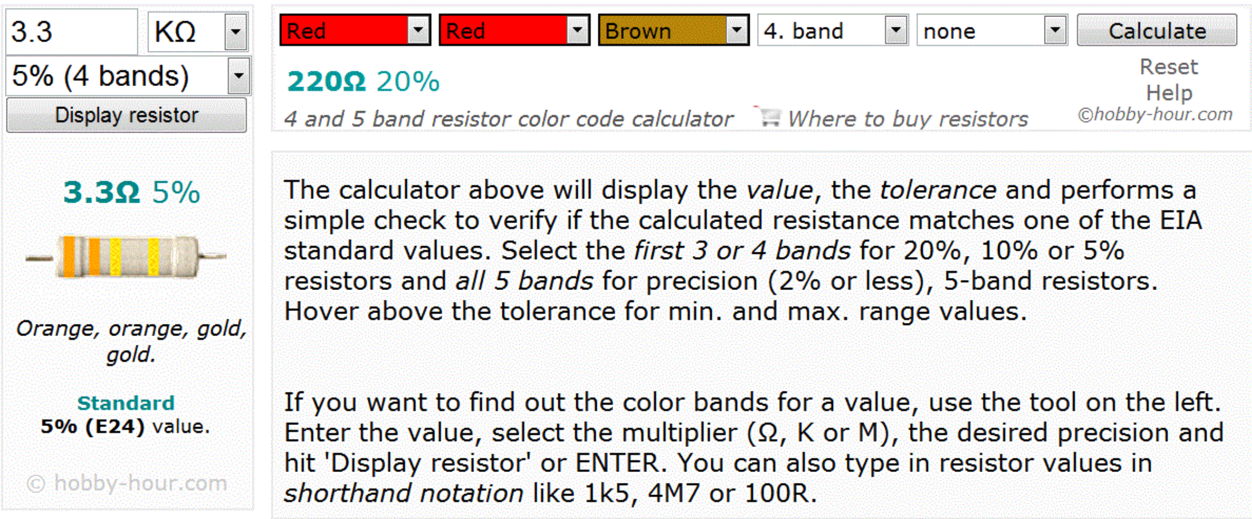
For example, the resistor shown in the illustration can have its value calculated thus:

1st band is brown = 1
2nd band is red = 2
3rd band is orange = multiply by 1,000 (simply add three zeros)

So the value is 12000 ohms, commonly known as a 12k resistor.
The fourth band is silver, which indicates that the resistor tolerance is + or – 10% (the actual resistor value can be greater or lower then 12k by as much as 10%).
If you prefer, you can use one of the many free software calculators that you will find on the internet.
This one can be found at

www.hobby-hour.com/electronics/resistorcalculator.php

Resistor color code calculator



Another calculator can be downloaded from
www.merg.org.uk/merg_resources/resistors.php

Wattage

The more current that flows through a resistor, the more it heats up (that's how your electric fire or soldering iron works). Of course, we don't use resistors as heating elements – its just a by-product of using them to reduce voltage or current in a circuit. In fact, excessive heat will eventually destroy a resistor. The wattage rating of a resistor tells you how much power it can dissipate over a long period of time without being damaged.
Resistors fitted to electronic circuits where only tiny signals are being carried may use 1/8th of a Watt, while heavy current usage requires higher wattage handling (e.g. a car lamp bulb or soldering iron may be 25W).
Most electronic circuits probably use resistors with ¼W or ½W ratings.

Tolerance

Resistors are manufactured in their millions. To keep their price down, they are not exactly precise, with their deviation between the actual resistance value and the stated value being known as its *'tolerance'*. Of course, you can buy resistors with very low tolerances – at a higher price. It depends on what you want to use it for. Most resistors are produced with tolerances of either ± 2%, ± 5%, ± 10%, or ± 20%. Close tolerance resistors are used where precision is important, such as transmission equipment (to work at the correct frequency) or measuring equipment, such as your multimeter. In most situations, the cheaper wider tolerance resistors are commonly used.
E6 series resistors have a 20% tolerance, while E12 series resistors have a 10% tolerance or occasionally 5%. E24 series resistors are mostly 5% but some are available in 2%.

Material

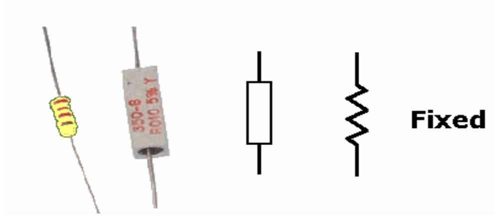
Most model railway applications use carbon resistors. They are made up of a rod of carbon with wire connections at each end. If you broke one in two, the inside would look like a pencil lead. They are the most common and are cheap and readily available.

Other types are produced for specialist purposes. For example, resistors made from metal film (nickel chromium) are used in audio amplifiers as they produce less noise than carbon resistors, and are less affected by heat and voltage changes
Where very high currents are being passed through a resistor, wire wound types are used. They are constructed in a similar way to the element that winds round a bar of an electric fire. They are often used for high precision measuring or where high power ratings are required (some handle up to 300W). Early locomotive controllers used large wirewound resistors to control loco speeds.
Surface mount resistors are now the standard for manufactured boards and are becoming more used for hobby kits. They are tiny, with no leads, and require practice to solder to boards.

Types

Fixed

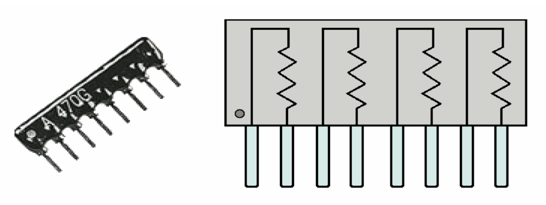
Most resistors are of the *'fixed'* type. That means that they are manufactured to a specific value and that value cannot be varied by the user. It is just a component with two leads.



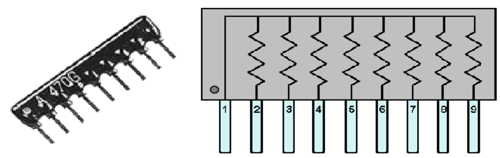
The illustration shows a couple of fixed resistors (one is carbon and the other wirewound).
The symbol for a fixed resistor that you see in electronic diagrams is either a zig-zag sawtooth-like shape or mostly commonly a simple rectangle.
Single resistors are not polarised. That means that they work whichever way they are connected. The only exception is for commoned resistor arrays mentioned below.

Arrays

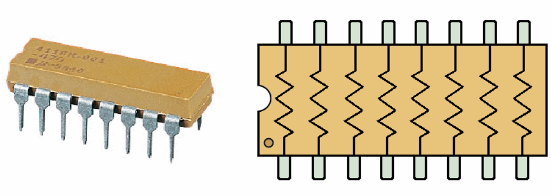
Most circuits use a collection of individual resistors in their construction. Sometimes, however, there are occasions when a collection of the same value are used in close proximity, usually for a common purpose (e.g. connecting to a set of LEDs or to a group of switches). In these situations, it is neater and saves space to use resistor arrays (sometimes shortened to *'resnets'*)
These illustrations show a type known as SIL resistor networks . SIL means single in line, a single row of pins.
As you will notice, some resnets are *'isolated'*.; each resistor has its one set of pins and has no connection to any other resistor.



The other, known as *'commoned'*, has one end of all resistors taken to a single output pin.
The dot on the resnet case indicates pin 1, which is the common pin for commoned resnets.



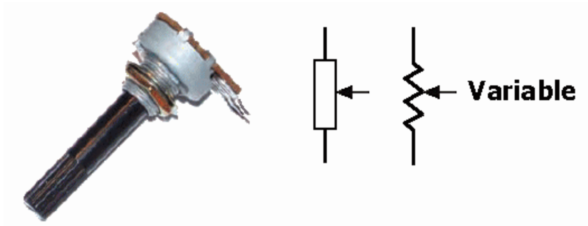
Another array is the DIL (dual inline) with two rows of pins. All eight resistors in the network will have the same value of resistance and there is no internal connection between any resistor.



Variable

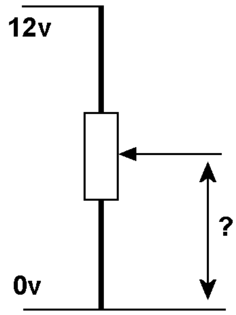
You will already be familiar with variable resistors as they can be found as volume controls in many audio systems and as a speed control for a model railway controller.

They are three-terminal devices and are often referred to as *'potentiometers'* (sometime shortened to *'pots'*)



Some are altered by turning a knob while others use a slider to change value. In both cases, a wiper slides over the resistive material and the result is taken out to a third terminal.

The symbol for a variable resistor is similar to a fixed resistor, with an arrowhead on the side to show that its value can be varied.



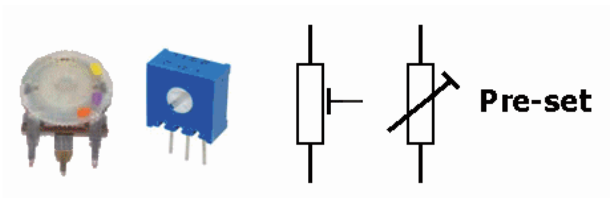
Take this example of a *'linear potentiometer'* with a value of 100k. That means that there will always be 100k across its outer terminals. Its the value seen at the wiper terminal that is changed when the knob or slider is varied.

In our example, the outer terminals have 12V across them.

As the inner material has a consistent coating, the voltage is spread evenly across the entire inner surface. So, the coating at the top terminal will be at 12V, while the coating at the bottom terminal will be at 0V. It follows that the middle of the coating should be at 6V. If we turn the knob (or move the slider) to mid position, the wiper should be at 6V. So, by moving the wiper, we can control the voltage that we use.

Note

There is another type of potentiometer known as the *'logarithmic'* type. They are specially made for audio purposes, as our hearing levels are logarithmic. Make sure you always use a linear type for model railway projects.



This illustration shows a couple of examples of *'preset potentiometers'*. These are much smaller than standard pots and have no external rod or slider to alter values. Instead, they have a small slot that allows the wiper's output to be set by a screwdriver.

These are common on printed circuit boards where there is no need for constant adjustment. The wanted value is set and then left alone.

The symbols used are slightly different from variable resistors.

Ohms Law

Lets have one last analogy with central heating.

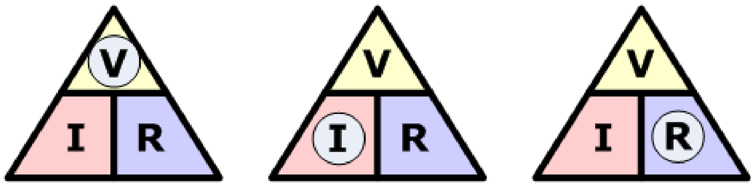
The amount of water flowing round your system depends on the amount of pressure being provided by the pump, and the stifling of the flow from valves, narrow pipes and clogged pipes. The three are interrelated. If you increase the pump pressure, more water flows round. If you turn down radiator valves less water flows in that radiator, and so on.

And so it is with electrical current.

The current flowing in a circuit is directly proportional to the voltage (greater voltage means more current) and inversely proportional to the circuit's resistance (greater resistance means less current).

So, if you know any of two properties, you can calculate the third. The calculations use V to represent voltage, I to represent current and R to represent resistance.

The illustration shows a simple way to remember the three formulae.



$V = I \times R$ $\quad I = \frac{V}{R}$ $\quad R = \frac{V}{I}$

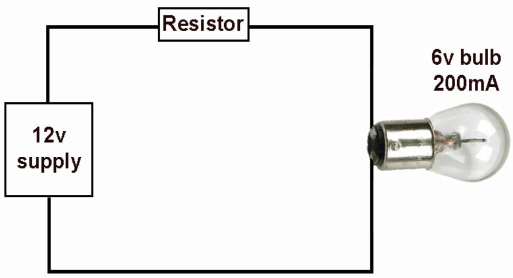
From the middle pyramid, we can deduce that I = V/R (i.e. that current flow *increases* with increased voltage but *reduces* with increased resistance).

These calculations are known as *'Ohms Law'* (remember Georg?).

Some examples

- If you place a12V battery across a 6V bulb, you get double the current. The excessive heat burns out the bulb filament.
- If you have a short in a circuit, you get as much current as the supply can produce – until the fuse blows - or worse.
- If you have a bad connection or use wire that is too thin, you have introduced excessive resistance. The diminished current flow will probably result in the device working badly, or not at all.

Some calculations



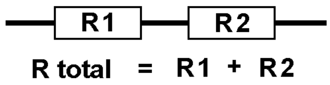
You have a 6V bulb but your supply is 12V. The bulb says it works on 200mA (0.2A).
We know that if we add extra resistance, the current flow will be reduced. We also know that the current flowing through the resistor we added will now drop some of the voltage across it.
So how do we calculate the value for the resistor?
Well, the resistance of the bulb filament when hot must be R=V/I = 6/0.2 = 30 ohms.
We want 6V to be dropped across the resistor, to leave just 6V across the bulb.
So the value of the resistor can be calculated thus: R=V/I = 6/0.2 = 30 ohms.
You could probably have guessed it would be the same value since we wanted to halve the voltage.

How about when we have a 4.8V 300mA bulb?
The calculations are only slightly more complicated.
We want to drop 7.2V across the resistor (12V-4.8V).
So, the resistance value must be R=V/I = 7.2/0.3 = 24 ohms.

> **Note**
> In these examples, we deliberately introduced added resistance to achieve our needs. In model railways, the introduction of unintended resistance is a source of problems. When your loco runs slower in one section of track compared to another; when your solenoid doesn't quite move the point as snappily as it used to; when the lights dim unexpectedly – these are all symptoms of unwanted resistance. This is covered in detail in the chapter on layout wiring.

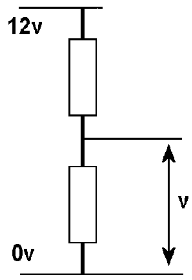
Series circuits

When you have more than one resistance, they can be in series with each other, or be connected in parallel.
The illustration shows a simple series circuit.
You simply add the resistance values together to get the total resistance between the two ends.



As the resistances are in series with each other; the same amount of current will pass through all the resistances – no matter what value of resistance they have.
But, according to Ohm's Law, the same current flowing through different resistances results in different voltages across each resistance.
This is used to good effect in many electronic circuits.

The illustration shows two resistors in series across a 12 volt supply.
The 12 volts is distributed across the resistors depending on their values, hence its name as a *'voltage divider'*. The resulting voltage at their junction depends on the ratio between the two resistors.
If both resistances were equal, there would be 6Vpotential between the junction of these two resistors and 0V potential.
All voltages are relative to the 0V line.

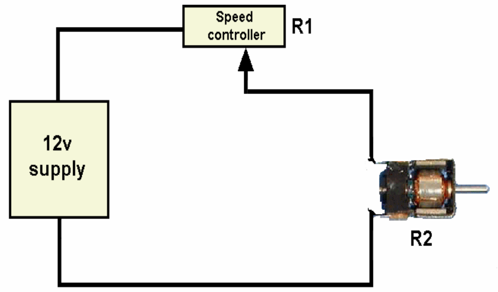


Other examples include:

10k on top, 1k on bottom results in 1.091V at the junction.
47k on top, 12k on bottom results in 2.441V.
47k on top, 33k on bottom results in 4.95V.

As always, there is an online calculator to make life easier.
Try: www.raltron.com/cust/tools/voltage_divider.asp

The voltage divider effect was widely used in early loco speed controllers such as the H&M models. Like many others, it had a large wirewound variable resistor. As the rheostats's knob was turned, the wiper rubbed across the windings and varied the amount of resistance placed in series with the locomotive's electric motor.
Turning the knob clockwise decreased the added resistance, increasing the voltage across the motor, thus increasing the current flow and the motor speed.



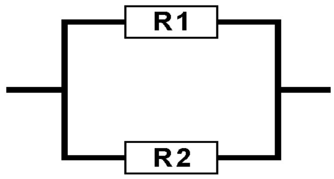
Nowadays, we use more sophisticated ways of controlling loco speeds (see later).

Tip :
If you need a value that you don't have in your spares box, make one up from other resistors in series. For example, wiring a 820 ohm and a 180 resistor produces a 1k resistor.

Parallel circuits

Not all circuits are wired in series.
Your house lights, domestic appliances, PC, etc. are all wired in parallel with each other.
They all receive the same 240V supply but each consume different amounts of current from the common supply.

In model railways, we use parallel circuits when we operate locos, points, lights, etc.
The illustration shows how to calculate the resulting resistance from two resistors wired in parallel.



$$\frac{1}{R\ total} = \frac{1}{R1} + \frac{1}{R2}$$

If there are only two resistors, the equation can be simplified to

$$R\ total = \frac{R1 \times R2}{R1 + R2}$$

To avoid awkward calculations, specially if you have many resistors, you can use on-line calculators such as:

www.sengpielaudio.com/calculator-paralresist.htm

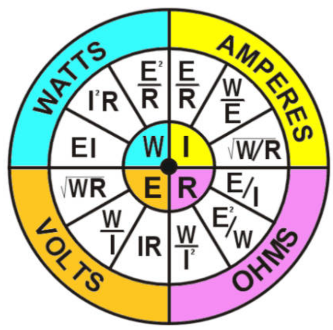
Tip : If you want a value that you don't have in your spares box, you can make one up from other resistors in parallel. For example, wiring two 1000 ohm resistors in parallel will give you a 500 ohm resistor.

Other examples include:

560ohms and 470ohms give 255 ohms.
100ohms and 150ohms give 60 ohms (or you could use 33ohms and 27ohms in series).

Wattage

We mentioned wattage earlier but how do we calculate the power being used by any given device?
You can calculate the amount of power if you know two of the three electrical values – voltage, current and resistance.
See the formulae in the wheel (E, like V, is used to indicate voltage).

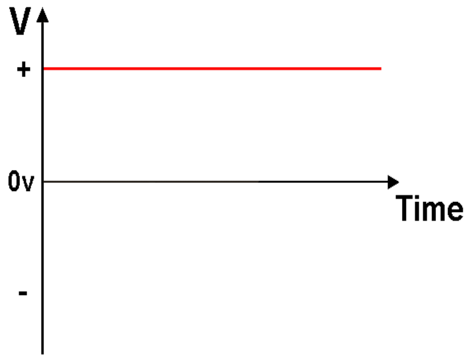


So, using W = E x I, we can make some calculations like those below:

If your loco uses 1Amp at 12 volts, then is consumes 12 watts.
A 240V light bulb uses 0.417 amps (470millamps) to be a 100watt bulb.
A 2 kilowatt kettle uses 8.333 amps (240 x 8.333 = 2000)

What is DC

DC stand for *'direct current'*.
A battery is a typical store of DC energy.   It has two terminals, with a potential difference between them.   One terminal (the -ve) has a surplus of electrons, compared to the other (+ve).   Place it in a device, switch the device on and the electrons start to flow.
All battery powered devices, such as your watch, mobile phone, iPod, etc. depend on a constant flow of direct current during their operation.
The graph shows the voltage unchanging over time.



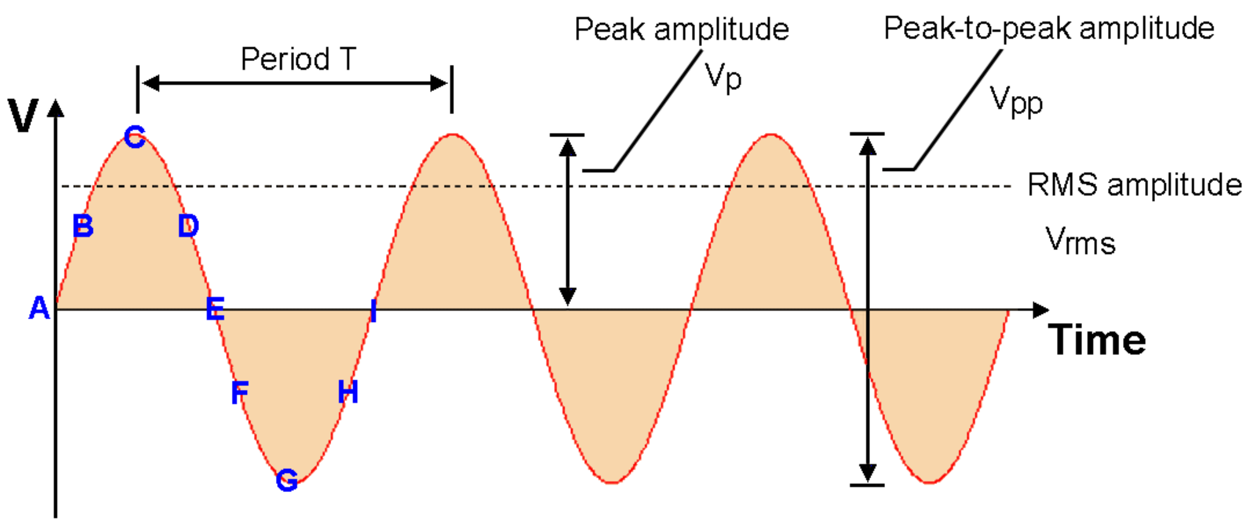
Electrons will always flow from the negative terminal to the positive terminal.
Its true that you can reverse the wires from a battery pack to a loco and train will go in the opposite direction.   However, although the current is flowing through the motor in the opposite direction, the current itself is still flowing in the same direction between the battery terminals. The voltage from a DC supply may vary and the current flow may vary (either intentionally or through a battery weakening) but its polarity never reverses.

DC is used inside nearly all electronic apparatus, including industrial, domestic and hobby devices.
DC is used extensively in model railways for loco motors, LED lights, most point motors, DCC decoders and more.
Even devices that are connected to the mains use DC for their internal components (transistors, microchips, logic circuits, etc.).   The incoming mains is converted from AC to DC as will be shown later.

What is AC

With DC, the current flow is unidirectional; it always flows in one direction.
Alternating current (AC) is completely different.
If you could look at the two pins of an AC supply, you would see that the voltage is constantly changing, getting larger, then smaller then changing direction.
The image shows a *'figure of 8'* mains plug and below is a typical AC waveform, like the sine wave type you receive on your mains supply.

The illustration shows three complete cycles of the waveform as time passes and the stages of supplying current to a circuit during a single time cycle are:



- At the beginning **A**, there is no voltage difference between the pins, therefore no current will flow.
- As time passes, see **B**, pin 1 has a higher voltage than pin 2, so current will flow out of pin 1, round the circuit and back to pin 2.
- Later, at point **C,** pin 1 is at its greatest compared to 2, with maximum current flowing round the circuit.
- At point **D,** the voltage difference has started to decrease, with reduced current flow.
- At point **E**, there is again no voltage difference between the pins and no current flows.
- At point **F,** the voltage difference is ***reversed***. Pin 2 is now at a higher voltage than pin 1, Current will again flow in the circuit – but in the ***opposite*** direction.
- At point **G,** pin 2 is at its greatest compared to 1, with maximum current flowing in the circuit from pin 2 to pin 1
- The AC mains supply then reduces again, see **H**.
- At point **I**, there is again no voltage difference between the pins and no current flow.
- The same cycle starts all over again... and again... and again...

You may wonder why such a complicated system is used instead of normal DC.
AC is more efficient to generate and distribute and cheaper and more reliable. AC in the UK is supplied at 240 volts AC as the standard domestic mains supply. It is used directly, without any modification, to power your cooker, electric fire, incandescent light bulbs, etc.

Frequency

The time taken to complete this cycle is measured in time as its *'period'*.
In the domestic supply for the UK, Australia, New Zealand, most of Europe, this cycle happens 50 times every second (60 times per second for the USA, Canada and parts of South America).
The number of times the polarity reverses is known as the *'frequency'* and is measured in *'cycles per second'* (CPS) or *'Hertz'* (Hz). So the UK mains frequency operates at 50Hz. That means that the mains goes through a complete cycle every 1/50th of a second – i.e. 20milliseconds. Consequently, the frequency = 1/ period.

Voltage

The total movement from zero volts to full voltage in any direction is known as the *'peak amplitude'*, with the *'peak-to-peak amplitude'* measuring the total swing.

You will have noticed that AC power does not supply a constant voltage and current; it is constantly varying. We need some way to measure the effectiveness of the power. This is done by measuring the heating effect of AC compared to DC.
Obviously, 200 volts of AC would not produce the same heating effect as 200 volts DC. In fact, the equivalent heating effect of a sinusoidal waveform compared to DC is 0.707 of its peak value. This value is known as the *'root mean square'* or effective value and just written as RMS.

When you read that the UK household electricity supply is 240V, it is really saying that it is 240V RMS. It is providing the same heating value as if it was 240V DC.

Note

If we measure the mains voltage with a multimeter set on the AC range, we will read 240V, as the meter reads RMS values.
If we measure the same mains supply with an oscilloscope (which displays the actual waveform), we will see that the UK mains has a peak amplitude of 340V and a peak-to-peak amplitude of 680V.

The RMS voltage is 120V in the USA and Canada and 230V for most of Europe, Australia and New Zealand.
The UK voltage is also meant to 'harmonise' with EU standards but still supplies at 240 (being within the 230V ±10% rule).

As a quick method, multiply the RMS value by 1.414 to get the peak value, or multiply the peak value by 0.707 to get the RMS value.
So, for example, a 15V AC socket on the rear of your loco controller has a peak voltage of 21.21V.

The difference between RMS (the value printed on the label of a supply) and the peak value becomes important when we start to look at building a regulated power supply later in this chapter.

Analogue and digital

Electronic circuits operate with either analogue or digital signals and both can be found in use in model railways.

Analogue

Also known as 'analog' in the US
A useful definition of an analogue signal is:

"A nominally continuous electrical signal that varies in amplitude or frequency"

Consider, for example, a simple loco controller with a potentiometer to control the output voltage. Turning the knob will result in an output between 0V and 12V – and every possible value in between. Depending on the pot's physical construction, and the steadiness of your hand, you could have an infinite number of outputs values.
The *'nominally continuous'* part means that after you set the loco speed, the output voltage will stay at that same level forever – or at least until you change the speed again.
Analogue technology is also widely used in audio devices such as a microphone, a tape recorder, a DJ's turntable, loudspeakers and headphones.

Digital

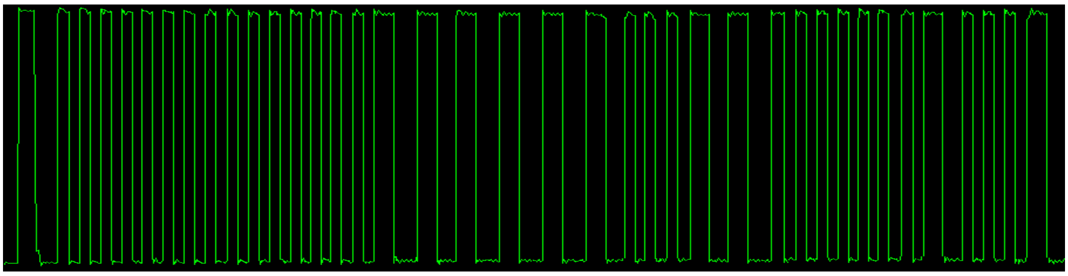
Digital technology takes a signal and converts it into a series of high and low voltages.
It has been around for a long time (Samuel Morse invented the Morse Code back in 1837) but has really taken off in recent decades. The supreme example must be computers, where every word, every picture, every sound or video clip is reduced to a long set of zeros and ones (i.e. OFFs and ONs).
Other domestic examples include digital radio, digital TV, CD players, etc.

Sometimes, we even have both at the same time. If you have broadband internet in your home, you probably are using ADSL technology (digital) on the same incoming telephone as your normal telephone calls (analogue as far as the exchange).

In model railways, digital signals are increasingly being used to enhance layouts.
Some examples are:

- Operating points using servos.
- Pulse width modulation (PWM) controllers to run locos.
- DCC controllers and locos fitted with digital decoders.
- The CBUS system to control accessories and traction.

The image below is the actual digital waveform from a DCC controller.



As you can see, the commands sent to the loco use a series of pulses and it is the pattern and length of the pulses that contain the information.

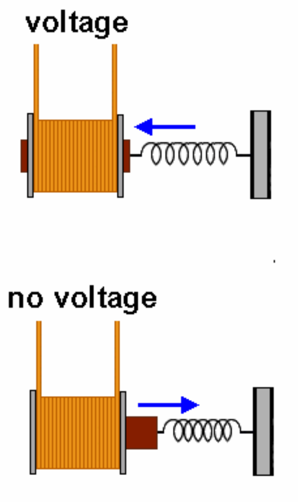
Electro-magnetism

If you pass current through a wire, magnetic field lines are formed around it.
If the wire is formed into a coil, this effect is magnified and concentrated; you have created an electro-magnet with similar properties to a permanent magnet.
This is put to use in a huge number of appliances including shavers, doorbells, buzzers, etc.
In model railways, it was the basis of the majority of *'solenoid'* type electrically-operated points and for magnetic uncouplers.

Let's look at some examples of electro-magnetism in action.

How a solenoid works

When the voltage is applied to the ends of the coil, the current flowing through the coil creates a magnetic field that pulls the metal rod into the coil. When the voltage is removed, the magnetic field dissipates and the spring pulls the rod back out of the coil.

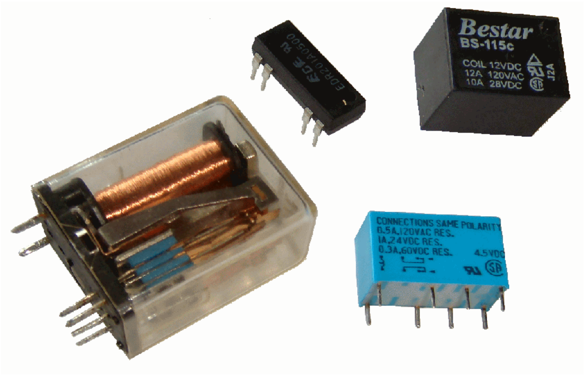


Although a very simple mechanism, it can be found in many everyday appliances such as washing machines, dishwashers, car starters, door locks, electric typewriters, circuit breakers, chime type doorbells, vending machines, automatic tellers, etc.

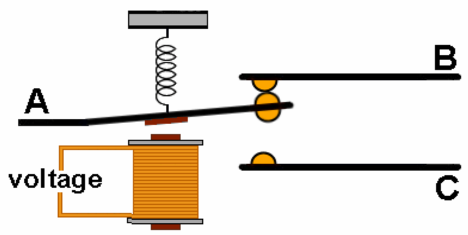
A variation is to have two coils – one that pulls the rod in one direction and another coil that pulls the rod in the opposite direction.
For model railways, it is the basic technique used in Seep and Hornby point motors (see later).

How a relay works

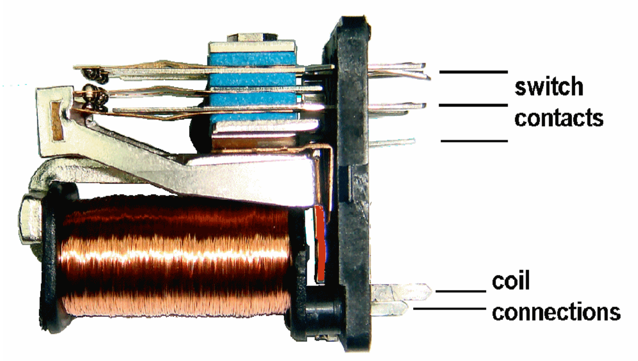
If the wire is wound round a metal core, the magnetic field is strengthened and this is the most effective way to create electromagnets.
A relay is just a switch that is operated by an electromagnet instead of your finger.
As the image shows, they are available in a range of different shapes and sizes, with different specifications for voltage and current handling and the number of switch contacts.



The illustration shows the principle of operation.
A voltage is applied to the ends of the coil and the current through the coil creates a magnetic field that attracts the small block of metal (called the 'armature') towards the coil. The armature is attached to the moving part of the switch and results in an electrical connection between switch contacts A and C.



Removing the voltage collapses the magnetic field and the moving contact is pulled back to its original position by the spring. If the relay is of a changeover type, there will be another contact (B in the illustration). In that case, contacts A and B are made when the relay is not energised.



This image shows a close-up of one of the relays. The moving part is L-shaped, with the longer section being used to move the switch contacts. The short section is close to the coil (shaded red just for the sake of clarity).
When the coil is powered, the shorter (red) section is attracted towards the coil and this forces the longer section to pivot upwards. This movement breaks the connection between the middle and lower contacts and the middle contacts now touch the upper contacts.
While some relays use a spring to return the armature when power is removed, this relay has its spring built in to the contact blades.

A main feature of relays is that there is complete electrical separation of input and output. This provides a number of advantages:

- A small current to energise the coil can result in large currents (say 5A) being switched.
- A small voltage across the coil (e.g. 5V) can result in higher voltages of 12V or even 240V being switched
- A relay can switch AC or DC.
- A relay can switch audio or data.
- A relay can have multiple contacts that switch at the same time.

The main disadvantages of relays are:

- Being mechanical, they can be quite large compared to other electronic components.
- Like all moving parts, they could be affected by dirt, rust, moisture, etc.
- They are slower to operate than electronic components.
- Some need more current than can be supplied directly be a module. This often means that they need extra components to interface relays to electronic circuits.

Relays are relatively cheap and are easily understood.
Although they don't need any special handling precautions (they are not prone to damage from static for example), they should always be used within their specifications.
The relay with the blue case shown in the earlier image is designed to be operated at 4.5V. While it will still operate happily at 5V, it should not be operated at a greatly higher voltage; the coil will overheat and burn out. On the other hand, the relay in the black cube is designed to operate at 12V. While it may still operate at a slightly lower voltage, too low a voltage will lead to unreliable operation.

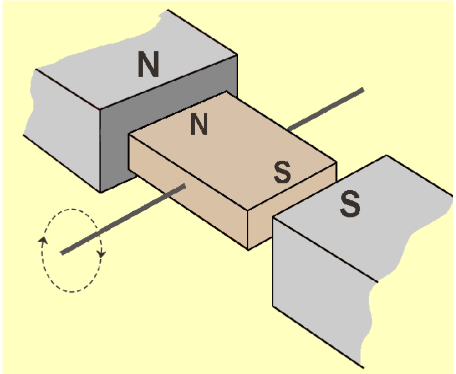
It is also best to avoid switching higher voltages or currents than the relay is rated for. This can easily lead to pitting of the contacts or even welding the contacts together.

Operating relays within their ratings usually provides many years of trouble-free operation.

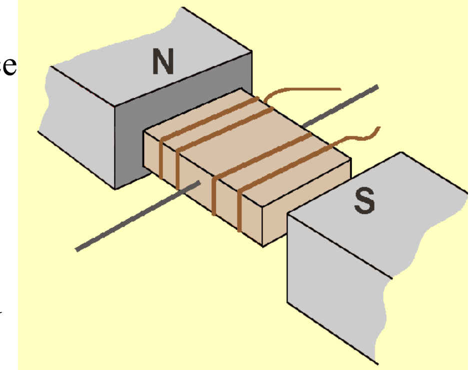
In model railways, relays are used to great effect for section switching and frog switching (more later).

How a motor works

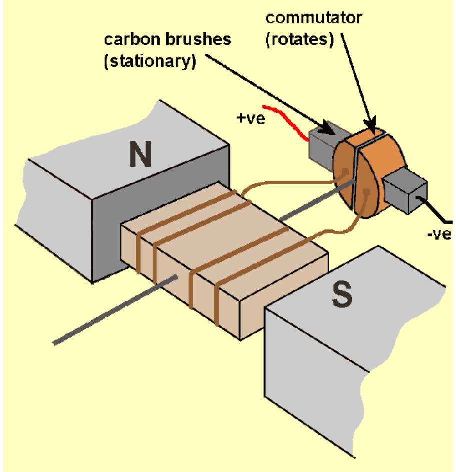
The most common type of motor used in model railways is the *'Brushed DC motor'*. Although AC motors exist, they are mostly run at mains voltage for use in domestic and industrial appliances. A huge amount of research has gone into motor design but we can see how a simple DC motor works without delving too deeply into the physics of it all.
These illustrations take us through the stages of creating a DC electric motor.



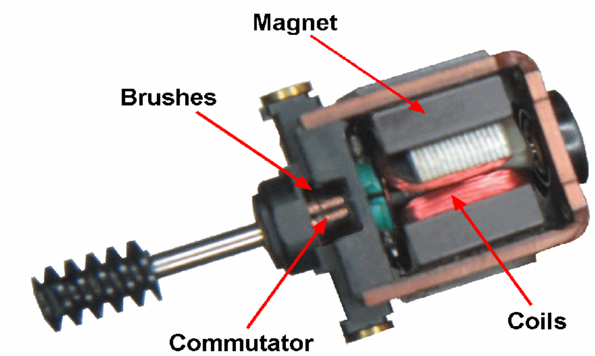
Everyone knows that if you try to place two magnets together with their North poles facing each other, the magnets will repel each. Similarly, if two South pole faces are placed close to each other. Opposites attract and likes repel.
In the first illustration, a magnet is fitted on a rotatable shaft and positioned between a pair of magnets.
Since the two magnet's faces have opposing poles, the middle magnet will be repelled and make a half rotation to align opposite poles.



The second illustration shows the middle magnet being replaced with an electromagnet (a coil wound round a piece of metal), usually called the *'rotor'*
When power is applied to the coil, it creates an electromagnet whose magnetic poles are identical to the magnet in the first illustration (i.e. N pole on the left, S pole on the right). This magnetic field results in the same 180° rotation (the N pole now has moved to the right-hand magnet). If the polarity on the coil is now reversed, the electromagnetic magnetic field would produce a magnetic field identical to the first field, the N pole is once again on the left-side – and the rotor will again spin 180°.



Since the rotor is continually turning, the power to the coil is fed as shown in the third illustration. The coil is wired to separate sections of a *'commutator'* which rotates along with the rotor. It has conductive sides, usually copper, and a bar of carbon is held in touch with each of the copper areas. These brushes are connected to the DC supply.



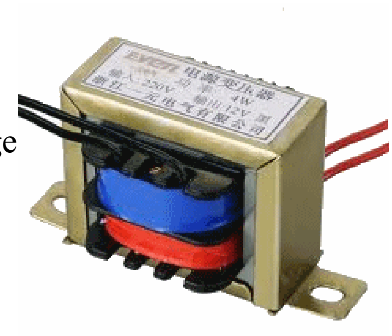
So, as the rotor spins, the voltage being fed out of the commutator into the coil is reversing regularly. This ensures continual rotation while the power is switched on.

The image shows an actual DC motor.
A motor may have multiple windings (3-pole, 5-pole or 7-pole) and multiple segments on the commutator, to ensure the smoothest rotation with maximum torque.

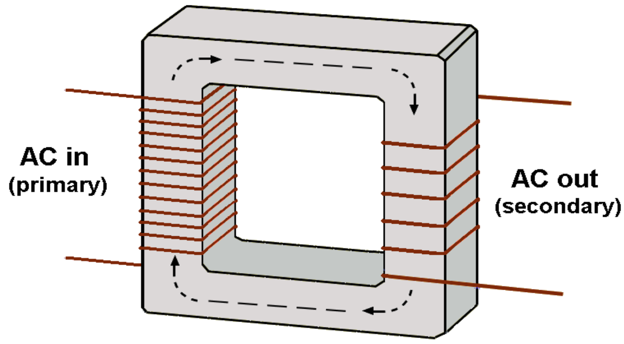
How a dynamo works

The dynamo is basically a motor being operated in reverse.
With the motor, movement was created by passing current through a coil.
With a dynamo in a circuit, current is created by passing a coil through a magnetic field.
So, if you take a DC motor and manually rotate its shaft, a DC output can be taken from its brushes.
Current is only generated by the movement of the coil through the magnetic field; stop rotating the shaft and the current will stop; turn the shaft faster and a greater current will be generated.

How a transformer works



A transformer takes an AC voltage input and produces an AC voltage output.
Sometimes the transformer is wired to produce an output voltage that is the same as the input voltage. This is called an *'isolation transformer'* as the device connected to the output has no direct connection to the mains supply. This helps minimise electric shocks and unwanted electrical interference.



Many domestic devices, including model railways, run on much smaller voltages than the 240V mains supply.
In these cases, the input is still at 240 volts but the output has to be wired to produce a much lower voltage (typically 6V, 12V, or 15V).

The illustration shows the basics of a transformer. There are two coils of wire wrapped round an iron core.
The input voltage is fed to the *'primary'* winding and the constantly varying voltage (in both amplitude and direction) results in a constantly varying magnetic flux around the core.
The varying magnetic field then produces an AC voltage on the output coil, the *'secondary'*.
Its like an electromagnet (on the primary) and an alternator (on the secondary) sharing the same iron core.

The actual output voltage depends upon the *'turns ratio'* of the coil on the secondary compared to the primary. If the two coils had the same number of turns, the output voltage would equal the input voltage (ignoring any losses).

Transformers can be *'step-up'* or *'step-down'* types. Step-up transformers have more turns on the secondary and are used to produce the high voltages used in the old cathode-ray-tube TV sets, and in car ignition systems.

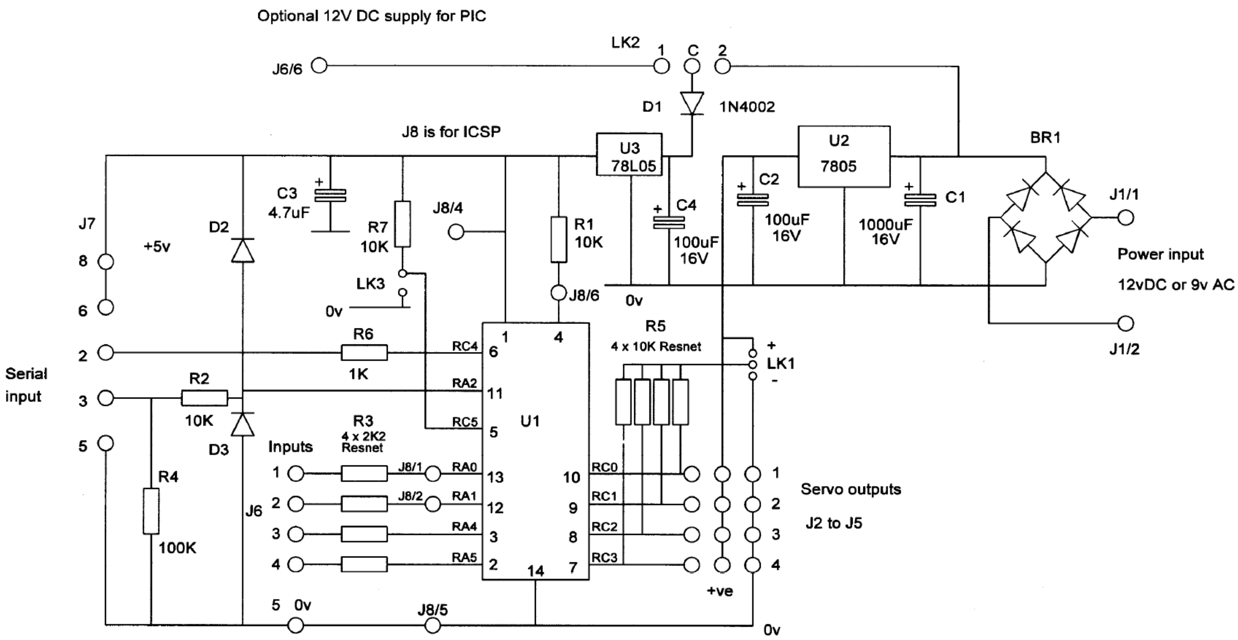
For model railways, we use step-down transformers. So for example, if we wanted a 12 AC output, the transformer secondary would only have a single turn for every 20 on the primary. The 1:20 ratio would result in a 240V input producing a 12V output.

Since a transformer has an AC output, it has to be converted to DC for use by most devices.

Reading circuit diagrams

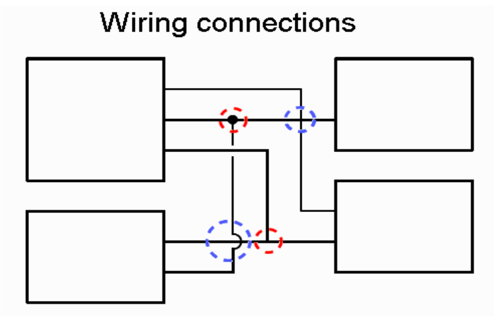
When a project is being designed, it is all about what components to use and how they should be connected together. This might be done by scribbling on a piece of paper or might be achieved through sophisticated software.
Either way, the result is a *'circuit' diagram'* or *'schematic'*. This does not try to show how the project would look when built – just a guide to what connects to what. This allows someone to see how the stages of a circuit work and trace inputs and outputs.
The working project is then developed, showing the printed circuit board tracks and where components sit on the board. This layout is used when constructing a kit.

Circuit diagrams can often be overwhelming at first sight, with lines darting all over the page. Here is an example of a module that is used to control four servo motors.



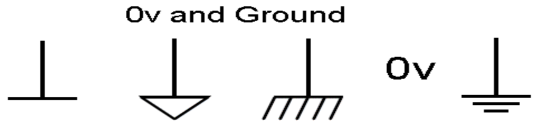
Before looking at individual components and their symbols, it is useful to know that there are different ways of indicating the same thing. Although we will look at various options, it soon becomes clear which symbols are being used on a particular diagram, as they are not mixed on any one diagram.

Wiring symbols



The illustration show four devices that are wired together. Some wires are connected to each other, while others cross each other without connecting. The red circles show two ways of connecting wires together (you will see both being used in magazines). In the top example, a blob is used to show that they connect to each other, while the other example shows wires connecting at a 'T' junction to indicate that they connect.
The blue circles show two ways of wires crossing each other. In the lower example, one wire 'humps' over the other, while the other two simply cross over. Just below the upper red circle, the wire seems to 'break' and this is just another way of showing that the wires cross each other without connecting.

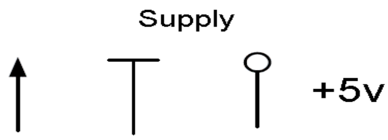
Ground symbols



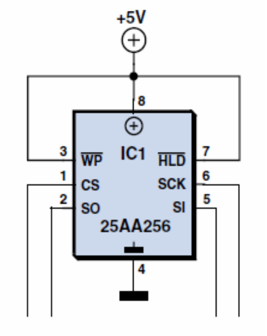
This illustration shows different ways of showing the common 0V point in a circuit, as many components often have one leg wired to 0V.
This is used as a common reference point for taking voltage readings (e.g. a pin on a chip may be 3V positive with respect to ground).
Most diagrams show the 0V line along the bottom of the diagram, with the +ve line along the top.
The diagram on the previous page just used the inscription “0V” to indicate the 0V line. The other symbols may also be found in diagrams, depending on how old the diagram is or whether it a UK or US circuit. Some circuits simply use the inscription “Gnd” to indicate the 0V line.
The symbol on the far right is the *'Earth'* symbol which is sometimes used, although it dates back to a time when the connection was physically taken to real earth by burying a metal rod in the ground. If you look at component C3 on the top left-hand side of the diagram on the previous page, you will see that one end goes to the positive supply (see the +5V) and the other goes to the first symbol in the illustration (the *'Circuit Ground'*). This just means that one end of the resistor is connected to the 0V line. It saves having to draw the line all the way down to the horizontal 0V line and thus makes the diagram look less cluttered.

The second symbol from the left is that of the *'Signal Ground'*. This is the common return line for any signals travelling along the circuit. Often, this is used in the same way as the first symbol.
The third symbol from the left is that of the *'Chassis Ground'*. This is a common point for components using the metal chassis of a device. It is often at 0V but not necessarily so.

Power symbols

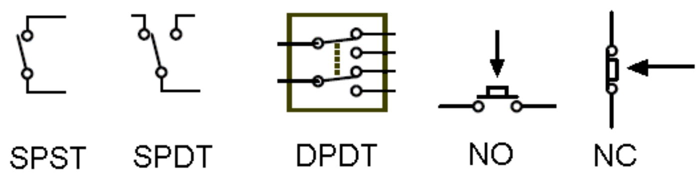


In the same way as you can simplify the drawing of a diagram by using symbols for ground, there are various symbols that indicate the supply voltage, whatever level that may be.



The diagram on the previous page used +5V to indicate the positive 5V supply line.
The three symbols shown are also used to indicate that the connection should be taken to the positive supply line.
This snippet from a circuit shows both the supply and ground symbols being used.

Switches



The symbols shown are for three varieties of lever switch and two push buttons.
SPST stands for *'single pole single throw'*. It has two terminals that are connected together when the switch is operated. The 'pole' indicates how many moving wipers are in the switch, while 'throw' indicates how many contacts the wipe can connect to. In this case, it is a single wiper that can either be connected to the other terminal or not – a simple ON-OFF switch.

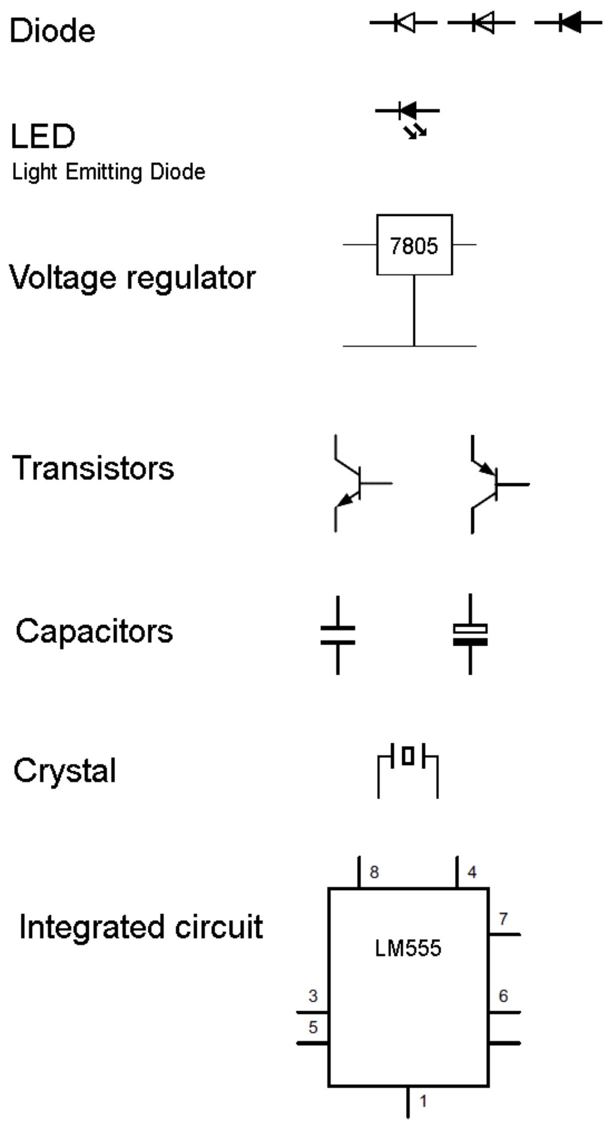
The SPDT switch stands for *'single pole double throw'*, so the one wiper can touch two possible terminals depending on which direction the switch lever is operated – a changeover switch.
A variation on the SPDT switch is the SPCO variety. It stands for *'single pole centre off'* and this switch can be moved to one of three positions. In the middle position, the wiper touches neither of the terminals.
The DPDT switch stands for *'double pole double throw'* and is really two SPDT switches in the one switch body. It has two input terminals and four output terminals. The dotted line indicates that both wipers move at the same time.

The NO pushbutton stands for *'normally off'*. The two terminals are only connected to each other when the button is pressed.

The opposite type is the NC pushbutton. It stands for *'normally connected'* so the connection is broken when the button is pressed,

Electronic components

Most of these components are covered later.
In the meantime, here are their symbols.



Power supplies

Now that we have covered some basic principles, let's look at how they can be applied to model railways. Most layouts need at least two sources of power:

- A variable voltage to run the locos.
- A fixed voltage to run lights, points, etc.

Sometimes, both power needs are built in to the one controller, with the controller providing a separate power socket for accessories. Other controllers need a separate power supply.

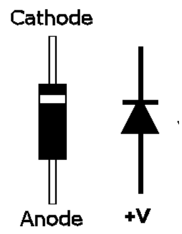
The remainder of this chapter shows how a power supply is constructed.

Building a DC supply

Most electronic devices use low-voltage DC to operate their components. However, the mains supply is at 240V AC.
We already covered how a transformer can reduce the 240V AC mains supply to a more manageable low voltage AC.
So, the next stage is to convert the AC into DC. This is achieved using diodes.

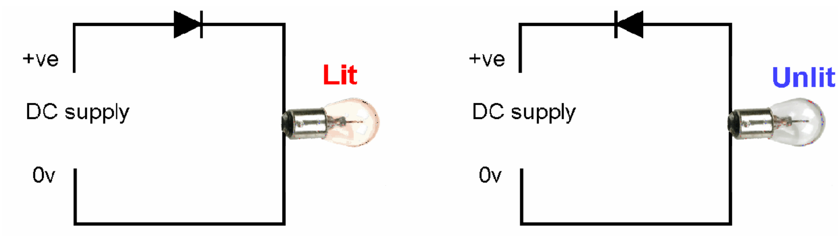
Diodes

A diode, also known as a *'rectifier'*, is a semiconductor device that allows current to pass in one direction only.
The illustration shows the shape of a common diode and its symbol.
Diodes are available in a range of sizes and shapes, depending on the maximum current they are able to handle. Other factors are operating speed and maximum operating voltages.



Diodes and DC

Consider this simple circuit of a power supply (battery or power unit) wired to a bulb via a diode. If the diode's anode is connected to the positive side of the supply, it is said to be *'forward biased'* and it allows current to flow.
Note that the electron flow is *against* the direction of the arrow.

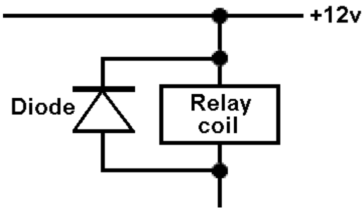


If the diode is reversed, with its cathode towards the positive supply, it is said to be *'reverse biased'* and no current will flow.
Incidentally, the bulb will not glow quite as bright as would without the diode, as 0.65V is dropped across the diode.

DC uses for a diode

Hobbyists sometimes connect power to a module the wrong way round, which can have grave consequences for the module. If a diode is wired in series with the power input, the module will only draw current when the power is connected the right way round and no current when reversed.

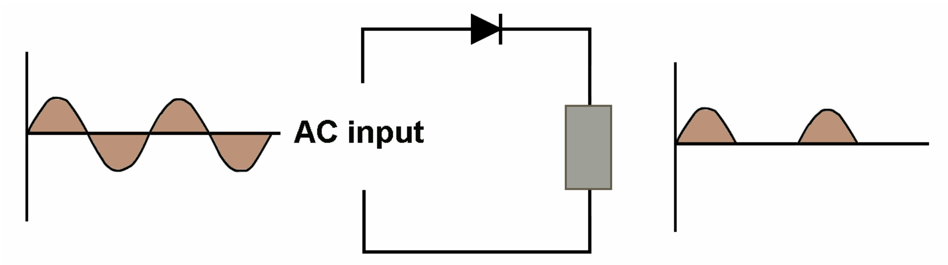
It is also common to see a diode placed across a relay's coil terminals, as in this portion of a diagram.



We know from earlier in this chapter that passing current through the coil results in a magnetic field being developed around it. When the voltage is removed from the coil, that magnetic field collapses and this produces what is known as a *'back EMF'* – a high voltage in the opposite direction to the original voltage across the coil.
This, if left untackled, could lead to destructive voltages being fed back into the rest of the circuit. The solution lies in fitting a diode across the relay. While the relay is held in, the diode is reverse biased and acts as if was not there. However, when the relay is released, it is forward biased to the back EMF, the diode conducts and the current circulates round the diode and the relay coil until dissipated.

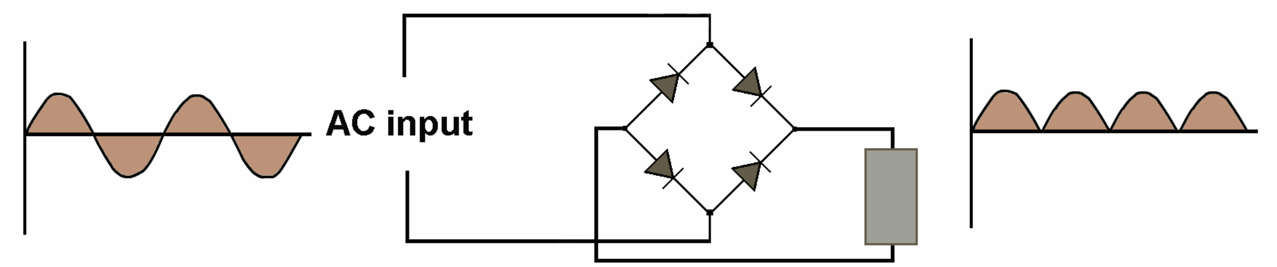
Diodes and AC

This diagram shows an AC supply being connected across a diode and a resistor.
During the first half cycle, the diode is forward biased, current flows and a voltage is developed across the resistor. During the next half cycle, the diode is reverse biased and there is no current flow, therefore no voltage across the resistor.
This process is repeated for every cycle of the AC waveform. This is known as *'half wave rectification',* as only half of the input waveform reaches the output.



The result is that the voltage across the resistor is a series of 'bumps'. It is not AC but is not a smooth DC either – it is a rippled DC.
This is sufficient for some basic purposes but is very inefficient. The shaded areas show that only half the available power is available after rectification.

A much improved supply uses a *'bridge rectifier'* circuit as shown below.

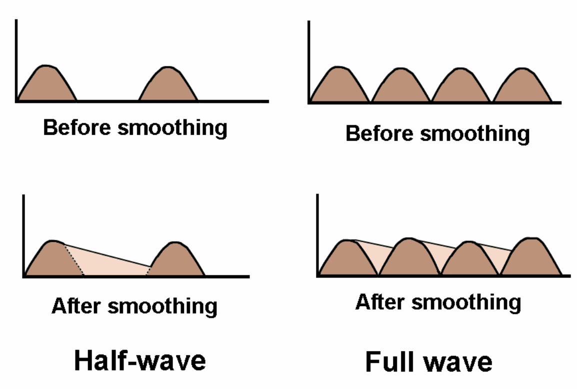


This results in *'full wave rectification'* as both half cycles are now used in the final output. This has greatly reduced the ripple effect and doubled the output power.

A bridge rectifier is available as a component with four connections, two for the AC input and two for the DC output.
Alternatively, a bridge circuit can be made up from four separate diodes.
Like diodes, bridge rectifiers are available in a range of sizes and shapes, depending upon the current and voltage they need to handle.

Smoothing out the ripple

The power supplies so far have very large voltage ripples on them, making them unsuitable for many circuits. The voltage level consistency can be improved with the use of *'smoothing capacitors'*. These are large energy stores that are explained later. The capacitor is fitted across the DC output and is charged (like a battery) while the voltage is high. When the voltage starts to drop, the circuit gets its current out of the capacitor. This causes the capacitor to discharge but it is then recharged during the next voltage surge.



The diagrams show the effect of the capacitors in half-wave and full-wave situations.
The darkly shaded areas are the voltage to be expected without smoothing and the lightly shaded areas show the additional voltage lift from using capacitors.
The rate at which the capacitors discharge depends on the storage ability of the capacitors and how much current is drawn from them by the circuit.

Note

The stated AC voltage from of any transformer's secondary is an RMS value. However, after rectification and smoothing, the DC level is quite different from the secondary's AC level. Remember, the AC peak level is 1.414 times the RMS level. For example, a 14V AC secondary output is really 19.796V peak. Since, after rectification, the smoothing capacitor maintains this level, the DC voltage across the capacitor might be thought to be 19.796V. Unfortunately, current passes through two of the bridge diodes during any one half cycle and each diode drops 0.65V across it. Subtracting these two voltage losses (a total of 1.3V) from the 19.797V , the final voltage across the capacitor is 18.5V – higher than the expected voltage.

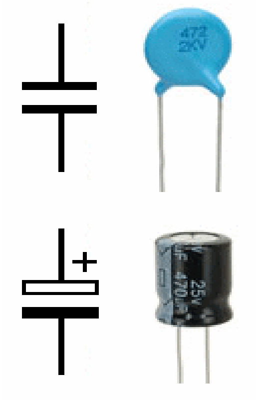
Capacitors

Capacitors are two wire components that are used in almost all electronic circuits, industrial, domestic including model railways.
They are available in a range of sizes, types and specifications. Some are built for high-frequency equipment, some are designed for high stability, some for close tolerance, some for high-voltage working and some for high storage capacity.
They are all, however, doing the same job; they store electrons for later use.
Their storage ability is measured in Farads and values range from picofarads (10^{-12}) through nanofarads (10^{-9}) and microfarads (10^{-6}) up to Farads.
In audio and radio frequency systems, coupling capacitors are used to connect stages in the system; it passes the audio but blocks any DC component. In radio systems, they are used for tuning and filtering.
In model railways they can be found in:

- Capacitive discharge systems to move point solenoids.
- Power supply smoothing.
- Stay-alive capacitors for DCC decoders (e.g. for the Lenz Gold Mini decoder).
- Wave shaping DCC and CBUS signals.
- Timing for PWM controllers, flasher circuits, and other timers.

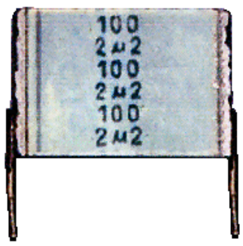
Back to Contents Page 5

Capacitors are made up from two metal plates that are separated by a conductive material, known as the *'dielectric'*. Like a battery it can store electrons; unlike a battery, it cannot create electrons. A capacitor builds up a potential difference across its plates when a voltage is placed across it. One plate accepts electrons from the circuit, while the other plate loses electrons to the circuit.
Depending on their use, capacitor dielectrics can be made from polystyrene, polyester polycarbonate and more, but the most used capacitors have ceramic or electrolytic dielectrics. Ceramic capacitors have low storage capacity and are used in timing and decoupling roles. Electrolytic capacitors have much greater storage capabilities and are used for supply smoothing and capacitor discharge systems.



The upper picture shows a ceramic capacitor. These are 'non-polarised' which means they can be fitted either way round. The symbol next to it is for any non-polarised capacitor.
The lower picture is of an electrolytic capacitor, which has a chemical dielectric. It is polarised and has to be fitted the right way round or it will be destroyed. They have a strip down the side with minus signs printed on them so that you can identify the positive and negative leads. Its symbol has a white rectangle on top (the positive lead) and a filled rectangle below (the negative lead). Some diagrams will also show a positive sign next to the positive end, but not always.

Identifying values



The letter 'p' is used to represent values measured in picofarads, while 'n' represents values measured in nanofarads.
Microfarads use the Greek letter mu 'μ' to represent microfarads as shown in this image. However, it is commonly replaced by the letter 'u'. So, 2μ2 and 2u2 both represent 2.2 microfarads.
Sometimes, a value is shown without the 'F'. So 50μF, 50uF, 50μ and 50u are all the same.

Similarly, a 10 nanofarad capacitor could be found in a magazine or on a diagram as 10n or 10nF. Equally, a 15 picofarad capacitor could be identified as 15pF or just 15p.

As with resistors, the letter is used as a decimal point.
So, a 6.8pF capacitor might be marked as 6p8.

If a capacitors marking has exactly three numbers, without any letters, then the value is measured in picofarads. If you look at the blue ceramic capacitor above, you will see that its value is 472 (i.e. 472pF).

The capacitor on the right is marked as .01, which means that it is .01 (one hundredth) of a microfarad. Where decimal places are displayed, it describes the capacitor's value in microfarads.

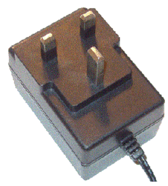
Lastly, diagrams might use different ranges to describe the same capacitor value. For example, this capacitor, although marked as .01, could equally be described as 10nF. A .001 capacitor is the same as a 1nF or 1000pF capacitor.

Regulation

So far, we have been able to convert a 240V AC mains supply to a lower voltage AC, rectify it to obtain a rippled DC voltage, and smooth out the worst variations with a large value capacitor. While this is sufficient for some basic uses, a much more smooth and consistent supply is preferable.

Note

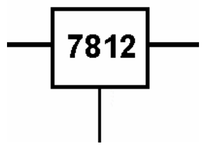
You cannot always rely on the stated voltage outputs of *'wall warts'*, the units with an integral mains plug, that are inserted straight into a mains socket. The cheaper ones do not match up to their performance description. For example, a unit rated as 12V may produce 18V, dropping to only 9V when handling a load. Check before connecting to your equipment.



So, a supply unit may state that it supplies 5V, but is it really at 5V and does it always maintain a 5V output? These doubts can be eliminated using a *'voltage regulator'*
A voltage regulator is a semiconductor device that offers these benefits:

- It provides a guaranteed specific output voltage
- It provides a constant voltage, over varying current demands.
- It eliminates all ripple.

It is a three-pin device:

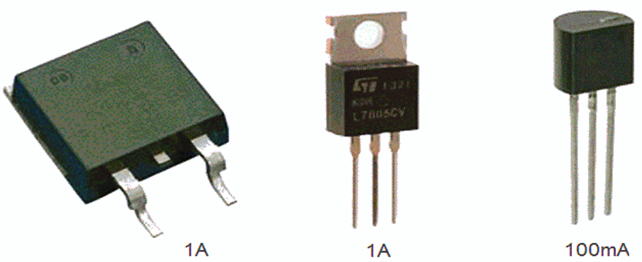


- A pin for the unregulated input voltage
- A pin that is connected to 0V
- A pin that outputs the regulated voltage

It has a very simple symbol and the output voltage can usually be understood from its markings. The most commonly used series start with a 78 or 79, followed by two other number that indicate the output voltage.
78 indicates that it produces a positive output voltage (the most commonly used type) while 79 indicates that it produces a negative output voltage (this one needs to be fed with a negative input, of course).
The voltage regulator shown in the symbol illustration is a +12V regulator.
If the first two numbers are followed by the letter 'L' it means that it is the smaller 100mA type. Some examples of regulator markings are:

7805	+5V 1A regulator
7806	+6V 1A regulator
78L05	+5V 100mA regulator
7912	-12V 1A regulator

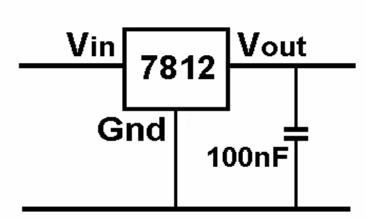
Some typical voltage regulators are shown in these images. They are not to scale; the one on the left is a surface mount component which is smaller than a grain of rice. The one on the right looks like a transistor and is used for low current circuits as it has maximum current handling rating of 100mA. The middle image is of a type that can handle larger currents and is intended to have a heat sink (a chunk of metal with fins to dissipate unwanted heat) attached to it, or to be bolted on to the metal case of a module.



The unregulated DC input is connected across the regulator's input and ground pins, with positive lead from the supply connecting to the regulator's input pin.

The input voltage should be at least 2V or 3V higher than the required output voltage. So, a 5V regulator should be fed by at least 7V and a 12V regulator should have at least 14V or 15V.
Any less and the regulator cannot be guaranteed to work consistently.
Although the 78xx and 79xx range of regulators have a maximum input voltage of 35V, there should not be too high a difference between the input and output voltage. This is because any unwanted power is dissipated as heat in the regulator and if the heatsink cannot conduct it all away, the regulator will shut down.
It all depends on the current you expect the regulator to handle. Remember, power is voltage times current. A high input voltage with a small output current will produce less heat than a lower input voltage with a large output current. If in doubt, use an input voltage that is not excessive compared to the regulator's output voltage.

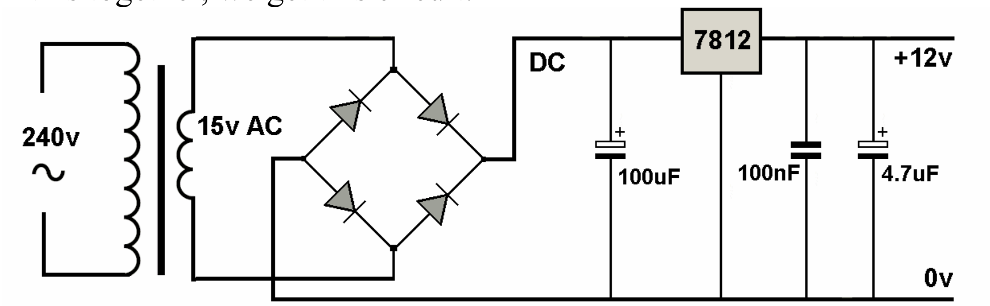
This portion of a circuit shows the voltage regulator in use, with a 100nF (0.1uF) capacitor across the output. This would be in addition to the smoothing capacitors that would be fitted.
It is fitted to shunt away any unwanted high-frequency voltage component that may have got through the regulator.



The benefits of regulators explains why MERG kits use a +12V input and convert inside the module down to 5V. Any temporary drop in the 12V supply, such as might be caused by a sudden high current (e.g. operating a solenoid or servo) would have to be very dramatic before it would affect the 5V inside the module. This ensures that the internal components of the module (e.g. logic circuits or PIC chips) are always supplied at the required operational voltage.

Our final supply circuit

Bringing all this together, we get this circuit.



To recap, the circuit functions are:

- The 240V AC mains supply is connected to the transformer's primary winding.
- The transformer's secondary winding connects 15V AC to the bridge rectifier.
- The rectified output from the diode bridge has a ripple that is partially smoothed by the 100uF capacitor.
- This is fed to the voltage regulator, which produces a consistent +12V DC output.
- The 100nF capacitor takes care of any unwanted high-frequency components.
- The 4.7uF capacitor carries out some final smoothing.

Chapter 2

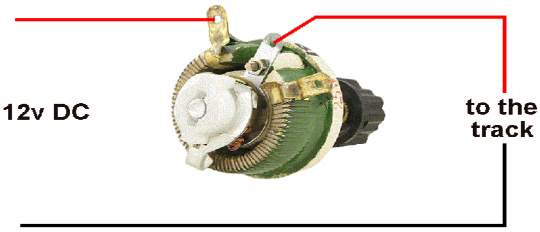
Loco controllers

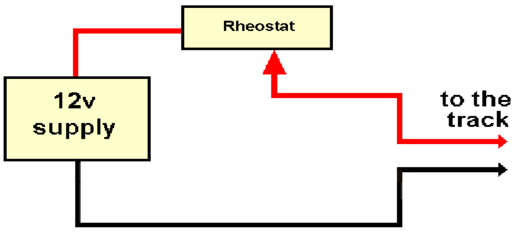
Loco controllers

As you have probably found out, loco controllers are available in all shapes, sizes, prices, with different facilities, accessories and so on.
But let's start at the beginning.

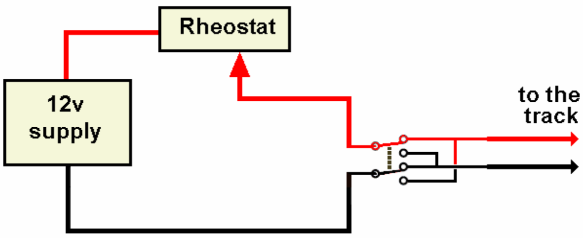
The simplest controller

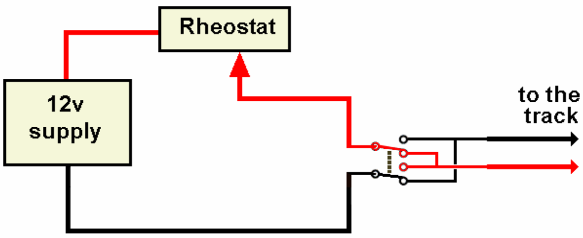
Here is the simplest DC speed controller – one that was used extensively in early model railway controllers. It uses a *'rheostat'* (a large wirewound variable resistor capable of handling large currents). As the knob is turned, the wiper rubs across the windings, varying the resistance placed in series with the locomotive's electric motor.
Turning the knob clockwise decreased the added resistance, increasing the voltage across the motor, thus increasing the current flow and the motor speed.





This works well on a basic level but it only allows the loco to travel in one direction – there is no facility to change loco direction.
Reversals are achieved by adding a DPDT (double-pole, double-throw) switch to the circuit.





Overload protection

No matter how well you build and run a layout, there are always unexpected derailments with wheels shorting across points or the track.
Our circuit above would not be able to cope with short circuits. If the rheostat was turned to minimum or no resistance, then there is almost no resistance across the power supply. Unless the power supply had its own protection, there would a huge amount of current (limited only by the current rating of the transformer). The wire windings of the rheostat would then act more like the bar of an electric fire – unless the diodes blew in the supply.
You could fit a simple fuse between the power supply and the rheostat but that would be a bad idea, as it would need to be replaced regularly.
It is best to fit a circuit breaker. The one on the left is a thermal breaker (it breaks and resets itself when the short is gone). The one on the right needs to be manually reset, like in your domestic fusebox.

If you want, go to eBay or an electronic supplier, buy a rheostat, a DPDT switch and a circuit breaker – and you can build your own basic DC loco controller!
Rhesostats are expensive so many circuits replace them with power transistors which are cheaper and smaller. Apart from that, they function in the same way – reducing or increasing the resistance between the supply voltage and the loco's motor.
Of course, you will not have the most refined controller like those that are now available.

What are problems with a basic controller?

Jerky starts

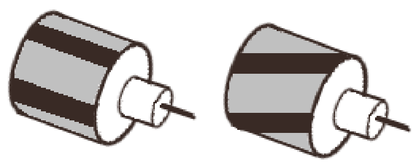
This is an all too familiar problem. The loco is stationary and you increasingly turn up the speed knob to find nothing happens – until the loco suddenly speeds away (a condition sometimes called *'jump starting'*). The effect is unrealistic and unsatisfactory.
The sources of the problem can be both mechanical and electrical.

Static friction

The motor and its gears have their maximum friction when they are stationary, a phenomenon unsurprisingly called *'static friction'* or *'stiction'* As the loco's speed control is turned up, the increasing current through the motor has no effect – until it is sufficient to overcome the static friction. By this time, the speed setting is well advanced and the loco flies off.

Cogging

The metal of an iron cored motor is not uniformly present round the armature. There is a gap between each pole on the rotor, as shown in the illustration (and look again at the picture of a motor in chapter 1). This results in parts of the rotor being attracted to the nearest permanent magnet and being reluctant to be pulled away. It is the same effect you will have noticed when pulling a magnet off a sheet of metal; its stays put despite increasing pulling, before suddenly being jerked off the material.
Better quality motors use *'skew wound'* construction.
The illustrations show the standard motor on the left and the skew wound motor on the right. The non-metal gaps run the length of the standard rotor, while the gaps run at an angle with skew wound. Skew wound rotors have less pronounced gaps, when viewed on the horizontal axis, and this minimises cogging.



Motor resistance

This is the main cause of jerky starts.
First, a quick look at the phenomenon called *'back EMF'*. We described earlier how a motor can act like a generator as well as a motor. Well, when a motor is being turned by an external voltage, it is also generating a counter voltage (the back EMF) which opposes the supply voltage. It is the resultant voltage that determines the speed of the motor. The faster the motor spins, the greater the back EMF; a stationary motor produces no back EMF.
You will recall that there are two resistances in series across the power supply – that of the controller and that of the loco's motor. The motor has a much lower resistance and a higher current when stationary than when it is turning, due to the lack of any back EMF.

The increased current means more voltage being dropped across the rheostat and less across the motor. However, once the motor starts turning, back EMF occurs, the current drops, less voltage is developed across the rheostat and more across the motor. The sudden voltage increase makes the loco surge forward. Although this problem is worse when using rheostats, it occurs with most DC controllers.

Loco slows or stalls

When an electric motor is presented with a greater load, its speed will slow. This reduces the back emf and results in it drawing a higher current and effectively lowering its resistance. Since this current flows through the controller as well as the motor, the voltage available to the motor is actually *reduced*, as the controller takes a greater proportion of the available voltage.

This explains why your loco slows on gradients, with heavy loads, or the friction of going round tight bends in the track. In the worst case, the motor voltage will drop to a level that prevents it from turning – a stall situation.

Jerky slow running

Some of the problems might simply be caused by dirt on the track, or the wheels or the power pickups. It might also be caused by poor mechanical transmission (motor gears, valve gear, coupling rods, etc.).

Otherwise, it is similar to the above problem. At slow speeds, most of the circuit resistance is in the controller and current changes affect the motor voltage less than the controller's resistance. There is insufficient torque (i.e. turning force) being created to maintain a smooth smooth rotation of the rotor.

Racing downhill

This is the opposite effect to slowing on inclines. When going down a gradient, the motor draws less current due to increased back EMF. The voltage across the motor increases and the loco speeds up.

Unrealistic performance

Unless you are very careful, it is difficult to achieve the realistic simulation of train inertia and momentum. It would be much better if the controller could simulate acceleration and deceleration effects.

So, if you suddenly turn the speed knob fully up you don't jerk away; you gradually build to maximum speed. Similarly, a quick turn of the knob fully down gradually slows the loco to a stop. In fact any speed changes should reflect the authentic performance of the loco.

Approaches to solving these problems

We can easily tackle mechanical issues with some cleaning, adjustments and lubricant.

The electronic problems mostly concern achieving and maintaining sufficient torque (i.e. turning force) in the motor.

If we can achieve sufficient instantaneous torque from the motor, we can overcome the retarding effects of friction and cogging – achieving smoother starts.

If we can maintain torque levels throughout the twists, turns, gradients and declines in the track, we minimize stalling and jerky slow running.

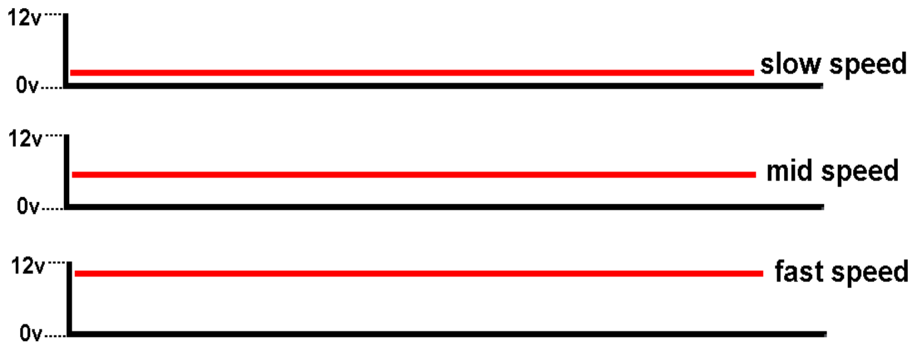
The two main ways that controllers tackle this is by:

- PWM – Pulse Width Modulation
- Back-EMF Feedback

PWM

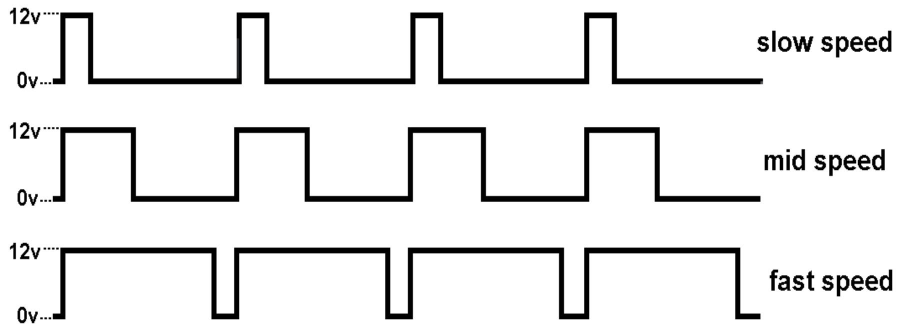
In most of the above situations, we wish we could just give the motor a bit of a 'kick' to get it started or to avoid jerking or stalling – and that's just what PWM aims to do. In fact, Pulse Width Modulation (PWM) is nothing other than giving the motor continual 'kicks'.
With normal DC, we change the voltage from 0V (stop) to maybe 12V (full speed), with every voltage in between.
At high voltages, we avoid some of the problems of stiction and cogging, due to the momentum of a high spinning motor.
It is at low motor voltages that the problems are mostly evident.

If you could look at the voltage across a motor when using ordinary DC, it would like these three charts which show the voltage at different speeds.



You can see how difficult it is to maintain torque at the slower speeds, with such small voltage levels.

A PWM signal, on the other hand, looks like this for the same three speeds.



At the faster speeds, the output looks similar to a normal DC voltage. In fact, at the very fastest speed, it would be a continual 12V just like normal DC.
It is the slow speed on that chart that is of the greatest interest. Instead of sending a small voltage (say 2V) to the motor, a succession of 12V pulses are sent to the motor. Over a period of time, its effective voltage is the same as the small DC voltage. If you tried to measure the signal with a meter, it would look like a low voltage, as the meter would average out the pulses.
However – and this is the big advantage – the 12V kicks provide the instantaneous torque that tackles stiction and cogging. Its increased torque also minimises problems with stalling and jerky slow running.

Back to Contents Page 5

The PWM signal has two factors:

- How often the 'kick' happen – known as the *'clock cycle'*.
- How long the pulse lasts – known as the *'duty cycle'*.

The clock cycle is measured in Hz and the duty cycle is measured as a percentage of the available time.

In the example, the duty cycle's slow speed is running at 10%, the mid speed at about 40% and the fast speed at around 90%.

Clock cycle

The clock cycle cannot be too long, as the pulses would appear too infrequently to be effective. If the clock cycle is too short, the pulses would be arriving at a much faster rate. At excessively high frequencies, the coil in the motor behaves more like an inductor and its impedance (resistance to current changes) inhibits the passage of high frequency pulses and reduces the current flow through the motor.

There is no one frequency that can be applied to all motors. ZTC Controls produces a useful chart that shows the best PWM frequencies to use for different motor types; it varies from 31Hz to 32kHz.

CV9 Value	PWM Frequency(Hz)	Notes	CV9 Value	PWM Frequency(Hz)	Notes
255	31	Large	143	327	Low
251	33	Motors	139	357	Power
247	35	only	135	393	Precision
243	38		131	437	Motors
239	41		127	490	Only
235	41		123	523	
231	49		119	561	
227	55		115	604	
223	61		111	654	
219	65		107	714	
216	69	**DEFAULT**	103	786	
215	70		99	874	
211	75		95	980	
207	82		91	1046	
203	89		83	1208	USE
199	98		79	1309	EXTERNAL
195	109		75	1429	FILTER
191	123		71	1572	WITH
187	131	MOST	67	1748	SMALL
183	140	SMALL	63	1961	MOTORS
179	151	MOTORS	59	2092	ZTC150
175	164		55	2242	ZTC151
171	179		51	2415	&
167	197		47	2618	ZTC156
163	219		43	2857	
159	245		39	3145	
155	262		35	3497	
151	290		31	3922	
147	302		27	4184	
			23	4484	
			19	4831	
			15	5236	
			11	5714	
			7	6289	
			3	6993	
			1	16Khz	
			0	32Khz	Whisper drive,
					Coreless Motors
					RG4, RG7 etc
					Mini Motors, etc

As a general rule, the older 3-pole open frame motors, with straight slotted poles, benefit most from running at lower frequencies.
Modern, 5 or 7-pole can motors, with skew wound poles, use mid frequencies.
Coreless motors can be run happily at higher frequencies, although slow speed running might not be enhanced.

Buzzing

The average human ear, depending on your age, can hear frequencies between 20Hz and 20kHz, with maximum sensitivity between 1Khz and 4kHz (your landline telephone only uses a band from 300Hz to around 3kHz).
With exception of the very highest frequencies in the ZTC list, all are within the human hearing range.
Feeding a motor with pulses in the audible range usually results in an distinct buzzing or humming sound from the motor. This is not normally harmful to the motor but some find it annoying.

To counteract buzzing, a loco controller can increase the frequency of its clock cycle, to take it beyond the range of human hearing. 16kHz usually is beyond the hearing of a middle-aged male, with frequencies up to 32kHz being beyond everyone's hearing ability.
The downside to a higher frequency is often poorer performance at slower speeds.

Note

If you are already using DCC, then your decoders are using PWM to control motor speed, set by the values in CV9 to match your individual loco's motors (see the chapter on DCC). This CV is sometimes referred to as the *'repeat rate'*. The value entered into CV9 is the period (the reciprocal of the frequency). So, higher values result in lower frequencies and vice versa. Decoder manufacturers like to promote their decoders' by describing their abilities to produce high frequency PWM. So, we have Silent Running(from NCE and Zimo), Supersonic (from Digitrax) and Quiet Drive (from TCS).

Duty cycle

The duty cycle decides the effective running speed of the motor. It is a ratio of the voltage's 'on' time to its 'off' time. A 0% ratio means the pulse has no width, hence zero volts, while 100% means the pulse has no 'off' time, resulting in the full supply voltage across the motor.
While the clock cycle is a fixed value, the duty cycle will vary as the operator speeds and slows the loco.
If you go back a page and look again at the PWM waveforms, you can make the following observations. In the first chart, the 'on' time was 10% leaving an 'off' time of 90%. This would be known as 1:9 mark-space ratio and would be the equivalent of 1.2V DC for a 12V supply. The second chart would equate to 4.8V DC and the bottom chart would equate to 10.8V DC. If the operator tuned the control to full speed, there would be no 'off' period and the voltage on the motor would a continuous 12VDC (just like a straight DC controller).
While high mark-space ratios are broadly similar to normal DC working, the real benefits come at lower ratios. While a constant 1.2V DC supply exposes the motor to stalling, etc., the short bursts of full power produce greater torque and are much more likely to keep the motor turning.

Heat dissipation

In our earlier example rheostat controller, if we wanted to reduce a 12V supply to only 3V for the motor, we would have to drop the other 9V across the rheostat. If the loco consumed 1A, then the power developed in the rheostat would be a full 9W. The same power would be developed if we had instead used power transistors as the drive electronics. In both cases, the power would be dissipated as unwanted heat.

PWM working reduces the harmful overheating effects. With PWM, power is totally switched off during the 'off' times and the power developed during the 'on' times is nearly all used by the motor. There is very little wasted power in a PWM controller.

PWM is also very handy for the controllers built into DCC decoders, otherwise the excessive heat would melt the plastic bodies of the locos!

Lower clock cycles however, can have an effect on increasing the heating of the motor coil; this is usually worse for small motors with little metal to absorb the heat.

Conclusions on PWM

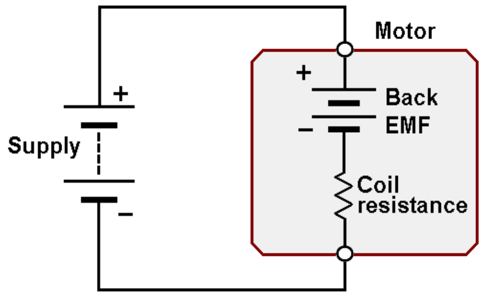
Despite the noise and heat problems, PWM offers a great step forward in obtaining much smoother running of locos, particularly at slow speeds.

Most DC controllers use PWM for that reason, and it is almost universally used as the motor control system built into DCC decoders.

For best results, the PWM clock cycle should be matched to the motors in use.

Back-EMF

We looked at back-EMF (often shortened to *'bemf'*) a little earlier and described it as the voltage that is generated as a motor spins, its voltage being in opposition to the supply voltage to the motor.



This diagram shows a battery supplying a motor. The smaller battery, marked as 'Back EMF' represents the voltage which is effectively in series with the main supply, but with a reverse voltage. The motor, therefore, is being run from the difference between the two voltages.

Back-EMF is the voltage generated by the motor windings moving in the magnetic field of the motor's magnets. Consequently, the faster the motor spins, the greater the amount of back-EMF, and vice-versa. Put another way, the back-EMF is directly proportional to the loco speed at any one time.

When a loco comes to an incline, its speeds slows and its back-EMF reduces.

When a loco hurtles down an gradient, its back-EMF increases.

You could think of the back-EMF as a variable resistance with a value that changes with loco speed.

In both cases, the current being drawn from the supply alters with the loco's speed – even though the operator has not touched the throttle control.

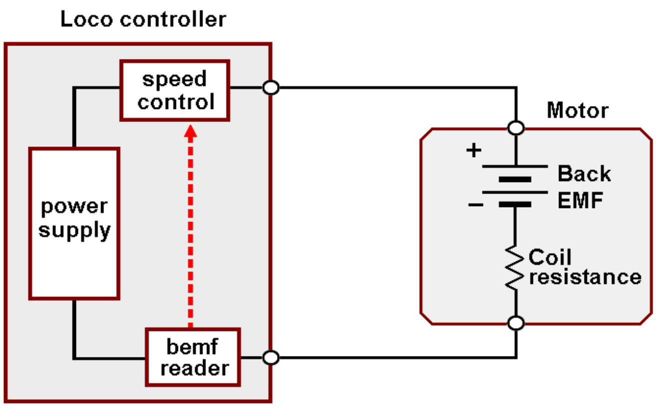
Feedback

If we could detect these changes, we would know if our loco was running slower or faster than the speed set by the operator. The controller's circuitry could then take steps to automatically adjust the loco's speed back to where it should be.

This approach is known as *'feedback',* as the changes inside the motor are fed back to the circuit in the controller.
Of course, it is not possible to read this feedback at the same time as powering the loco.
So, the steps are:

- Power the loco.
- For a very short time, cut off the power.
- Read the feedback (the unpowered motor is still spinning due to momentum).
- Restore the traction power.

The illustration shows the principle of how it works.
The voltage from the controller is taken to the motor via the tracks.
When the feedback is to be measured, this voltage is cut off.
The back-EMF generated by the freewheeling motor is now on the tracks and fed back to the controller, but in the *opposite* polarity.
The detector circuitry measures the amount of back-EMF and compares it with previous readings.



If the readings are identical, no adjustments are made.
If the back-EMF reading has altered, either greater or smaller, this difference is used to alter the speed control's output (see the red dotted link) to compensate.

Three different control systems use feedback techniques.

- In a normal DC feedback controller, the circuit has to make periodic power breaks to allow the bemf to be measured. These breaks may only last less than a millisecond and need only occur infrequently (maybe every second or less).
- In a PWM-based controller, the power is being switched on and off continuously anyway, so it is easy to read the feedback during some of the 'off' periods of the duty cycle.
- In a DCC system, the bemf is read within the decoders, with no involvement on the part of the controller.

Controllers which combine PWM and back-EMF feedback, and also DCC decoders, provide much more consistent running than normal DC controllers.
You can expect:

- Elimination, or a huge reduction, of jerky starting.
- Improved slow running.
- Reduction in stalling.
- Consistent speeds up and down gradients and on track curves.
- Speed calibration on DCC, so that all locos run at the same speed on the same controller setting.

However, a balance has to be struck between improved running and having a train that runs at the same speed under all conditions (like a car's cruise control). After all, you would expect a train to slow somewhat on an incline.

Back to Contents Page 5

Other features

Other useful features for a loco controller are the ability to introduce acceleration and deceleration. The controller converts an operator's sudden large speed change into a gradual increase or decrease in speed.
No more jerking away from a standstill, or screeching to a sudden halt, or lurches as the speed is altered by substantial amounts.
The controller's circuitry ensures gradual pulling away and slowing to a halt.
In DCC systems, this facility is saved as user-defined values inside in the DCC decoders.

DCC - Digital Command Control

For many years, model railway systems ran on DC. You varied the DC voltage to the track and the loco slowed down or speeded up.
Modellers observed that things worked differently in the real world. The voltage supplied to electrified railway systems or underground systems did not vary their voltage. The overhead lines or rail tracks were supplied at a constant rate; the varying of the voltage to the electric motors took place within individual locos. That observation started the development of DCC, notably byBernd Lenz, which has now become widespread in model railways.

Some find DCC a little complicated but we are discussing it here because the choice of loco controllers, and probably some layout wiring, will depend on which system (DC or DCC) you decide to use.
If you are not using DCC, you could skip to the next chapter.
If you are thinking of looking at DCC, then this is good place to start.
If you already have a DCC system, it is best to know how it works before considering how it should be wired.

With DCC, the loco is always provided with full track voltage and it is up to the operator to control how much of it is provided to the motor. This is explained in more detail later but the main thing to recognise is that:

- All sections of track are live at all times.
- The electronic devices (known as *'DCC decoders'*) that can be connected to the track provide great flexibility.

These provide a great many benefits to model railway enthusiasts, such as:

Simplified wiring

- DCC does not require any complex wiring systems for normal operation.
- A single power bus can feed the entire layout, connecting to every piece of track.
- A DCC layout does not need block/section switches or relays.
- A DCC layout, with less complicated wiring, results in fewer wiring problems.
- A DCC control panel has a cleaner look, with far fewer switches.
- A DCC layout need have no control panel at all, running everything from a handset.
- Wiring reversing loops, wyes and turntables are simplified.

Simplified working

- With all sections live at all times, there is no need to remember what sections are live or dead.
- You can run multiple locos on the same section of track, at the same time.
- You can run consists (one or more locos pulling a train).

Improved running

- Non-polarised rails attract less dirt.
- DCC locos use PWM, with its benefits of improved slow running and less stalling (see earlier).
- Given its constant high voltage track supply, it is less troubled by voltage drops in cables.
- Each loco can be set to have a constant stopping distance.

Motor tuning

- You can set the loco's minimum working voltage, so that the motor is fed with this voltage as soon as the operator advances the speed control to the very first notch.
- You can set the loco's maximum voltage. So, for example, a shunter will be prevented from racing round the yard.
- You can set the rate of acceleration for an individual loco.
- You can set the rate of deceleration for an individual loco.
- You can set bemf (see earlier) for maximum performance.
- You can tweak the reverse running to be identical to forward running.
- You can match the performance of multiple locos, to allow smooth-running consists.

Each loco's electronics has its own individual settings that you set up and it remembers.

More than loco motors

DCC offers more than running motors. It is designed to handle a range of accessories, both in the train and at the trackside.

Loco features

Since there is power available to all locos on any section of track, at any time, even when they are stationary, the loco decoder can be configured to switch off and on various loco features such as:

- Sound (horns, whistles, bells, diesel engine and steam engine effects, brakes, etc.).
- Lights (headlights, beacons, Mars lights, firebox flicker, ditch lights, etc.)
- Uncouplers
- Cranes

Lineside features

There are two types of DCC decoders- those intended for handling loco motors and those intended for handling accessories. Accessory decoders obtain their power (and their instructions) from the track and are used for:

- Points.
- Signals.
- Crossing and barrier gates.
- Uncouplers

Simplified computer control

Computer control of DC layouts is perfectly possible, with the computer turning relays on and off to switch the power to various track sections on and off. This is in addition to controlling the track voltage levels for speed control.
Computer control is made easier with DCC, as the computer only has to tell a loco's DCC decoder to start or stop the motor, or change the motor's speed. There is no elaborate and complicated track power switching to worry about.

Each loco has to have its own electronics board fitted – called a DCC loco decoder.
The trackside decoders are known as '*Accessory Decoders'*. They don't have the electronics for controlling the loco's motor, but provide switchable outputs for signals, lights, etc. Most of these provide 4 or 8 switchable outputs on the one decoder. Other decoders are designed to drive servos for point operations; others handle stall point motors or capacitive discharge point motors.

Downsides

As usual, benefits come with downsides namely:

Cost

- The cheapest decoder costs around £12 and can be as much as £70 for large gauge locos that require higher current. To equip a sizeable fleet of locos would be an expensive undertaking.
- While a bottom of the range DCC command station costs little more than a good DC controller, the top of the range command station could cost over £500.
- A DCC sound decoder costs around £100.
- Railroad & Co Train Controller Gold is at the top end of computer automation software and sells at 499 Euros. (fortunately, there are cheaper versions – including free programs).

Complexity

- You can only run one loco straight out of the box. Each loco decoder has to be allocated a unique identification number. So, if you bought two locos with pre-fitted decoders, they would both be supplied with the same default address. You would have to change one decoder's address to get both working on the same layout. Similarly, if you bought decoders to fit to locos, you have to ensure that they have different address numbers before trying to run them both at the same time.
- Decoders can be awkward to fit, specially in smaller gauges.
- All decoders require some initial tweaking to get the best from them. This is covered in a later chapter.
- DCC decoders don't make poorly running loco run much better. If your motor and mechanism is dirty, sticky, or badly adjusted, the electronics inside the decoder cannot compensate for the lack of basic maintenance.

DCC shopping list

If you are considering starting a DCC layout, here are the items to look for.

Minimum

- A DCC command station. The image shows a budget model – the Bachmann E-Z Command.
- A separate power supply, if one is not built into or supplied with the command station.
- A decoder for each loco.

Additions

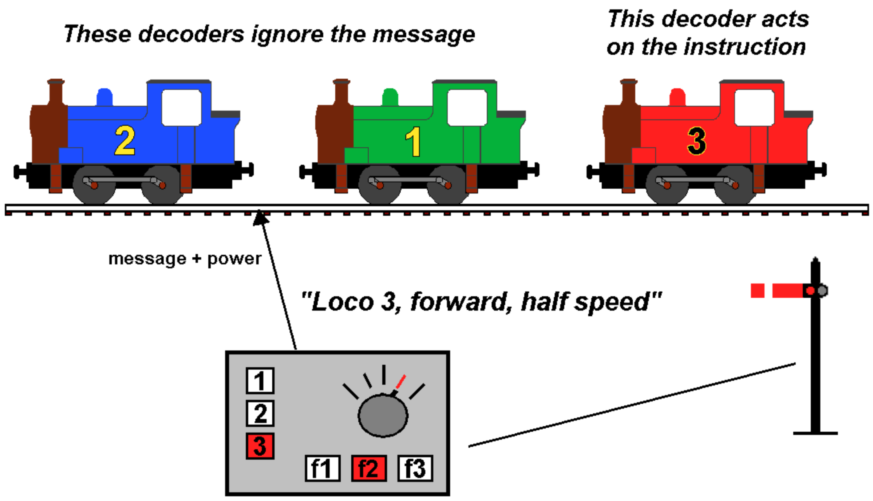
- A booster unit, if you need to run more locos than your command station can cope with at any one time.
- DCC function decoders, if you want to operate points and accessories via DCC. The image shows the Lenz LS150 accessory decoder which can switch six separate outputs.
- A hand held cab controller, if you want to walk around the layout while operating locos, etc. These can be wired to the layout via a plug and socket, or can use a wireless system.

How DCC works

This is a simplified explanation of DCC working, as there is a full chapter on this later on.

A DCC controller is called a *'Command Station'* because it send out commands to the various decoders that are connected to the track – either in the locos or at the trackside.

These commands are sent on to the track as part of the power supply to the track (this is explained in the later chapter on DCC).



The illustration shows three locos on the same section of track, being operated from a single command station. Although they are shown next to each other, they can, of course, be anywhere on the layout.

Each loco is fitted with its own decoder. In the example, the decoders have been given the

unique addresses of 1, 2 and 3. There can be no two locos with the same decoder address on the layout at the same time.
Each decoder is constantly looking for any incoming messages for itself. Messages for other addresses are ignored. When a message is received that matches its address, the instruction from the command station is acted upon. The message could be to go faster, go slower, stop, reverse, etc.
In the example, the operator wants loco 3 to drive forward at half speed, so he/she selects loco 3 on the command station and turns the speed control to half way forward. The command station converts this setting of the speed control into an electronic message that is sent out along the track.
Although the message is received by all the loco decoders, only the decoder with address 3 will respond to the message.

A similar process takes place when the operator wants to move a point. The accessory decoder for that point will also have its own unique address. In this case, the message is not picked up from the loco's wheels and pickups. The decoder is permanently wired across the track near where it is to be used. Although the decoder will be receiving messages via the track, it will require additional power for running the accessories it is attached to.

Back to Contents Page 5

Chapter 3

Layout wiring

Layout wiring

Some layouts are planned and some layouts just 'develop' with time.
While a lot of thought often goes into designing the tracks, the wiring of the layout is often looked at afterwards. This can easily lead to ad-hoc wiring schemes involving duplication of work, greater fault liability and even ripping out some parts to start again.
This is all avoidable if the layout wiring is considered along with the rest of planning of the layout. Don't lay the track then think about the wiring; plan your wiring requirements before laying the track.

This, in turn, means that you should give some thought about what type of layout you are aiming to produce - from an electrical standpoint.
As we will see later, there are a variety of ways to operate and control a layout, with each requiring different wiring systems (e.g. DC, DCC, CBUS, etc.). You may even decide to buy particular brands of track and points that best match your final layout configuration (e.g. do you need insulfrog or electrofrog points).

Layout Wiring Goals

The wiring of a model railway should have as its main aims:

- Reliability (not prone to breakdowns, overheating, fire or shock risks)
- Maintainability (easily looked after and corrected when things go wrong)
- Alterability (the existing layout can be easily altered or added to)

These goals can be established by implementing some 'standards' with respect to the wiring.

Reliability complications include

- Long wiring runs (voltage drop, induction, interference)
- Multiple baseboards / connections
- Multiple operators
- Mixing DC and DCC
- Signalling
- Feedback / automation

Maintainability / alterability complications include

- No / incorrect documentation
- No wiring scheme
- No diagrams
- No fault history
- Inaccessible mountings

Safety

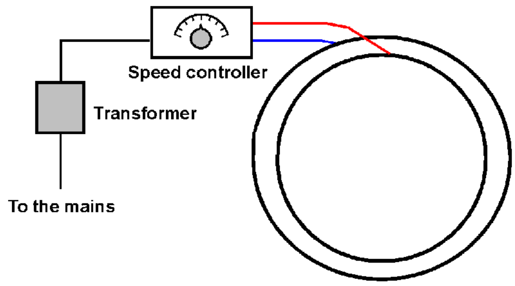
A short, but important, word before we begin.

> All the wiring that runs above and below your baseboard should be at low voltage.
> Mains voltages should never be allowed on the baseboard.

There are a number of circuit diagrams available on the Internet for constructing home made loco controllers. Most use mains transformers to feed the rest of the circuit. Please don't attempt this unless you really know what you are doing. An alternative is to use 'wall warts' or 'power bricks' – power supplies that are sold for use with consumer goods or computers. These are double insulated and safe to use.

Different operating approaches

Model railways are a hobby. That means that there is no 'correct' way to design a layout. It's about what you want to achieve. Some may aim to run an exact timetable for a particular railway line in a particular era. Others may prefer a more eccentric and chaotic narrow-gauge layout based on imagination. Yet others may enjoy the challenge of constructing a fully-automated, computer-controlled layout. Last, and not least, some may be content with a simple *'Thomas the Tank'* layout running round a loop.
All these, and others, are legitimate layouts. They only differ in the complexity of the layout wiring.



A simple loop of track, with a point or two, is easily wired by even the newest beginner, only requiring two wires from the controller to the track.
Moving beyond a simple loop, to the medium and large sized layouts, there are different attitudes to how a layout should be run. These differences lead to different wiring requirements.

Firstly, there is the ongoing discussion as to whether the train is controlled by the driver or by the signalman. In most layouts, the trains are controlled by the operator (driver) with the signals being purely ornamental (doing nothing) , or an optional extra (manually operated) or simply tied to the trackside detectors. The other view is that to emulate the real world, train movements are controlled by the setting of the signals and therefore the hobbyist should have the primary role of signalman. The train movements are then dictated to by the signals, either manually or automatically. That leads to different wiring approaches.

Purely manual

Everything is left to the operator.
There is no interlocking.
Trains can run into the back of each other.
Its up to you to control everything.
Everything is wired back to the control panel.

Manual with train detectors for signals

Train detectors are used – but only to operate the signals.
They have no effect on the trains.
Trains are still controlled by the operator.
It is still up to the operator to ensure trains don't run into each other.

Manual with some interlocking

Train detectors are used to operate the signals.
They also operate relays that ensure power is switched off in the preceding section.
The operator runs the trains but is prevented from running trains into each other.

Manual with computer assistance

The operator still runs the trains by hand.
The train detector information is fed back to the computer.
The computer controls the power to the sections.
The signals can be operated directly by the train detectors or by the computer.
Useful for 'glass panel' operators - those using mouse controls on a monitor screen.

Fully automatic.

The train detector information is fed back to the computer.
The computer controls train movements based on this information.

Layout options

The 'train set' of our youth was simple. A pair of wires from the controller was attached to the track of the oval and the train ran in circles. One operator, one controller and one loco on the track at a time – layout control at its simplest.
When a more complicated layout is planned, that involves having several locos on the track, or several operators, a number of different wiring methods are available.

There are two main approaches to layout wiring:

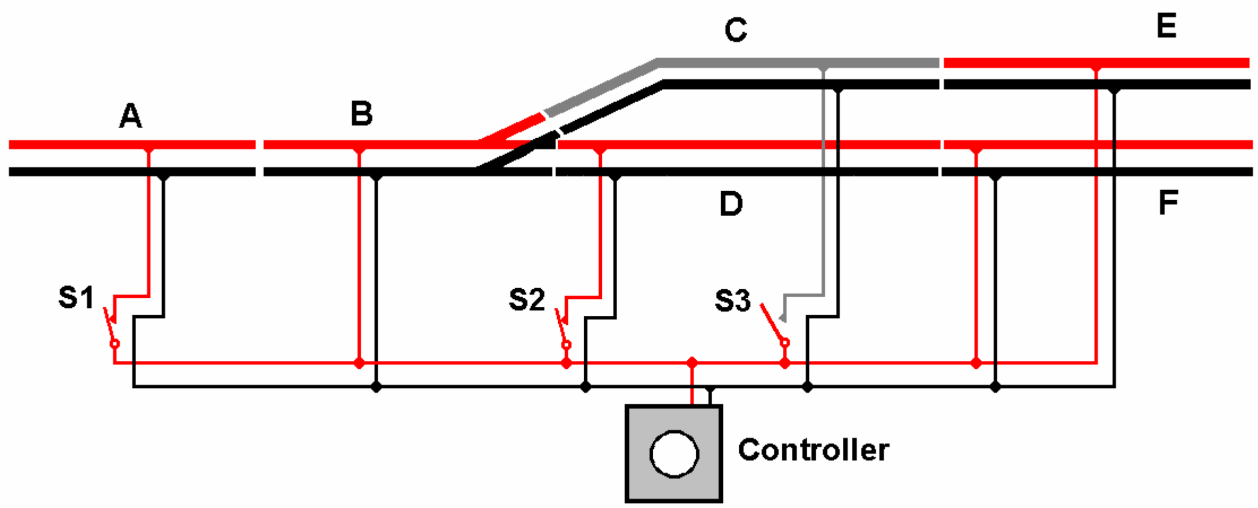
- Block or cab control
- Bus-based layouts

Block control

Real-world railways divided up mainline track into long blocks, with entry to each block controlled by signals. This prevented trains from running into the back of each other.
This concept can be adopted for model railways in various options. We can carry out all operations manually, through electronic modules, or through computer control.

For us, block control is a common means of controlling multiple DC locos on a layout. If you divide your layout into sections of track that are electrically isolated from each other, you can place multiple DC locos on the layout, each in a separate track section.
If you are installing track, you can use plastic rail joiners to isolate adjoining rails. If the track is already laid, the rails can be cut with a fine-toothed saw or with a cutoff disc in a Dremel drill (don't forget to wear protective glasses).
Since each loco sits on an isolated section of track, it can be operated in that section, independently from any other locos on other sections.
You would create a block for each section of track on which you are likely to run, or park, trains.   Some sections will be long (e.g. a stretch of main line) and others short (e.g. a siding or a passing loop). The more blocks you create, the more flexible the layout becomes – but the complexity of the wiring and switching also increases.
Since each block requires to be powered independently, they need their own separate power wiring back to the loco controller.

This illustration shows a basic block system (sometime referred to as a *'Star'* or *'Point to Point'* system).



All the track sections shown are isolated from each other but are all connected to a central point (the controller), usually via switches or relays.
The operator decides which locos to run by powering up that particular section by throwing its power switch.
In the example, sections B, E and F are permanently wired to the controller so these track sections are always powered.
Because switches S1 and S2 are thrown, sections A and D are also powered but section C is unpowered as switch S3 has not been thrown.

So, throwing these two switches allows the operator to drive the train from block A, through block B, to block D. If switch S2 is now turned off, block D has no power and the train sits stationary in the block.
Now, if switch S3 is thrown, a train in block C is powered and could be driven from block C, though block B to block A.

Pros:

- Each section is independent, so a failure in one section need not prevent the rest of the layout from working.
- Each section only draws a proportion of the total layout current.
- You can concentrate the entire layout's device controls (circuit breakers, block detectors, light switches, point control modules, etc.) in one accessible location.

Cons:

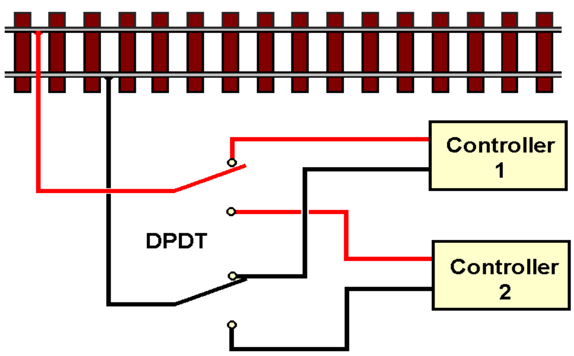
- The huge amount of wiring that is involved produces cable management problems, unless meticulous records are kept.
- The long runs to the remotest parts of the layout may cause voltage drops in the cable and connectors, reducing the efficiency of the devices at the far end (e.g. locos running more slowly, LEDs dimmer, CDU's being less punchy).

Multiple operators

The simple illustration above shows three tracks sections that can be switched between being powered and unpowered. It is designed for one operator, running one train at a time, using a single controller. That is enough for many home layouts, with the switches to power the blocks mounted in a central control panel.

For larger home layouts and club layouts, there is a need to have multiple operators, each running their own locos – while still maintaining electrical and physical separation.
For every loco running on the layout, there is an operator with a separate controller and two controllers should not be connected to the same block at the same time.

In some cases, the operation is simple: one operator can be handling the main station while another handles the mainline, another the marshalling yard, and so on. To maintain movement, there would have to be a common handover point for each operator's domain. So, for example, the operator of the main station and the operator of the main line would require control over a common block (e.g. entry to the station throat).



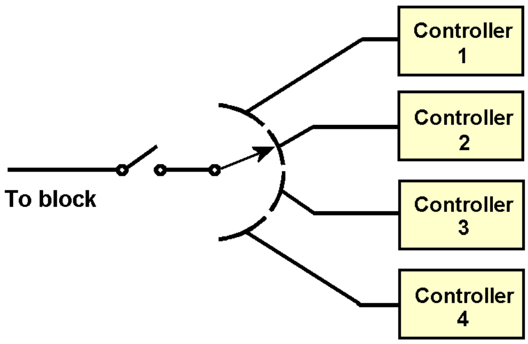
The illustration shows a DPDT (double-pole double-throw) switch. When set one way, it connects controller 1 to the block; when set in the other direction, controller 2 is connected to that block. You might prefer to have a 'centre-off' version of this switch so that, in mid-position, the block is left unpowered.
This approach suits layouts where operators have defined layout districts to control and eventually have to hand over the loco to another operator. It is less suitable if you want to control a train as it goes all round a layout.

Cab Control

The most flexible approach is to allow all operators and all controllers access to all blocks. This is known as *'cab control'*. It gives any one operator the ability to control his/her train anywhere on the layout. Naturally, this raises problems of *'interlocking'* – making sure that no two controller access the same block at the same time. This can be achieved through lots of switches, or through more complex arrangements using electronics and relays.

We could have a separate switch for each controller to every block, but this quickly creates problems. Imagine a control panel that allowed 6 controllers access to 80 blocks – a total of 480 switches! However, there is an even bigger problem. To ensure that one controller had sole access to a block, we would have to ensure that the other five controller's switches were switched off – accidents waiting to happen! Luckily, there is an improved switching system that drastically cuts down the needs to a single switch for each block.

Let's consider a single block for the moment. If there were four operators, for example, we could have an switch arrangement like that shown in the illustration. It uses a 1-pole, 4-way rotary switch. If controller 2 wants access to the block, the knob is turned to position 2. This prevents the other three controllers from connecting. The on/off switch can be left in the 'off' position when the block is not being used, preventing unintended train movements.



This works well for a single block and can be replicated again and again for other blocks, with a rotary switch and on/off switch for each extra block added.

Switch panels

With a small layout, cab control is usually achieved by fitting all the switches to a central control panel.
It is still possible to have a central switching panel in a larger layout, as long as the operators are happy to have a main co-ordinator in charge of all layout block switching.
The main co-ordinator decides which operator can take over a block and provides access to the block to the chosen controller. The co-ordinator informs the chosen operator that he/she is ready to go – verbally, through lights, or through switching track signals. This method may be thought useful for running timetables, with the co-ordinator as the signalman, dispatcher, or 'Fat Controller'. It still allows some degree of flexibility. So, if block 12 should never be available to controller 3, then controller 3 will not be wired to block 12's rotary switch.

Central control might be considered too restrictive and a common alternative is to wire smaller local control panels throughout a large layout. The blocks of a goods yard may be given their own control panel, as may a station or a fiddle yard.
This system is usually accompanied by portable handheld throttles that plug into the layout close to the control panel. This way, any operator can walk up to a control panel, plug in a handheld throttle, dial his controller number on a block switch, and start running the train in that block. If required, a number of block switches can be similarly set, allowing the operator to run the train across multiple blocks.
Moving from control panel to control panel, an operator can run a train anywhere on the layout.
Another advantage of using local panels, is the shorter wire runs between the track and the panels.

Electronic switching

While switches are the simplest way to ensure cab control, power to blocks can be controlled electronically. When a relay switches one controller's power to a block, it prevents other relays from switching until the block is released by that operator.
Many home-brew designs have appeared to meet the needs of particular layouts.

Ready-made modules are available for automatic block control.
Heathcote Electronics have a range of modules that implement interlocking and also include acceleration and deceleration of the loco.

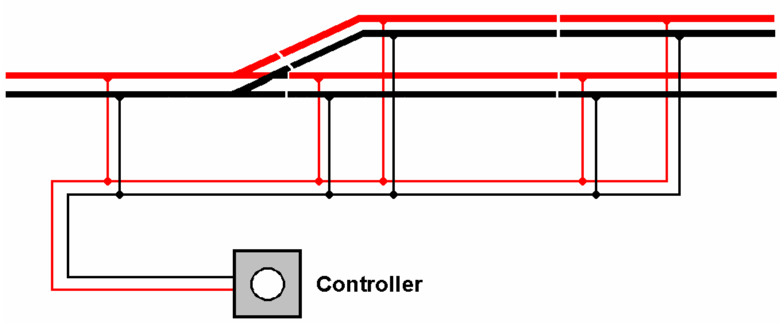
MERG's SuperBloc system produces electronic kits that ensure there is never more than one train in each section at any one time. Trains can be controlled manually by setting signals in each area, or run automatically via interconnected block controllers. It is designed for DC motors.

Finally, control of blocks could be achieved by linking electronic boards to a personal computer.

CTI Electronics market a range of boards, under the 'Train Brain' range for controlling DC locos, via a PC and MERG produce a range of kits for DCC in its CBUS range for standalone or PC working.

Bus control

This approach dispenses with lots of block switches
This illustration shows that a single pair of wires is taken from the controller and routed round the layout.



Wires are then run from each piece of track to connect to this 'bus'. So, all sections of the track have power at all times. Control of individual locos have to be controlled by other means – this is the basis of DCC-based layouts.
In some cases, a bus could be wired as a loop, although not for DCC.

Pros:

- The huge potential reduction in layout wiring.
- Fewer wires means the wiring is cheaper/quicker to install.
- Fewer wires means reduced maintenance problems (fewer wires to break, fewer connections to corrode, fewer wires to have to trace/colour code/document.
- Fewer wires means less interference between longs runs of parallel wires (noise, crosstalk, induction).
- Easy to extend. Adding an extra baseboard, or adding a terminus, does not require lots of extra wires to be run back to the central point. The devices are simply tapped on to the existing bus which is extended to the new work area.

Cons:

- Vulnerability to breakdowns. Since the entire layout relies on a single pair of wires, any breakdown in the bus could stop the entire layout from working. A system of cutouts to various sections is necessary to restrict the breakdown to the section that is causing the problem.
- Without using additional boosters (see below), the bus has to handle the entire current demands of all attached devices.

DCC (Digital Command Control)

DCC is becoming ever more popular due to its many extra facilities, including track wiring simplification. At the level used by most small layouts, a single controller is wired to all sections of track. The entire track system is powered at all times and special decoders are fitted to each loco. That allows the controller to send messages to specific locos on the layout, instructing them to move, change speed, change direction, switch on loco lights, make sounds, etc.
Locos are free to move around the entire track system without the need for track breaks, isolated sections, section switches or relays. Movement between stretches of track is seamless.
Pros:

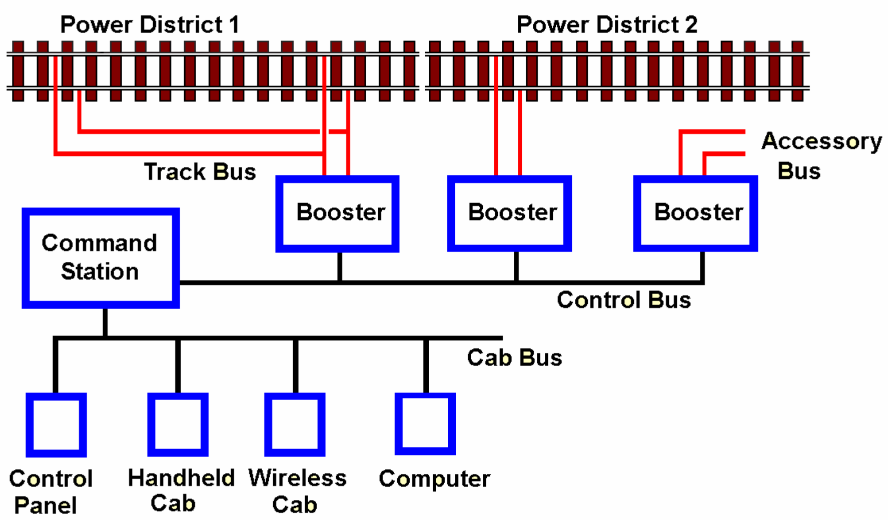
- No block/section switches or relays.
- Less wiring / fewer connection problems.
- Can run multiple locos – even on same stretch of track.
- Automation software, commercial and public domain, is easily integrated.
- Stay-alive capacitors can be fitted to locos. This minimises problems with insulated frogs and dirty track (particularly welcome for smaller gauges with short wheelbases).
- Points and accessories can be controlled via trackside decoders. These attach to the DCC bus and get their power and instructions via this single bus. No need for a separate accessories bus.

Cons:

- Expensive. Savings on wiring costs are wiped out by high cost of decoders.
- Requires initial tweaking of decoders (setting the CV, motor tuning).

Large scale DCC

Large DCC-based layouts create additional problems that need extra wiring.
If many locos, or locos that consume heavy current, are used then the total current demand may be greater than that available from a DCC controller. Up to a point, that can be satisfied by adding a booster to the controller.
Beyond that, the layout would have to be divided into separate isolated sections with each section having its own power booster.
Also. large club layouts usually have multiple operators, each controlling their own work area on the layout. The wiring system has to take that into account.
These problems are overcome at the expense of more complex wiring.
This illustration shows a large DCC-based layout that has locos with high current demands and multiple operators.



A single DCC command station still controls the entire layout and locos can move freely throughout the layout without any block switches. The layout is separated into a number of isolated sections that connect to the command station through separate boosters. The boosters have their own power supply and that means that the command station and its wiring does not have to carry the total current of the entire layout. Each booster powers its own district. The commands to control locos and DCC accessories come from the command station and are sent to each booster via its own bus – the Control Bus.
Another requirement for multiple operators is a connection between their hand-held controllers and the command station. This wiring connection is known as a 'Cab Bus' and may be specific to a particular commercial DCC system, as there is no NMRA standard for CAB buses. The Cab Bus (or Throttle Bus), as shown, also allows computers and/or control panels to communicate with the command station.

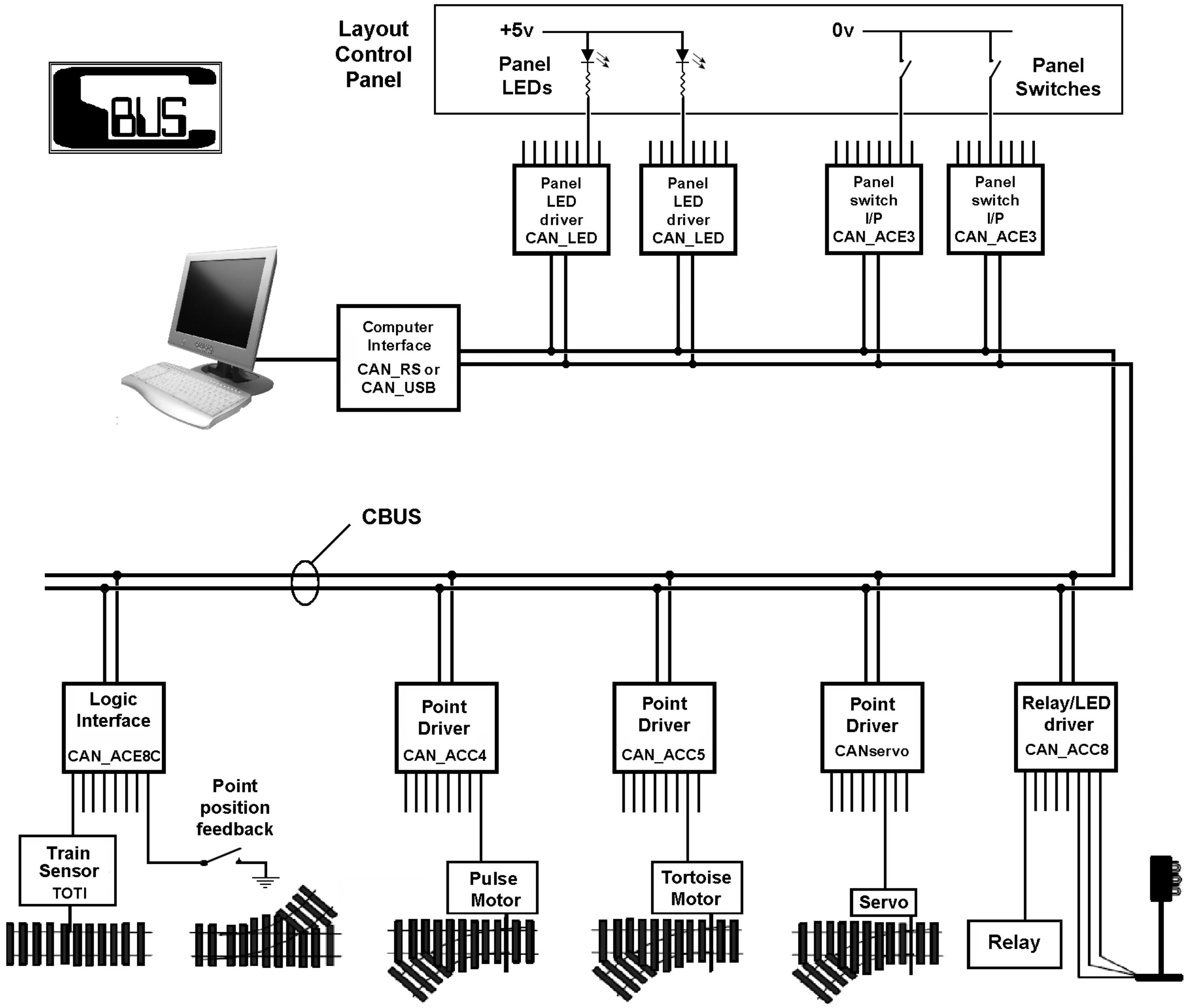
Other buses

Star and Bus wiring is not restricted to loco traction. It is also applicable to point control, signalling, lighting, etc. For example, if a factory used 20 LEDs for internal illumination, it would be very messy to have 40 wires routed back to the power supply. A single 'accessory bus' could take the power to the building, where is would fan out to individual LEDs. In other words, a bus to the building and a star formation inside the building.
Where heavy current peaks are expected, it is best to have a separate loco traction bus and an accessories bus. That way, if a loco suddenly pulls away, the lights don't dim in the

station; or if a group of solenoids are operated for setting a route, the loco does not suddenly jerk down to a slower speed. In fact, for larger layouts, it is usually advantageous to have a separate power supply for the accessories.

CBUS

MERG's own 2-wire layout control system has huge implications for layout wiring.
It is based on the distributed bus method and is extensively covered in MERG's Technical Bulletins (there are 10 different TBs in the G32 list).
There are also many articles in back copies of the MERG Journal.



This diagram, slightly adapted from previous MERG material, shows the versatility of the system.
Apart from power, all the devices that are shown interface with each other over a single pair of wires – the CBUS.

The devices include switches, LEDs, various point motor controllers, track occupancy indicators, relays, etc.
To that can be added the CAN_CMD, which is a DCC command station that be controlled by MERG hand controllers (the CAN_CAB) or by computer.

Back to Contents Page 5

This 2-wire bus can replace much of the traditional layout wiring.
For those beginning to plan a new layout, the CBUS system should be given consideration. Even for existing layouts, CBUS can be used to implement new additions such as signal and point operations, or track occupancy feedback to a mimic/control panel. This can be achieved without altering the existing track power arrangements.

Summary of buses

This chart looks at the wiring demands of the most popular layout control systems.
The simplest DC system only needs track power – a single bus. If you need to control points or accessories from a control panel, you need to run wires directly from the panel to each device. This includes the signal to operate the device and the power for that device. Similarly, if you want to have panel lights indicating track occupancy, you have to run signal wires back from every detector to the panel (plus the power for the detectors). In total, you would need a track bus, a power bus to take the 12V power round the layout, an accessory bus and a feedback bus – a total of four buses.

	DC	Simple DCC	Large DCC	CBUS
Track Bus	Direct from controller to track (maybe via switches or relays)	Directly from controller to track or via a booster	Via track boosters and power districts	Use CBUS with CAN_CMD and CAN_CAB
Accessory Bus commands	Direct from switches etc. to devices	Carried by the DCC track bus	Carried by the DCC track bus	Uses CBUS with CANACC8, etc.
Occupancy and other feedback	Separate bus needed	Separate bus needed	Separate bus needed	Uses CBUS with CANACE8C
Accessory power	Needs 12V feed	Via track bus	Via accessory bus booster	Needs 12V feed
Control Bus	Not required	Direct from controller to booster (only if booster is used)	From controller to multiple boosters	Not required
Cab bus	Not required	Not required	From all controllers to command station	From all controllers to CBUS

Note that the CBUS system can control traction and accessories and feedback from the one bus, with an additional power bus for the accessories and points. For an average layout, two buses are sufficient; an extra CAB bus would be needed if a larger layout had multiple operators.

Chapter 4

Track Wiring

Track wiring

The previous chapter looked at running a single train on a simple loop of track with a single power feed. A place where most of us started. Later, however, you add points and extra track, to allow for more varied running.
So that chapter discussed methods of laying the track as a collection of separate isolated sections, making possible more complicated running operations. These included Cab Control systems and large DCC systems using multiple boosters to feed many power districts.

Another reason for having isolated sections, both with DC and DCC, is to allow for track detectors to detect and act upon track occupancy (knowing that a loco is sitting on that section of track, whether moving or stationary). For example, the MERG TOTI-4 and TOTI-12 kits detect tiny currents when a loco is sitting in an isolated section. How the feedback from these detectors can be used is covered in a later chapter.

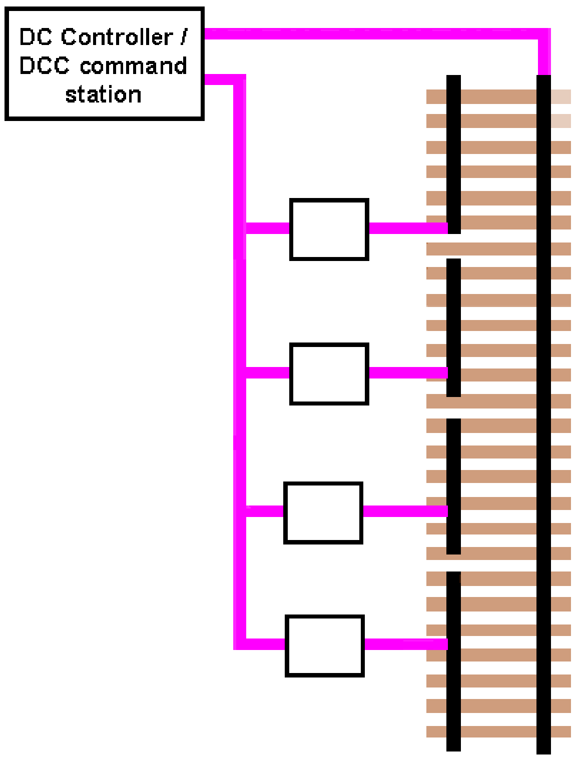
There are two approaches to wiring isolated track sections:

- Common return
- Two wire feed

If you have not yet laid your track, you may want to look at these two options and decide what is best for you.

Common returns

One wire from the controller is fed to one rail as shown in the illustration.
The other rail is cut into sections and the pieces are joined with insulated rail joiners. Each section is connected back to the controller through a device shown here as a blank box.
For a cab control system, these boxes represent switches. This means that individual sections are only powered when their corresponding switches are thrown.
Alternatively, the boxes might represent track occupancy detectors, such that the only detector triggered is the one that is passing current (through the loco's motor).



Pros:

- Simplifies wiring. Half the wiring of the twin feed approach.
- Simplifies fault-finding. If one section stops working, there is only a single wire to trace for a fault.

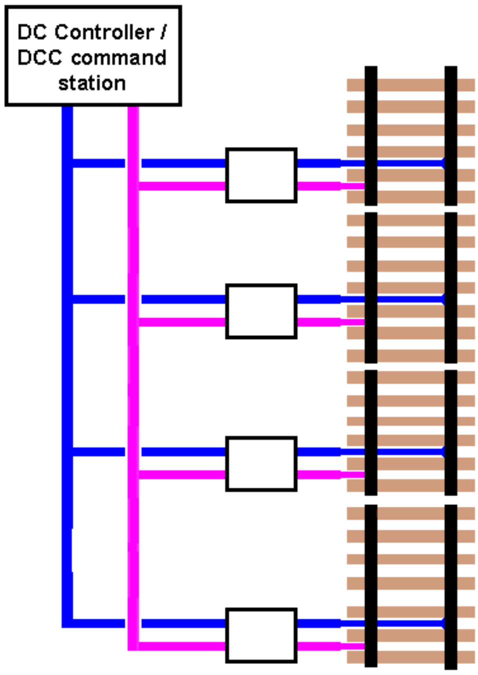
Cons:

- If using multiple boosters, the common return must have a very high amperage rating, as it has to carry the sum of all the boosters' currents. For example, if six of the sections were consuming 1A of current, the return wire has to carry 6A.

- If a single common wire is used to connect the return wire from the track, solenoids, detectors, etc., there is an increased risk of interference (e.g. a pulse from the solenoids affects the detectors resulting in false triggering).
- Reverse loops still need to have isolated sections, to avoid shorts.

An improvement could be using separate returns for different high and low current devices. So, for example, the power bus, traction bus and solenoids could share a common return. These are all high current wires but are relatively immune to interference.
Low level signal wires, such as track occupancy detectors and logic signals for motorised points could then share a common return that is unaffected by the large current pules on the other return wire.
Note that although the diagram shows a single feed to the common return rail, multiple feeds to different parts of the common rail are used to minimise voltage drop.

Twin feeds



To isolate a piece of track, both rails are cut, instead of one. Every piece of track has both rails isolated from adjacent track sections and each section is is fed with two wires, one for each rail.
For block control, this means use a double-pole double throw (DPDT) switch, so that both rails are connected when the switch is thrown.

Pros:

- Reduces unwanted interactions. Since no sections share a common return wire, the chance of changes in one section affecting another section is greatly reduced.
- Easier to make future alterations, as all rails sections are already wired under the baseboard. There is no need to cut track and insert insulated joiners. You simply rewire your under board connections to what your new needs are, without touching anything above the baseboard.

Cons:

- Many more wire runs. Twice as many wires to connect.
- More scope for faults. Twice as many connections to go faulty.

Positioning track breaks

As we have seen, there are a number of circumstances when we want to introduce isolated blocks and sections:

These include:

- For DC Cab Control.
- For DCC power districts.
- For DC/DCC track occupancy detection.
- For programming the CVs of DCC decoders.

DC Cab Control

The layout's track is cut into isolated sections that meet the operational needs of the user. Sometimes a long stretch of track will be considered as a single block. If, however, you want to run several locos on that stretch at the same time, the track has to be further sub-divided. On most occasions, a single siding has to be isolated (or at both ends for a passing loop. The plan has to cover all track sections where a loco might be reasonably expected to stop, and also take into account the likely maximum length of the train on each particular isolated section.

Power to the track sections is routed through switches/relays.

DCC power districts

Power districts are introduced to spread the heavy current demand in the layout. This is usually caused by concentrated activities in distinct 'work areas' such as main line stations, goods depots, marshalling yards, fiddle yards, etc. These work areas can be isolated from each other and be fed from separate boosters, each booster taking a share of the overall current demand.

Power to the track sections is routed through DCC power boosters.

Track occupancy detectors

The isolated sections should be chosen according to the function they are to fulfil. For example, you may only want a single occupancy detector, to operate crossing gates. Another option is allocate a detector to each of a layout's hidden sidings, to indicate which are currently occupied. You might even consider having many occupancy detector sections all round a layout, to provide feedback for mimic panel updates or computer automation.

Power to the track sections is routed through the current detectors.

Programming track

DCC-equipped locos have their decoders programmed using a SPROG programmer, or the programming output from a DCC command station. The loco has to sit on a stretch of track that is dedicated for programming locos and is isolated from the rest of the layout, to prevent accidental re-programming of other locos. Some users prefer using a small piece of track on their workbench, to avoid any possible mistakes.

Power to the track comes from the SPROG programmer or a DCC command station's programming output.

Feeds and droppers

Track buses carry power all round the layout but they don't connect directly to the track rails. The track bus has to carry all the current needs of an entire layout, or an entire block or section. To minimise track resistance and connection problems, the track bus is best wired as a pair of continuous heavy current conductors that makes its way along the layout under the baseboard.

Short wires are then taken from the track bus and fed through holes in the baseboard to connect to the track rails above. These wires can be of lower current capacity, since they don't carry the total layout current.

These short wires are known as droppers (because they drop through holes drilled in the baseboard) or feeders (because they feed the power from the track bus to the rails).

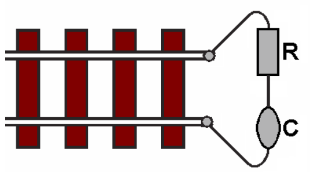
There are conflicting opinions on how many droppers should be used for any one section of track, with advice varying from 18” to 12'.

However, there is general agreement on:

- If a section is made up of two pieces of track, take a dropper to each piece – do not rely on getting a good contact from the rail joiners. Alternatively,some solder both rail ends to the rail joiners, although this can lead to problems when the track expands/contracts under changing temperatures.
- Droppers should be soldered to track rails, not to joiners.
- For long stretches of isolated track, use more than one dropper. This ensures continuity should one dropper fail. It also improves performance (less voltage drop, since track buses have lower resistance than most track rails).
- Use the thickest dropper wire you can manage, taking into account aesthetics and any soldering difficulties that may occur with smaller gauges.
- Droppers should be as short as possible. This prevents the introduction of a high resistance section into the current loop, which might prevent the short circuit protection system of a controller or booster from operating (the current may be sufficient to cause damage while not being high enough to trip the protection).
- Smaller gauges have track rails with much smaller mass and therefore higher resistance. It follows that a long run of z gauge track would need more droppers than the same length of O or G gauge track.

DCC bus terminators

DCC systems generally run perfectly happily, with small distortions of the DCC waveform being handled by the decoders.
However, where the bus is over 10m, or where you are experiencing decoder efficiency problems, terminators can be fitted to each end of the bus.
These devices are simple filters that prevent voltage spikes appearing on the bus and upsetting (or damaging) decoders. It is intended to improve the DCC waveform and thereby improve DCC reliability.



Terminators can be bought or can be made from two components as shown. The resistor can be around 100 to 150 ohms 1W or 2W, and the capacitor is 100nF (0.1uF) ceramic at 50V working. The two are wired in series then wired across the ends of each bus.

Wiring suggestions

These are drawn from the experiences of others and should be given due consideration.

Routing wires

The track bus should be as short as possible (to minimise voltage drop), while following the main track as closely as possible (to keep droppers as short as possible). Sometimes, you may have to run spurs off the main bus to avoid long snaking runs for the main bus.
Of course, the track bus (or buses) are not the only wires on the layout. Under the baseboard, you may have buses for accessories, lighting, point operation, track occupancy feedback, etc.
Since every piece of electrical equipment has the potential to generate magnetic fields and also to pick up these unwanted radiations, consideration has to be given to preventing mutual interference.

This means that wires from feedback devices (e.g. track occupancy detectors) should not be routed close to DCC buses or, even worse, point solenoid wires or decoupling magnets. Similarly, wires to heavy current motors should not be run close to DCC buses, to minimise motor commutation spikes from upsetting DCC signals.

Where there are multiple DCC buses (from boosters), they should be spaced apart to prevent 'crosstalk' – the signals from one booster being picked up on the neighbouring bus.
All this suggests that you should avoid running all the cable together in trunking, or bunching to pass through the same hole drilled in the baseboard frame.

Twisting wires

DCC system rely on having clean signals that the decoders can understand. Any external interference (from loco motors, point solenoids, radio transmitters, etc.) can modify the signal and prevent reliable operations.
The longer the track bus wires, the greater is the danger of interference. While this may pose few problems for small layouts, precautions are recommended for large layouts. A bus wire acts like an aerial, picking up external interference and converting it into unwanted voltages on the wire. So, if the two bus wires are twisted together, the unwanted voltage on one wire is cancelled out by the voltage being induced in the opposite polarity in the other wire.
There is no need to buy special twisted pair wire. Just twist the two wires loosely round each other, approximately two to four twists per foot.

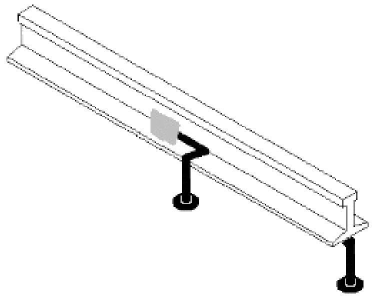
Connecting wires

Good, low resistance connections are essential for trouble-free operations, with reliable metal-to-metal contact. This requires making 'gas tight' connections to prevent the build up of oxides (poor conductors) between the wires and the terminals.
Soldering is the best way to ensure an air-tight connection.
Screwed connections can be successful as long as precautions are taken to ensure that the wire ends and the screw surfaces are clean, shiny and free from dirt, corrosion, rust, oils, etc. If the wire end or terminal needs cleaning, the connection should be completed as soon as possible, before surface oxidisation starts.

Fitting droppers

The illustration shows two methods of soldering droppers to the track rail.

On the left, the dropper comes up through a hole in the baseboard and is bent as shown before being soldered (don't forget about cleaning the track area, tinning that area and the wire end before soldering). This method is the only option if the track is already laid.



On the right, the dropper is soldered to the underside of the rail <u>before</u> the track is laid. The other end of the wire is then inserted through the hole and soldered to the track bus. This method is used where some thought has been given to the locations of track blocks/sections prior to laying a section of track.

What current does a layout need?

The table shows typical current demands for one of each device.
These values will vary depending on different manufacturers of the same product.
The current demand can vary from as little as a few milliamps (e.g. LEDs) to 5A for a point solenoid. For locos, it can be under 100mA for a Z gauge loco and as as high as 20A for a long consist of O gauge locos.

Typical current demands

Locomotive motor: Z scale and N scale TT, 00/H0 scale O scale	100-300mA 250-500mA 400mA-1A
Twin-coil turnout (point) motor e.g. Peco, SEEP	5A pulse
Motor-actuated turnout motor e.g. Tortoise, Conrad	15-20mA
R/C servo turnout motor	100mA
Memory wire actuator (turnouts and signals)	200mA
Light Emitting Diode (LED) (signals, train headlights / taillights)	5-20mA
Incandescent lamps (signals, train headlights / taillights)	40-80mA
Relay coils	20-50mA
CBUS modules	20-70mA
Track occupancy detectors	2-70mA
Signal wires (from TOTIs, to Servos, to/from CBUS modules)	2mA

We have to ensure that the layout's wiring can meet the demands of all the locos and other devices that may need to run/operate at the same time.
Let's see what wire types are available and then decide what's best for a layout.

Wire standards

As usual, why have one standard when you can have several? Why learn about them ? Because you will often find them referred to in model railway magazines and on the internet. More importantly, they are used as descriptions of wire being sold by Rapid, etc.
The American Wire Gauge (AWG) is a U.S. standard for wire gauges. Higher gauge number indicate smaller wire diameters, with thicker wires having lower gauge numbers. Although an American standard, it is often quoted in UK and European descriptions.
There is also an Imperial Standard Wire Gauge (SWG) and its values are not directly equivalent to AWG values. SWG is not often used.
Probably the most understandable is the metric measure, which describes the wire in terms of its diameter or its cross-sectional area (CSA) measured in square millimetres. Unlike AWG, higher metric

AWG	Dia (inches)	Dia (mm)
10	0.1019	2.59
11	0.0907	2.30
12	0.0808	2.05
13	0.072	1.83
14	0.0641	1.63
15	0.0571	1.45
16	0.0508	1.29
17	0.0453	1.15
18	0.0403	1.02
19	0.0359	0.91
20	0.032	0.81
21	0.0285	0.72
22	0.0254	0.65
23	0.0226	0.57
24	0.0201	0.51

measurements indicate thicker wires.
Wires are sold either as a single solid conductor, or as a multi-stranded cable.
Multi-strand cables are described in the metric system by giving the number of strands, followed by the diameter of each strand. So, for example, a wire described as 7/0.2 is made up of seven strands of wire where each strand has a diameter of 0.2mm.

AWG, on the other hand, does not make a distinction between solid and multi-strand.
So for example, an AWG value of 24 could be a single strand of 24 AWG, or it could be made up of seven strands of thinner 32 AWG wire (which together result in the same current-passing capacity as a solid 24 AWG conductor).

Current ratings of different wires

As expected, thicker wires can carry greater current than thinner wires of the same material. This tables describes some the wires that are available from Rapid Electronics and other suppliers.

Number of wires/ diameter of each wire	Current capability	Available colours
10/0.1	0.5A	11
Ribbon cable – 7/0.127	1A	11
7/0.2	1.4A	11 (also 16 bi-colour)
Security alarm cable (7/0.2)	1.4A	4,6 and 8 core
1/0.6	1.8A	11
16/0.2	3A	11
32/0.2	6A	11
Twin and earth 1mm^2 (lighting)	10A	2
Twin and earth 1.5mm^2 (lighting)	15A	2
Twin and earth 2.5mm^2 (power)	20A	2

Notes:

The current ratings shown are a good indication of what can be expected under normal conditions. Some manufacturers are reluctant to stipulate an exact value, as the current handling can be affected by external conditions such as the ambient temperature and problems with heat dissipation.

The picture shows a piece of 'twin and earth' cable. It has three solid conductors, two being insulated and one being bare. Normally used for house wiring, it is often stripped of its outer plastic sheath and used for carrying heavy currents round a layout. Although often specified for higher currents, the table shows a more conservative estimate to allow for the inevitable voltage drop at such high currents.

Some cheaper wires (e.g. bell wire) may not be made from copper, instead being iron with copper plating. This increases the wire's resistance significantly (perhaps a factor of 5 times) and lowers its current carrying capacity.

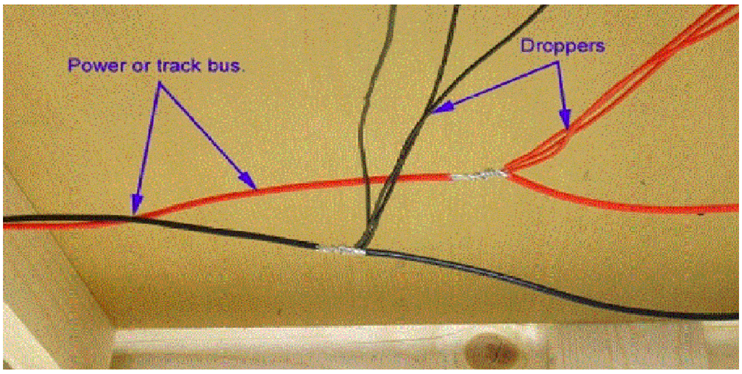
Selecting wires for layouts

The earlier chart showed the current demands of various elements of a layout.
The wire chosen for a particular job should be able to supply the current needs of the maximum amount of devices that could be operating at the same time – plus an extra margin.

Track bus

This is likely to have the heaviest current and the one most likely to suffer from shorts (e.g. locos derailing). So, the wire should not be chosen for the maximum loco currents, but for the maximum current in the event of a short.
You cannot always rely on a controller's cutout.
Consider a 12V, 5A controller, where the bus wire's resistance is such that the circuit current can never reach 5A. In a situation where the supply can provide 5A but the short only results in 3A , the controller cutout won't trip. However, since power is calculated as I x V, the circuit (the track bus and probably part of the track) will dissipate 3 x 12 = 36W!
Clearly it is necessary to ensure the lowest possible resistance in a track bus, particularly in the larger gauges where loco currents can be many amps.
Since the track bus runs the length of the layout and is hidden under the baseboard, it pays to use the thickest twin and earth wires that you can afford.
One solution lies in sub-dividing the layout into blocks (DC) or power districts (DCC), where the current in any one section is reduced.

Droppers



The track bus carries the controller's power across the entire layout.
It does not normally appear above baseboard level.
To connect the track sections to the track bus, small connecting wires are soldered to the track, poked through holes in the baseboard and then connect to the track bus.
To prevent unsightly connections between the droppers and the track in smaller gauges, it is common to use thinner wire than that used for the track power. This is not a problem, as long as the dropper lengths are short. Although these wires are thinner and therefore have a higher resistance, the fact that they are short means that they do not much alter the overall resistance of the circuit.
The picture shows part of the insulation on the track bus being bared back to allow droppers to be soldered to them.

Points

Solenoid-operated points are designed to pass a large current for a short time and are usually used in conjunction with CDU (capacitive discharge units) modules such as MERG's PD3 kit.
I measured the coil resistance of a Peco solenoid coil at just over 5 ohms and an old H&M coil at just 2 ohms. In theory, a CDU's pulse of 20V should result in a pulse of around 4A through a Peco solenoid and a whopping 10A through an H&M solenoid. Where such small

coil resistances and such large currents are in play, the presence of even a small unwanted resistance can have a major impact on performance.   Wiring for solenoid power, therefore, should use either 1.5mm or 2mm wire. The same considerations apply to electromagnets used for uncouplers.
For these reasons, it is best to locate CDUs as close to the points as possible, to minimise long runs carrying high currents.   After all, the signal that controls the CDUs only requires small switching currents.

Motorised switch machines, such as the Tortoise, operate with much less current (from 4mA to around 20mA) and will happily work with 7/0.2 wire or even ribbon cable – unless you have particularly long runs to the point motors.

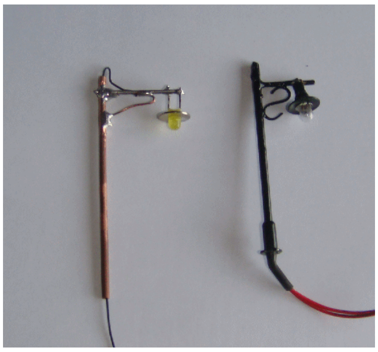
Servo operated points are becoming more popular for a variety of reasons (cost, speed, controlability).   An individual servo may only demand a few hundred milliamps while operating.   On the other hand,, moving multiple points (as in route switching) increases the current demand substantially during movement and that has to be taken into account when running power wires to servos.

Accessories

The connecting wires to each accessory depends on the current draw of that accessory.
In most cases, 7/0.2, alarm wire or ribbon cable is more than adequate.
If you have a large motorised accessory, you would use 16/0.2 or higher.

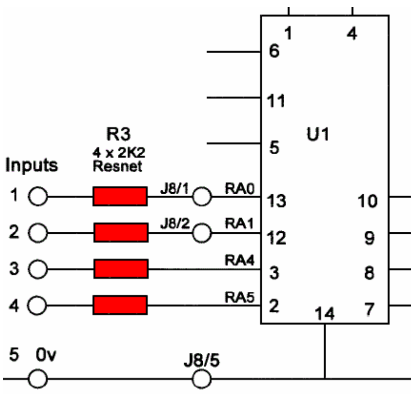
Lighting

Grain of wheat bulbs use around 80mA at 12V.
Lighting a row of buildings or a chain of street lamps soon totals a substantial current draw.
To reduce wire thickness and reduce power requirements, most modellers are turning to LED lamps for illumination
The picture shows a commercial OO lamp using a grain of wheat bulb that uses 60mA.   On its left is a home-built equivalent using an LED that uses just 5mA.
Even the thinnest of wires can handle the current requirements of a chain of LEDs, particularly if they are wired as groups (where a group may consist of 3 LEDs in series).



Signal wires

Layouts may use a range of different signal wires such as feedback from occupancy detectors, signals to drive servo boards and the inputs and outputs of CANBUS modules.
This diagram shows the input signals to a MERG Servo4 board.   Each input has a 2k2 resistor in series. Since the voltage to any input is either +5V or 0V, the maximum current that can be drawn is 5/2k2 (a couple of milliamps).
Similarly, the outputs of a CANACC8 have 10k resistors in series, limiting their current draw to half a milliamp.
Consequently, 10/0.1, 7/0.2, alarm cable or ribbon cable can be used for signal wires.



Suggested wire usage

10/0.1 or ribbon	0.5A	Colour-light signalling, individual motor-actuated switch machines (Tortoise Conrad, etc.), LED lighting, signal wires to Servo4s, from track occupancy detectors and to/from CBUS modules.
7/0.2	1A	Track feed droppers.
16/0.2	3A	DC traction feeder, DCC power Bus (booster output up to 3A), twin-coil solenoid point motors
32/0.2	6A	DCC power Bus (booster output up to 5A), DC Bus, twin-coil solenoid point motors
Twin and earth	10A , 15A, 20A	DCC power Bus (booster output up to 10A), DC Bus

Stranded v solid wire

Copper wire is available as a single solid conductor or as a set of thinner strands twisted together. To some extent, it is a matter of personal preference, although some factors have to be considered.

Stranded wire, by its construction, is much more flexible and is really essential in situations where the cable will be flexed (e.g. hand-held controls, or trailing plugs and sockets for joining multiple baseboards) or subject to excessive vibration. In these situations, solid conductors would suffer from stress fatigue and probable eventual failure. The thinner the strands used in the cable, the more flexible it is.

That is not to say that stranded wire may not eventually suffer from stress fatigue (I have replaced my multimeter leads three times due to broken wires inside the insulation or where the wire meets the plug). Solid wire, specially twin and earth wire, is often used for power buses under the baseboard where it is a permanent installation. Stranded wire can be used here too, although many prefer the rigid, non sagging, nature of solid conductors.

Solid wires take longer to corrode than stranded wires (less surface area) and can be useful for garden railways.

Skin effect?

DC current travels along the wire using the total cross sectional area of the copper.

When high frequencies travel along a conductor, they tend to travel close to the outer surface of the conductor – known as *'skin effect'*. The higher the frequency, the thinner the skin depth (how far into the conductor is capable of carrying the current).

This has led some to advocate the sole use of stranded cable for DCC signals, since they have multiple surface areas. Of course, since multi-stranded wire do not insulate the individual strands, this benefit is hard to realise.

Although DCC is not an AC waveform, it changes its logic state at the approximate equivalent of 7KHz. According to the online calculator at

http://chemandy.com/calculators/skin-effect-calculator.htm

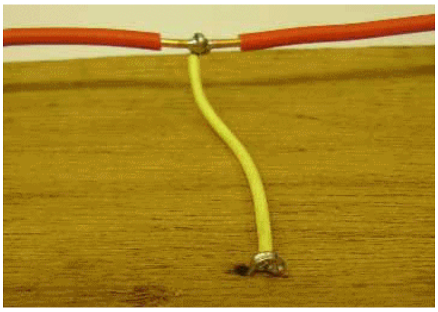
the skin effect depth for DCC signals is 0.779mm. This is over 1.5mm across the conductor, so the only wires directly affected should be those with a smaller diameter. On the other hand, as the DCC power bus is required to pass many amps, it will not be using any wires as thin as that in the first place. The skin effect for CBUS messages, which operate at 125KHz is given as 0.184mm (or a diameter of 0.368mm. Again, this is greater than our most likely wires in use.

It can be argued, however, that the DCC and CBUS signals are square waves that generate multiple harmonics of its basic frequency, hence making the problem worse than first expected. However, over 98% of the current is able to flow within a layer that is four times the skin depth, counteracting any potential problems.
Skin effect, therefore, is really only a problem when working at much higher frequencies than those used in model railways.

Types of connection

The layout will have many connections between wires, between wires and devices and between wires and various connectors. Users have employed and advocated a large range of connection types, with some controversy being generated as a result.
The descriptions that follow explain what connection components are available.
Users can decide which of the following methods most suits their needs.

Soldered



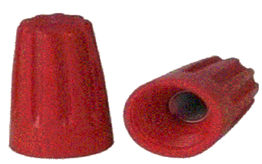
For many places, this is the most secure, most permanent, most trouble-free connection, if done correctly.
You can expect many years of problem-free use from soldered joints, assuming there was a good clean mechanical joint and soldering was carried out efficiently. The picture shows a dropper wire being wrapped round a power bus before being soldered.

WAGO 222

These terminal blocks are designed for handling 240v domestic lighting at 32A. They are intended for permanent installation in homes, where a robust and reliable connection is essential.
They are not straight through connectors. They are available in strips of 2, 3 or 5 lever connections, with all the sockets joined internally.
They are useful for power fanouts, with the main feed and the auxiliary feeds sharing the block.
Their use requires no special tools, just strip the wires and insert their bare ends into the holes before closing the levers. It is a very quick assembly process which is helpful when used in awkward and hard to reach locations. Wago terminal blocks can accommodate up to 2.5mm solid wires, although it also works with stranded wires.

Of course, not all connections lend themselves to these two methods, since they may have to allow for devices to be plugged/unplugged or baseboards to be attached/detached on a regular basis.
There, a more flexible approach is required.

Wire nuts

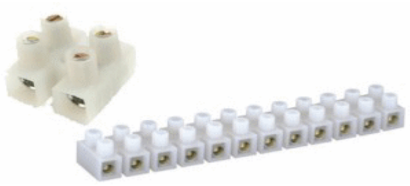


These are also known as cone connectors, thimble connectors or Marrette connectors and are sometimes used as an alternative to terminal blocks. They are plastic cones with a coiled metal insert. They are specially effective for connecting stranded wires.
They are available in five sizes and are colour coded to indicate the

gauge of wire they are intended for.
To install, just bare the wires, insert them into the cone and twist the cone. The wire-to-wire and the wire-to-insert contact make a reliable connection that can be readily undone by untwisting the cone.

Screwed

The most common non-soldered connections make use of terminal blocks (also known as 'chocolate blocks').

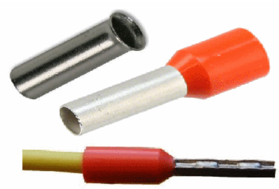


These are straight thorough connectors, with screws on either side of each connection. They are often sold as strips of ten and can be easily cut into smaller blocks with a knife.
There are holes between each connector, allowing easy screw mounting to a baseboard. They are cheap and come in different sizes, to provide 3A, 6A, 10A, 15A, 16A and 30A versions.
There are a couple of things to watch out for when using these blocks.
As Howard Watkins, MERG Publicity Manager, recently pointed out:

> “Some people prefer to solder the wire, especially if several wires are to go in the same terminal block position. This should be avoided because solder is a soft metal and tends to squeeze out from between the strands. As time progresses the joint can become loose and the connection can become unreliable.”

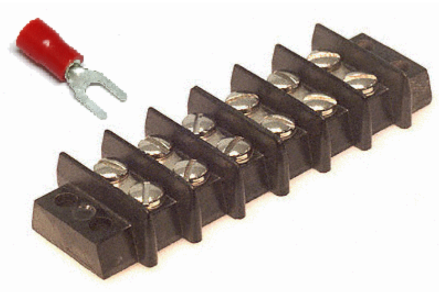
Another danger is not having every strand of wire securely held under the screw. If, for example, three strands of a 16-strand wire ended up as stray unconnected wisps, you have lost almost 20% of the connections current carrying capacity.
One way to prevent this is to use crimp ferrules.

Crimp ferrules



Where multi-strand wire is used with terminal blocks, consideration should be given to the use of crimp ferrules. These are simply metal tubes into which the bared end of a wire is inserted. The tube is then crimped to tightly grip the wire, as shown in the bottom example.
The ferrule can then be inserted into the terminal block and screwed down tightly. This provides a stronger, more resilient termination than using bare wire alone.
The illustration shows both bare and insulated versions (the bare version is also useful for chimneys on buildings or locos). The insulated version provides additional protection against stray strands of wire causing shorts or increased resistance.

Terminal blocks



These chunky connecting blocks were developed for high current connections in a range of industrial devices and power supplies.
They comprise two rows of screws and are produced in block from 3-way to 12-way.
Often referred to as *'barrier blocks'* because of the protective wall between each pair of screws.
They are available in ratings from 15A up to 45A and some have a protective cover to prevent accidental shorting, spillages, etc.

They are popular among some hobbyists due to their ease of use and their large contact surfaces. Wires can be connected by baring the ends, curling the bare ends into hooks and inserting under the screws. An alternative method, shown in the picture, is to terminate the bare wire end in a crimped spade connector.
One way to help ensure a gas tight connection is to place a star washer between the wire/spade and the head of the screw. As the screw is tightened, the star serrations bite into both the wire/spade and the screw head. This cuts through any existing layer of oxide and ensures a good contact.
This is particularly useful where the connections may be disturbed or altered, as the washer ensures a good connection each time.

PCB terminal block

These blocks provide a simple way to connect wires to a printed circuit board module and are commonly found on MERG kits and other PCBs.
Each connection uses a pin to solder to the printed circuit board and a screw terminal for connecting a wire. The bared end of a wire is inserted into the hole and the top screw is tightened.
Despite their small size, they have large current carrying capacity, with some handling as much as 57A. Even the smallest in the range can handle 6A.
They are available in different pitch widths (the distance between the pins) as the larger current carrying versions uses wider pitch widths. Pitch widths range from the commonly-used 3.5mm and 5mm up to 10.16mm. Stripboard has a 2.5mm pitch and can accommodate a 5mm pitch, although not on adjacent tracks.
The blocks are available in a range of sizes, from 2 way to 7 way and many types can be interlocked to create custom widths.
PCB terminal blocks provide for the quick and easy removal or replacement of a module, requiring only a screwdriver.
Many of these blocks use a *'rising clamp'* method of securing the wire. As the screw is tightened, the wire is gripped between two plates – one fixed to the top of the insertion hole and the other which rises as the screw is tightened. The two flat plates provide an even contact over all of the wire's strands. This results in better contact with the wire and less chance of stray wires causing problems.

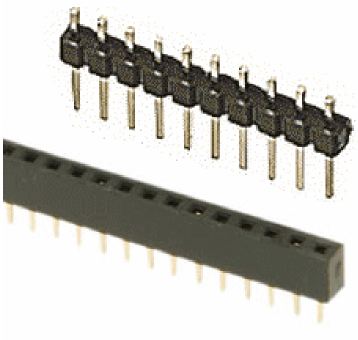
Note, however, that with some blocks, there is no physical contact between the top and bottom plates until the screw is tightened. This can sometimes lead to mistakes when taking meter readings, when the meter prod is touched the electrically unconnected screw.

Crimp connectors

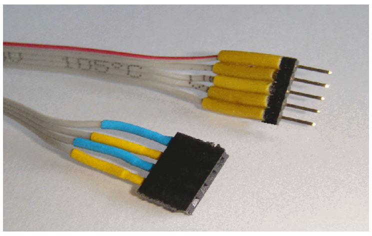
The pictures shows a male and a female connector.
They have a large surface area and make a tight fit when connected.
This ensures a reliable connection and they can be found in domestic (boilers, hot tubs), car audio wiring (resistant to vibration) and industrial machinery.
The bared end of a wire is inserted into the connector tube and the tube is then crimped to tightly grip the bare wire. Although pliers and wire cutters are often used, crimping is best done with a proper crimping tool. These connectors handle solid conductors and work at up to 10A.

PCB interconnect

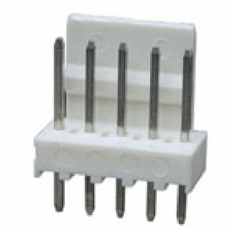
As the picture shows, these are produced as header plugs and matching PCB sockets.



They are used on MERG kits, as jumpers on Servo4 modules and as ICSP connectors on CBUS kits, etc. The sockets are available in a range of sizes from 2 way to 25 way, and in single or double rows. They can be cut or sawn to create any length of strip. They have a 2.5mm pitch, making them ideal for use with home-grown projects built on stripboard.
The picture shows both types with ribbon cable soldered to them and protected by heatshrink tubing.



This provides a very neat way to achieve cable runs of multiple wires. When only three pins/sockets are used this way, it also makes a cheap alternative to commercial servo extension leads.
These connectors were designed for light duty and are only suitable for running signal wires and low current devices.



Another PCB connection is the Molex plug and socket system.
As you can see, it designed such that the plug can only be inserted into the socket one way. This polarisation prevents the accidental insertion of power leads or output leads in the wrong polarity and is commonly used for connecting power to PCBs.

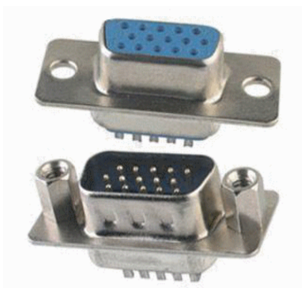
DIN

DIN plugs and sockets have been around for a very long time and have been used for audio connections, MIDI connections and the old AT computer keyboard cable. They are available in 3-way, 5-way, 7-way and 8-way versions and are keyed so that the plug can only be inserted one way (no accidental polarity reversals).

They are also available as mini-DIN versions as used on the mouse and keyboard cables of older computers (known as PS/2).
Standard DIN connectors are cheap but only have a maximum rating of 2A. The socket can be mounted to the baseboard and the plugs can be wired to flying leads. An old keyboard can have its cable cut off and then you have a ready wired DIN plug and cable.

D-type

D-type plugs and sockets were designed for use with computers and data communication equipment. The various versions include the 9-pin and 25-pin (RS-232 serial ports), the 15-pin (monitor VGA ports) and the 37-pin (RS-449 high speed serial interfaces). They are rated as either 3A or 5A.

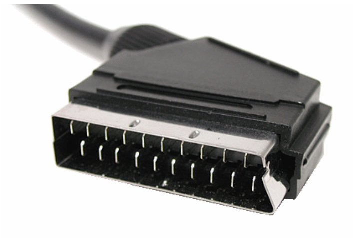


They can be found interconnecting model railway baseboards, due to their large number of contacts. Indeed, the larger versions can wire together two or more pins for a single connection – increasing the current carrying capacity and providing redundancy (if one connection fails, there is still one or more others to carry the current). However, the rear pins of the plugs and sockets can be difficult to solder wires to, particularly if the wires are thick (7/0.2 seems to be about the thickest that's usable).

The pins are bunched very close together, requiring skill/patience to avoid shorting adjacent pins (with solder blobs or stray strands of wire posing the greatest problems)
Alternatively, eBay sell a 2 metre female to female 9-pin serial cable for a couple of pounds. Simply cut it in half and you have two ready made cables. For around the same cost, buy a 25-pin male to female cable, cut it in half, and you have two trailing leads for connecting two baseboards. Make sure all the pins are wired and check for any wires crossing over.

SCART

Before HDMI emerged, most connections between DVD players, satellite boxes, TVs, etc. used SCART cables and many are still in use. SCART connectors offer 21 pin connections, although not all pins wired through in some cables.
They are low voltage low current connections and are not suitable for carrying high currents. The large pin count makes them a possibility for carrying multiple signal wires to and between baseboards (for LEDs, servos, etc.).
Like the D-type cables, a SCART cable can be hand wired or you can buy a ready-made cable and cut it in two.



Circular locking

If you want to ensure that you have a high-current connection that won't shake loose or drop out, a screwed connection is available.
As the picture shows, the socket is designed for panel mounting, although it could be mounted in a small case with a trailing lead if required. The ends of the plug and the socket are threaded and can be screwed together for a sound mechanical joint.
The picture shows a 3-pin 5A version, which might be considered for taking power from power supply or controller to a layout.
Bulgin market 32A versions, varying from 2-pin to 10 pin.
These connectors are bulky but merit consideration for 'must work' connections.

Durability

An important issue when choosing connectors is their durability- measured in *'mating cycles'* (how many times you can plug and unplug the connection before good performance is no longer guaranteed). If you have a home layout made up from several interconnected boards, this may not be an issue, as you will only occasionally disconnect a board for maintenance/upgrading.
For a club that carries out a lot of exhibiting, it can be an important consideration.
It pays to check a particular dealer's product specification, as quality varies. D-sub connectors can have mating cycles from as low as 50 to hundreds, DIN connectors vary from 1,000 to 5,000 cycles, SCART connectors from 750 to 3,000 and Molex from 1,500 to 10,000.

CBUS

Some of the connectors described above allow for huge amounts of wires to be connected together. With even the best of care, there is much scope for faults to develop.
If you find you need such large numbers of connections, you might wish to consider the merits of the 2-wire CBUS interface.

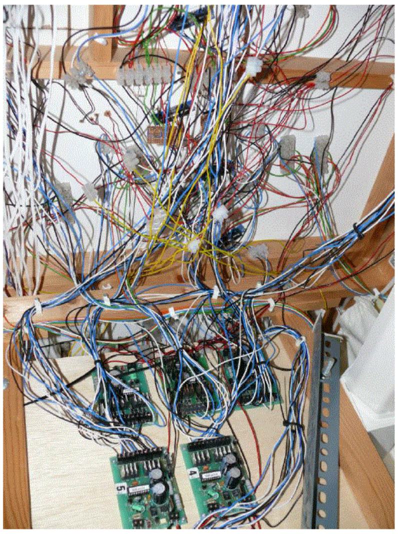
Maintenance aids

The greatest aid to maintaining a model railway layout is not a multimeter – it is <u>knowledge</u>. To be able to test a layouts' wiring, you have to know where wires are supposed to go and what conditions should be expected on them. As you build a layout, more and more wires are run, more components are added and configured. Unless this is approached in a methodical manner, fault-finding will be much more troublesome. Problems will start to occur long after you have forgotten what you laid where and for what purpose.

The time spent in developing and sticking to a plan for rolling out a layout is more than repaid when the layout has to be maintained.

MERG's Technical Bulletin LC01 provides details.

Colour coding



Imagine a single wire running through a rat's nest of cabling.
Consider the problem of tracing it along the length of a layout.
Life is much easier if you settle on a colouring scheme for wires
For example, every time you see a red wire, you know it carries +12v, an orange wire carries +5v, a yellow wire carries the output from a track occupancy detector, and so on.
There is no universally recognised 'standard' colour coding scheme, although red and black are widely used for main +ve and -ve wires.
Digitrax have their own suggested wiring scheme, see:

http://www.dccwiki.com/Wiring_color_code

For a home layout. The colour scheme is entirely up to you, as long as you record it and stick to it.
For a club layout, where many different people will be working on a layout and developing and maintaining it over a considerable period, an agreed colour code needs to be developed and recognised by all members.
Fortunately, wires are available in a range of colours.

Labelling

Unfortunately, there is not an inexhaustible range of coloured wires available, so you are liable to find many wires of the same colour running under a baseboard. Although it is useful to know what it is used for (e.g. wires to Servo4 boards), it does not tell you where the wire originated from and where it is heading. Attaching small labels to wires during installation can provide very valuable information at a later date. For example, a label that says 'ACC84/3-Ser3/2' would let you know that the wire ran from your fourth CANACC8, pin 3 to your third Sevo4 module, pin 2. Some people prefer to use a numbering scheme for their labels, with a printed table explaining what a particular wire number is used for. Others use a lettering system for labels while some prefer adding coloured tape or short lengths of coloured shrink wrap to wires.
There is no standard for labelling; just use what makes sense to you.

Schematics

Consider a layout that has many electronic modules, connectors, etc.
By the time a fault occurs, you will have long forgotten that output 7 of this module is connected to input 5 of that other module. The only way is to manually trace it. Every time you have a fault, you have to work out what goes where all over again.
Much better to spend a little time making up a simple diagram or chart that shows how modules,

connectors, etc. are interconnected. That way, when you have a fault, you know exactly where to go to make your tests.

Consider documenting a station throat as being served by number 5 Servo4 board, with points being noted as 1,2,3,4. If point 2 fails to operate due to no input signal, your chart tells you what drives that particular board input. So, for example, you can put your meter across pin 5 of a CANACC8 or across switch 7 on your control panel – or whatever. There is no need to even trace the wire initially. You can check whether the control output signal is being produced. A quick check of your schematic tells you exactly what connects to what, no matter how far apart they may be. The schematic can be further refined if the wire passes through several connecting blocks or plugs and sockets.

Documentation

Clearly, documenting your layout pays dividends in the longer term.
To be useful, the documentation must always be readily accessible - in a folder and/or wall chart.
A typical folder might contain:

- Tables, showing the colour coding and labelling of wires used on the layout.
- Schematics showing the interconnection between modules and connectors.
- Development history, detailing who added what when.
- Fault history, so that recurring faults can be identified and dealt with.
- CBUS details, listing the node numbers of each module, what events do what, etc.
- DCC details, listing the CVs of locos, accessory decoders, etc.

Important note

Documentation is a chore and most people are reluctant to keep updating it. However, it is important to recognise that documentation that has not been updated can sometimes make things worse. The key is to make the documentation easily alterable. For example, you don't want to have to redraw a diagram every time a change is made. Those with computers have an obvious advantage, as lists, tables, etc. are more easily updated and reprinted.

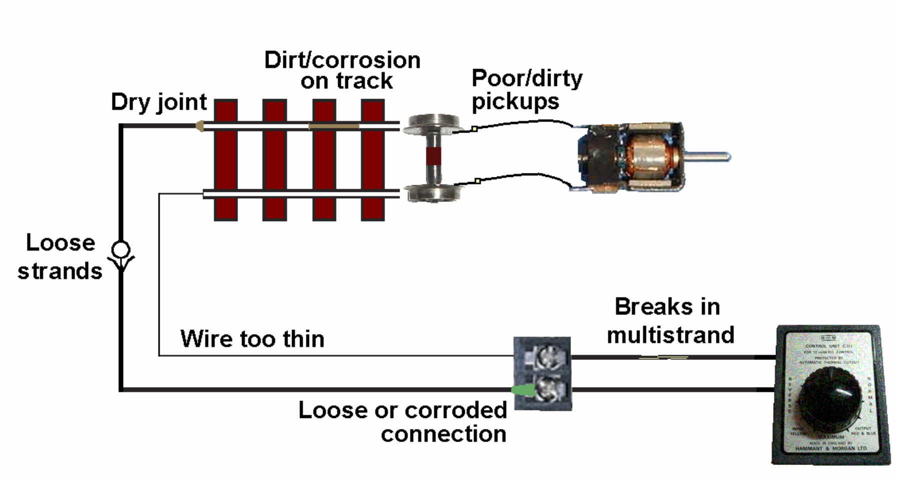
Chapter 5

Voltage problems

Voltage problems

In an ideal world, the only loads across your controller would be ones that you wanted, such as locos, lighting, point motors and so on.
In practice, the layout wiring, connections, plugs and sockets, switches, etc. all add unwanted additional resistance into the circuit.
Some of the power is lost in the bad connections, long wire runs, etc.
This chapter looks at what causes voltage drop problems.

This illustration shows just some likely causes of unwanted resistance in loco/track wiring.



These examples all happen and you may suffer from one or more of these problems.

- Some are the result of poor installation (poor soldering, using wires that are too thin).
- Some are the result of poor maintenance (dirty track, dirty loco pickups).
- Some just happen over time (breaks in cables, corroded or loose connections, power supplies that get overloaded with ever increasing numbers of locos being run).

If the unwanted resistance is very small compared to the wanted resistance, this causes little change to the layout performance. All layouts have some voltage loss, although mostly too small to be noticed.
However, if the unwanted resistance becomes excessive, too much of the supply voltage appears across this resistance and reduces the voltage to loco, relay, point motor or whatever. These effects are predictable and can be calculated using Ohm's Law.

Its Ohms Law time

If you don't know Ohm's Law, you really ought to get familiar with it, as it is the cornerstone of many voltage, current and resistance calculations that we will come across.

Put briefly, electric current is the flow of electrons round a circuit. The electron flow is measured in Amps (or mA – thousands of an Amp). The force that moves the electrons round the circuit is known as the voltage and is measured in Volts (or mV – thousands of a Volt). The physical properties of different materials act to resist the flow of current. This is called Resistance and is measured in Ohms (or kohms – thousands of ohms or Mohms – millions of ohms).

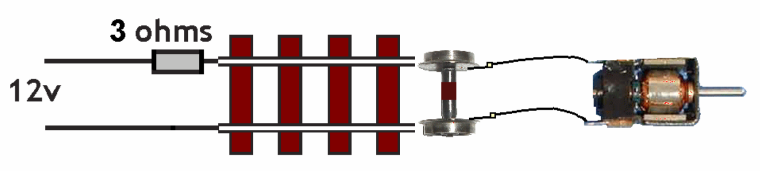
There is a direct relationship between the voltage, the resistance and the resulting current flow.
This is expressed in Ohms Law as three simple formulae:

I = V / R	- increased voltage results in increased current.
	- increased resistance results in reduced current.
V = I x R	- increasing current through a fixed resistance results in an increased voltage across that resistance.
	- increasing the value of resistance, while maintaining the same value of current through it, results in an increased voltage across that resistance.
R = I / V	- resistance is proportional to the current (i.e. increases) and inversely proportional to the voltage (i.e. decreases).

On our model railway, therefore, voltage drop occurs when current flows through our wires, our loads and any unwanted resistances. It obeys Ohm's Law, where the amount of voltage drop is equal to the resistance times the current.

Some worked examples

Consider this illustration. It shows an accumulation of 3 ohms of unwanted resistance from all the possible sources shown above. From this, we can calculate the expected voltage losses for different locos.



The best case

Let's assume we have an 'N' gauge motor which runs at 250ma (0.25A) when supplied by 12V. In an ideal world, free from unwanted resistance and other effects, the resistance of the motor must be:

R = V / I = 12/ 0.25 = 48 ohms

Working backwards, to check our formula:

V = I x R = 0.25 x 48 = 12V

Now, if there is 3 ohms of additional resistance, the total resistance has risen to 51 ohms, while the supplied voltage is still 12V.
The increased resistance results in less current:

I = V / R = 12 / 51 = 235mA (instead of 250mA)

The Voltage drop across the motor is now

V = I x R = 0.235 x 48 = 11.3V

The other 0.7V is lost across the unwanted 3 ohms of resistance.

Worse cases

The same calculations for an 'OO' gauge loco using 0.5A at 12V shows that the current drops to 440mA, the voltage across the motor is reduced to 10.7V

For an 'O' gauge loco drawing 1A, the current is reduced to 800mA and the motor voltage is reduced to only 9.6V

The worst, worst case

Now, consider a US-style consist of ten 'O' gauge locos, with a total current draw of 10A. The unwanted resistance is large in relation to the resistance of all the motors running in parallel.

The current in the circuit is reduced to a mere 2.8A and the loco motor has only 3.4V across it. The other 8.6V is dropped across the unwanted resistance.

Notes

You won't get these exact results in practice as Ohms Law calculations are for DC and doesn't take into account effects of inductance, back emf, etc. from motors. However, it does give a fair representation of the dramatic effect of voltage drops in some circumstances.

Voltage drop is not just an issue for loco motors. Solenoid point motors can be an even bigger problem as they have very little resistance, even lower than loco motors. The Peco solenoid measured at just over 5 ohms, while the old H & M solenoid measured just over 2 ohms. With the MERG PD3 CDU kit supplying a 20V kick, that results in current surges of 4A and 9A respectively. With such low resistances, point motor coils are even more vulnerable to unwanted resistances.

Power and heat

Power is measured in Watts and can be calculated by multiplying voltage by current. So, for example, if the N gauge loco took 0.25A when supplied by 12V, its wattage is

0.25 x 12 = 3W.

As we go up in gauge, with larger motors drawing more current, the wattage is also increased.

With motors, most of the power is used in turning the motor. With an unwanted resistance, the power will be dissipated as heat.

In the above example for an 'O' gauge loco, the voltage across the unwanted resistance was 2.4V and the current through it was 0.8A. That represents a power dissipation of

0.8 x 2.4 = 1.9W.

In the worst case example, the consist of 10 locos, the voltage across the unwanted resistance was 8.6V and the current through it was 2.8A. That represents a power dissipation of

8.6 x 2.8 = 24W,

generating more heat than many soldering irons!

Wire resistance

The current carrying capacity table shown earlier does not fully tell the whole story.

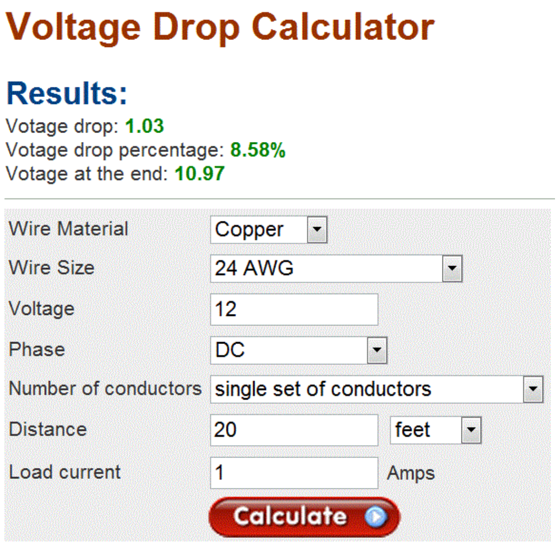
Even the best conductors have some resistance and the resulting voltage drop in the cable can be calculated from its CSA (cross sectional area). The CSA can be expressed in mm^2.

The internet provides a number of calculators that let you see what voltage drop to expect from different gauges over different lengths of wire runs. Many of these the AWG sizes but you can get some conversion information at

http://www.calculator.net/voltage-drop-calculator.html

For example, 7/0.2 means 7 strands of wire each 0.2mm diameter, This results in CSA of $7x(\pi x0.2x0.2)/4 = 0.22mm^2$.

The image shows a calculation for 24AWG, the nearest equivalent to 7/0.2 stranded wire.



Remember that a circuit uses two wires, one feed and one return. So, a 10ft layout would have a 20ft run.

Some calculated results

- A 10ft layout using 7/0.2 for N gauge (250mA) locos would suffer a drop of just 0.13V.
- The same 10ft layout for OO, using 0.5A locos would suffer a 0.26V drop. If that layout was 20ft long, a drop of around half a volt would result.
- When an 'O' gauge loco at 1A is run on a 20ft layout, the voltage loss is just over 1V.
- However, if a 5 ohm solenoid coil was operated by 20V over a 10ft cable, the drop would be 4.11V. This could be reduced to 1.62V or 0.81V by using 16/02 stranded wire or 1mm solid conductors respectively.
- Our earlier example of a 10A consist running over a large 60ft layout would drop a whopping 6V using 1mm twin and earth and even 2.5mm twin and earth would result in a 2.4V loss.

Clearly, voltage drop in the layout wiring is less of an issue with smaller gauge locos, although it still pays to use as thick a wire as you can afford (specially for solenoids and other heavy current devices).
As the layouts get larger and the loco currents get larger, the problems of voltage drop become an issue that needs attention.

Note:
The calculations for stranded wire assume that all strands are intact and correctly connected.
It is not uncommon for stray strands of wire to fan out as wisps and fail to be connected or soldered.
It is also common for wrongly-adjusted wire strippers to cut or nick some strands.
If, for example, a 7/0.2 wire had only two unconnected strands, the final resistance of the wire has increased by a significant margin, having lost two-sevenths of its current carrying capacity.
Even a nick in a strand can lead to an eventual break in that strand, if placed under stress or vibration.

 Back to Contents Page 5

Finally, note that these calculations are for voltage drop in the wire alone, while other resistances such as the bad connections, etc. mentioned earlier may have to be added to the calculations.

Rail resistance

A track bus with multiple droppers may seem excessive when we might simply rely on the nickel silver rails to carry the current. However, nickel silver is an alloy made from copper (60%), nickel and zinc. The difficulty is that the resistance of nickel silver is nineteen times greater than that of copper.

Since not all nickel silver alloys use the exact same proportions and not all rail sections have the same cross sectional area, the final resistance is not precise – although it is very significant.

That is why a length of track has a corresponding length of copper wire running underneath, with frequent connections between them. The poorer conductor (the rail) only carries current over a small section of track to the loco, while the current is mainly carried by the better-conducting copper track bus. The effect is to greatly reduce the voltage drop.

Track is sold as code 70 or code 100, etc..

The code number is the height of the rail expressed in thousands of an inch.

So, code 75 rail is 0.075” in height while code 100 is 0.1” in height.

Here are some rail types:

Code 40 – Micro Engineering N and HOn3
Code 55 - Peco N, Micro Engineering N and HO
Code 60 – Peco OO9
Code 70 - Micro Engineering N and HO, Shinohara N and HO and HOn3
Code 75 - Peco OO Finescale
Code 83 – Peco HO, Atlas, Walthers/Shinohara HO, Micro Engineering HO
Code 100 - Peco OO, Atlas, Walthers/Shinohara HO, Micro Engineering HO
Code 148 - Micro Engineering O

The rails with the smallest height and the small cross sectional area present the greatest resistance.

The DCCWiki displays a chart of measured resistance of various brands of track.

Brand	Code	Rail Resistance, Ohms/foot	Equivalent Wire (AWG)
Atlas / Shinohara	100	0.038	26
Walthers / Shinohara	83	0.052	27
Atlas / Micro Engineering	83	0.078	28
Shinohara	70	0.077	28
Micro Engineering	70	0.116	30

And adds this note:

> Note the difference between two brands of Code 83 rail. One has about 50% more resistance. That can be explained by the rail profile, which has an effect on the cross sectional area of the rail. The cross section of 27AWG wire is 0.102 mm2, 28AWG is 0.08 mm2.

From this table, we could calculate that an N gauge layout with a 20ft track run would have a resistance of 40 x 0.116 = 4.64 ohms, resulting in just over 1V of voltage drop in the track. A track bus using 1mm wire, on the other hand, would only drop 0.05V.
Now you know why lots of droppers are recommended!

You can never eliminate voltage drop entirely; you have to decide what amount of drop you can live with. You may simply prefer to increase the supply voltage to compensate, after considering whether there are any unacceptable overheating risks as a consequence.

Overload and short circuit protection

There are two dangers that need guarding against: with special protection circuitry.

short circuit protection

Imagine buying a controller with a 10A capability. You only run a single 1A loco at the moment but are looking to the future. Now, if there is a short on the cable, there is almost zero resistance and the maximum current possible will flow. In the example, the current is not 1A but 10A. This produces 120W which will manifest itself in heat that will destroy electronic components, melt plastic and potentially causes fires. Speed controllers and power boosters have systems that detect the surge in current and switch off the supply.

current overload protection

Short circuits produce a sudden surge in current and this is easily detected. Not all current increases need be that dramatic. Running more and more locos on a layout, switching on lighting and other devices, may incrementally increase the current demand to a level greater than that for which the supply was designed. A good controller/booster will allow the user to set the current at which the circuit will cut off the power. In the above example, a user may decide to set the 10A supply to trip at just 3A, as that is his/her current maximum usage.

Protection devices

Fuse

A fuse is simply a short length of wire that melts if too much current flows through it, thus breaking the circuit. The fuse is disposable and needs replacing after it blows.

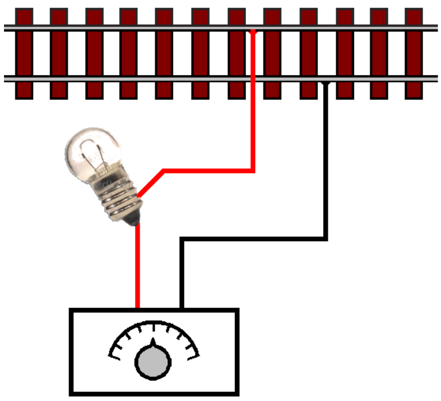
Circuit breaker

Like a fuse, a circuit breaker stops the flow of current when it exceeds its set rating. They are thermal breakers, which means that the heating effect of the current passing through it eventually bends an internal element that switches off the power.
Some, like the one shown in the picture are automatically reset once the element cools down again. These are available with trip ratings of between 1A and 5A.

An alternative is a thermal circuit breaker that needs to be manually reset after it trips (like the ones in your house mains box).
The picture shows a panel mounting circuit breaker, with trip ratings from 3A to 20A .
Many push to reset breakers can only be reset after a minimum time has passed (e.g. one minute). This can be a good things, as it forces you to look at the source of the problem.

Car bulb

Perhaps one of the oldest, and most controversial, protection methods is the use of a 12V 25W car tail light bulb.
A car bulbs is wired in series with the supply to a track section and provides almost no resistance when unlit. The locos run happily with their full voltage. When a short occurs, the current increases rapidly, causing the bulb to light. A lit bulb has a much greater resistance than when unlit. This increased resistance then limits the amount of current flowing in the circuit, This increase occurs very rapidly, around 4mS. When the short is removed, the bulb cools down and its resistance lowers again.



Speed controller

Car bulbs are cheap and you may like the idea of an immediate visual indication when a short occurs.
However, it is important to remember that they are not cutouts, only current limiters. If a short occurs the current round the wire and the track is limited to a maximum of 2A, which still has potential for causing overheating damage.
If using car bulbs, choose their position carefully, as they are very hot when lit.

DCC protection

A DCC controller or booster has its own cutout circuitry which will cut off the power and most check periodically to see whether the short is removed before restoring the power.
If the whole layout is fed from a single controller/booster, then a short will disable the entire layout. Troubleshooting is then a problem, as the fault could be anywhere on the layout.
This is usually prevented by dividing the layout into power districts, as previously discussed.
Of course, that means that each power district would have to use a separate circuit breaker.
Now, when a short occurs, only the section affected will be cut off, with the rest of the layout working normally.
Circuit breaker modules are available as standalone devices that are wired in series with the wire from the bus to the section track. These include:

- The PSX from DCC Supplies.. Handles up to 20A per block, with 1 to 4 blocks on a board.
- The OnGuard! OG-CB, from DCC Specialties, rated at 4A .
- The NCE EB1 single circuit breaker.
- The MERG DC05 block cutout.
- The MERG DCC District Cutout Kit 57

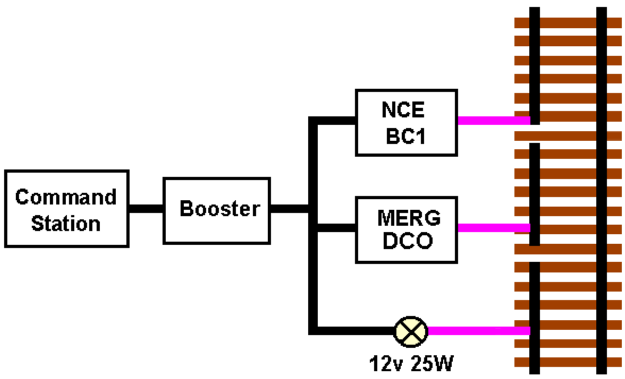
The MERG NB1B DCC booster has own internal cutout with 3.3A and 5A options as standard.

Wiring cutouts

For DC, one controller is allocated to every running loco. If a layout has many blocks, there is still only one controller powering one loco in any one block. Given that controllers have cutouts, a DC layout is already reasonably protected. If a short develops in a block, that controller's cutout will remove the power from the offending block; the rest of the blocks carry on normally, as all the block are electrically isolated from each other.
If a block is large, it may be sub-divided, with all sub-blocks still being wired to the one controller. If each sub-block was fed via a circuit breaker or car bulb, the whole block still stops working, but you can quickly identify the problem sub-block. This is an aid to troubleshooting.

For DCC, with a single command station powering the entire layout, a short anywhere on the layout cuts the power off from the entire layout. All locos stop running.
This can also be overcome by sub-dividing the layout into independent power districts, each with their own booster. A short in one power district only stops running in that power district.
Alternatively, for smaller layouts, a single command station, with or without an added booster, can feed power to power districts through a collection of DCC circuit breaker modules or other protection devices.

While circuit breaker modules are more efficient, they are also much more expensive compared to car bulbs. Since most shorts are caused by derailments at points and crossing, it makes sense to cover the blocks that contain many points with circuit breakers. Other less-frequently used blocks could be more cheaply covered with car bulbs.
In the illustration above, three power districts are fed from a single command station/ booster combination. One is protected by a commercial cutout, one from a MERG district cutout, and one from a car bulb. It is important to note that the bulb may limit the current below the level where the booster would be expected to cut out. Most modern boosters may activate before the car bulb illuminates.



In an ideal world, a DCC layout would be divided into many power districts, each supplied via its own circuit breaker module. This approach may be unsuitable for financial reasons. However, since car bulbs are substantially cheaper, they allow more sub-districts to be individually protected.
A combination of circuit breakers for point-intensive districts and bulbs for more sparse districts may be considered. Sometimes dogma has to be replaced by pragmatism. If we can't have all that we want, we can at least get what we need.

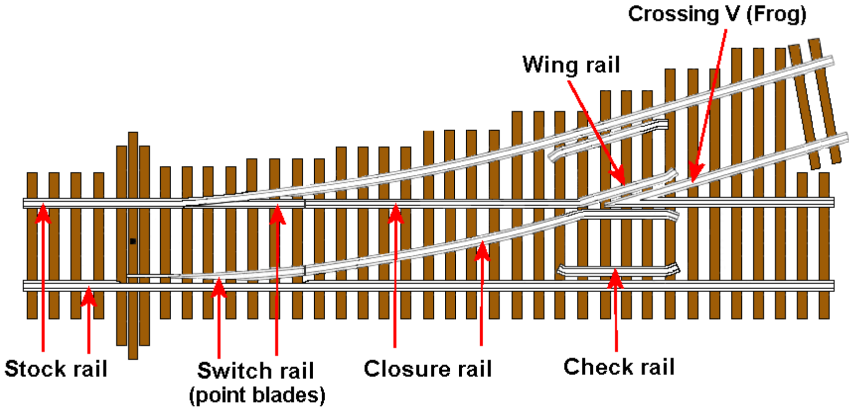
Prudent operators may wish to test that all the protection features work,before an operating session (particularly at exhibitions). This is usually achieved by placing a coin across one section of track at a time.

Chapter 6

Wiring points

Wiring points

Before looking at optimising the performance of points, eliminating possible sources of poor performance or shorts, it is best to remind ourselves of the names of the various parts that make up a point (according to www.britishrailways.info). Other countries may give parts different names (e.g. even calling it a turnout instead of a point) but using this illustration ensures that the rest of the material is understood.



In most points, all or part of the closure rails remain fixed and stationary, while the point blades (switch rails) are the moving parts that guide the loco and wagon wheels to the required track.
Most magazines talk about 'frogs' and points are described as 'insulfrog' or 'electrofrog'. Sometimes, they refer to the area shown in the illustration as the 'crossing V', while other times it refers to the area that encompasses both the crossing V and the wing rails.

Point varieties

Commercial manufacturers offer points with many different constructions and wiring differences.
It is common to find 'power routing' points. These are constructed so that power is always connected to the direction set by the point. For example, a train in a siding would be disconnected if the point was set to the main line – and automatically reconnected when the point is set back to the siding.
In many cases, all the point blades, closure rails and wing rails are bonded together by wires under the trackwork (even the crossing V might be similarly bonded).
Also, some points have both point blades bonded together, while some have each point blade bonded to its nearest stock rail. Some points have the points blades bonded to the closure rails with wire, while others use rail joiners.
In most cases, commercially-produced points have plastic check rails.

Notes

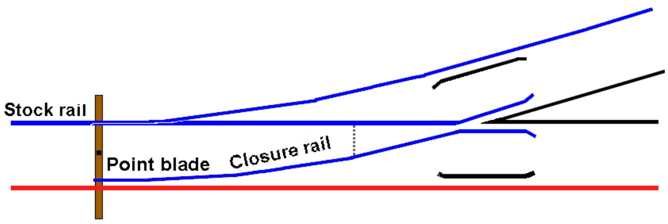
If your pointwork runs happily for you, that's fine.
If you are experiencing electrical problems with your points, then you might wish to consider the following alterations to your point wiring.
If you want to achieve the best future problem-free running with your points, you might wish to implement some, or all, of the alterations that are outlined.
If you are planning a layout, you might want to consider the following pages before

selecting your points.
Not all the following may apply to your points. If in doubt about how your points are wired, use a multimeter to check your point setup.
A multimeter will soon tell you whether you need to consider making any alterations.

Point problems

The biggest problem experienced with points is shorting of the track power as the loco runs through the point trackwork.
If you are using DC, this may not even be noticed, apart from an occasional hiccup of the loco on the way through. That is because the loco's controller can handle the temporarily increased current for a sufficient time for the loco's momentum to take it past the point of the short. While this may seem to be acceptable, bear in mind that every time the short occurs, the spark is pitting the surface of the trackwork. So, even if your loco does not come to a complete halt on points, it may still be worth looking at improving the point's wiring.
With DCC, shorts are a bigger problem because the command station will close down when it detects an unexpected current surge. Not only will the loco come to a stop – the whole layout may come to a stop, if it is controlled by a single command station.

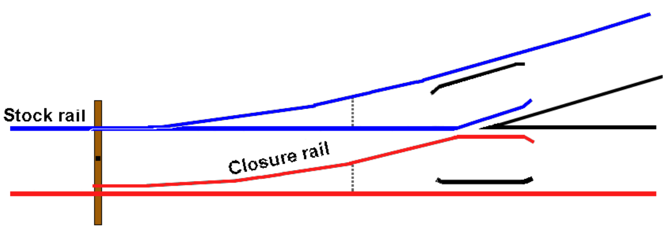
Shorts at point blades



Consider the case shown in the illustration. The point is set to straight-ahead, with the upper point blade touching the upper stock rail.
Both closure rails are bonded together which results in the lower point blade sitting only a small distance from the lower stock rail – but at the opposite potential.
In an ideal world of fine tolerances, this might not pose any threat.
However, with many locos having coarse scale wheels with wide wheel treads, there is a possibility that the metal wheel will touch both the lower point blade and the lower stock rail. This produces a short across the track bus. This effect can also be caused where a wheel is out of gauge or has slack bearings. Another cause may be taking a loco round a small radius curve, such that the back of a wheel flange touches the open point rail.
The solution is to cut the link that bonds the two point blades together.
Left this way, a point blade/closure rail combination will only be 'live' when it is touching its own corresponding stock rail.

Dirt at point blade

It is commonly accepted that it is best not to solely rely on a good contact between the point blade and the stock rail to provide electrical connectivity. While a new, out-of-the-box point may work happily for a while, the eventual build up of oxidisation and grime on these points of contact could lead to high resistance joints, or complete loss of contact.

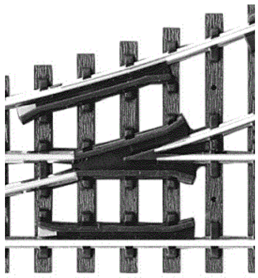


The solution is to have a permanent link between each point rail and its

neighbouring stock rail.   Both point rails, and possibly their closure rails and wing rails, are permanently powered similarly to their stock rails.
A couple of links of Kynar wire or similar fine wire should be soldered between the point rails and their neighbouring stock rails.
As always, it is much easier to do this before a point is laid.
This modification ensures the best long-term running for both DC and DCC.   Since the point rails are always at the same polarity as their nearest stock rail, DCC shorts are minimised and track pitting due to DC arcing across them is eliminated.

Shorts at frog

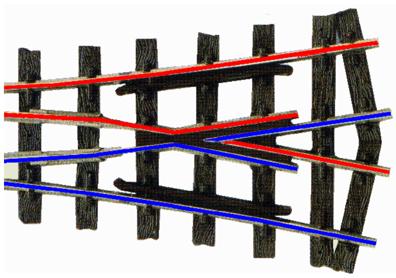
If you look at the previous illustration, you will see that there may now be a problem at the 'frog' area, depending on how a point's closure rails, wing rails and crossing V are currently connected.
As shown, there is little space between these track components, which again provides the possibility for shorts.
One solution from manufacturers is to manufacture that problem area from plastic, preventing possible shorts.   These are often referred to as 'insulfrog' points and an example is shown in this picture.



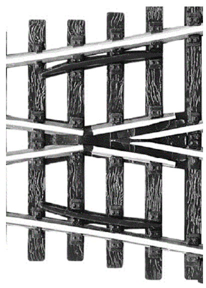
This avoids any wiring modifications and is a suitable solution for long wheelbase locos that have pickup wheels that can span the dead gap, so that at least one set of pickups is on live track at any one time.
It is more of a problem for short wheelbase locos, or longer locos where one set of pickups is not functioning.
That is why many user prefer 'electrofrog' points, where the wing rails and the crossing V are conductive.   Also, plastic frogs are more prone to wear than their metal equivalents, causing an eventual increase in derailments.

Wing rail shorts

Live frog components provide the most reliable point operation, but using electrofrog points is not without its own problems.



The picture shows just how close the wing rails and the crossing V are to each other and – more importantly – at different polarities.
Also, the junction between a wing rail and a closure rail is close to the other wing rail/closure junction.
Potential shorting can be overcome by breaking the electrical connection between the closure rails and the wing rails.
This may require cutting the junction with a cut-off disc in a Dremel drill, or something similar.



This picture shows a point that has gaps between the closure rails and the wing rails.   Other prefer to cut further back down the closure rail, to further reduce any risk of shorts during derailments.
In some Peco points, these gaps may already be provided, with just wire links under the trackwork between the wing rail and its closure rail., In that case, the links should be cut.
Left this way, this would be a 'dead frog' point, which is little better than an insulfrog point.   Being metal, it allows the frog area to have its polarity switched as the point is switched (see later).

Shorts at crossing V

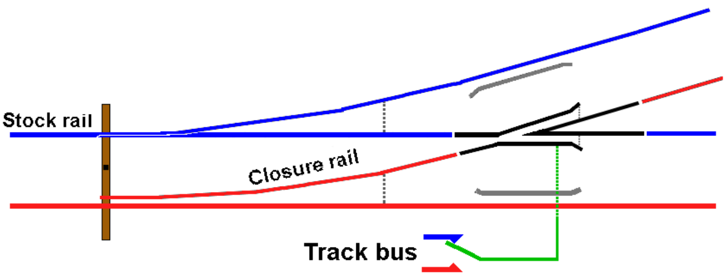
Another source of shorts is the tiny gap between the two halves of the crossing V on most points. As the above pictures show, the two rails are chamfered down to create a fine point, with almost no gap between them at the tip of the V.
The wide tread on some metal wheels are able to bridge the gap, causing shorts if both rails of the V are powered (if they are powered then they will be at opposing polarities).
A number of solutions have been offered for this.

- Applying a small touch of paint on one of the rails, at the tip of the V. Most report that this wears away and is not a permanent solution.
- Applying a small touch of glue, such as SuperGlue, to the same area. This wears better than paint.
- Both of the above are only semi-permanent remedies and both cover up a part of the conductive surface of the rail. An alternative approach is to widen the gap in between the two rails of the V. A cut-off disc fitted to a Dremel drill, or similar, can be used to grind a small section off the inner edge of each rail of the V. It need not be too wide or even too deep; simply enough to prevent wheel to rail contact. The same effect can be had using a small round jewellers file to gradually widen the gap. If required, the gap could be filled with epoxy then trimmed.
- Users have success without modifying the V at all. If the flangeway of the points (the gap between the check rail and the stock rail) is too wide, it may allow a loco's wheels to drift over the gap in the V. Gluing a thin plastic strip to the check rail narrows the gap and prevents excessive loco sideways movement. This solution is dependent on the points and locos in use.

An alternative approach does not modify the point's rails.
As shown in the illustration below, it treats all the frog components as a single condutive unit. In other words, the crossing V rails, wing rails, and maybe even part of the closure rails, are wired together. In some points, this may be how they are supplied as standard; if not, they can be wired together with Kynar wire or other fine connecting wire.
This requires additional electrical switching as below.

Frog switching

The first step is to ensure that the V is isolated from the rails that feed out from it, using insulated joiners. Similarly, the wing rails need to be isolated from the closure rails.
The illustration shows the crossing V, the wings rails, and a small part of the closure rails are wired together.
Left that way it would be a 'dead frog' point which is little better than an insulfrog point, apart from the fact that the metal frog will wear less.
Clearly, the frog area needs to be powered and, since all parts of the frog area are at the same polarity at any one time, its polarity must always corresponds to the direction to which the point rails are switched.



So, the frog area needs its own electrical feed whose polarity changes dependent on whether the point is set or unset.

Point switching can be achieved in a number of ways:

- Using the changeover switches available on some point motors (Tortoise, Hoffman, Cobalt, Fulgurex, Seep) , or purchased as point add-ons (Peco). These are often regarded as not fully reliable in the longer term.
- Using the Tam Valley Frog Joicer. This module works by detecting a short if a loco enters when the frog is at the wrong polairty to match the direction of travel. The surge in current is detected and the frog polarity is reversed. These have a number of qualifications. They are expensive (currently £57 for a module to handle 6 frogs); they cannot be located close to all 6 frogs simultaneously, requiring wires to be run from the module to distant frogs; they rely on a current spike to function (i.e. arcing and pitting have ocurred before the frog is switched).
- Using a changeover microswitch that is mechanically operated by the movement of the point's tie bar. This keeps the switch close to the point and independent of any other point wiring. Could be a little tricky to adjust for the correct switching point. Also, a low operating force microswitch is preferable, to prevent the back pressure from the microswitch moving the point blades away from the stock rails.
- Using a non-mechanical detection system such as attaching a small magnet to a location on the point operating mechanism, such that it activates a Hall Effect switch that controls a relay. An alternative might be a small shutter that breaks an infrared beam detector. These are more finicky to set up but have the advantage of placing no backlash effect on the point blades.
- Using a relay with SPDT (single pole double throw) contacts. The relay could be operated by a simple trackside or conrol panel switch, or by the output from a CBUS module such as the CANACC8. For example, the signal that operates a servo in a Servo4 board could also be wired to a relay to simultaneously changes the polarity of the frog.

There are plenty of options and their relative merits are a matter of debate.
It all depends on what you already have, how efficiently you expect it to work (now and in the future) and the lengths you are prepared to go to ensure trouble-free point operation.

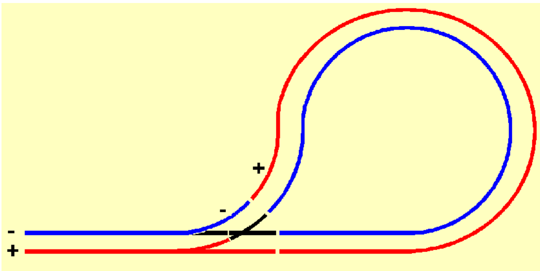
Point feedback

Throwing a switch on your control panel, or clicking a button on a screen display, does not ensure that your point has operated correctly. Any number of problems could prevent a point from moving. There may have been a cable break, bad connections, a mechanical failure, a motor coil burned out, or a chunk of ballast or other dirt stuck between the stock rail and the point blade.
Ideally, the operator should be informed that a point movement has been successful, by lighting a lamp on the control panel, or illuminating a part of a screen display. That involves taking feedback from the point back to the operator. This could be via a length of wire, or via a CBUS CANACE8C module.
If a point is fitted as supplied (with none of the above modifications), it can add a microswitch that is operated by the movement of the point's tie bar; the output from the microswitch is switched either high or low depending on the tie bar's position.
If a point has wiring for frog switching, then the most of the above wiring methods used for frog switching could also be used as the feedback.

These methods are suitable for DC or DCC, conventional wiring or CBUS.
Most of the time, your points work dependably. For more critical situations, such as at exhibitions or for automated layouts, feedback is much more useful if not essential.

In most situations feedback from a single source (e.g. point motor switch, microswitch, Hall Effect switch, infrared detector) will be perfectly adequate.
However, it has limitations. If the frog polarity has been successfully switched, without the point blades moving, you will soon know about it as there will be a short when the loco reaches the frog. The 'feedback' will take the form of the power being cut off from that section/power district.
If the blades partially move, as with dirt blocking full movement, there is likely to be a derailment with possible shorting once again.
In both cases, this can be prevented by having a system that detects the tie bar has moved the point blades fully in either direction. At its simplest, this might mean having a detector at either end of the tie bar – one to ensure a fully open point and one to ensure a fully closed point. Circuits can also be devised that will only power the frog when a tie bar is fully in one direction. This results in a dead frog but prevents overall shutdown, and the operator is notified of the problem.

Reverse Loops



Although not common on actual railways, reversing loops are popular on many model railways.
As the illustration shows, if your track reverses back on itself, you will have a short.
Solutions require that a 'dead' section be created in the loop, using insulated joiners at both ends.
This unpowered section is often called the reversing section.
The only connection to the reversing section of track is through some switching system, either manual or automatic. The switching system (see below) is fed from the track bus and feeds out to the reversing section – usually two wires in and two wires out.

Before looking at different switching approaches, it is important to ensure that the length of the dead reversing section is longer than the longest train that will be on the loop. Anything that has track pickups must be included in the train length. Apart from the loco, take into account carriages with lighting, with sound, or with extra power pickups, or a train that has a loco at each end.
All of the trains pickups must be sitting in the reversing section, otherwise there will still be a short.
If a train has only a single loco, and no other pickups, it is possible to have a short reversing section.
However, that is only possible if you know that a loco will always be pulling the train and never pushing it. Otherwise, where do you position the short dead section – near the entrance to the loop or near the exit. To play safe, and to allow for a range of different train lengths, it is best to have a long reversing section.

Different switching approaches are adopted, depending on whether you are using DC or DCC.

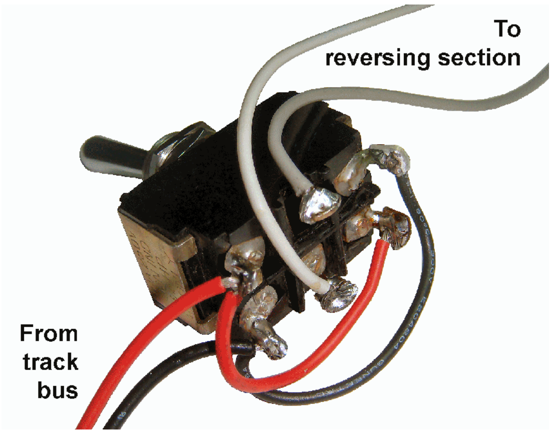
DC

The usual way of handling reversing loops in DC is manual and uses these steps:

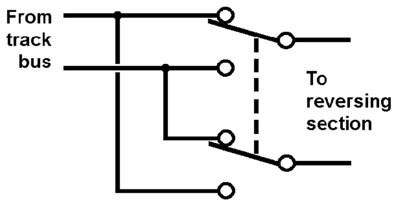
- Make sure that the reversing section is powered to match the direction of the point blades (otherwise you have a short).
- Drive the train into the reversing section.
- Stop the loco.
- Change the point.
- Reverse the mainline track power.
- Start the loco again.

As you can see, this involves the operator to be in attendance and to carry out a sequence of tasks in the correct order.
There are a number of methods of handling DC reversing loops.

Switch

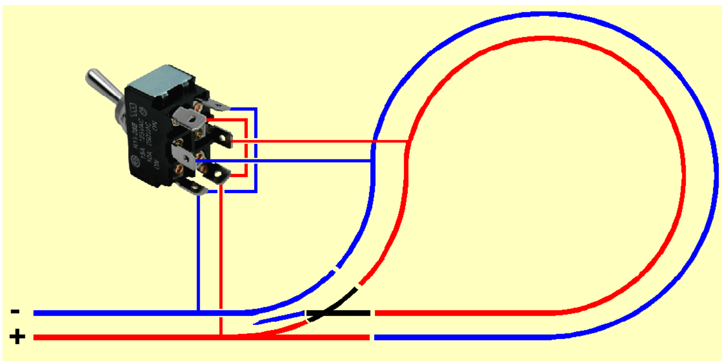


This is the most common method of manually switching the section polarity.
The picture shows the use of a DPDT (double-pole double-throw) switch to reverse the polarity of the reversing section. When the switch is thrown to the left, the two middle tags are connected to the two leftmost tags. When the switch is thrown to the right, the middle tags connect to the rightmost tags. By cross-wiring the outer tags, the polarity of the output on the middle tags depends on the direction that the switch is thrown.
The illustration shows the schematic representation of a DPDT switch; this is how it will look in a circuit diagram. The dotted line indicates that the two switches are always operated at the same time,

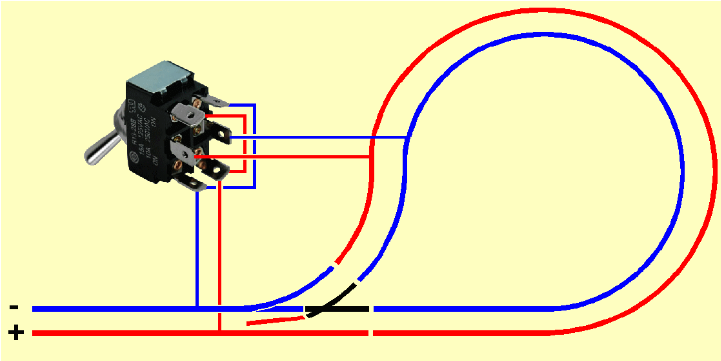


Consider the case of a train entering the loop via the 'branching' track of the point and leaving the loop via the 'main' track of the point.

This illustration shows the point being set to branch. The switch is thrown such that the polarity of the reversing section matches that of the branching track.
The train can safely enter the reversing section, since the rail polarities match.
The track where it exits the loop has a polarity that does not match its adjacent rails; this does not matter since they are isolated from each other.



The train is run into the reversing section and stopped.



This illustration shows the point being set to main. The switch is thrown in the opposite direction from that above, making the polarity of the reversing section now match the main track of the point.
At this point, the mainline track polarity is reversed, as the loco is entering the mainline in a different direction from before.
The train can now be safely run out of the reversing loop into the main line.
The track where it originally entered the loop has a polarity that does not match its adjacent rails; this does not matter since they are isolated from each other.

Manual operation is very simple to wire but relies on the operator carrying out the sequence correctly.

Relay

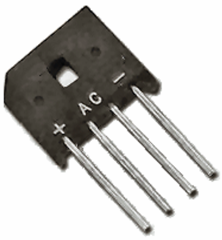
This can be improved on with the use of a DPDT relay to switch the reversing section, instead of a switch. When the point is switched, the relay is also switched, eliminating some manual operations and ensuring that the polarity of the reversing section always matches that required. This eliminates the possibility of shorts caused by the operator carrying out the sequence incorrectly.
The relay can be operated by the switch on the control panel, or by contacts on the point motor. You must use a latching relay with solenoid points, as there is nothing to keep the relay switched after voltage pulse is gone.

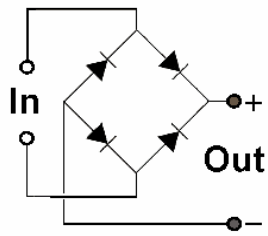
Automatic methods

Automating DC loops is possible, although complicated. The electronics has to detect when the train is fully in the loop, possibly slowing the train to a halt before reversing the track power to the reversing section. These are not commonly used but an example is the MREV Track Polarity Controller by Azatrax. This method uses track detectors to know when the train has sufficiently entered the reversing section.

Bridge rectifier

This method simplifies even further – provided you are happy to have the train always run round the loop in the same direction (always clockwise or always anti-clockwise).
Many reversing loops are hidden, so is often OK if loco only every goes round one way.
A bridge rectifier is just a single component that houses four diodes in a special formation (the MERG Kit Locker sells a WOO5 bridge rectifier rated at 1.5A)
If you prefer, you can use use four separate diodes instead (the MERG Kit Locker sells a single 1N4001 diodes and 1N5401 diodes, rated at 1A and 3A respectively)



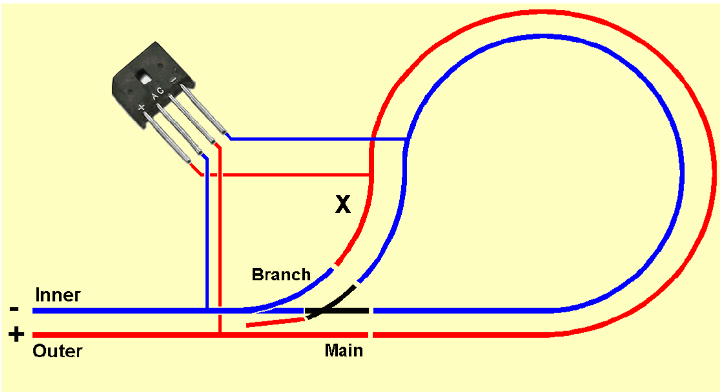


The picture shows a typical bridge rectifier and its equivalent seen as a schematic symbol in a circuit diagram.
Bridge rectifiers come in all shapes and sizes, depending on their voltage handling and current ratings.

The chief purpose of a bridge rectifier is to convert AC into DC. It has four leads. The two marked with a ~ symbol are the input connections and the other two are the positive and negative outputs.
However, a bridge rectifier can also accept a DC input. Because of the way the diodes are wired, the + and – wire will always provide a positive and negative polarity respectively – regardless of the polarity of the DC input.

This allows us to connect the reversing loop section such that it always has the same polarity on its inner and outer tracks, regardless of the polarity of the incoming track and point.

This example assumes that the loco always enters from the 'main' track and exits from the 'branch' track after going round the loop.
The operator moves the point to 'main' (i.e. straight ahead),
The illustration shows the conditions when entering the loop. The feed track has its inner rail is at negative and the outer rail is at positive. This is also the conditions fed to the reversing section.



The train will run from the main line into the reversing section and carry on round the section, even when the operator reverses the controller's output.
When the point is changed, the inner and outer rails of the feed track now match that of the reversing track, so the the loco can carry on out of the section into the main line.

This allows three options;

- The operator still stops the train in the reversing section, either for timetabling reasons or to allow time to change the feed polarity and move the point.
- The changing of the feed track polarity is wired to automatically change the point. Instead of six steps, the operator only has to change the track polarity at the correct time (when the train is fully in the reversing section).
- Using automation modules that do the job for you. Automatic methods involve electronic modules to control the points and the main line polarity, normally through computer interfacing and track detectors (such as the one marked as X in the illustration).

DCC

With DCC, there is still a need to have the reversing section isolated and connected to power in accordance with the direction of the point. The problem with shorts is still prevented by having an isolated section but there are operational differences.
With DC, locos are dependent on the polarity of the rails. Reversing a track's polarity results in the loco changing direction.
This is not the case with DCC. The loco direction is controlled by sending commands to go forward or in reverse. Once given that command, the loco keeps going in that direction, even when the track polarity is reversed.
This is handy for reverse loops. The loco drives forward into the loop, drives round the loop and is still going forward as it exits the loop – all without any change of the incoming track's polarity. Of course, you still need to change the polarity of the reversing section track once the train is fully in, to prevent shorts.
Stopping the train in the reversing section is optional. You may want to stop the train for timetabling reasons, or to give you time to change the polarity of the reversing section and to move the point. However, if you want, you can allow the train to keep driving through the loop as long as you change the section polarity and point direction before the loco exits the loop.
As before, the polarity changeover can be achieved by a variety of means;

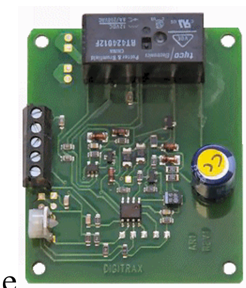
- By manually throwing a DPDT switch.
- By linking a DPDT switch to the point's tiebar.
- By having a relay change the section polarity; the electronic signal that moves the point also operates the relay.

The last two options simplify the activities of the operator.

Automatic methods

Like DC, special modules are available in DCC to automate the process. Indeed, there is a greater variety of electronic modules available in DCC.
These circuits do not need track occupancy detectors. Instead, they work by detecting a short when the loco bridges the gap between the reversing section and the point track, as the loco leaves the section. Instead of closing down the track power, these circuits quickly reverse the section polarity to remove the short. At the worst, there may be a slight jerk in the loco movement.
These circuits may be built into a booster, such as the SPROG SBOOST DCC Booster, or are available as stand-alone modules.
Standalone auto reversing units include

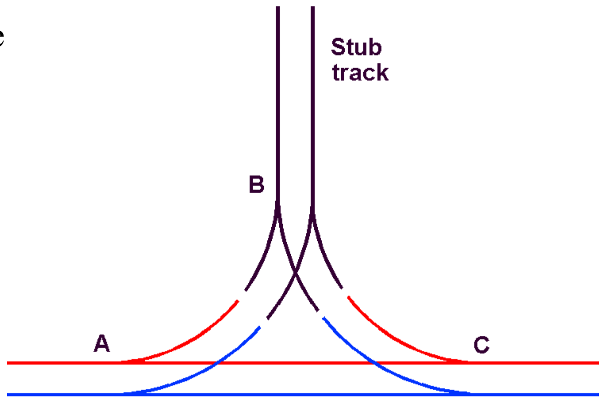
- Digitrax AR1 DCC40 Auto Reversing Module (see picture)
- Gaugemaster DCC40 Auto Reverse Module
- Lenz LK200 Auto Reverse Loop Unit
- Hornby R8238 DCC Reverse Loop Module
- Bachmann E-Z Command Reversing Loop Module



All of these modules save you the job of switching the polarity of the reversing section polarity. You still have to switch the point – unless you use track occupancy detectors to switch the points.
So, with a combination of an auto reversing unit and a track occupancy detectors operating the point, the whole process can be automatic. The operator simply has to control the loco speed.

Wyes

This formation is much more common on USA rail systems and is rarely found in Europe. Like the reversing loop, it allows a train to enter a track configuration facing one direction and exit facing the opposite direction. The wye uses three points and a single isolated section (the 'stub end'). The train comes in to the isolated section from one mainline point and reverses out through the other mainline point. It is now back on the main line but facing a different direction. The wye uses more points than a reversing loop and is less popular, although it may better match the physical locations of some layouts. Like the reversing loop, the stub end of the wye must be long enough to accommodate your longest train.



DC

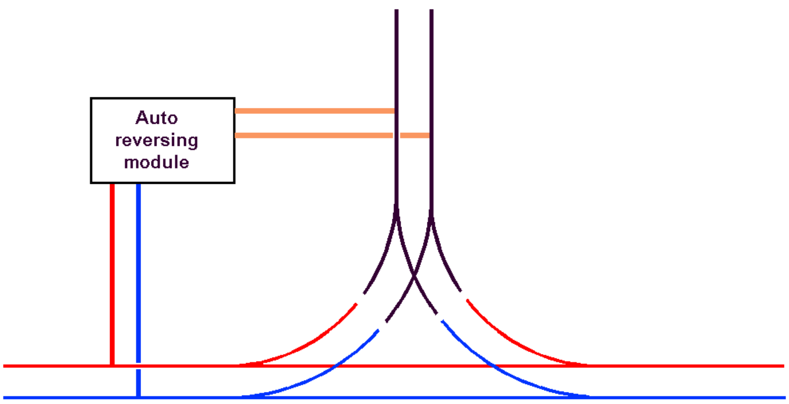
The operator's sequence might be:

- If necessary, switch point A and point B to allow the train to enter the stub track.
- Move the train into the stub track through point A and point B.
- Stop the loco.
- Change the direction of point B and point C.
- Reverse the polarity of the reversing section.
- Move the loco out of the stub track into the main line through point B and point C.
- Stop the loco.
- Switch point A and point C.
- Reverse the polarity of the track power.
- Drive the loco along the main line through point C and point A.

The train is now travelling forwards, in the opposite direction to that when it approached the wye. It is possible to have the power to the isolated section reversed by a DPDT switch or relay linked to the point operation. Nevertheless, operating a wye is a detailed process that needs to be carried out in the correct order – which is maybe another reason why it is less popular then reversing loops.

DCC

Once again, operations are somewhat simplified when using DCC. The illustration shows an auto reversing module ensuring that the stub track is always kept at the correct polarity to prevent continuous shorts.
You still have to switch points, unless automated with track occupancy detectors, etc.



Important note

The illustrations for reversing loops and wyes ignore any consideration on frog or point rail wiring or switching. This is to make the diagrams more clear. Consideration has still to given to how you wire the points themselves, as discussed earlier.

Chapter 7

Point motors and servos

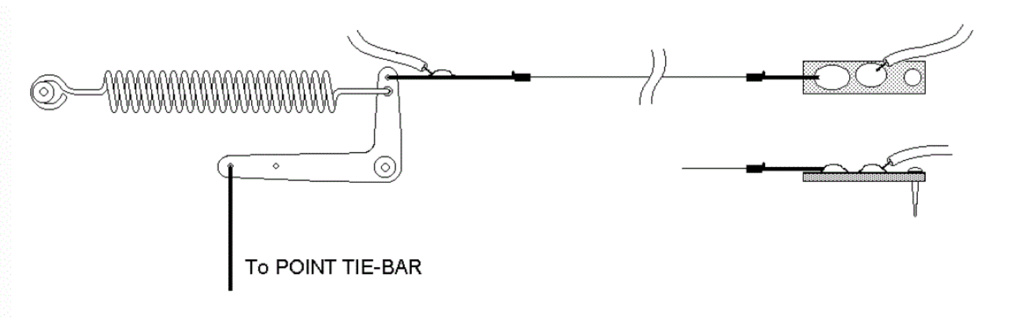
Point motors and servos

All layouts have points; some have a few points while others have scores of points.
Over the years, many different approaches to operating points have been used.
They all have one thing in common, the need to move the point's sliding tie bar back and forth to change the point's routing.
Some operators are happy to just lean over the layout and flick the point by hand, although this becomes more troublesome on larger layouts with points dispersed across greater distances. The need to move distant points led to a range of mechanical solutions, ranging from simple push/pull dowel rods to wire in cable systems, often connecting to lever frames.
However, since this book is about model railway electronics, we will concentrate on their contribution to point operations. The available options are:

- Memory wire
- Solenoids
- Stall motors
- Servos

Memory wire

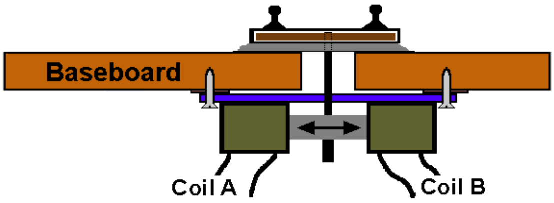
Memory wire has an interesting property – it shrinks when heated and restores to its original length as it cools again. It is heated by passing current through it (around 200mA depending on the wire gauge used) and shortens by about 5%. So, for a 4mm movement, we would need a length of 80mm (a 1/2" movement would need 10").
We could use levers to get greater movement, at the expense of pulling power.
This illustration, taken from the MERG Technical Bulletin G19/1, shows how memory wire can be used to operate a point.



Power connections are made to each end of the memory wire, normally using bolts with washers, crimps or terminal blocks. You can't solder directly to memory wire, as the excessive heat will result in the wire losing its special properties.
When sufficient current passes through the memory wire, it shrinks and pulls the lever rightwards. In so doing, it moves the point's tie bar. When the current is withdrawn, the wire slowly restores to its original length and the spring ensures that the lever moves back to the right. The shrink time takes around 1 sec and the restore time is around 1-3 secs.
The wire's power is generally fed from a constant current source (e.g. an LM317), as excessive current leads to overheating and shortens the life of the wire. Power to the wire can be fed through switches, power transistors or Darlington arrays (e.g. the 2003A).
It is not a very popular method, as it needs constant current to keep a point set in the operated mode, and it takes up a lot of space under the baseboard.
See MERG Technical Bulletins G19/1 and G19/2 for further reading on memory wire.

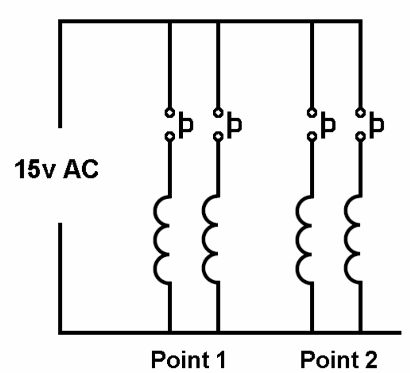
Solenoids

Solenoid-operated points was by far the most common system in use for many years, and is still popular today due to its simplicity.
A solenoid was described in the first chapter as a wire coil that acted as an electromagnet to exert a pull on a metal rod.
We can use two such coils, slightly spaced apart, with a common moveable metal rod.
If coil A is supplied with current, it pulls the rod deeper leftwards into its coil. Supplying current to coil B produces a pull in the opposite direction.

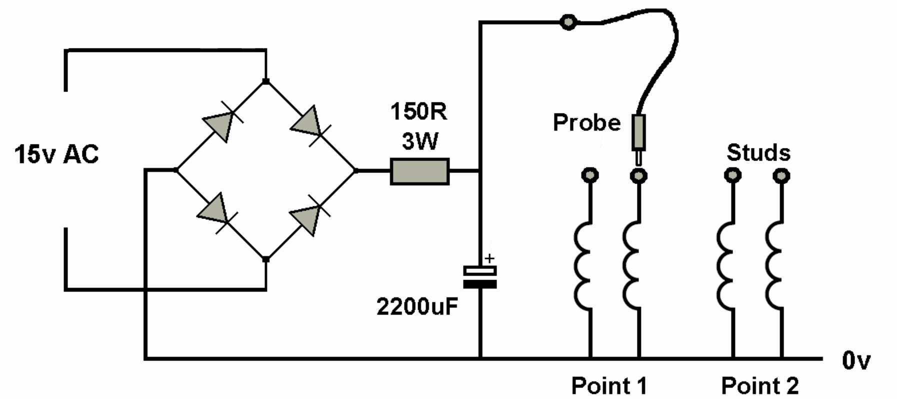


In this example, an operating pin is attached to the rod and it too slides left or right. As it does so, it moves the point's tie bar.

The illustration shows the simplest possible method of operating a couple of point motors.
Each point motor has two coils and each coil is connected across the power supply when its push button is pressed. It is common practice to join two ends of the coils together, requiring only three wires to to be connected to each point motor.
Push-buttons are used instead of switches, as the coils would overheat and burn out if power was applied for too long.

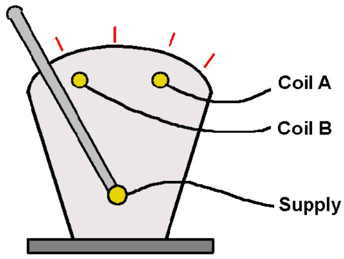


Although this systems works, solenoid point motors are more commonly used with capacitive discharge (CDU) systems. They avoid burnouts and provide a more reliable performance.



This basic CDU circuit rectifies a 15V AC supply and uses it to charge up a high value capacitor. The connections from the point motors (apart from those that are commoned) are wired to studs on a control panel. The positive end of the capacitor is wired to a floating lead with a probe on the end (an 'electric pencil'). When the probe is touched to a stud, the capacitor provides a short duration burst of high current into that particular coil. This ensures that the solenoid snaps over successfully. The capacitor's charge is soon exhausted and so cannot provide a damaging continuous current. The resistor limits the current coming from the supply, if the probe is kept on the stud for too long.
The capacitor then needs to be recharged and this is done through the resistor, ready for the next operation. The charging delay is usually between 0.5 sec and 1 sec. This delay slows up setting up routes of multiple points. You only need one CDU circuit, no matter how many points you have on your layout.

While operation by push buttons produces a neat control panel, the sparking across the switch contacts and the high current surges they handle can lead to short life and unreliable operation. For this reason, manufacturers have produced *'passing contact switches'*.



The switch has a lever that can rest in three positions – in the middle and at each end of the travel. The lever is connected to the power supply. As it is moved between the middle position and any of the outer positions, it rubs against a contact that is connected to one of the solenoid coils. As long as the operator pushed the lever all the way, the contact is only momentary and enough to discharge the capacitor through the coil.
Passing contact switches are useful in replicating a signal box lever frames, but are not suitable for operator's control panels.

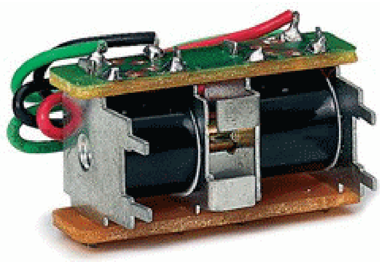
Over the years, this basic system was improved upon by using power transistors (high-power devices acting like switches) to direct current through solenoid coils. This allowed ordinary pushbuttons and switches on control panels to be used, letting the transistor do all the heavy switching. It also allowed other devices to switch the points (any electronic device such as train detectors, logic circuits and computers).
A further improvement is using more sophisticated electronics to recharge the capacitor more quickly.

Solenoid point motors

Firstly, get used to the fact that solenoids are described as point *'motors'* when they don't actually contain any motors. Later, as we shall see, there are point-operating devices that do use motors – but all the main brands of solenoids are described as motors (or sometimes as *'point machines'*). Solenoid motors are all pretty much alike in the way they provide linear motion. The main differences, apart from price, are:

- Some also include changeover switches, so that when the point is moved the switch can set the point's frog polarity (more on this later), light panel LEDs, etc.
- Some are designed to clip onto the underneath of the point, making installation easier, assuming you buy the same brand of point and solenoid.

Hornby

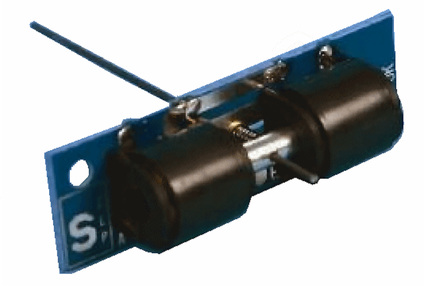


The images show two Hornby solenoids that have different styles for mounting. They operate on 12V DC or 15V AC and are recommended to be switched with Hornby passing contact switches.
The first is the R8014 standard point motor. You can clearly see the two coils. The four metal lugs are designed to be plugged into the underside of their points; a rectangular slot is cut in the baseboard to allow the point motor to fit. Alternatively, they can be mounted above the baseboard and hidden in a Hornby trackside hut that includes an extension arm to move the point.
If fitted below the baseboard, an extending rod transmits the mechanical movement from the solenoid to the point's tie bar.

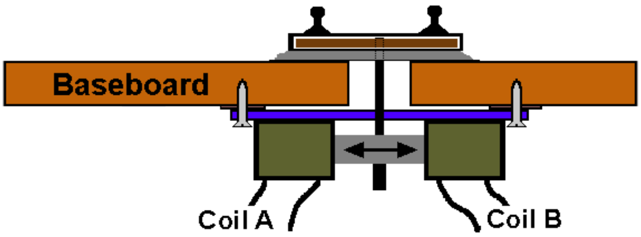
The second image shows their R8243 surface mounted point motor. This is designed to connect directly to a Hornby point. Neither model includes a changeover switch.

Seep

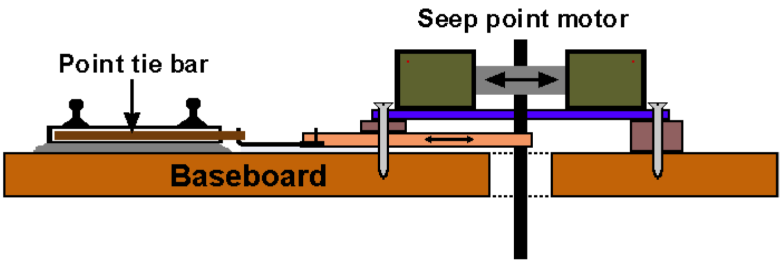
This solenoid operates from 12V to 25V, AC or DC.
They make two models; the basic PM-2 and the PM-1 which is physically identical but has a built-in switch. However, the switch uses a sliding metal washer to rub against the tracks of a PCB (printed circuit board) which limits its useful working life.



As the image shows, it has a very simple design and is expected to be screwed underneath the baseboard, as shown in the illustration. The operating pin protrudes through a hole/slot in the baseboard and slides the tie bar.

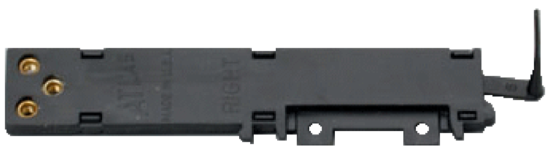


The Seep point motor can also be fitted above the baseboard, if preferred, as shown in the second illustration.
The motor has to held above the baseboard using spacers, to allow room for an extension arm to slide back and forth.



Atlas

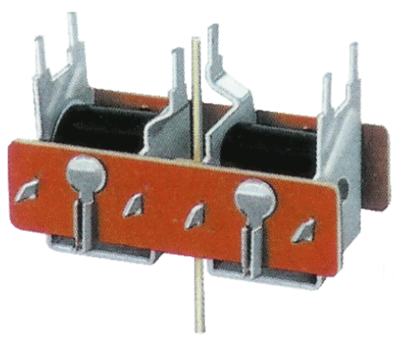
The basic Atlas point motor has its coils encased, with just the three connection points visible.
It has a slim profile but, unlike other point motors, only has a flimsy plastic operating rod.



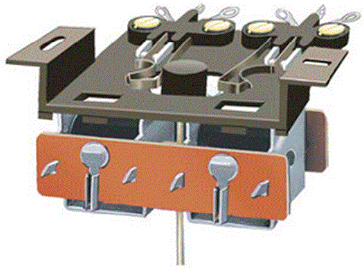
The second image is of the Atlas Deluxe Undertable Switch Machine.
It has a similar mechanism but embeds a DPDT (double pole-double throw) relay for use in frog wiring, panel lights, etc.

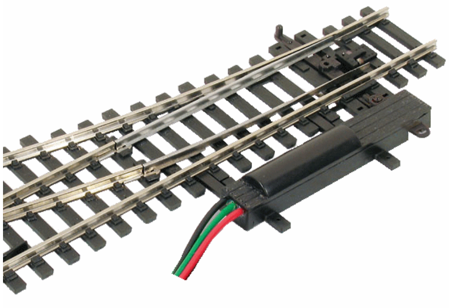
Peco

Like Hornby, Peco manufacture solenoids that plug directly underneath their own brand of point, resting in a slot cut out of the baseboard.
The image shows the basic PL10 model, operating on 16v AC.
Peco also sell a PL-9, which is a plastic mounting plate to make it easy to fit the solenoid above or under the baseboard.
If fitted under the baseboard, they recommend the PL10E, which is identical to the PL-10 but has a longer operating rod.
The other item sold for the PL-10 is the PL-13 accessory switch. It fits under the PL-10 and is operated by the lower part of the solenoid's operating rod. It provides a single changeover switch. Again, it uses a sliding metal washer contact with possible long term problems.



If you need two changeover switches, or simply more reliable switches, they sell the PL-15 accessory switch, as shown in this image.



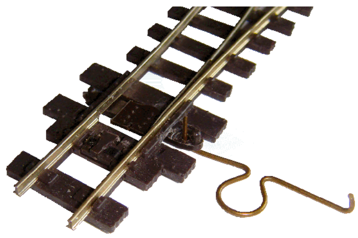


This image shows the PL-11 model, which is designed to snap on to the side of a Peco point. Due to its smaller size, the use of a CDU is recommended to ensure positive point switching.

Non-solenoid point motors

Although a proven generally reliable system, the solenoid/CDU combination suffers from a number of problems:

- The long runs of wire between the switches and the solenoids requires heavy gauge wire to avoid voltage drops, due to the heavy current they have to carry.
- The fast operation of the solenoid armature produces annoying noisy thumps and clicks.
- The fierce operation of the points can loosen and damage point parts.
- The fierce operation of the solenoid's armature can hit the end so hard that it slightly rebounds, resulting in the point blades not lying flat against the stock rails.
- The movement of the point blades is too fast to look realistic.
- Most solenoid point motors have no adjustment – the armature simply travels from one extreme to the other. The user has to employ mechanical corrections to make the required point travel match the solenoid's actual travel. This usually involves fitting an omega loop between the solenoid and the point's tie bar, to absorb any unwanted movement (the shape of the loop allows it to compress or stretch). It may also involve fitting levers to adjust between the difference in tie bar throw and solenoid throw.

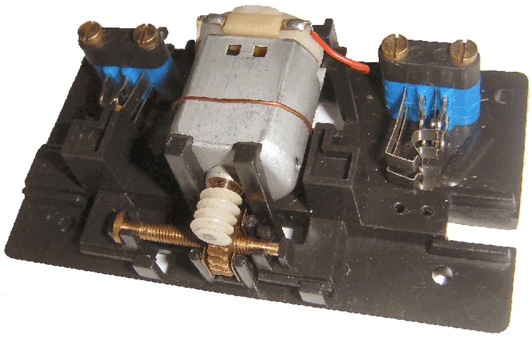


Most of these problems are solved by using motor-driven mechanisms.
They have the following advantages:

- They don't rely on short bursts of heavy current, so wiring is less crucial.
- They are much quieter when moving the point blades.
- There is no thump at the end of travel, so less wear on point parts.
- They produce a more definitive end stop, with no bounce back.
- The slow movement of the point blades is much more realistic.
- In some cases, you can set the end points, eliminating the need for omega loops.

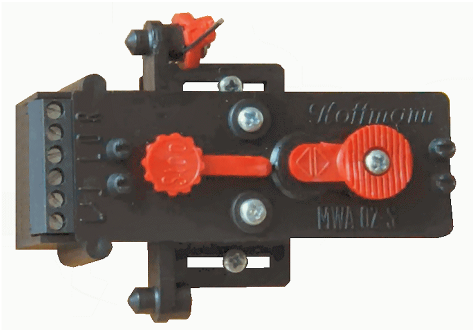
Note
If your point has an over-centre spring fitted, so that the blades snap to either side once they pass the mid position, this should be removed to allow the smooth movement that motorised mechanisms provide.

Fulgurex



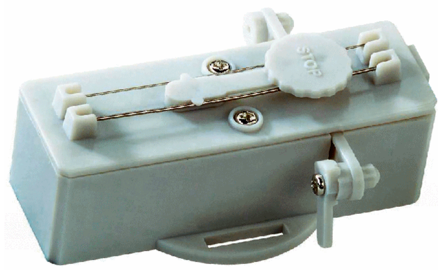
The Fulgurex point motor is not the most popular seller but we start with this one because you can clearly see how it works (the other brands have their mechanisms encased in plastic).
A DC motor drives a worm and worm gear to produce a linear movement.
A rod to operate the point's tie bar can either be attached vertically to the moving bar (for under the baseboard mounting) or horizontally to the moving bar (for above board mounting). The motor runs on 10V-14V DC at around 200mA. There are two layers of switches on each side of the motor. The bottom layer is used to switch off the motor power when the bar reaches the end of its movement. The two upper sets of changeovers are available for frog switching, lighting LEDs, etc. It is fairly large at 85mm x 50mm.

Hoffman



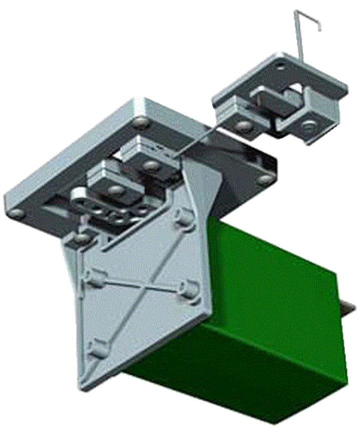
This is the Hoffman MWA02-S Point Machine.
It works on 16V at a fairly high current of 1A.
It has built-in limit switches that cut off power to the motor when the operating lever reaches the ends of its travel. The operating wire rests on on a moveable cam that allows the lengths of travel to be adjusted.
It also contains an internal changeover microswitch.
It is substantially smaller at 68mm x 24mm.

Conrad



Like the Hoffman, the Conrad point motor works on 16V at 1A.
This motor moves faster than other motorised mechanisms but is still slower and quieter than solenoids.
Part 219999-62 is the basic mechanism and Part 219998-62 includes a changeover switch (using copper brush contacts which are not as reliable as microswitches in the longer term.

Tortoise



The Tortoise switch machine works different from the others above, as its motor is continuously powered. It is known as a *'stall motor'* because it stalls at the end of each throw, avoiding the need for cut-off switches.
An unloaded tortoise draws 4mA while running, rising to around 15mA at stall. The constant current means that there is constant tension between the point blades and the stock rails.
It is intended for mounting under the baseboard and is quite large at 60mm x 55mm x 85mm.
It has two built-in changeover switches.
MERG Technical Bulletin A06/2/2/2 cover the Tortoise in more detail.

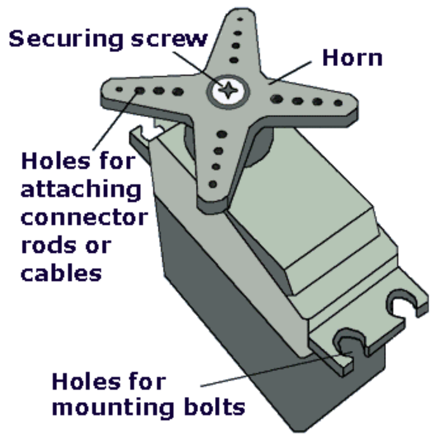
Servos

Servos have become very popular in model railways in the last few years.
Servos are available in a large range of sizes, for industrial/military uses such as automatic machine tools and satellite tracking to domestic hobby uses such as autofocus cameras and radio controlled planes and cars.
It is the smaller, lighter servos that are common in model aircraft that are now found in railway layouts. Apart from their use in operating points, they can be found in semaphore signals, crossing gates and barriers, uncouplers, cranes, and a variety of lineside animations (moving people, animals, cars, aerials, radar heads, guns, etc.).

Solenoid mechanisms and most motorised mechanisms only move an operating rod between two fixed positions. This leaves the user to make mechanical adjustments to ensure that the point motor movement is accurately reflected in the required movement of the point's tie bar.
With servos, this task is eliminated. A servo can be made to stop at any position in its mechanical range – it does not have to go from absolute end to end. So, the servo can be made to move the point tie bar by the exact amount needed each time.
Not only can you decide where its end stop points are – you can decide what speed you want to move the point blades (from super-slow to the snap speed of a solenoid).

As the illustration shows, a servo is a plastic box with some mounting lugs, an output shaft that partially rotates. Apart from a connecting cable, there is nothing more on the outside.

Think of a servo as a motor that turns an output shaft (called a 'horn') through about 180 degrees (see later). With appropriate external electronics, the horn can be rotated to a precise point in that rotation.



Benefits

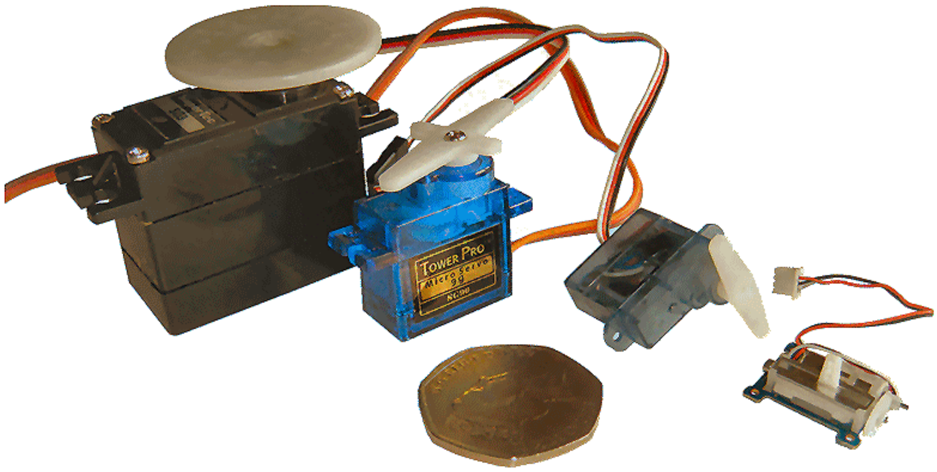
- Readily available from many sources.
- Cheaper than point motors.
- High torque.
- Accurate repeatable positioning.
- Can provide a fast movement like a solenoid, or can provide a slow smooth movement, which is more realistic and more gentle on point parts.
- Speed can be varied for each direction (e.g. to mimic a semaphore signal arm).
- Adjustments are electronic or computer based, rather than mechanical. So, point settings can be carried out remotely from the point.
- Unlike a solenoid that is either fully one direction or the other, servos can be rotated to many different positions (e.g. for a turntable, or flying a flag at half-mast).
- No annoying ‘thump’ as with solenoids.
- Works for all gauges
- Works with DC, DCC, CBUS, etc.

Drawbacks

- Normal point motors can be fitted ‘straight from the box’. Servos need additional mounting arrangements and additional control circuitry.
- Unlike solenoids (with fixed distance movements) or Tortoise-like motors (with microswitches to set end stops) servos need to have a source of electrical pulses to determine their positions. Hence the need for extra electronics.
- Servos don’t come with mounts for baseboards, unlike devices designed for model railway use. Modellers have to make their own mounting brackets, or buy commercial mounts that often cost more than the servo.
- Requires a little more technical expertise to construct and set up.
- Most servos work at around 5V, which is not a common power output from loco controllers – hence the need for extra electronics.
- Cheaper servos wear out after only 25 hours of use (on the other hand, that is 30,000 point operations based on 3 seconds per point movement).

Choosing a servo

With such a variety to choose from, its best to look at their specifications.
This image shows some of the smaller servos that are available. In most cases the middle two servos are employed for most tasks.



Servo specifications cover these factors:

Torque

A measure of the ‘turning power’ of a servo. It measures the amount of 'push' or 'pull' that the servo can handle before the motor stalls. Torque is measured in ounces*inch or kilograms*centimetre. Consider a servo that is rated at 2kg*cm, and has a horn with a hole drilled through it that is exactly 1cm distance from the horn centre. The servo would stall if it had move a mass of 2kg at that point; any mass less than 2kg can be moved by the servo. If the hole in the horn had a distance of 2cms, it would stall at 1kg.
It makes sense – the further out from the centre of the motor, the less able the motor is at handling mass. The horns that are supplied with servos have a number of holes at different distances out from the centre (see the earlier drawing). These allow operating rods, cables, etc. to be attached to them.
Servos are available in a wide range of operating torques. For example, the DYS0206 has a

torque of 1.5Kg.cm, while the mighty Multiplex RHINO digi 4 has 24.47Kg.cm.
For comparison, the smallest servo in the photograph has a torque of 1.5Kg.cm, while the largest has a torque of 3Kg.cm. For operating lightweight fittings (e.g. signals or crossing gates) even the least-endowed servo will have no problem.
If you intend to keep using centre-sprung points, a little extra torque is required, compared to other points.

Operating speed

Retailers describe the servo's speed (or transit time, or servo turn rate) as the time the servo takes to rotate a certain amount. Thus, the Protech B305's speed is quoted as taking 0.2 seconds to rotate 60 degrees (0.25sec/60) and the tiny DYS0206 has a performance of 0.12s to rotate 60º. Servo speed is important when used to fly a model aeroplane but even the slowest operating speed is fast enough for model railway applications.

Weight

Again, this is very important for model aircraft but not normally a significant factor for railway modelling.

Size

This is less of an issue for operating points, as there is usually sufficient space under the baseboard. Smaller servos might prove useful inside small buildings and structures (e.g. people moving inside a signal box, or where a point has to be operated from above the baseboard). The most commonly used servos are significantly smaller than point motors.

Type	Case size	Possible uses
Large-scale	60mm x 51mm x 28mm	Robots, heavy work. Mostly non railway
Standard	38mm x 35mm x 19mm	Heavy work on layout
Mini	28mm x 25mm 16mm	Medium / light work
Micro	22mm x 16mm 10mm	Light work, small places
Ultra	20mm 15mm x 11mm	Linear (sliding door, semaphores)

Performance

The construction of the servo and the materials used have an effect on how smooth and reliable the servo operates. Compared to model aircraft and boats, servos in model railways would not likely suffer the same amount of constant use during a session. While an aircraft operator is constantly altering servo settings (i.e. many thousands of changes) during a flight, points or signals will probably enact much fewer changes in a session. Nevertheless, reliability can be enhanced by purchasing servos that offer improved construction such as higher-quality motors, double ball races, long life bearings or long life motor brushes.
The higher quality servos are designed for sudden acceleration and fast sequences of movements.
Servos with metal gears and metal output shafts have the most strength, handle the highest torque and are used where heavy or jarring loads are being handled. However, the gradual wearing away of the metal results in accuracy being slowly lost over time.
Nylon gears are cheaper, suffer much less wear but are less durable in adverse conditions, being prone to snap the shaft or a gear wheel when over-worked.
Karbonite gears are stronger and even more long-wearing than nylon gears and are found in servos such as the Hitec HS-6965.
For most railway applications, the cheaper nylon geared-servos are quite sufficient.

Here is a typical advert, providing specification details of a commonly used servo:

Tower Pro SG90 Heli/Aero Micro Servo A

Tower Pro A Grade Servo - A tighter and better manufactured SG90!

The weight of 9g is including the horn, this servo is only 8.5g without horn. Fast 0.9/Sec/ 60 degree speed and a whopping 1.6kg torque put this servo into a performance bracket of Servos twice the price and more ! Featuring an excellent quality casing and nylon gears this servo will suit Futaba, JR, Hitec, New Sanwa and many other radio systems with the same polarity.

This servo comes complete with fittings - when comparing prices check the servos are with fittings - many competitors charge extra for fittings!

Specification

Dimensions (mm): 11.0 x 22.0 x 22.0
Weight (grams): 8.5
Speed (sec): 0.11
Torque (Kg.cm): 1.60
Ball Raced: No

Price

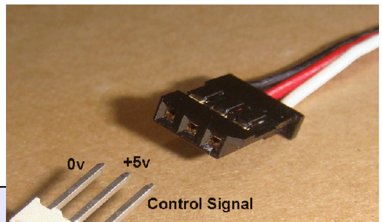
£2.99 Each
4 for £11.00
10 for £25.00

Analogue / digital

Servos are available both as analogue (by far the most common) and digital versions. Both use the same motors, gears, etc. The difference is in the electronics inside the servo. The analogue servo checks its position 50 times per second and adjusts the rotation if required. The digital servo checks and adjusts 300 times per second. This results in faster operating speeds and a more constant torque. Digital versions are more expensive and provide no real added benefit for the tasks performed in model railways.

Connections

Most servos have three wires exiting the casing and terminating in a 3-pin socket connector as shown in the image. The wires are colour-coded and their uses are:



Wire Colour	Signal	Description
Red	+5V	Draws anything from 10mA to over 1A, depending on how powerful the servo is (read each servo's specification)
Black (sometimes brown)	Ground (OV)	Common connection for both the supply and the input signal.
White (sometimes Yellow or Orange).	Signal input	The incoming signal to the servo, to set the servo's position. The input can be fed from transistors, ICs (both TTL and CMOS), PICs, etc.

Most servos seem to use red, black and white wires but this varies between manufacturers. Also, most servos use the same order of wires on the connector with the red wire in the middle. Again, this is not always the case, so check before connecting.
The socket has square holes spaced 2.5mm apart, to accommodate PCB header plugs. These are available from MERG, Rapid, Farnell, etc. They can be purchased as 3-way

single row PCB header plugs but is more economic to buy a 36-way strip and cut it into twelve 3-way pieces.

Note

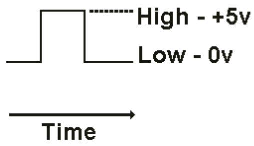
The socket is not polarised, so the plug can easily be inserted the wrong way round. It is up to you to ensure that the plug is wired to match the servo socket – and is inserted the correct way round.

Control signals to the servo

Assume for the moment that the three wires from the plug are connected to some electronic device that is capable of controlling the servo (more on servo controllers later).
The servo, in addition to its motor and gearbox, has a small printed circuit board with its own electronics, including some noise suppression. The internal electronics translate the state of the incoming signals into the rotational position of the output shaft.
The servo operates on an input signal of 5 volts.

- Low signal, also known as logic 0, is 0 volts
- High signal, also known as logic 1, is +5V

The ground wire (usually the black wire) is held at zero volts. The red wire supplies the power to the servo and is always held at +5V. In fact, various servos can operate over a range of supply voltages from 3V to over 7V, depending upon the model. However, its best to work on a standard of a +5V supply, since this is a voltage that all servos can expect to operate at.
The remaining wire sends +5V pules from the controlling electronics to the servo. Many people place a 10k resistor in series with the lead to the signal input, to limit the signal current into the servo.
The input signal sits at low level and is taken high for a certain period of time before being returned to low. This looks a square wave pulse as shown in the illustration.

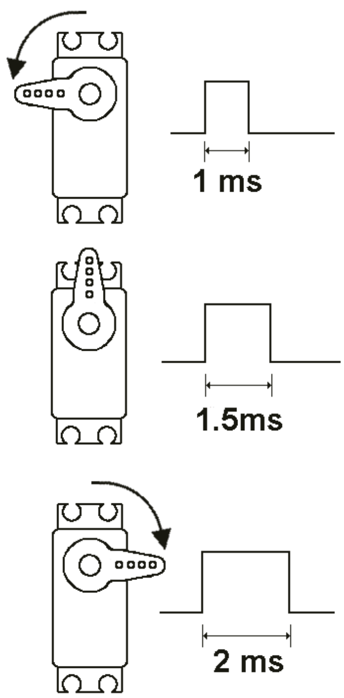


Timing and rotation

The control of the servo depends on this simple rule:

The width of the input pulse to the servo determines where the output shaft rotates to.

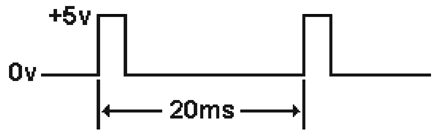
So a particular pulse width should always result in the same movement of the servo horn.
All servos are designed so that a pulse width of 1.5ms rotates the output shaft to a mid position.
For the purpose of the illustration, a pulse of width 1ms (a millisecond – a thousand of one second) makes the output shaft rotate to its most counter-clockwise position and a pulse width of 2ms rotates the shaft fully clockwise.
This relationship between pulse width and rotational position is shown in the diagram.
This supposes that the servo has a 180-degree maximum rotation. In practice, servos can be found with maximums that vary from around 90 to 180 degrees.
Some servos allow an input pulse to be as short as 0.75ms and as long as 2.25ms.



Note

While all servos of a particular make and model are identical, the range of available rotation and the pulse width required for that range could vary between different models. Modellers might consider sticking to one model for identical operations (such as points), or ensuring that replacement/additional servos have the same specification as those currently in use.

Other, physical, alterations to the servo are possible to obtain continual rotation with no ends stops. This is not normally needed for our purposes (if you need a rotating motor – buy a motor!) but interested readers can find details on this via Google.



Normally, pulses should be sent to the servo 50 times per second (i.e. every 20ms) to keep the shaft in any fixed position.
So, for example, to keep the shaft in mid-position, pulses of 1.5ms duration would be sent every 20ms. Many experiments using different servos has shown that while desirable, this is not strictly necessary.

Where a servo is under strain, such as supporting a weight or acting against a spring, repeated pulses are required to ensure that the load does not move the shaft from its chosen position. In other situations, such as centre-sprung points, a few pulses are sufficient to move the blades, which are then kept, in place by the spring. Also, light loads such as signals and swing gates do not require constant refreshing, as the friction in the servo gears is sufficient to maintain position.

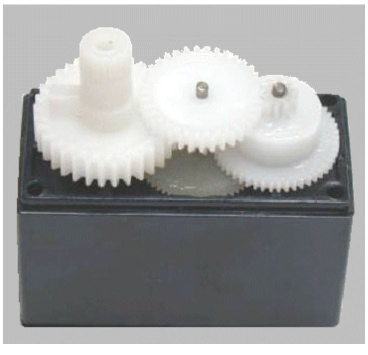
Where a stream of pulses is used, the pulses do not have to be at the exact 20ms rate. Anywhere between 15ms and 40ms works quite happily.
Repeating the pulses too frequently (e.g. 10ms) may cause servo jitter/chatter, produce a buzzing sound, or make uncontrolled movements. Repeating too infrequently (e.g. 70ms) may cause the servo to shut off and the shaft might lose its rotational position.

Also, the pulses do not have to be repeated at exactly the same rate. So, there could be a 25ms between some pulses and 35ms at another point in time.
This is because the length of the pulse is the most significant characteristic of the input signal

How a servo works inside

This raises two questions:

- How does the servo motor know when it has reached its required position?
- How does the servo automatically maintain its position, assuming a regular pulse stream?



The answers lie in the internal workings of the servo.
The image shows a servo with its upper casing removed.
On the left is the output shaft and this is rotated via nylon gears to a motor whose shaft can just be seen on the right of the servo.
Inside the case, there is a small potentiometer (variable resistor) connected to the output shaft. As the shaft rotates, the potentiometer's value is altered and the resulting change is fed back into the servo's circuitry.
This allows the circuitry to act as a 'closed loop'. It knows

what value should be fed back from the resistor for any given incoming pulse width. If both values correctly match, the motor is not energised since the servo is already in the correct rotational position.
If the value of the requested position (given by the pulse width) doesn't match the current actual position (given by the value from the variable resistor), the motor is turned in the appropriate direction until the two values match. At that stage, the motor stops drawing power.

More technical explanation

The servo's circuitry uses a pulse width comparator. It compares the width of the incoming pulse with a pulse from a one-shot timer whose period is set by the potentiometer value. The difference signal feeds a flip-flop circuit whose outputs determine the motor direction and drive an H-bridge circuit to provide the high current to supply the motor
The in-built feedback mechanism will handle any external force that tries to alter its position. So, if a heavy load begins to turn a servo from its designated position, the movement is detected and the motor is re-energised to bring the shaft back to the chosen position. Just try turning a horn with your fingers when the servo is being sent positional pulses!

Servo accuracy

To be really useful, you have to be able to depend on the accuracy and repeatability of the servo's rotation – for any given pulse width, it must always rotate to the same position every time.
However, tiny overshoot adjustments must be minimised, otherwise the servo would be continually drawing current as it makes tiny jittering movements back and forth.
For most servos, therefore, there is a trade-off between absolute resolution and ignoring tiny insignificant signal changes (known as the 'deadband').
Lets look at an example.

> A servo's incoming signal can vary from 1ms to 2ms – a difference of 1ms, or 1000µs (microseconds). There are no occasions when you would require the servo to take 1,000 steps to get from one end of travel to the other. For the sake of the example, let's say we are happy with using 50 different steps (still quite a lot). That means the width of the incoming signal would change by 20µs (1,000µs divided by 50 steps) to move to the next step. If the servo had a deadband value of 10µs, it would ignore all pulse width changes of 10µs or less – resulting in greatly reduced unwanted small movements, less current draw and less overheating of the servo motor and its electronic components.

In practice, servo deadbands vary between 1µs and 8µs.
In precision military and scientific applications very high specification servos are used with a very narrow deadband and this allows for greater resolution for the servos.
In a hobby situation, where the incoming signal cannot be guaranteed to be stable, or where cheaper servos are in use, a wider deadband allows for greater stability at the cost of resolution.

If we look at a servo with a 90º maximum rotation and a 2µs deadband, the resolution that is available would be 90 x 2 / 1000 = 0.18º, which is more than sufficient for all model railway activities.

Servo problems

Servo systems are generally very dependable but, like all electronic and mechanical systems, there are potential problems that may need addressed.

Buzzing

The servo's deadband prevents unwanted servo movement for tiny pulse width changes. But there are other problems that might prevent the servo reaching its deadband area.

- Sticky / blocked points – if the servo cannot move the point blades to the required position, due to debris on the track or a gummy tie bar, the servo will carry on trying to finish its task. The servo motor will buzz as it tries to reach its final stopping point.
- Incorrect alignment – the mechanical linkage between the servo's horn and the tie bar are improperly aligned, preventing the servo from reaching its final destination.
- Incorrect settings – the mechanical arrangements are fine but the servo has been given an incorrect final stopping point and is trying to move the tie bar beyond the stock rail.
- Poor power regulation – the glitches in the power supply to the servo, or its controller, or both, affect the voltages fed to the servo's internal comparator circuit. A servo only consumes a few mA when stationary but this can rise to 500mA when rotating, depending on the servo size and model.
- Interference – pickup of unwanted signals (noisy loco motor, DCC messages, CBUS messages, etc.) appear to the servo's signal wire input as movement instructions.

These problems can be resolved by ensuring that the points are kept clean and clear, the mechanical link between servo and tie bar are aligned properly, the servo is correctly configured and that the power and signal wires are routed to avoid interference.

Kicking

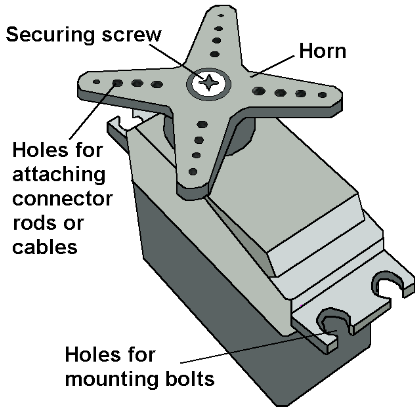
It is well-known that most servos produce a 'kick' (an unwanted movement) when they are first powered up. This is a feature of their internal electronics and is cleared in almost all cases by soldering a 10k resistor between the servo's signal wire and the +5V supply line. In some cases, it only effective if wired across the signal line and the 0V line. Configuration of the control electronics can also minimise this effect.

Twitching

Unlike the kick, which only happens during power up, servo twitching occurs during the running session.
This is almost always a result of external interference and can tackled by careful routing of layout wires, or placing a capacitor between the signal wire signal and 0v to shunt away any high frequency spikes on the signal wire.
Alternatively, the problem could be tackled at the servo controller end. If the servo needs to keep pressure on, it requires the constant repetition of the incoming pulses. Otherwise, the servo controller could be organised to stop sending pulses after a few seconds.

Although this looks like a long list of problems, in practice they are all either avoidable or are usually easily remedied.

Fitting servos

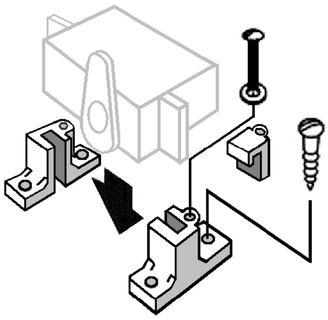
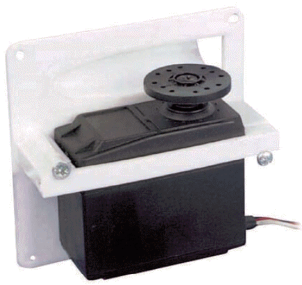
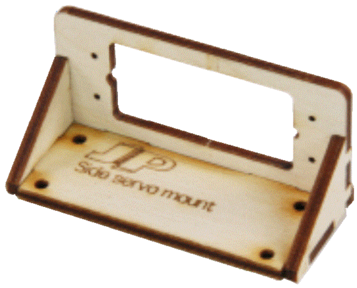


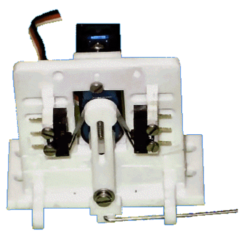
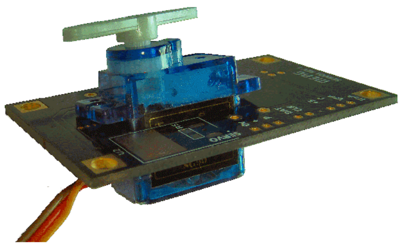
The illustration shows the key elements for fitting servos. Usually, the lever (called a 'horn') and fitting screw is supplied with the servo. Often, a selection of horn types is supplied. The horn has multiple holes drilled in it. The choice of horn, and the way it is used, allows a servo to provide either a rotational or linear motion.
There are mounting lugs lips on the side of servos; larger servos have four holes, two each short side and the smaller servos provide a single hole at each end. Some retailers also supply rubber grommets. These fit in the holes to provide a degree of suspension, which is useful in vibrating environments such as model aircraft.. They are often left unfitted to ensure accurate repeatable positioning of the servo horn.
There are two problems facing modellers:

- Mounting servos to baseboards, inside buildings, etc.
- Fitting operating arms, cables, etc.

Commercial mounts

These images show a range of the mounting brackets that are available from dealers.

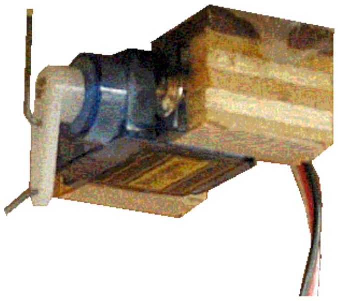
Top row: Small plastic mounts (eBay)
Solid one-piece plastic mount (SLEC Ltd) – plastic version from Conrad
Plywood self-assembly bracket (J Perkins)
Bottom row: Aluminium bracket (Active Robots)
Plastic self-assembly bracket, with two microswitches (MERG)
Servo1 PCB that acts as a mounting bracket as well as providing the control electronics (MERG)

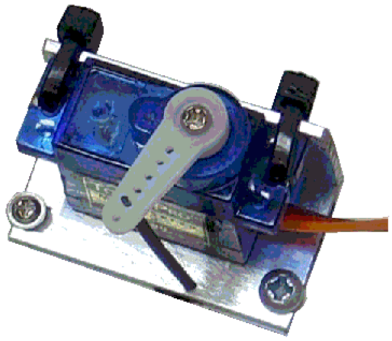
The images, which are not to scale, show that a variety of materials have been used in their manufacture.
Unfortunately, in most cases, the cost of the brackets exceed the cost of the servos themselves, although the MERG version is cheapest and provides additional facilities.

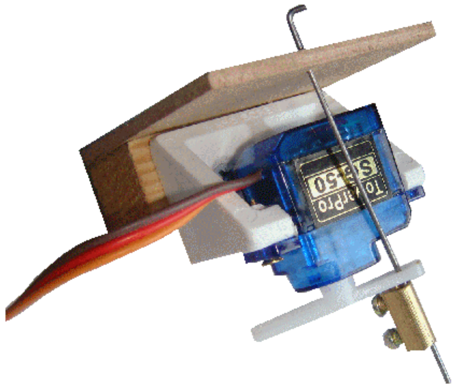
Home brew mounts

The relatively high cost of commercial mounts has led to hobbyists developing their own cheaper alternatives.

Modellers will provide endless alternative means of fitting servos and using them to operate devices. These images show three such approaches.







The servo on the left is screwed onto two small wooden cubes of the type that you buy as a bag's worth in a craft shop. The blocks are then glued to the underside of the baseboard.

The middle servo uses a piece of aluminium angle that you can buy in DIY stores. Two holes are drilled in the side and the servo is held in place by plastic ties. Two other screws hold the angle piece to the baseboard

The servo on the right uses L-shaped plastic brackets that you buy in DIY stores for fixing shelves and cabinets. A slot is cut in one side and the servo sits in the recess and two screws hold the servo to the bracket. The plastic bracket can then be screwed to a block for fixing to the baseboard (screwed or glued).

All three methods allow the mount to be fabricated before fitting to the layout.

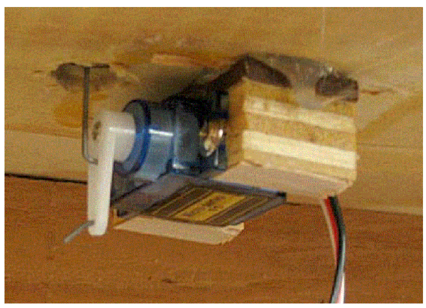
Attaching points to the servo

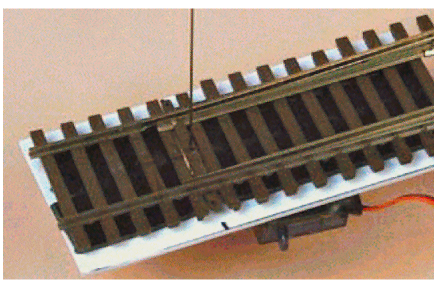
You can still use a servo to operate a point using the traditional indirect methods such wire in tube, omega loops, cranks and levers. In fact, this might be the best way for points that are not easily directly accessible through the baseboard.

For most servo users, however, the servo operates the points using a length of piano wire. There are two ways to achieve this:

Direct

This is the most simple method and is shown in these images. The piano wire is connected to the servo horn, bending the wire through the horn holes (some even securing it with additional wire binding and enclosing with epoxy).

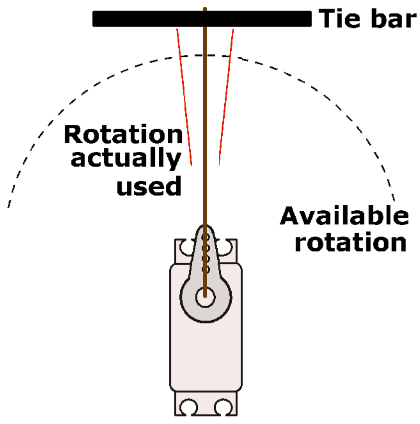




The piano wire passes through a slot in the baseboard and protrudes through a hole in the point's tie-bar. It needs no other mechanical linkage; as the servo rotates, the piano wire slides the tie bar.

This method works happily in many situations but has some limitations.

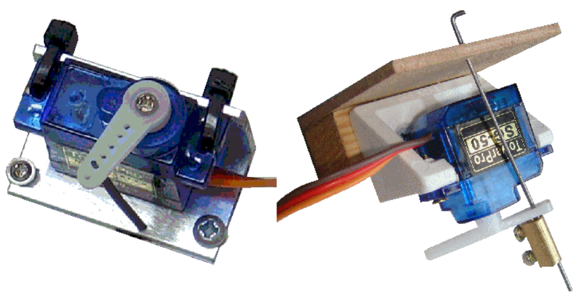
Consider this illustration showing the direct movement of the point tie bar with servo rotation. As you can see, the servo is capable of a much greater range of rotation than that being used to move the tie bar.
In fact, as the piano wire length is increased, the servo will rotate less and less to achieve the same amount of lateral movement of the point.
This has two effects:



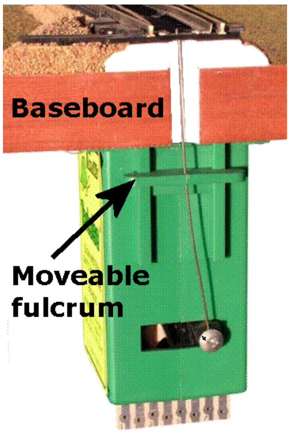
- Although the servo can complete its arc of travel using many different step (e.g. 256 steps) the tie bar is moved end to end with very few steps (e.g. 8). This greatly affects the movement resolution, as a slow motion point change would finish in just 8 steps, resulting in a more jerky movement.
- Small changes in servo rotation result in large swings at the end of the piano wire. This turns small unwanted kicks into potentially large and destructive jolts of the tie bar, unless stringent steps are taken to minimise start-up kicks, interference, etc.

Pivot

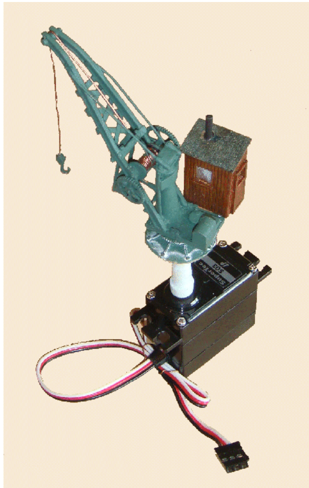
This method is safer and more realistic.
Looking again at these two homebrew mounts, we can see that they also have long piano wire actuators – but the fulcrum is not at the servo end, it is near to the tie bar.
So we have the opposite effect of direct linkage. A large rotation of the servo results in a smaller lateral movement of the tie bar.



Thus, we have the horn rotating almost from end to end. The many steps to move the point provide smooth slow motion. Also, any 'kicks' have no effect, since the servo works at its extremes anyway.



There must be a small hole at the fulcrum – no wiggle room.
The pivot method is the most configurable, with plenty of options for tweaking the movement of the wire actuator.
The piano wire can be inserted in any hole in the horn arm, providing different amounts of leverage. You can also set the distance from the horn to the tie bar and the position of fulcrum. Finally, you have the ability to change the servo's end stop positions in software.
The image on the left shows how the Tortoise point motor using a sliding fulcrum to adjust the amount of leverage.



Moving accessories

Servos are commonly used for point control but can be used for a wide range of lineside accessories, such as crossing gates, sliding doors, cranes, semaphore signals, uncouplers, grade brakes, etc.
The image on the right shows a model crane epoxied to a plastic tube that is run through a hole in the baseboard and is a push fit on to a servo's output shaft.
Servo use is only limited by your ingenuity.

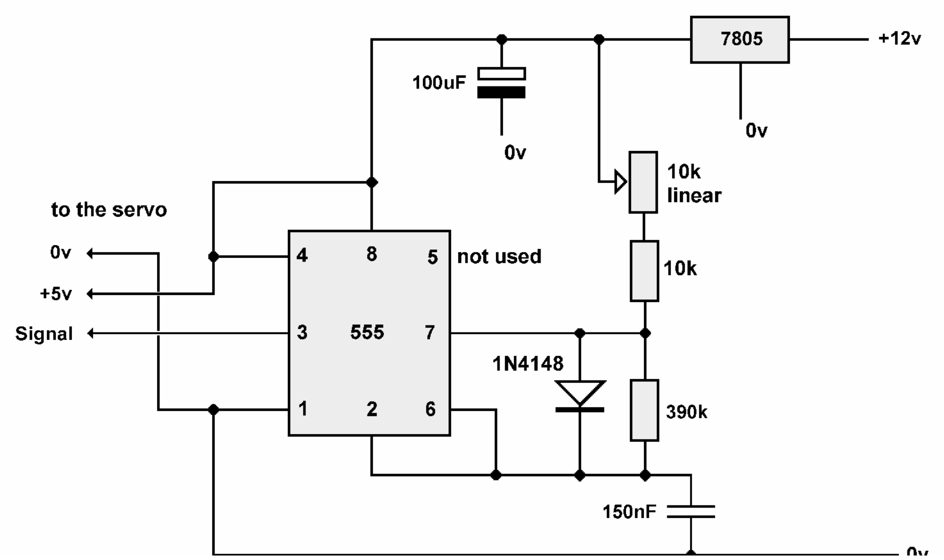
Back to Contents Page 5

Controlling servos

To control a servo, we need a source of pulses as described earlier. The circuit is called a 'servo controller' and can be built from transistors, 555 timer chips, flip-flop ICs or PICs. The circuits can be controlled from mimic boards or from computers.
They can be built from scratch, from MERG kits and from commercial control boards.

A basic homebrew controller

The circuit below is ideal if you want to experiment with servos without committing too much cash. The only costs are a cheap NE555 IC, a handful of components and a single servo. It is a tried and tested circuit.



The 555 chip is a timer IC which is just what is needed for creating timing pulses.
Servos expect to receive regular pulses on their *'Signal'* wire (the other two wires on the servo are for our 5V power). The pulses on the signal wire should vary in duration between 1ms and 2ms and be repeated every 20ms.
The 390k resistor and the 150nF capacitor are used to control the time between pulses, while these same capacitors are used in conjunction with the two resistors and 10k potentiometer to control the length of the pulses. So, varying the resistance of the potentiometer controls the servo arm's rotation.
The circuit works from a standard 12V DC supply and uses a 7805 voltage regulator to bring the voltage down to 5V for the chip and the servo.
For more details on 555 timers, read John Matthew's article in the MERG Winter 2006/7 Journal. Another very simple servo circuit is printed in the Elektor magazine Jul/Aug 2008 edition. This circuit uses a 7400 IC.
Although simple, its uses include:

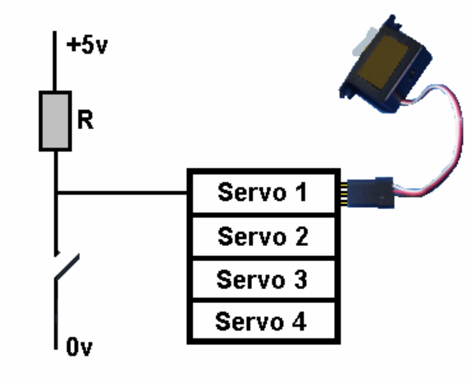
- A cheap introduction to playing with servos.
- A module to test servos before fitting them on a layout.
- Manual control of aerials, radar heads, guns (great for kids).
- When you need to move to intermediate positions (e.g. a flag at half mast)
- Mounting a cam on the arm, so you can scan a hidden area of the layout.
- An asymmetric turntable for smaller gauges (the pivot point is near one end rather than the centre).
- Any manual control for occasional use.

For many purposes, a more comprehensive controller is required.

So, what are the requirements for a model railway servo controller:

1. Easily operated by switches (control panel)
2. Easily operated from other railway modules (e.g. track detectors)
3. Easily operated by a computer
4. Easy to set a servo's endpoints (the limit of travel in each direction)
5. Easy to set the speed of servo rotation, preferably with each end point being able to be set independent of the other (e.g. able to move fast in one direction but slow in the other direction).

The first three requirements are usually met by having a controller that has a separate input pin for each servo. Pulling the pin to 0v makes the servo rotate in one direction, while pulling the pin up to +5V makes the servo rotate in the opposite direction. By operating from a simply change of voltage level, the servo controller input can be attached to anything that produces that voltage change. In most cases, this would be a simple on/off switch in the control panel, as shown in the illustration. Otherwise, any circuit that switches logic level (between +5V and 0V) can be attached to the servo controller input pin.



The fourth requirement, setting the end points, is either accomplished by a computer software application, or by using switches / pushbuttons on the servo controller board.

The last requirement, altering the speed of servo travel, is provided by some controllers, while others simply switch at a set speed.

Commercial controllers

The benefits of these boards are that they are supplied ready-built and tested.
There are three categories of servo controller:

- Industrial / military / scientific / medical – High specification, high price, modules that are not covered here.
- Hobbyist robot builders.
- Model railway controllers.

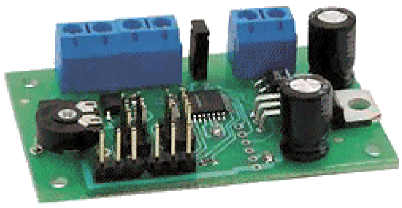
Model Railway Controllers

These three models are specifically designed for the model railway market.

Viessmann 5268 Servo-Control

This servo controller uses a supply of 10V to 16V, AC or DC and converts it down to 5V to operate the servo.
Unfortunately, it only operates a single servo.
This is fine if you only want to have the odd feature here and there on your layout (e.g. a signal or a crane).
It is too bulky and is not cost-effective if you want to operate dozens of points.
The end stops are set using two separate adjusters.
There is no control over servo speed.

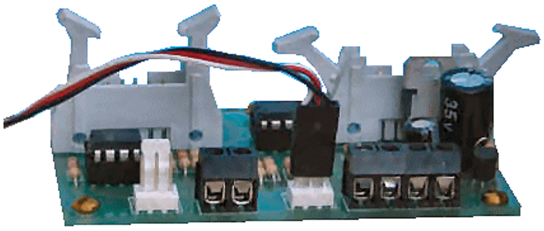
Conrad Train Module 73726



This controller is a little better, as it handles two servos.
It uses a supply 7V to 24V, AC or DC and converts it down to 5V to operate the servo.
Like the Viessmann, it is really only suitable for occasional servo use on a layout.
There is no control over servo speed.

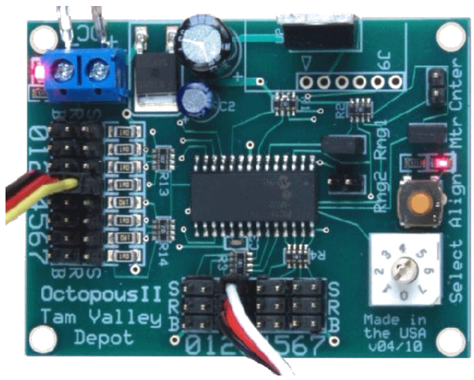
Heathcote Dual Servo Motor Controller

It uses a supply of 9V to 16V, AC or DC.
It also only handles two servos, limiting its appeal for point operations.



The endstops are set using two pushbuttons.
Seven different speeds can be set, using another pushbutton.
The sockets that can be seen at the back of the board are for connecting a remote setting box that replicates the boards pushbuttons. This saves you having to adjust point servos while crouching under a baseboard.
It has an interesting sequence facility, allowing you to control one servo from another. This can be used for crossing gates, as the second servo will only start to move when the first has finished rotating.

Tam Valley



The image shows the Tam Valley Octopus II model.
It uses a supply of 7V to 18V, DC only.
The rows of pins along the bottom are for connecting 8 servos.
The rows of pins on the left are for connecting 8 toggle switches, or digital inputs.
There is no control over servo speed.
It provides automated endpoint adjustments. You turn the rotary switch (the white one in the image) to the servo you want to set, then press the pushbutton above it. The controller automatically finds the end points for you.

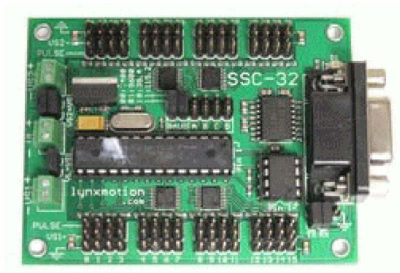
Robot servo controllers

A larger range of servo controllers has been developed for designers of robots.
They are more comprehensive and more complicated.
Some of the earlier controllers we looked at handled too few servos to be useful on a large layout. Robot controllers, on the other hand, can handle up to 48 servos from the one module. However, this is not as great an advantage as might be initially considered, since it would mean having a central servo controller with many extension leads snaking all over the layout to reach each servo.
The biggest drawback for most robot controllers is that they do not have a simple on/off switch approach to moving servos, relying instead on computer serial or USB port connections or special microprocessor signals.
Although they often provide extra facilities beyond simple servo switching, they require a degree of programming knowledge.

 Back to Contents Page 5

Here is a list of some robot servo controllers.

Website	Order code	Description
www.proto-pic.co.uk	Pololu Micro Maestro 6 Pololu Mini Maestro 12	Drives 6 servos via a USB connection Drives 12 servos via a USB connection
www.milinst.com	1-904 1-131 1-132	Drives 8 servos via a serial connection Drives 16 servos via a serial connection Drives 16 servos via a USB connection
www.active-robots.com	MR-Servo Board Simple Servo Lynxmotion SSC-32	Drives 16 servos. Programmed using ICSP Drives 1 servo using a serial connection. Drives 32 servos via a serial connection.
www.robosavvy.com	Red Back Spider Robot Controller	Drives 48 servos via a I2C connection

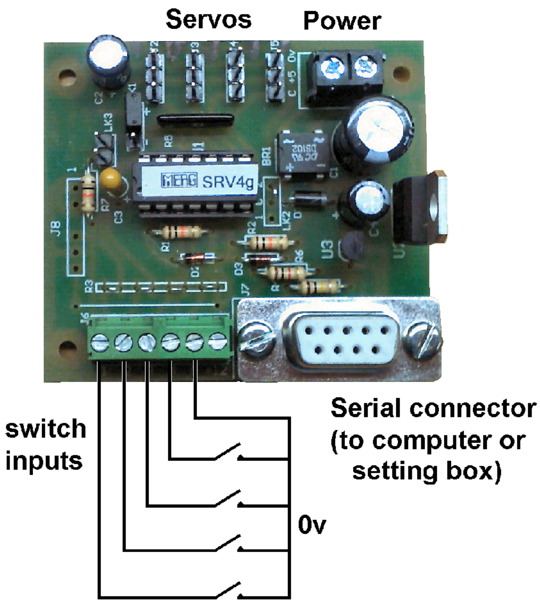


Here, for example, is the Lynxmotion SSC-32 controller
It can handle 32 servos and supports servos that can rotate a full 180º. Its resolution is an impressive 1µs (0.09º) – it can rotate a servo's horn in steps of only 0.09º.
It is operated via a computer serial cable

MERG controllers

MERG members have developed a number of controller options. One of these, the Servo4, is available in kit form and another, the Servo1 is sold as a PCB and pre-programmed chip.

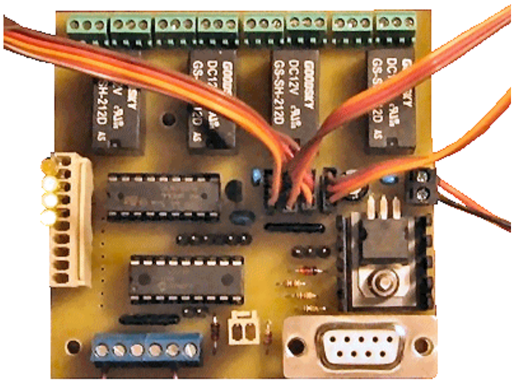
Servo 4



This image shows an assembled Servo 4 kit.
It uses a supply of 12V DC or 9V AC.
There are four servo connections at the top of the board. The four left-most connections in the lower screw connector block are the inputs that control the servos.
They are held at +5V internally but can be brought to 0v when they are wired to switches that are connected to the module's 0v line (the fifth connection). These switches would most likely be on a mimic board or control panel.
This module allows both a servo's endpoints and speed to be set, via the serial connector, using MERG's software utility (free to members).

You can program the board directly, if your computer has a serial port, or through a USB-to-serial adaptor if your computer only has a USB socket.
If you prefer, you can plug MERG's setting box into the serial connector, instead of using a computer.

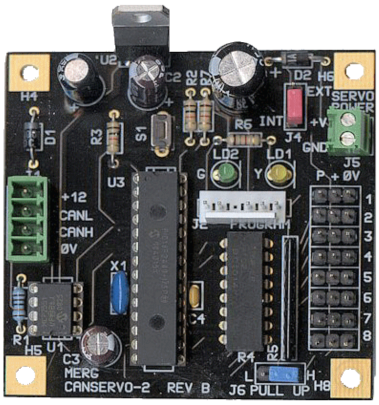
Servo4R



This image is of the Servo4R, which is similar in most ways to the Servo4 module above – except that it adds four relays to the board.
It uses a 12V DC supply.
These are used for point frog switching, etc.
This is a member's modification and is not available as a kit or as a PCB, although the schematic diagram, parts list, PCB layout, etc. are available on the MERG website.   Like the Servo4, it is set up via the serial connector.

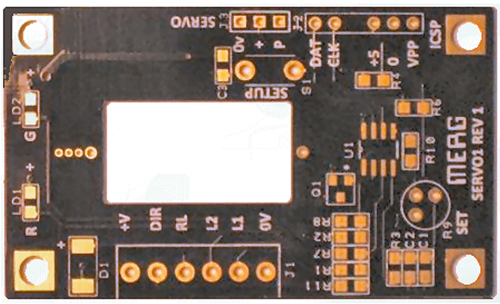
CANSERVO8

This image shows a module that is a combination of a CBUS board and a servo driver board.



It only works with other CBUS modules, where it controls up to 8 servos.
This is a member's design and is not available as a kit or as a PCB, although the schematic diagram, parts list, PCB layout, etc. are available on the MERG website.
If preferred, an existing CANAC8 module can be converted to do the same functions.   The module is programmed over the CBUS cable and both endstops and servo speed can be configured.   There is no provision for relay switching.
The CBUS system is covered in a later chapter.

Servo1



Like the Viessmann 5268 Servo-Control module looked at earlier, this design from MERG member Trevor Stockill only operates a single servo.
However, it has a number of additional features.
The PCB has a rectangular slot with holes either side, as shown in this image.   This allows a servo to be be screwed or bolted to the PCB, which has holes in each corner for mounting on a baseboard.   This saves the expense of a separate mount.
It has two output connections that change state at end of the servos' rotation – ideal for LEDs.
It has another connection for operating a relay – ideal for frog switching.
The servo's endstops are set with a pot and pushbutton.
Unlike the Viessmann or Conrad controller, the Servo1 lets you choose the servo's speed (using a single pushbutton).
It operates on either 5V or 12V DC.
It uses surface mount components (some as tiny as a grain of rice) which requires good soldering skills and good eyesight/large magnifier.
It is not available as a full kit, but the PCB and a pre-programmed PIC chip are available.

Installing a servo

This depends on the exact nature of the installation – where it sits (above board, under board), the method of actuation (piano wire, levers, wire-in-tube), etc.
Lets assume for this explanation, that we have servo screwed on a mount, wired to a controller for under baseboard operation using piano wire.
Before screwing the mount to the baseboard:

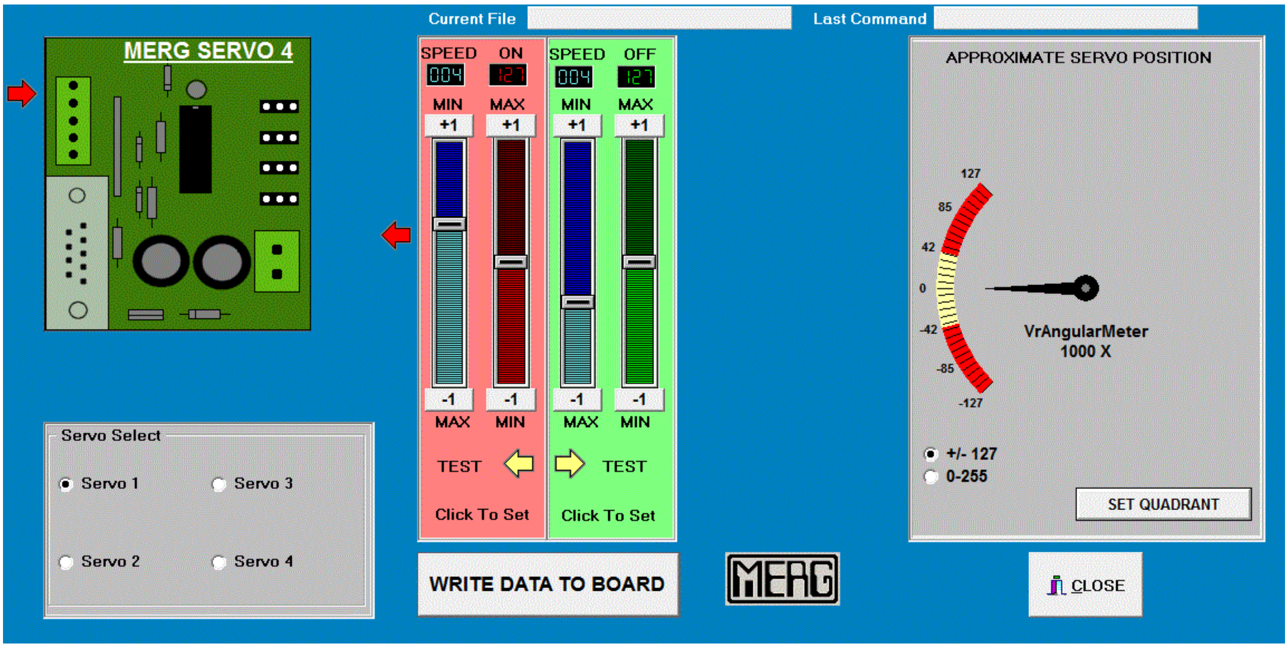
1. Use the servo controller to rotate the servo to its mid-position.
2. Fit the servo horn so that the hole being used for the piano wire lines up with this mid rotation.
3. Move the point tie bar by hand to mid position between both stock rails. Secure in that position with a piece of card/balsa/ foam.
4. Slip the piano wire through the hole in the baseboard from underneath and gently insert into the hole in the point's tie bar.
5. Move the servo and its mount to a position where there is no side pressure on the piano wire (i.e. the entire servo/wire/tie bar are in alignment).

Then:

6. Screw the mount to the baseboard.
7. Adjust both servo endpoints using the servo controller (this may be using pushbuttons, switches or software depending on the controller being used).

Configuring the Servo4 module

The Servo4 controller is MERG's most popular kit and the software used to configure the servos can be downloaded from the MERG website (MERG Software Download Pages).
When installed on your computer and run, it shows the following menu screen:



You choose which servo to configure by clicking in the "Servo Select" options.
The two left-hand sliders set the endstop and servo speed when the controller's chosen input goes high. The two right-hand sliders set the endstop and speed when the input goes low.
As you move the 'ON' and 'OFF' sliders, the servo position is instantly altered. Also, you can click the "Test" arrows to check that the servo swings are satisfactory before selecting another servo to configure. When finished, clicking the "Write data to board" saves all the settings and the connection cable can be removed.

Chapter 8

Scenic lighting

Scenic lighting

So far, we have concentrated on the technology that makes the layout run.
Now, lets have a look at some features that can enhance the look and feel of a layout.
Many layouts look good and run well but still seem to lack some sense of life.
This can be partly caused by the lack of movement – the only things moving being the trains. The previous chapter on servos opens up possibilities to bring other movement on your layout – people, doors, crossing, semaphores, etc.

This chapter looks at lighting your layout, to provide that lived-in look.

Lighting is often overlooked in otherwise excellent layouts. Yet it mostly easy to add to a layout, whether under construction or already existing.

Benefits

These are just some of the lineside features that could be illuminated to good effect;

- Street lamps
- Traffic lights
- Station lights
- Yard lights
- Crossing lights
- Semaphore signal lamps
- Lamps on buffer stops
- Buildings and factories
- Lighthouse
- Flickering fires / welding flashes
- Neon signs
- Flashing beacons in police cars/ambulances/fire engine/breakdown trucks
- Vehicle headlights
- Night scenes / sunset behind the mountains / skyline

These lights would certainly add appeal to trains:

- Loco cab lights. headlights, Mars lights, ditch lights
- Guards van / caboose lights
- Steam loco fire box
- Coach lighting

To these could be added cinemas, concerts, roadworks, crash scenes, searchlights, and so on – limited only by your imagination.
Clearly, having the layout looking more like a Xmas tree would not be helpful, but the judicious use of lighting can greatly enhance a layout.

Many of the features require little electronic knowledge and you can buy ready built units for some of the more complicated lighting features.
If you are willing to build your own circuits, you can make substantial savings and build lighting features that perfectly match your layout's character.

Filament bulbs

Before the coming of LEDs, all layout lighting used tiny incandescent bulbs. These are just much smaller, lower voltage, versions of the house light bulbs in common use.
Sufficient current is passed through the bulb's wire filament to heat it to the point where it gives off light.
The bulbs are available in two sizes, known as 'Grain of wheat' and 'Grain of rice'.
Grain of wheat bulbs are 3mm in diameter, while grain of rice bulbs are 1.8mm in diameter, with some as small as 1.4mm. They produce a natural light but can be purchased with coloured glass bulbs or lenses.
They are available in 3V, 6V and 12V versions.
They work on AC or DC and typical bulb currents are:
Grain of wheat bulbs

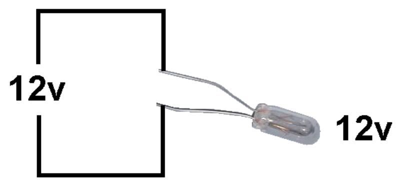
12V at 50mA or 80mA
6V at 50mA or 60mA
3V at 120mA

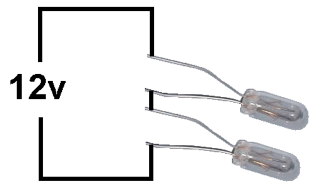
Grain of rice bulbs

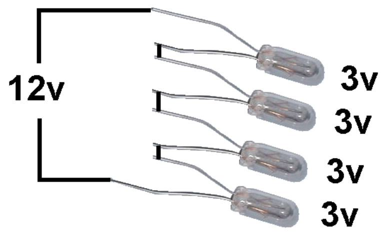
12V at 30ma

Series and parallel wiring

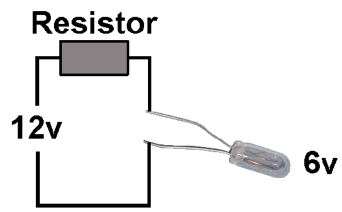
Since 12V is a common supply on a layout, a 12V bulb can simply be wired across the 12V. You can fit as many 12V bulb across the supply as you want, subject only to the supply and the wiring being able to cope with the current demands.
If you use 6V bulbs,you can wire two in series across the 12V supply, giving each bulb the 6V it needs. Similarly, you can wire four 3V bulbs in series across the 12V supply.







If you only want to light a single bulb, and it has a lower voltage rating than the supply, you have to put a resistor in series with the bulb to lower the current through the bulb.



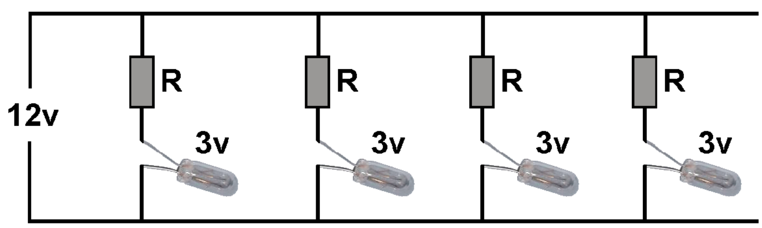
Take, for example. Using a 6V 50mA bulb on a 12V supply.
The value of the resistor can be calculated as:

R=V/I

where V is the 6V we want to drop across the resistor (leaving 6V for the bulb) and I is 50mA. So the value of the resistance would be:

R=V/I = 6/0.05 = 120 ohms

That works fine where a single bulb is being lit in a single location. If you want to have multiple bulbs lighting a building or a station platform, then you would not use multiple bulbs in parallel.
If we look at wiring four 3V



120mA bulbs as in the illustration, then 120mA is flowing through each resistor/bulb combination. The total current draw is 480mA (nearly half an Amp) and the power consumption is almost 6W. Wire a dozen inside a mainline station and you have used up 18W – more power than many soldering irons.
Compare that with the earlier example of wiring four bulbs in series across the 12V supply. The current drawn is 120mA (the same current flows through each bulb) and the power consumption is a smaller 1.4W.
Since filament lamps heat their elements to produce light, around 90% of the power consumed by the bulb is dissipated as heat – a clear fire risk. So, you have to be careful that you don't fit too many into a closed location (e.g. inside a building) without adequate ventilation.
Some users deliberately run their bulbs at less than their quoted voltage rating. This results in less heat, a warmer glow and an extended life.
Here is a comparison of incandescent bulbs to LEDs:
Pros

- A more natural white light
- Colour more suited to lamps of period layouts,
- Very thin leads – great for positioning, compared to stiff leads on LEDs

Cons

- More expensive
- More fragile
- Run hot
- Consume too much power

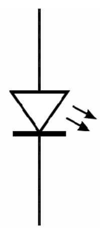
Light Emitting Diodes

LEDs have been mentioned a few times in previous chapters, since we have all come across them in many different domestic devices.
They have many advantages over using filament bulbs:

- Longer life
- Consumes far less current (some as low as 3mA)
- Run cool
- Available in many colours
- More robust
- Cheap
- Available in high intensity versions
- Rapid switching on and off does not decrease their life.

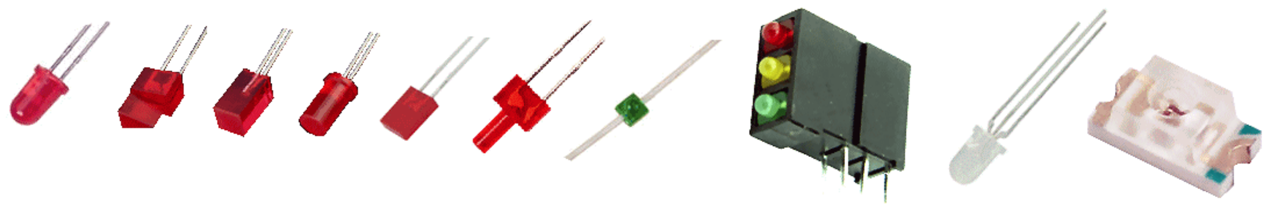
To make best use of them in a layout, we should look at how they work and how they can be wired.

The symbol for a LED is shown in the illustration. It looks similar to the symbol of a diode because it acts like a diode – it passes current in one direction only. In addition, it emits light when an appropriate amount of current passes through it. Inside the epoxy resin case, there is a small piece of semiconductor material that emits photon energy. This is called *'electroluminescence'* and consumes very little power and generates very little heat, compared to filament bulbs.

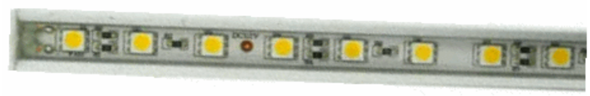


LED types
LEDs are available in a range of sizes, colours, shapes and features.

Shapes
This images shows some of the types you are likely to use.
They are not shown to scale. The last image on the right, for example, is the smallest of all.
They are produced in diameters of 1.8mm, 3mm, 5mm. 8mm and10mm

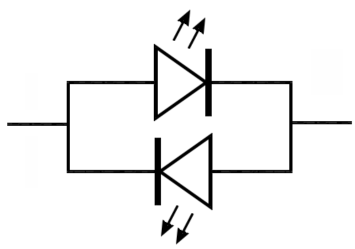


From left to right:
Standard, triangular, square, cylindrical, rectangular, lighthouse, sub-miniature, PCB mount, tri-colour, surface mount.
Most of the ones shown are known as 'through hole' LEDs, as they are intended to be mounted on some kind of panel (in our case, a mimic panel). Others, like the one with three LEDs and the one on the far right, are intended for mounting on a PCB.
This image shows a LED strip which can be bought by the metre or in 5m rolls. It runs on 12V and can be cut into smaller lengths.

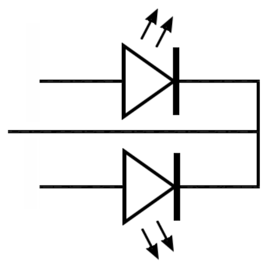


Colours
A LED's colour comes from the semiconductor material used, not from additional filters or a coloured epoxy casing.
The colours include red, pink, orange, yellow, amber. Green, blue, white and warm white.
They are available in versions known as diffused (a frosty casing that results in a wider viewing angle) or clear encapsulation (produces a viewing narrower angle but a more intense brightness).

Bi-colour LED
This is just two different-coloured LEDs built into the one casing, wired in reverse of each other. It has only two connecting leads. If you connect the power one way, the top LED lights. If you reverse the power, the lower LED lights. To the user, the LED seems to change between the two colours.



Tri-colour LED
This also has two differently-coloured LEDs inside a single casing but this time they are not wired in reverse. It has three wires, with two of them commoned to the negative lead of the supply. Connecting any one of the remaining LED leads to the positive supply lights that particular LED – giving two possible colours. Now, if the positive supply is connected to *both* leads, the colour mix produces a third colour. The most common combination is red and green LEDs that produces a yellow colour when both are lit at the same time.



Brightness

The brightness of LEDs is measured in millicandelas (mcd), being one-thousandth of a candela, so higher values of mcd mean brighter output.
As mentioned, LEDs can be designed for different viewing angles, depending on whether you want a spotlight effect of a general illumination effect. Viewing angles can vary from 15º degrees to 120º degrees
The unit of measure commonly used to describe LED intensity is the *'mcd'* – the millicandela. 000 millicandela equals 1 candela.
Here are some typical specifications taken from suppliers material.

Colour	Max voltage	Max current	Viewing angle	Brightness
Blue	3.2V – 3.8V	20mA	20º	10,000 mcd
Green	3.5V	20mA	20º	6,000 mcd
Red	1.8V – 2.2V	20mA	60º	5.000 mcd
Orange	2V	20mA	20º	4.000 mcd
White	3.5V	20mA	45º	100,000 mcd

The 'max voltage' is known as the *'forward voltage drop'* and is the voltage across the LED when it is running with its maximum current, providing its maximum brightness and specified colour.
It is important to recognise that these are *examples,* intended to stress that the specifications vary between different manufacturers and different products. For example, while one LED might use a 3.8V potential across it to illuminate, that might be enough blow a different LED. If in doubt, error on the side of caution or, even better , read the specification.

Surface Mount LEDs

These LEDs are very small and are useful for signals, traffic lights, lamps and lanterns, etc. They are available in a range of sizes, with their code describing their dimensions.
LEDs in the smaller range are described in thousands of an inch. Take, for example, the 0402 LED. The 04 indicates that it is 0.04” long (i.e. 40 thou) and the 02 indicates that it is 0.02” (i.e. 20 thou) wide. This equates to a tiny 1mm x 0.5mm.
Confusingly, the 3000 and 5000 ranges describe the measurements in mm.
So, for example, a 3528 LED measures 3.528mm x 2.8mm.
Here are some of the more common surface mount LED types:

	inches	mm
4002	0.04 x 0.02	0.02
0603	0.06 x 0.03	0.03
0805	0.08 x 0.05	
1206	0.12 x 0.06	

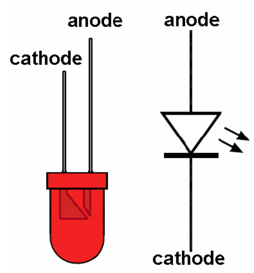
	mm
3014	3 x 4
3528	3.5 x 2.8
5050	5 x 5
5630	5.6 x 3

Other LEDs

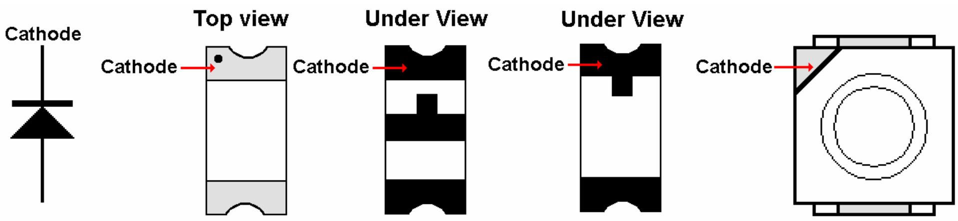
Apart from those already mentioned, LEDs are available that produce infrared or ultra-violet radiation.
There is even a LED with a built-in flasher, so that it flashes on and off at a rate of around 2Hz when powered. No need to build electronic circuits to flash your LEDs, unless you need a different rate of flash.

Wiring LEDs

A LED has two leads, called the anode and the cathode.
The cathode lead is the shortest and goes to the negative side of a supply.
The epoxy casing often has a flat edge along the cathode side.
The anode connects to the positive side of the supply.
If you are using a pre-used LED, the leads will have been cut. If your LED does not have a flattened surface, you can always look inside the LED, as shown in the illustration. The larger, wedge-shaped, piece is the cathode. If all else fails, you can just use the trial and error method – it will only illuminate when wired the correct way round.



Surface mount LEDs have various possible markings to identify the cathode as shown.



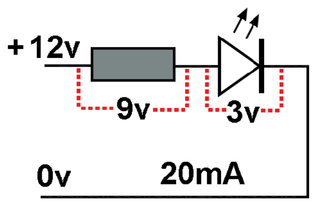
LEDs are current driven devices, which means that their brightness level increases as the current through it increases, with two provisos:

- A LED will only start to illuminate after it reaches certain critical voltage level.
- A LED cannot be given more than its maximum rated current without affecting its performance and eventually destroying it.

From the earlier table of typical LEDs, we can see that they work from voltages that are not common on model layouts. So, we have to reduce the voltage by placing a resistor in series with the LED.

Resistor calculations

To work out the required resistance, you need to know:



- The voltage of your supply
- The forward voltage drop of the LED
- The maximum current of the LED

Consider the example shown in the illustration.
It uses a 12V supply and the LED needs a maximum voltage across it of 3V. That means the resistor must drop 9V across it. Since we know the maximum current of the LED, we can calculate the resistor value by Ohms Law thus:

R = V/I   R=9/0.02   R= 450 ohms.

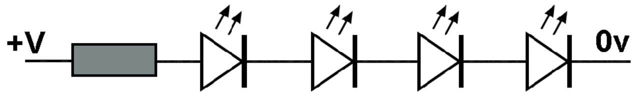
Since you can't buy a 450 ohm resistor, you would always choose the nearest *higher* value of resistor in the preferred series (to avoid excessive current). You would use a 470 ohm resistor in this example.

If you are technical, the formula is

$$R = (V_S - V_F) / I_F$$

where R is the resistance being calculated, V_S is the supply voltage, V_F is the forward voltage of the LED and I_F is the maximum LED current.
If you are not technical, you can use an on-line calculator to do the sums for you. Look, for instance, at `http://led.linear1.org/1led.wiz/`
These calculations are based on running the LED at its maximum specification. In practice, LEDs can often be run at much lower current levels (as low as 5mA) and still produce satisfactory illumination.

LEDs in series



LEDs can easily be wired in series, as shown in the illustration.
The same current flows through all the LEDs and the resistor, but the LEDs may have different forward voltages and this has to be taken into account when calculating the resistor value.
The formula for LEDs in series is:

$$R = (V_S-(V_F1 + V_F2 + V_F3)) / I_F$$

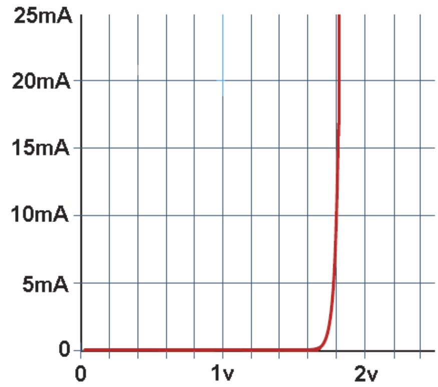
As an example, let's look at wiring four LEDs to a 12V supply.
All four LEDs run at 25mA but have different forward voltages of 3.5V, 2.2V , 2V and 3.8V.
The calculation is

$$R = (12 – (2.5 + 2.2 + 2 + 3.8)) / 0.025$$
$$R = (12 – 10.5) / 0.025$$
$$R = 1.5 / 0.025$$
$$R = 60 \text{ ohms}$$

Therefore a 62 ohm or 68 ohm resistor would be used.

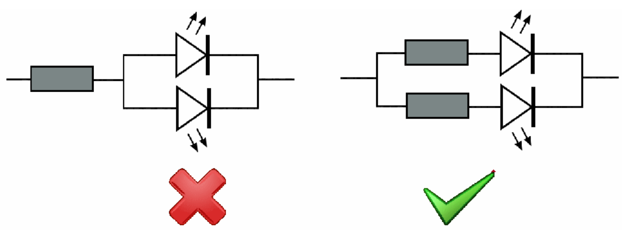
Parallel LEDs

Wiring LEDs in parallel should be avoided.
This graph shows how the current through a LED changes with increasing voltage.



If you placed a variable voltage supply across a LED and slowly increased the voltage from zero, for some time nothing would happen.
Then, suddenly, current would start to flow at an ever increasing rate for very small voltage increases.
The point at which this occurs depends on the characteristics of the LED. It is about 1.7V in the example but can vary between LEDs of different colour and even within the same batch of LEDs.

The problem with a bunch of LEDs in parallel is that they will all have different voltages at which current starts to flow. If you set the voltage sufficiently high to ensure that the LED with the highest characteristic voltage passes current, then the others are passing too much current. The LED with the lowest characteristic voltage then passes most current and finally dies. This results in the LED with the next lowest characteristic voltage passing even more current … and so on.



On the other hand, if you set the supply voltage to a level that is correct for the LED with the lowest characteristic voltage, it illuminates but the rest don't.
If you need to have LEDs close to each other, simply wire them in series. If there is not enough voltage to handle a couple of LEDs, then use a separate resistor in each LED, to match the individual LEDs, as in the right hand diagram.
The left hand diagram should be avoided.

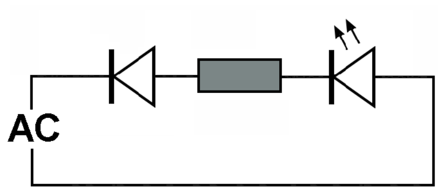
Fading LEDs

If you look at the earlier graph, you will see that eventually very small voltages increases result in large current increases.
This makes altering LED brightnesses very difficult simply by using voltage changes. So, for example, you could not dim a LED by placing it in series with a variable resistor.
In the chapter on loco controllers, we looked at pulse width modulation as a means of controlling the current through a motor. The same principle can be used to control LED brightness. Firstly, the resistor value is calculated for the required maximum brightness. Then power to a LED, or a string of LEDs in series, is delivered as a succession of on/off pulses. If the pulse period is long, with hardly any off time, the lamp illuminates brightly. Conversely, if the on period is short, with long periods off, the LED will be dim.
In other words, altering the pulse width changes the LED brightness. The rate at which the pules repeat has to be fast enough to prevent the LED simply flashing on and off.
The device that sends the pulses can be built from a simple 555 IC or a PIC chip. It can be a commercial dimmer (see Microminiatures) , or a complex DMX controller that handles all kinds of devices such as fog machines and moving lights.
Even at its simplest, the ability to simulate dusk, dawn, sunsets, etc., opens up ways to bring interesting effects to a layout.

LEDs on AC

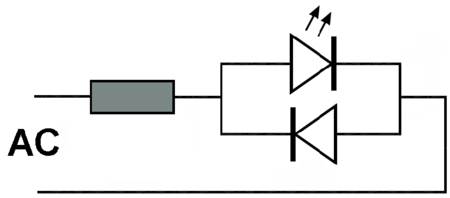
Some loco controllers provide a 15V AC socket for running accessories and this could be used to illuminate LEDS, with some precautions.
The diagram shows a single LED connected to an AC supply, with the usual dropper resistor to set the current level. The problem with AC in this case is:

- It is not always at a constant level
- It reverses polarity

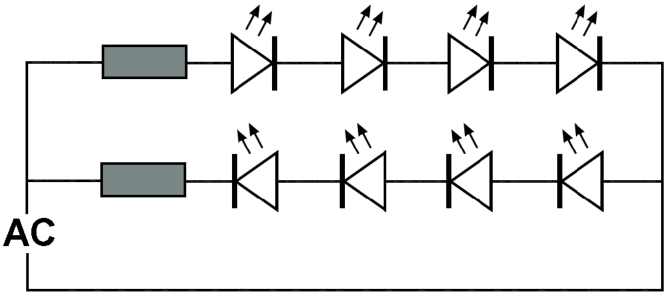


In chapter 1, we showed a AC supply as a sine wave, constantly changing both its amplitude and direction.
Now the LED is a diode, so it will ignore voltages that swing in the opposite direction to its conducting polarity. That means that the LED is not lit for half the time when it is forward biased. It is turning on and off at the frequency of the mains supply (e.g. 50Hz in the UK). Some users can detect the flicker.
Even when the supply is connected during the correct polarity, the amplitude is always changing. If we set the resistor value to switch on during the highest point in the voltage swing, it will not be illuminated for the rest of the half-cycle. If we lower the resistance value, the LED stays illuminated longer but we risk reaching currents at the peak which will destroy the LED. In short, we cannot achieve the optimum LED brightness with AC.

The diagram on the previous page shows a diode wired in series with the supply. This is fitted to prevent excessive reverse polarity voltage damaging the LED, as the diode can tolerate a much higher reverse voltage than a LED.



This diagram shows an alternative approach, using a diode (e.g. a 1N4001) being connected across the LED but in reverse polarity.
The diode and resistor then absorb the power when the voltage is reversed.



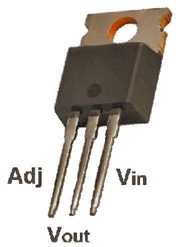
This diagram takes the idea a little further by wiring two sets of series LEDs, such that one set illuminates during one half cycle of the sine wave and the other set illuminates on the opposite half cycle.

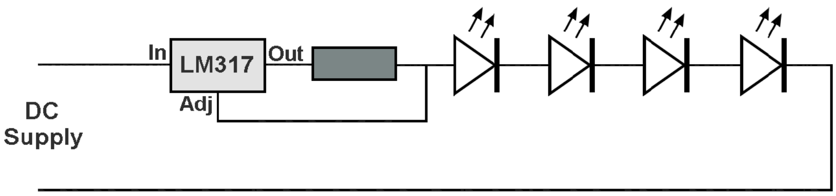
Another factor to bear in mind with a 15V AC supply, is that the value is stated as an RMS value (see chapter 1). The peak value is 1.4 times higher, so during the 'on' half cycle, there is actually a peak voltage of 21V being supplied.
This must be factored into the calculation for the resistor value, otherwise the LEDs will be destroyed.

The other approach is to take the AC supply, turn it into a DC supply using a bridge rectifier, then smooth it with a capacitor, before using it to power LEDs (look at the section on power supplies in Chapter 1).

Constant current supplies

Since LEDs are current-driven devices, and we know the current we want for any particular LED or set of LEDs, we can build a simple circuit that always provides that exact amount of current.
The circuit uses a *'constant current regulator'* and a widely used example is the LM317, which is a simple 3-wire device.
As the diagram below shows, it is inserted in series with one or more LEDs, and has its 'Out' and 'Adjust' leads connected across the resistor.





There is a very simple equation for calculating the resistor value:

R = 1.25 / LED current

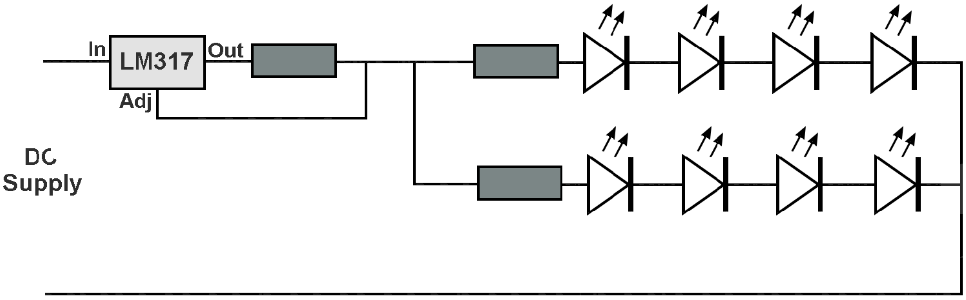
So, for example, if the required LED(s) current was 25mA, the equation is:

R = 1.25 / 0.025

R = 50 ohms

The resistor value is not dependent on the voltage of the supply.
However, you do have to ensure that the supply voltage is sufficient to handle all the LED forward voltages, plus an allowance of 3V for losses in the regulator.
If the above circuit used a 12V DC supply and the LEDs each had 2.2V forward voltage, then the circuit could handle four LEDs in series.

If you wanted to use multiple strings of LEDs, the following circuit is required, with each string having its own limiting resistor. The LM317 regulates the current to total current draw from each string. So, if the upper series used 25mA and the lower series used 20mA, the value of the resistor for the LM317 would be selected to provide 45mA.



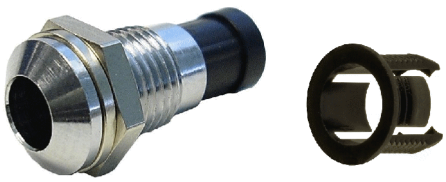
The specification for the LM317 regulator states that it can handle up to 1.5A, with an input voltage between 3V to 40V.
If you use this circuit with a much higher input voltage than that required by the regulator and LEDs, the excess voltage is dropped across the regulator.
Also, if you use the circuit to control high-power LEDs, or many strings of LEDs, the current through the regulator will be high.
In both these cases, you would have to ensure that the LM317 did not overheat.
Fitting a heatsink to the regulator, or bolting the regulator to a metal chassis (suitably insulated) helps prevent overheating.

Fitting LEDs

Once you start using lights on a layout, you find all sorts of uses for them.
These are only some suggestions on how to use them. You will find your own uses and your own ingenious ways to fit them.

LED holders

The images show two different methods of attaching LEDs to control panels.
The one on the left is popped into a hole drilled in the panel, the nut is fitted from the rear and the LED is inserted so that is appears on the panel. The other image is of a simpler plastic holder that is a push fit into a hole in the panel. It is a matter of taste, although many prefer the less intrusive look of the simpler holder.

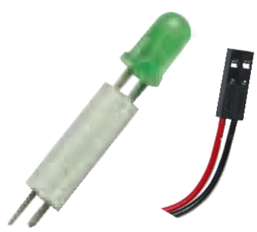


Panels lights can be used for:

- Indicating that the panel is powered.
- Confirmation of which direction a point is switched.
- Reporting back track occupancy information (see later chapter).
- Indicating which blocks / sections are switched.

LED mounts

Apart from control panels / mimic panels, LED holders are not required.
In many cases, the LEDs are hidden from direct view, illuminating buildings, etc.
The images shows a couple of approaches to attaching LEDs.
These are known as *'LED mounts'* and the LED is plugged into a holder that is connected to the rest of the layout wiring.
You can, of course, solder wires directly to the wires of the LEDs.



Bear in mind that LEDs are heat sensitive and avoid applying the soldering iron for more than about 3 secs when soldering.

Also, don't bend the LED's leads too close (say 3mm) to the case, to avoid damage.
Finally, don't forget that you can buy LEDs pre-mounted on strips with self-adhesive backing. They can be used in long runs, or cut into groups of three LEDs and pressed against a surface to hold it in position. Then, you run a 12V supply to it.

Attaching Lights

Before placing lights all over your layout, it is best to have a plan.
That plan should cover what switches, cables, fuses/cutouts, you might need. You might want to be able to switch different areas of light separate from others.
It should cover working out the current loads for different groups of lights and ensuring that the supply can meet your needs and there is no significant voltage drop in cables making long runs.
There is no 'correct' way to attach lights, as circumstances can be different in different parts of your layout.
A variety of methods have been used to attach lights to parts of the layout, depending on the space available, ease of access, etc. These include soldering to existing devices (e.g. relays), using a hot glue gun, double-sided tape, staples, homebrew brackets, choc blocks, terminal strips, tag strips, and so on.

Lights in buildings

While street lights, signal lamps, etc. add to a layout's appeal, it is the illumination of buildings and structures that most gives the impression of life going on.
It is particularly impressive when the house/overhead lights are dimmed and the glow and twinkle of the lights are most evident.

Power

The first task is to decide how a structure receives the power. This is almost always through a hole drilled in the baseboard. If the structure is removable (e.g. for protection during transporting to exhibitions), then a plug and socket arrangement will be most likely.
If the structure is permanently attached to the baseboard, the light can be permanently wired in. Remember, however, to allow for future access to the lights for future maintenance. This is most important when using filament bulbs with shorter lives. A removable back or lift-off roof is the most likely approach.

Sound structure

When you add lights to a building, the light should only be visible in the areas you want people to see. You have to ensure that the building's structure meet this need. Plastic kits and some cardboard kits are often sufficiently translucent to make the building look like a lampshade. Even home-built structures can allow light to leak from joints and seams.
This is easily checked by placing a light in the building and viewing it in a darkened room. If the walls and/or roof glow then you can either paint the interior walls/roof with black paint, or line them with extra card (leaving holes for the windows). Any gaps in joints can be covered with card or taped over.

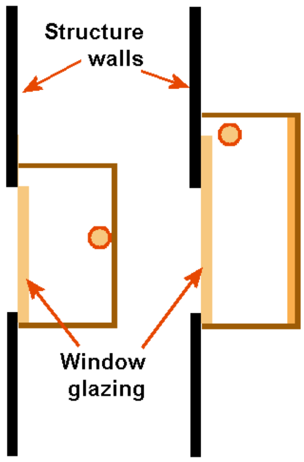
Windows/doors

The placing of lights should reflect their use in real life. So, you might expect all the lights in a factory/restaurant/car showroom to be illuminated but not all the windows in a block of flats. The lobby, hallways and stairways of a hotel would be illuminated but not every

room. You could use soft white, yellow or amber lights for period layouts and cool white for modern layouts.
If you have built a kit, it probably included glazing sheets. Often, they are transparent and some of the windows have to be rendered translucent while other are blocked off with card/tape. The Wills pack of plastic windows is supplied in a clear plastic 'blister' pack and you cut pieces from the packaging to glaze their windows.
If you have built your own structures, you can create windows from various materials. A popular option is the clear acetate sheets that are sold for overhead projectors or for laminating machines. To make them translucent, they can be lightly sanded with a fine grade grit paper. Alternatively, the light can be diffused by covering the window material with masking tape or 3M Invisible Tape, depending on the degree of dimming you want.

The picture shows a range of other materials that could be put to use. The 'frosty' sheet and the red plastic sheet at the top of the picture were recovered from the packaging of domestic goods. The orange and yellow crinkly sheets are sweet wrappers and the yellow sheet in the lower right-hand corner is from a plastic folder for storing files. On the right, you can see a piece of clear plastic with its corner sanded to make it translucent.

Care is required when fitting multiple lights in a structure, specially if you use particularly high brightness lights. You may have to fit card partitions between floors or between sections of a floor, to prevent light seeping into unwanted areas.
If a single light can illuminate a number of windows, that's good. Otherwise, you can fit lights to a single window or group of windows,by fitting a container over the window(s), as in the illustration. The box, usually cardboard, prevents light seepage and helps focus on the window.



The drawing on the left shows a LED being mounted at the rear of the box and this may be satisfactory where the window glazing is quite dense; otherwise the LED would be seen as a separate entity. The drawing on the right shows a LED mounted above the window and shining down on the rear wall. The LED is out of sight and all you see is the illumination of the real panel (which should be made from white or reflective material).

Note
You can widen and diffuse the narrow beam of light from a LED by gently filing down its dome end and polishing the flattened surface with toothpaste.

Another little touch is to cut out the silhouette of a person and stand it between the window glazing and the rear wall, casting a shadow. Don't place the cutout directly on the window glazing as this reduces the effect.

Remember, not all windows in all building should have the same degree of brightness – just as in real life. For added realism, you can place curtains or blinds on the windows. In industrial situations, some windows may be boarded up or covered in corrugated iron sheets.

Interior details

To provide additional realism, you can allow people to look into certain parts of a building and see its contents. This may be a room in a home, a shop front, a car dealership, a loading bay interior, a loco workshop, etc.
In these cases, the window glazing is clear and you have to take extra care to provide sufficient even illumination, without the lights being visible.
The visible contents are up to the individual modeller. You can print images, or cut them from magazines, and glue them to the inside walls of the room. You can go further, by adding internal details such as furniture or machinery.

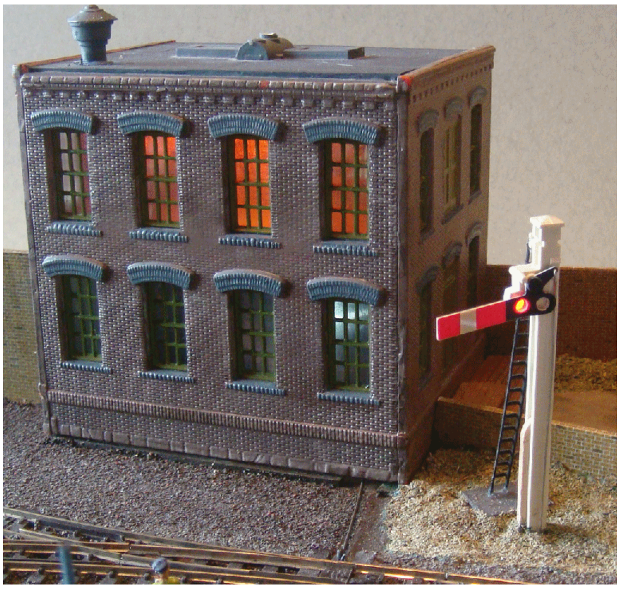
Exterior details

Buildings often have exterior lights in addition to interior lights. Domestic buildings may have porch lights and industrial buildings often have lights under canopies and arches, or over entrances.
Commercial buildings may have lights over their signs or even neon signs.
Business signs are easy to make if you have a computer but functioning miniature neon signs are best purchased (e.g. see Express Models).
Express Models also sell a multi-LED strip that looks like a fluorescent panel.

Other effects

Static lighting can be augmented by lighting that conveys a feeling of activity.
For example, commercial, MERG kits and homebrew circuits are available that simulate the flashing of a welding operation. They switch LEDs on and off in random bursts to achieve this effect. This can be used in various situations – loco shop, car repair shop, fabrication plant, roadside repairs, construction sites, and so on.
Another strategy to convey real life, is to have timers that switch various lights on an off in a random pattern – not too quickly of course.
There are also circuits that allow a LED to slowly increase then decrease its brightness – effective when built into a lighthouse.

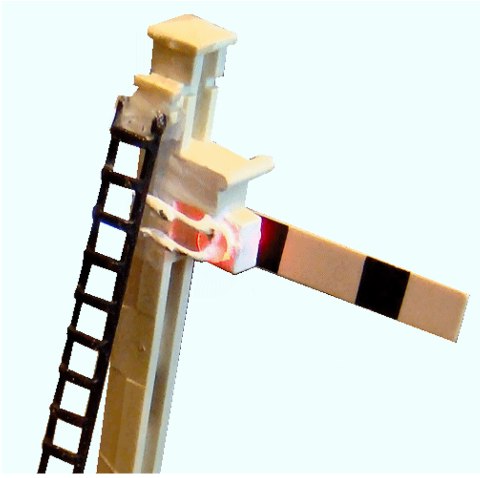
Lineside lights



Lighting should not be confined to buildings. There are lots of lineside situations where lighting can be effective. Consider station platform lighting, yard lights, lights in crossing gates and barriers, lamps in buffer stops, etc.
While platform lamps and yard lights are available commercially, you may consider it worthwhile fabricating your own.
This image shows a factory building and a semaphore signal. The building's top floor is permanently illuminated and you can just see a welding operation starting on the ground floor. The 'welder' is made from a MERG “Twinkler”

kit supported by extra brightness LEDs. The image shows a commercial semaphore signal with two modifications;

- The arm is operated by a servo.
- It uses a bi-colour LED for illuminating the lamp.

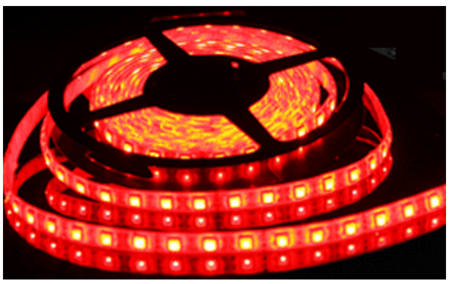


The image on the right shows how the LED is inserted into a hole drilled in the body of the signal. The two connections to the LED are soldered to short lengths of Kynar wire (very thin wire that can easily be hidden) which are fed through the base of the signal.
The LED is a red/green bi-polar type. When fed with its voltage in one direction, the green light illuminates; reversing the voltage illuminates the red light.
The reversal of the LED's polarity could be achieved by a relay, logic chips or a PIC chip.
In the example shown, a PIC chip was chosen to provide a delay before switching, to allow the semaphore arm to travel.

Scenic lighting

There are many occasions when you come across lights, other than in buildings or railways. Examples include street lamps, traffic lights, vehicles headlights, flashing beacons in police cars/ambulances/fire engine/breakdown trucks. Model railway versions can be purchased or built from simple circuits.
A variation on the welding flashes mentioned earlier is to achieve the flickering effect of an open fire. It uses much the same circuit but with different timings and different coloured LEDs.

For the ambitious, there is also the option to install 'mood lighting' on your layout. Imagine your scenic background being illuminated to simulate different parts of the day and /or different weather conditions.
Different coloureds lights and different light levels could be used to convey mood (bright sunlight, the glow of a sunset, cold or snowy conditions, moonlight, etc.).
It requires multiple lights whose light patterns must overlap, otherwise you simply get a row of light 'spots'). The lights could be hidden behind buildings or walls and angled to shine on to the layout background. If preferred, lights could be hidden behind a valance that skirts the top of the background. An even better approach is to have hills/mountains/low relief buildings set back slightly from the background board; the lights fit in the space and the resultant light appears to come from behind the hills.



The effect can be stunning but requires a lot of lights and electronic controls.
If you want more than one effect, you would have to fit separate rows of lights for each effect – an amber row for sunsets, a blue row for cold mornings, etc.
Fortunately, the LED strips mentioned earlier are available in 5m strips at a relatively moderate price.

Commercial lamps

The images below show a range of commercially-available lamps (they are not to scale). They include gas lamps and electric lamps, old lamps and modern. They are available as wall lamps, street lamps, platform lamps, yard lamps, etc. Most use LEDs but some are still produced using grain of wheat bulbs.

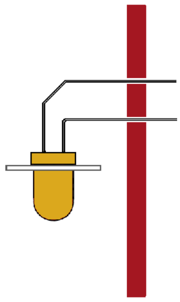
Manufacturers include Busch, Viessmann, Brawa – and unknown brands available on eBay. Double check before you buy, because some street lights, such as those from Faller and Ratio, are dummies. They are cheap but contain no bulbs or LEDs.

Homebrew lamps

Some of the commercial lamps are quite intricate and would take a lot of effort to replicate. However, they are quite expensive and it is quite easy to make some other types yourself.

The image on the right is of a commercial lamp using a filament bulb. The image on the left is a partly-made lamp using an LED. It is cheaper than the commercial version, uses a longer lasting lamp and its construction is quite straightforward.
The pole is a thin copper tube, which also acts as one of the conductors to the LED. The other conductor is a length of fine Kynar wire that is threaded down inside the tube. One lead from the LED is soldered directly on to the mast, with the other being soldered to the Kynar wire.
The mast is made from copper wire, or paper clips. Ornate swirls are associated with period lamps, while modern lamps are more plain.
Similarly, you may select a yellow, white or ultra-white LED, depending on the period in which the layout is based.
A washer is glued to LED to appear as a lampshade; the top of the LED would be painted to prevent unwanted light being seen above the lamp.
The lamp is left unfinished to make it easier to see how it is constructed.
A hole would be drilled in the baseboard and the copper tube would be inserted and secured in the hole. The wire and the copper tube protruding under the baseboard would be wired to the lighting circuit.

This image shows a wall light made from an LED.
Again, a washer is glued to the LED and the LED area above the washer is painted to exclude light.
The wires from the LED are bent as shown in the sketch, to represent the lamp's bracket.
Two tiny holes are drilled in the wall of the building and the two wires are inserted into the holes.



After soldering the two wires to the leads of the lighting circuit, the wires can be secured in place behind the building wall. A simple bend of the LED's leads will stop the lamp from moving but you can always use glue or a glue gun if you prefer.

For very small lamps, or for smaller railway gauges, you could consider using surface mount LEDs. These are very small indeed (e.g. 0.mm x 1.6mm from Bromsgrove Models). They are also suitable for fabricating railway lamps for wagons and coaches, buffer stops, guard's lamps, etc. Their extra small size allows them to be fitted into seriously small spaces that other lamps could not use.

Fibre optics

Fibre optic filaments are used to transmit light over distances as, for example, in carrying data over telephone phone lines. The light source is attached at one end of the fibre strands and the light bounces along the strands.
The separation of light source from the eventual point of light is very useful for railway modellers, as it allows lights to be fitted where filament bulbs or even LEDs cannot be located.
Example uses of fibre optic lighting include:

- Headlights in small gauge vehicles.
- Headlights in locos.
- Guards' lanterns.
- Miners' helmets.
- Signs.
- Fluorescent tubes.

As one light source can handle multiple individual strands, it is ideal for places where many small lights are required. How about a star-lit sky, or the the city lights of a background scene?

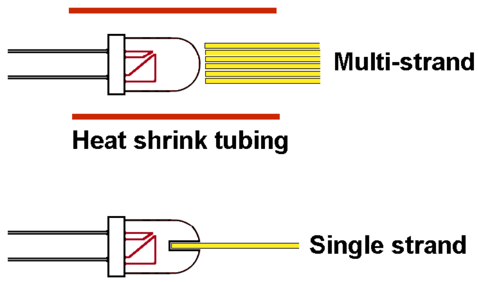
Fibre optics are produced in glass and plastic – we would use the polymer-based versions. These are sold either in single strands or as multi-strand cables.
You can buy single strands in different diameters ranging from 10mil to 120mil (e.g. 0.25mm to 3mm), depending on how much light you want to shift and how much space is available. You can also buy mulit-strand versions, where 32 or 64 strands are contained within a single covering.

While any light source can be used, ultrabright white LEDs are commonly preferred.

Unlike bulbs and LEDs, a single fibre optic strand does not project much of a beam of light; its light is meant to looked at, not used for illuminating other things (unless you use a bundle of strands).

At the LED end

The strand(s) to be used should be prepared prior to fitting. The ends should be flat for maximum light transmission. A scalpel will normally cut a length of fibre strand neatly. If not, the end can be gently rub on a piece of wet and dry paper, fine grit paper or emery board. You can take it further by polishing the end by rubbing it with toothpaste.
The strand(s) should be fitted close to the LED's body.
If using a single LED, a hole the same diameter as the strand can be drilled through the LED's epoxy casing (as deep as possible without damaging the LED elements).
Ensure that the hole is drilled as close to the centre of the dome as possible; alignment is more important than proximity.
The strand is inserted into the hole and secured using any plastic-friendly glues. This includes a glue gun, epoxy, silicone glue, polyurethane glues but not Super Glue or any Cyanoacrylate (CA) as these will attack the plastic of the strand(s).

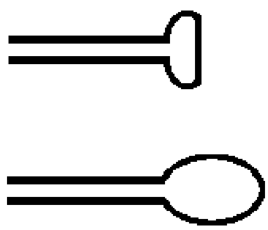


A multi-strand cable can be attached to the LED by covering both with a length of tape or, better still, a length of heat shrink tubing. If using heat shrink tubing, make sure you avoid overheating the fibre strands inside, as this will degrade their light transmitting properties.
You can make your own multi-strand fibre cables, using as many strands as you need. You would bundle them together using a suitable diameter length of heat shrink tubing. Then you would cut and polish for best effect before encasing the bundle ends with the LED in a larger diameter length of heat shrink tubing.

At the light end

Single strands are so thin that they can be easily routed in most layouts. The LED can be located under the baseboard or in a lineside structure. The strand can be routed through a hole in the baseboard or structure and is flexible. It can bent to a radius of around 9mm radius for thinner strands or 20mm for thicker strands, although over-bending may result in some loss of light. Because they are so thin, they can be routed above the baseboard; under grass, roadways, track (there are no mutual interference problems). Strands can be painted with no ill effects on the fibre strands.

Fibre optic strands do not a very wide viewing angle – you have to be looking directly at the end of the strand.
This can be improved by shaping the end of the strand.
The first drawing shows creating a 'bell end', by briefly touching the end of the strand against the stalk of a soldering iron and gently pushing into the strand. The end now offers an improved viewing angle.
If you heat the end of the strand and then, before it cools, flatten it with any tool. This produces a disc-like illuminated surface that can be viewed side-on instead of the usual end-on. See the second drawing.



Other effects

The light zig-zags along the fibre strand, bouncing off the smooth wall surfaces. If you scratch the strand surface, you will lose some of the light.
This can be a problem but is also an opportunity.
If you scrape the outer surface of a length of fibre strand, using wet-and-dry or fine grit paper, you create an illuminated length that looks very much like a fluorescent tube. If you want to take it further, you can light the strand from a LED that is driven by 'twinkler' circuit, and you get the effect of a flickering fluorescent light. The same approach can be used to create neon signs, using coloured LEDs.

How about making your own sign? Take a piece of plastic sheet, drill holes where you want the lights to be and insert strands into each hole. The sign can be your own design, using as many strands as are required from a multi-strand source. Even better, you could animate the sign by having two or more sets of LED-fed strands and switching between them.

Lastly, for those with the eyesight and necessary skills, how about lighting figures on your layout? Station staff holding lanterns, miners' helmets, etc. are possible.
Once more, your imagination can add to the effects listed above.

Chapter 9

Train lighting

Back to Contents Page 5

Train lighting

Now that we have our stations, streets, factories, houses, etc., all lit, we can turn our attention to our rolling stock.
Locos with headlights and tail lights could be a good starting point. Then, imagine the effect of a glow from the cab of a steam loco, or coaches illuminating their passengers and internal details.
This chapter looks at how this can be implemented both in DC and DCC layouts.

Of course, it is not quite as simple as scenic lighting, since the objects we want to illuminate keep moving. There are other restrictions such as avoiding excessive current draw, space restrictions, difficulty of fitting, and so on. Nevertheless, it is well worth the effort.

Choice of lights

There are three types of light that we can use for train lighting.

1.5V lamps

These are marketed as *'Grain of rice'* bulbs and are available from Rapid Electronics (part number 41-0186) or from many model railway stores such as Kytes Lights. They light at a lower voltage than LEDs but consume a higher current (15mA) compared to many LEDs. They also produce heat and have a shorter life. They are small, at only 1.4mm diameter and, being incandescent, they are non-polarised (they can be fitted either way round).
Generally, they are being sidelined in favour of LEDs. The examples used later all assume using LEDs.

Single LEDs

Single LEDs are mostly used where a single light is required, such as a loco's headlight or a train's tail light. They require a higher voltage before they illuminate – around 2V for yellow LEDs and 3V for white LEDs. They are more robust than grain of wheat bulbs, and produce usable light at currents as low as 5mA and sometimes even 2mA (well below the figures given as their maximum ratings).

LED strips

LED strips are purchased as a long continuous strip or roll of LEDs, with an adhesive backing. This makes them ideal for lighting coaches with a long LED strip; it saves wiring lots of individual LEDs and the strip can have its backing strip removed and be pressed on to the coach ceiling.
You can choose which LED spacing suits your project best. The strip/roll is available with either 60 or 120 LEDs per metre length.
Although continuous, the strip consists of a chain of individual sets of three LEDs with their own dropper resistor. Each section has its own solder pads, so you can cut off a group of three, a group of six, or whatever length you require.
You only require a single pair of wires to connect to a chain of any length.
LED strips rated at 12V, and the current consumption for a group of three LEDs is 20mA. They start to illuminate at 8.5V and provide adequate illumination at 9V, where they only consume 5mA per group of three LEDs. This means that a LED strip can be supplied by a 9V voltage regulator or even a PP3 battery.
At the time of writing, a 600-LED 5m roll cost £5. This makes home-brewed lighting a much cheaper option than commercial products.

Lighting features

The circuits for train lighting should meet some, or all, of these requirements.

Constant brightness

The intensity of the light(s) should remain constant regardless of the loco's speed or load. This is easily achieved on DCC systems because the track voltage level remains unchanged at all times. With DC, the track voltage varies from 0V up to 12V or more. At the lower levels (between 0V and around 3V) there is insufficient voltage to illuminate the LEDs and special approaches need to be adopted.

Flicker-free

The track voltage that is at a sufficiently high level to power the LED(s) may still be interrupted due to dirt on the track, the wheels or the pickups, or due to crossing a point's frog. This affects both DC and DCC systems. The circuit should maintain current to the lights during any track power interruptions.

Directional lighting

The use of headlights and rear lights varies with the era, the company and the country. For example, with double-ended diesels, there may be a white light at the end moving forward and a red light at the trailing end. This is easily achieved in DC and with DCC decoders that implement directional lighting.

Variety

Locomotives, in particular, may have various lights and lighting effects. These might include a flickering glow in the firebox of steam loco, flashing ditch lights, a rotary beacon, a gyra light, single and double pulse strobe lights, Mars lights and headlight dimming.

Switchable

Constant brightness lighting is fine, but we may not always want the lights to be constantly illuminated. When coaches or wagons are out of service and sitting in a siding, we don't want them to have any working lights.
With DC, its a bit easier, as the lights will only be lit if they rolling stock is sitting in a section of track that is powered. So, with block control, removing the power from a siding, etc. automatically switches off any lights. Otherwise, other means have to be used.
With DCC, some of the lights can be switched on and off through instructions issued by the command station.
For lights that are not controlled by decoders, there is a similar problem of how to switch them on and off.

Small size

Some of the circuits use very few components and these can usually be tucked away out of sight. In other cases, particularly where onboard power storage is used, space is an issue in the smaller gauges.

Powering the lights

There are two ways to supply power to train lighting:

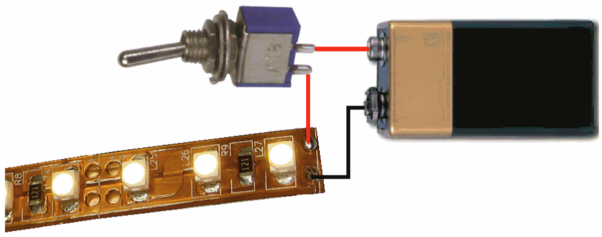
- Power on-board
- Power from the track

Power on-board

The idea of using battery power to run trains has been around for a long time and is still a common feature in cheaper 'toy' trains. Some modellers with garden railways also use battery power to run their locos. These use re-chargeable NiMh battery packs.

We can install batteries in our locos and rolling stock, dedicated to powering lights.
This provides reliable, flicker-free lighting. On the other hand, batteries need replacing or recharging and are bulky.

Here is a very simple homebrew circuit comprising a PP3 9V battery, a switch and a length of LED strip. Using an alkaline battery with a capacity of 550mAh, lighting a LED strip that uses 10mA gives a theoretical 55 hours of continuous running before needing replacing (120 hours if using a lithium version).



This is the basis of many commercial offerings.
For example, Layouts4u's coach lighting kit comprises a LED strip (white or amber), a coin cell battery and holder and a miniature slide switch.
For more details, see:
www.layouts4u.net/coachlighting.html

T&M Models provide LED strips, a switched battery box and either AA or AAA rechargeable batteries.
For more details, see:
www.tmmodels.net/Lighting%20Pages/Lighting_Main_Page.html

The 'O' Gauge Coach Lighting kit from First Class Trains comprises a white LED strip (Standard or Bright), a switch and a battery holder for AA NiMh rechargeable batteries.
For more details, see:
www.ukmodelshops.co.uk/catalogue/firstclass

All the above products use a switch to control the lights, which can be inconvenient. The need to replace or wait while batteries recharge is also a nuisance. This results in onboard power only being used where track-powered lights is impossible or inconvenient (e.g. with plastic tracks or unpowered tracks).

Power from the track

Most model railways run their locos from power picked up from the track rails.
This same power connection can be used to light loco and rolling stock.
Every loco has its own track pickups and these can also be used to provide lighting within or on the locomotive body.

Coaches and wagons do not have their own power pickups and therefore the power for lighting has to obtained in one of two ways:

- Getting power from the loco, using connectors between the loco and the first coach, between the first and second coach, and so on.
- Getting power directly into each coach or wagon by fitting it with its own track pickups.

Micro-connectors

Fitting your own pickups to coaches can be a bit of a challenge to some and they prefer to utilise the power that is already present in the loco.

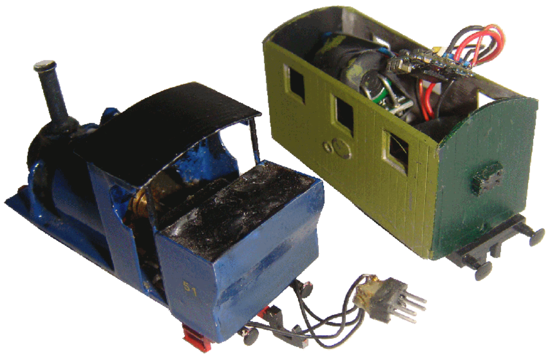
The image shows a range of miniature connectors that can be used to connect between the loco and items of rolling stock. With the two-wire version, you solder the socket's two wires to the loco pickups and the socket exits from the rear of the loco. The two wires from the plug are used to wire up lighting circuits in the coach. If desired, this can be repeated to connect each coach in a train.

The downside is having to plug and unplug the connectors every time you couple or uncouple a coach from a train. If you run trains of fixed length, this is not a problem. Although very small, the connectors are nevertheless visible, particularly when used with smaller scales.

For more details, see:

www.expressmodels.co.uk/acatalog/Micro_Connectors.html
www.digitrains.co.uk/ecommerce/electronics/micro-connectors/micro-connector-2-pin-pair.aspx
www.dccsupplies.com/shop/product_info.php?products_id=469

Home made



The micro-connectors are relatively expensive and a home-brewed version is shown in this image.

It uses PCB interconnect headers and sockets. You can just see a four-pin socket mounted on the side of the coach.

In this example, it is used to take track power from the loco's pickups into a coach which contains a DCC decoder and stay-alive capacitor, as these are too large to be hidden in the open-cab loco. The motor output from the decoder is then fed back to the loco's motor.

The header and sockets can be bought as strips and cut into the sizes needed – a 2-pin version would be less conspicuous.

For details of the connectors, see:

http://uk.farnell.com

The socket is Order Code 1577754 and the header is Order Code 1577743.

Electric couplers

In an effort to hide the electrical connection, Viessmann has built connectors into its own couplers. The image shows their 2-pole electric couplers and these are available from:

www.modeltrainsuk.co.uk
www.railroomelectronics.co.uk
www.gaugemaster.com
www.expressmodels.co.uk
www.dccsupplies.com

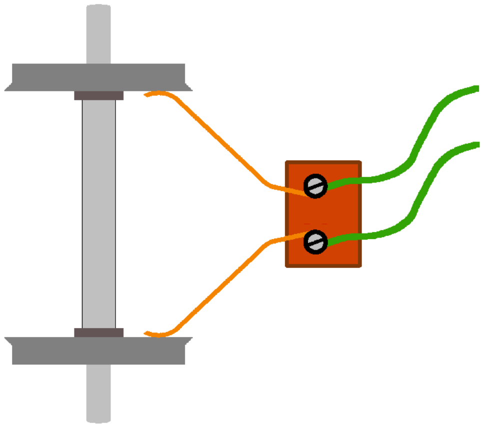
Gaugemaster also sell a four-pole version.

Unfortunately, they are fairly expensive and may not match an existing coupling/uncoupling system on a layout.

Using additional pickups

You don't have to depend on loco power and external connectors for coach lighting, if you fit your own track pick-ups to the coach or wagon you want to illuminate.
If your coach's wheels are plastic, you will have to replace them with metal wheels. These must be mounted on an axle that insulates the metal wheels from each other, to prevent a short across the track.
Often the axle is metal but the wheel at both ends are insulated from the axle. Other wheel sets have only one wheel insulated from the axle.
You can buy these ready made or can convert your own axles.

After replacing the plastic wheels, the coach's wheels are 'live' and we can pick up power from them.
The illustration shows one approach to this but enthusiasts will adopt a method that best suits their situation. In this example, two lengths of phosphor-bronze wire (maybe 28swg) are held in tension against the inner disc of the wheels. This has to be a balance between getting a reliable contact and avoiding excessive friction.

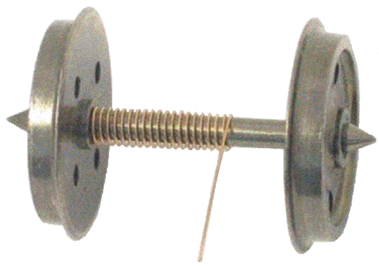


Suitable wire for this can be obtained from suppliers such as:

www.eileensemporium.com
www.mainlytrains.co.uk/acatalog/4mm-loc-pick-ups.html
www.dccsupplies.com/shop/product_info.php?products_id=2023

If you prefer, you can use phosphor bronze strips instead of wire.

Wipers introduce extra rolling friction and this will increase with every coach that is similarly wired.
To minimise friction, another method is to use coiled springs as shown in the image.
This relies on the metal axle being electrically connected to one of the wheels.
These are the DCCconcepts *'Flickerfree pickup springs'* and are available in N and OO/HO from



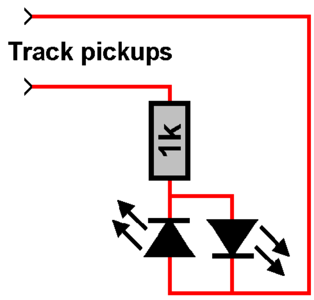
www.dccconcepts.com/index_files/DCCflickerfree.htm

Since there is much more surface contact between the phosphor bronze spring and the metal axle, a more reliable pickup is obtained with much reduced friction. You need less tension as there is always one part of the spring likely to be touching the axle at any one moment.
To fit the spring, you have to remove one of the wheels, slide on the coil and then refit the wheel to exact gauge. You repeat this with the other wheelset, ensuring that it is refitted to the chassis such that both springs wires are connecting to opposite track rails.

The great benefit of fitting coach pickups is that they are hidden, unlike the system of connecting plugs and sockets.
On the other hand, being separate from the loco's power, the lights will illuminate whether or not the coach is attached to a train.
Although there are similarities, there are also some differences between lighting on DC and

DCC systems. Therefore, they are considered separately. In each case, the reference to track pickups includes whichever method you use to get the power from the track to the coach or wagon.

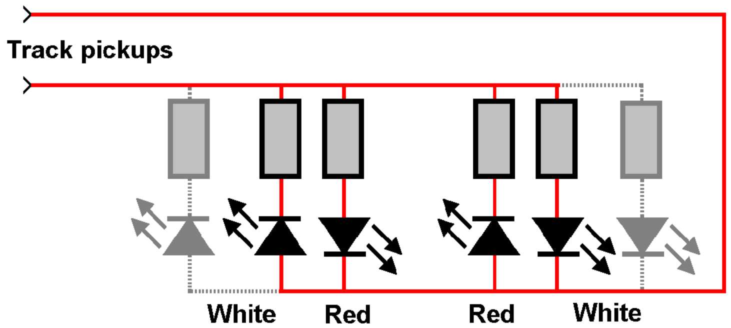
DC Track pickup



This first lighting circuit is very simple, using only three small components. The items cost pennies and will fit into small spaces. It uses two LEDs, wired in inverse parallel with each other. This ensures that one of the LEDs will light as long as the loco is travelling in either direction.
This would be ideal to illuminate a loco's firebox using two red or amber LEDs, or simply to illuminate a cab's interior.
The diagram shows a 1k resistor and this can be altered to suit the LED you are using. Being very simple, it has some limitations:

- It needs around 3V and higher to illuminate
- The LEDs brightness will vary with the loco speed.
- When the train stops, the lights will go out.
- The LEDs will flicker as the power suffers temporary interruptions.

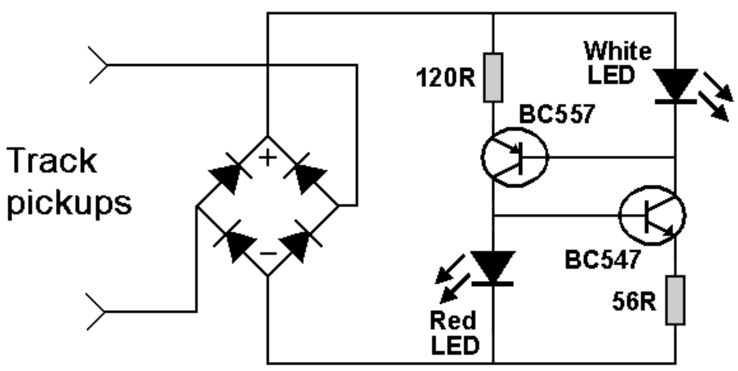
Later circuits show how to improve this.



This circuit takes us a little further forward. It provides directional loco lights for double-ended diesel locos, with either one or two white lights at the end that is moving forward and a red light at the trailing end.
In this case, separate resistors are used with each LED, as the white and red LEDs will have different minimum working voltages.
When the track pickup that connects to all four/six LEDs is negative with respect to the other pickup, the red LED on the left hand side and the white LED(s) on the right hand side illuminate. When the track polarity is reversed, the other two/three LEDs illuminate.

Constant brightness lamps

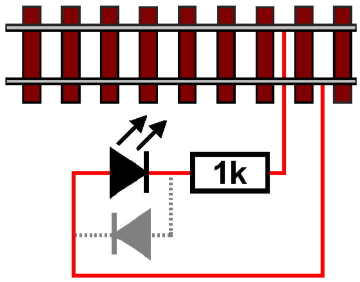
The problem of varying lamp brightness can be partially overcome with this circuit from MERG member Chris Stanforth. It ensures a constant brightness from the two LEDs, as long as the output from the bridge rectifier is between 3V and 15V. It uses a handful of cheap, easily obtained, components and the addition of the bridge rectifier ensures that the LEDs illuminate whether the loco is moving forwards or in reverse.



It also work with DCC power, where the LEDs will be lit – even when the loco is stationary.

DCC track pickup

With DCC, the track is always powered.
This is a potential constant source of power for lighting – both lineside and on rolling stock. For example, why run a long cable just to illuminate a single semaphore lamp, isolated cottage, or buffer stop lamp, when you can just tap into the nearby DCC traction bus?



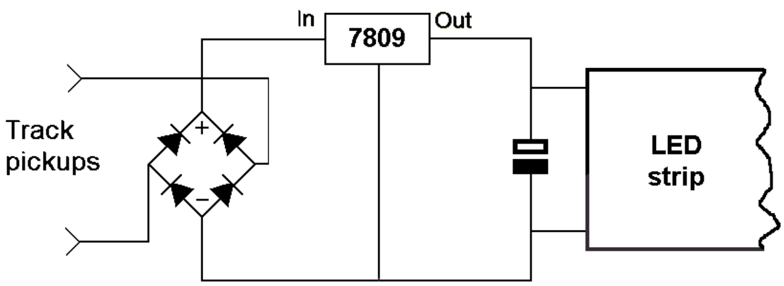
This circuit connects to the track and permanently illuminates a LED. As the chapter on DCC pointed out, the track voltage polarity is constantly reversing.
When the track that connects to the resistor is negative with respect to the other pickup, the LED will illuminate. When the track polarity is reversed, a diode (e.g. a 1N4001) may be fitted in inverse parallel across the LED. This protects the LED from any excessive reverse polarity voltages.
While useful for the occasional remote light, the DCC bus should not be used as the main power supply for all your lineside lighting – use a dedicated power supply.

This approach can be used for lighting locos and rolling stock, by using track pickups instead of direct connections. Since the LED only uses DCC power for half the time (during alternate half cycles), its light is dimmer than when fed by a DC supply.

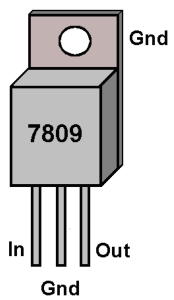
LED strips with DCC

As mentioned earlier, LED strips are purchased as a long continuous strip or roll of LEDs, with an adhesive backing. The strip/roll is available with either 60 or 120 LEDs per metre length. You can choose which LED spacing suits your project best.
The strips are rated as working on 12V. At this voltage, the strip consumes 20mA for every group of three LEDs. In practice, most strips first illuminate at around 8.5V and provide adequate illumination at 9V, taking only 5mA per group of three LEDs.

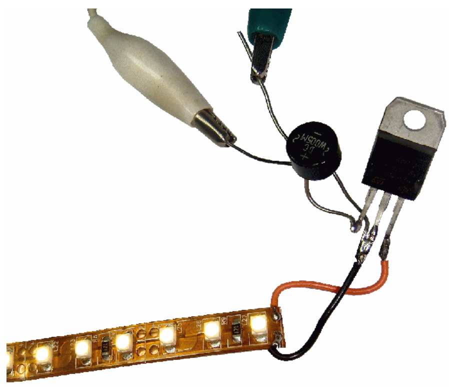
This circuit uses only two or three components and is capable of driving a strip of up to 600 LEDs, although you will more likely be using a strip of 3,6,9, or depending on the size of the coach.



The DCC is fed through a bridge rectifier so that both cycles of the waveform are used. This is fed to a 7809 voltage regulator, which ensures that its output is always at 9V.
To minimise flicker, a capacitor (4700uF 16V) can be wired across the regulator's output. If your coach is large enough, you might consider connecting two of these capacitors in parallel, to provide even more stable illumination.



Here is the pinout of the regulator.
Note that the metal tab with the fixing hole is actually connected to the Ground pin, so be careful the tab does not touch other components.

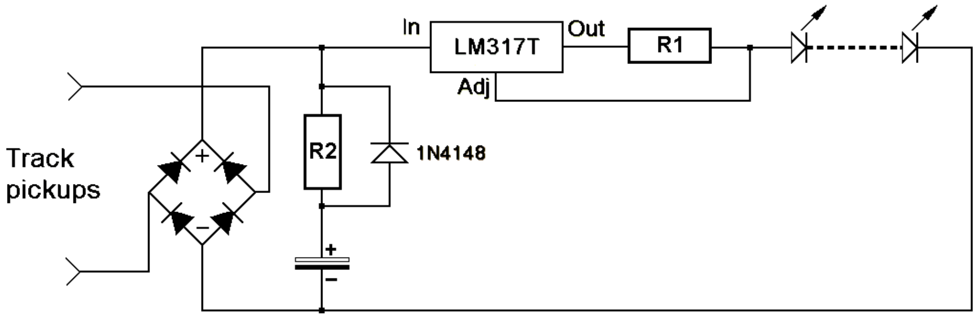


Using an LM317

The LM317 is a three-pin device that can be used as a voltage regulator but can also be used as a *'constant current source'.*
Here is the Wikipedia definition of constant current source:

> *"a circuit element where the current through it is independent of the voltage across it "*

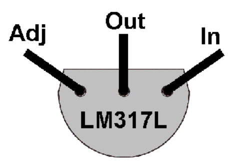
This circuit uses the LM317 as a constant current source to feed a chain of LEDs wired in series.



Since the same current flows through each LED, all the LEDs in the chain should have identical characteristics (don't mix LEDs that work on 5mA with LEDs that need 25mA).
The maximum number of LEDs in the chain depends on three factors:

- The DCC voltage from your command station.
- The voltage drop in the LM317. The output can be expected to be 3V lower than its input voltage.
- The forward voltage drop of the LEDs used. Different coloured LEDs usually have different voltage levels before they illuminate (e.g. yellow LEDs are usually 2V, while white LEDs are usually 3V)

So long as the output voltage from the LM317 is sufficient to illuminate the series of LEDs, there is no need for a series dropper resistor. A 1A 100PIV bridge rectifier is sufficient. The LM317T is a larger tabbed device (like the 7809) while the smaller LM317L is rated at 100mA, which is more than sufficient for our needs.



The value of R1 is calculated as 1.25 / current required; here are some examples:

- for 10mA, the resistance is 1.25/.01 = 125 ohms
- for 20mA, R = 1.25/.02 = 75 ohms
- for 5mA R = 1.25/.005 = 250 ohms

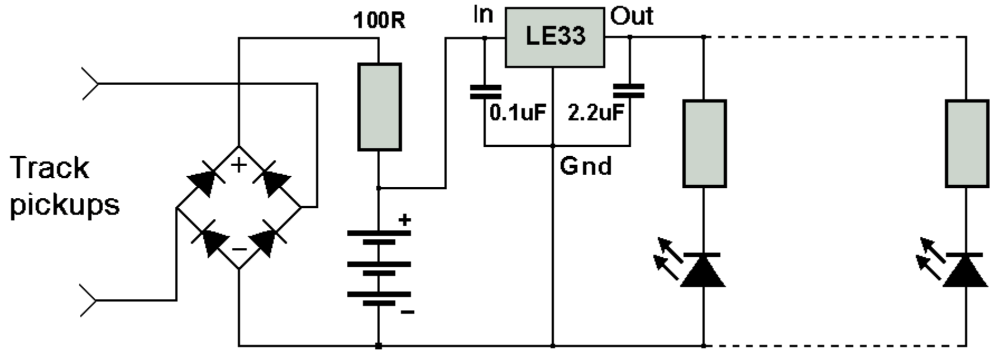
If you prefer, you can always use this calculator:
www.reuk.co.uk/LM317-Current-Calculator.htm
After calculating, you would use the nearest resistor value. In the above examples, the most likely choices are 120 ohms, 68 ohms or 75 ohms, and 240 ohms or 270 ohms.
The current you choose to feed to the LEDs need not be their maximum rated current. For example, many 20mA rated LEDs illuminate quite brightly at 5mA.

To minimise flickering, a large value capacitor (4700uF or better) can be used. R2 is a 100 ohm ¼W resistor. When you first place this circuit across the track, there would be a surge of high current to charge the capacitor and this would trip command station overload detector – it would see it as a short. The resistor ensures that the capacitor is charged more slowly. The diode allows the current to flow unimpeded from the capacitor to the LM317. We charge the capacitor gently but discharge it quickly.

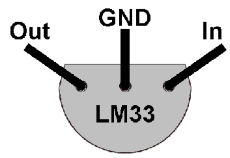
Rechargeable batteries

If you have the space in your coaches to store three rechargeable Nickel-Metal Hydride NimH cells, then you can combine on-board storage with track power. This will provide a light source that is completely flicker-free and will illuminate your coach for hours, even when track power is removed.

As before, the DCC track supply is fed through a bridge rectifier to obtain a DC supply. This is used to keep the three rechargeable batteries fully charged during normal running. The 100 ohm resistor prevents current surges during initial charging and provides trickle charging of the cells. At 1.2V per cell, the set of batteries can then supply 3.6V.



The batteries' positive connection is fed into the input of an LE33 chip. This is a 3-pin voltage regulator that provides a stable 3.3V output. With most voltage regulators a few volts are lost between their inputs and outputs. That would normally mean that an input voltage of 5V or 6V would be required to ensure a stable 3.3V output. However, the LE33 is a *'very low dropout voltage regulator'*. As the description suggests, very little voltage is lost across this regulator. So, three cells (3.6V) are sufficient and saves having to use extra batteries as would be required with a normal voltage regulator.



Here is the pinout of the regulator, viewed from underneath.
It has a maximum 100mA rating which is more than enough for our needs.
The values of the dropper resistors depends on what LEDs you use.
White LEDs operate from 3V upwards and you will probably find that they can be directly connected across the LE33 output without any dropper resistors. Yellow LEDs will need in the region of a 100 ohm resistor, depending on the brightness you require. You can connect as many LEDs across the LE33 output as you require, subject to the 100mA maximum current draw. In practice, this could be 20 or more LEDs running at 3mA to 5mA.

You can use AA or AA rechargeable cells. The AA cells are larger but have a greater storage capacity. This is measured in maH (mA per hour), with AA cells starting from 1700mAh upwards and AAA starting from 600mAh upwards. So, for example, a 1000mAh cell would provide up to 50 hours of light at 20mA.

An alternative is to use the battery packs for cordless phones. They are smaller (e.g. 50x32x11mm) but have less capacity (400mAh upwards).
Even smaller (e.g. 22x16x14) are backup batteries and these, although around 150mAh and 200mAH, should prove more than sufficient for most purposes.

Commercial options

If you don't want to try constructing your own DIY circuits, you can always purchase commercial products.

Some are simple plug-and-play units, where the loco or coach has already been fitted with pickups and connectors. Like the Dapol 156 and Mk3 coach units, described here:

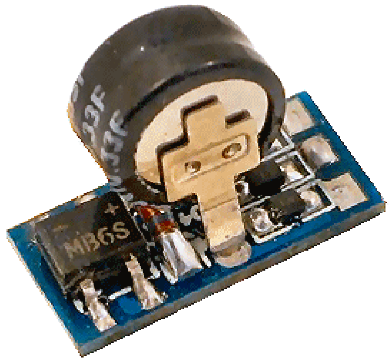
www.ehattons.com/25472/Dapol_Model_Railways_NC040a_Light_Bar_coach_lighting_unit_for_modern_coaches_multiple_units_Class_156_MkIII_/StockDetail.aspx

Some are designed for specific models and the kit supplies the necessary track pickups.

www.kato-unitrack.co.uk/kato-11-201-coach-lighting-kit-1959-0.html
www.topslotsntrains.com/topslotsntrains/final.asp?ref=KATO-11-210-N-GAUGE-INTERIOR-LIGHTING-KIT-WHITE-LED---6-CAR-PACK-&id=7268&manufacturer=KATO%20N%20Gauge'

Some are general-purpose, leaving it you to find a source of track power.

www.ukmodelshops.co.uk/catalogue/firstclass
www.digitrains.co.uk/ecommerce/electronics/leds/interior-car-lighting---incandescent--coach-lighting-kit-.aspx



This image shows the 'Flickerfree' module from DCC Concepts. It provides a set of coiled spring pickups and uses a super capacitor (a very high value storage capacitor) to ensure long-running, flicker-free lighting.
For more details, see:
http://www.dccconcepts.com/index_files/DCCflickerfree.htm

Mounting the lights

The actual steps in fitting lights will depend on your particular loco, coach or wagon. These are just some general suggestions.

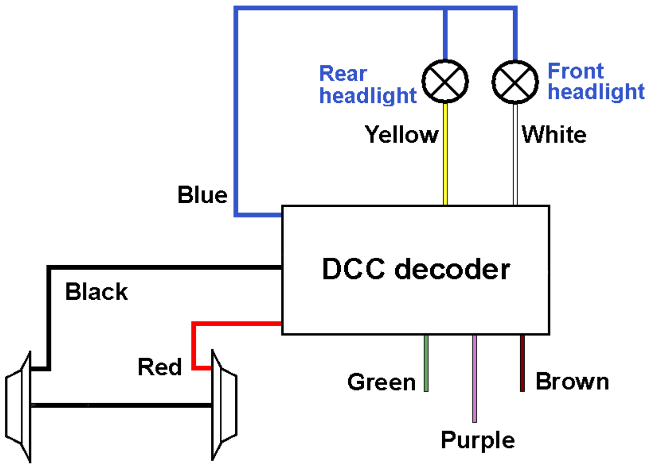
- Don't make any lights too bright, even though you can – it will not look realistic.
- Make sure that light does not escape through thin plastic walls or joints. If necessary, paint the inside of the body, or line the inside with a printed coach interior, such as:
  www.mousa.biz/downloads/coach_seats.html
  www.kitmaster.org.uk/PecoCoachInt.htm
- Aluminium silver foil self adhesive tape can be stuck to the ceiling of coaches, prior to fitting LEDs, to achieve maximum even reflection.
- Sometimes it can be more effective to hide the LEDs on the coach floor and bounce the light off the ceiling.
- Consider using 30AWG Kynar wire to wire the LEDs and connectors, as it is very thin and comes in a range of colours.
- There are some situations where even Kynar wire is awkward to use, for fear of the soldering iron melting plastic or the difficulty in making connections to tiny surface mount LEDs. In these instances, consider using conductive silver paint. It is available from Rapid Electronics, Maplin, Farnell, etc. It comes in a small bottle but a little goes a long way.

Using a DCC decoder

The circuits covered so far are easy and cheap to build but they all suffer from the same deficiency – switchability.
The lights are either only lit when the train is moving – or they are often permanently lit. This is not realistic, since not all locos and coaches are constantly illuminated.
Of course, we could lean over and throw switches on the locos/coaches, or we could introduce automatic light switching using reed switches, Hall Effect switches and the like. Neither of these is particularly satisfactory.

That is where using a DCC decoder offers advantages. We, as controllers, can switch lights on and off using our command station or handheld.
That involves sending specific instructions to the decoders.

The chapter on DCC included an illustration of a DCC decoder. For clarity, it only included the motor, headlight and rear light wiring.
Here it is again, stripped of the motor wiring.
As a minimum, a decoder will have two function outputs, as shown. Often, manufacturers provide additional function outputs.
In this example, the decoder provides three additional function outputs. Each of these can be connected to a light or light strip.



The blue wire is the common return wire for all function outputs.
These additional outputs can be used to provide any effect you require such as a firebox glow, hazard lights, ditch lights, beacons, etc.

The output voltage of a function connection is approximately the DCC track voltage level. There is usually a current limit for each individual function output – often 100mA.
There is also a total current limit that the decoder can handle (e.g. 500mA or 1A). This can be found in your decoder specification, but you would not expect to approach this limit since LEDs consume so little current.

Programming the lights

Your DCC controller will have buttons marked as 'f0', 'f1', etc.
You want to be able to press one of these buttons, knowing that it will switch a particular decoder output. Deciding which button should switch output and then setting it up is known as *'function mapping'*.

There are three ways to do this:

- You can look up the manual for your particular decoder, work out what you want and then use your controller to program the decoder. This can be a fiddly and error-prone process.
- You can us an on-line calculator to work out the values for you. Have a look at:
  www.digitrax.com/support/cv/calculators/

This image shows an extract from the calculator and you tick the boxes to match the function number to the wire to be switched. It then displays the values that you need to enter into the appropriate CVs in your decoder.

Throttle Function Key	Function Mapping CV#	CV Value	White 0F	Yellow 0R	Green 1	Violet 2	Brown 3
F0 (F0F)	CV33	1	✓				
F0 (F0R)	CV34	2		✓			
F1	CV35	4			✓		
F2	CV36	8				✓	
F3	CV37	16					✓

This image shows an alteration being made such that f1 switches the violet (purple) output and f2 switches the green output.
As you can see, the values to be entered for CV35 and CV36 have been altered as a consequence.

Throttle Function Key	Function Mapping CV#	CV Value	White 0F	Yellow 0R	Green 1	Violet 2	Brown 3
F0 (F0F)	CV33	1	✓				
F0 (F0R)	CV34	2		✓			
F1	CV35	8				✓	
F2	CV36	4			✓		
F3	CV37	16					✓

- You can use Decoder Pro. Like the above calculator, it lets you match functions to outputs.

However, it is more intelligent. It only shows the white and yellow options because the decoder being programmed does not have any other function outputs. Even more useful, you don't have to calculate or remember any values; they are all hidden from view.

Use this sheet to determine which functions will control which outputs

Description	Output wire or operation 1 A - White	2 B - Yellow
Forward Headlight F0(F)	✓	
Reverse Headlight F0(R)		✓
Function 1		
Function 2		
Function 3		
Function 4		
Function 5		
Function 6		

In addition, your decoder may have other facilities than can be set, such as directional lights, headlight dimming, flashing, flickering, etc. Read your manual, or see if Decoder Pro offers you those options.

Coach decoders

Using the decoder's function outputs gives you control over the loco's lights – both internal cab lights and externally lamps.
That still leaves coaches and wagons to be made switchable.
Of course, you could extend the wires from the function outputs into the train's coaches, using the plugs and socket previously mentioned.

Alternatively, you can fit track pickups to a coach and install a separate DCC decoder into the coach. You would simply use the function outputs for controlling the lights in the coach, ignoring the loco outputs. The decoder would have its own unique address, separate from the loco and this allows the coach to be used on a loco with no other modifications.
DCC accessory decoders can also be used in coaches. Although larger and more expensive than most DCC motor decoders, they can carry out activities that are not readily suited to loco decoders.

Using accessory decoders

The accessory decoder is really intended for use at the lineside, where it can operate points, signals, etc. Consequently, they are designed for higher currents and are on large PCBs. These can still be used inside coaches, if you have enough room to fit them.
The big advantage of using these decoders is the larger range of outputs and/or features.

The Tam Valley LEDecoder only has three outputs but can handle a bi-colour LED and flickering effects.
For more details, see:
www.tamvalleydepot.com/products/lighting.html

The ESU LokPilot Fx v4.0 has six function outputs with a maximum current of 250mA per output.
For more details, see:
http://www.esu.eu/en/products/lokpilot/lokpilot-fx-v40/

If you fancy building your own, have a look at:
http://atoomnet.net/circuits/tiny-dcc-decoder-for-leds/

The MERG Kit 53 is a DCC accessory decoder with 8 steady state outputs

Special features

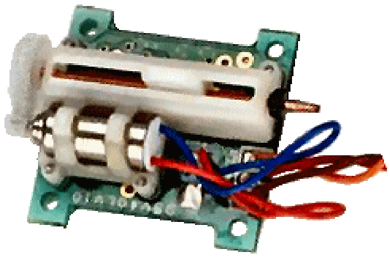
Why stop at just lights? Accessory decoders can be used to operate any animation in your coach or wagon that you can think of – space permitting.
You can now consider modelling:

- A maintenance or track-laying crane.
- A guard waving his lantern (another LED?) from the guard's van.
- On-board uncouplers.
- A moving pantograph or trolley pole.
- Sliding carriage doors.
- A remote-operated smoke or steam unit.
- And so on and so on.

We know from a previous chapter that we can animate things using servos – and some accessory decoders are able to also control servos.
Check out:
www.dccsupplies.com/shop/product_info.php?products_id=2063
controls one servo and one relay
www.dccsupplies.com/shop/product_info.php?products_id=2058
controls four servos and four stall motors
www.digitrains.co.uk/ecommerce/accessory-decoders/multi-function/mx82v-turnout-accessory-decoder.aspx
The Zimo MX82V controls four servos and four LEDs
If you are working in larger scales, the commonly used servos such as the Tower Pro SG90 could be installed.

For smaller scales, the tiny Spektrum linear servos can be used. These come in standard throw, as in this image, and in long throw (just under 1").
For more details, see:

http://www.4-max.co.uk/servo-015a.htm
http://www.indoorflyer.co.uk/spektrum-15g-linear-servo----no-packaging-572-p.asp
http://www.rcpitstop.co.uk/spektrum-29gm-linear-tail-servo-spmsh2040t---in-stock-3141-p.asp



Another feature that can be controlled via DCC are the smoke units and steam units marketed by Seuthe. These are available from:

www.gaugemaster.com/seuthe.html
They have a handy chart detailing all models and their current requirements
www.ukmodelshops.co.uk/catalogue/firstclass
Includes great advice from Harvey Godber on installing and using these units.

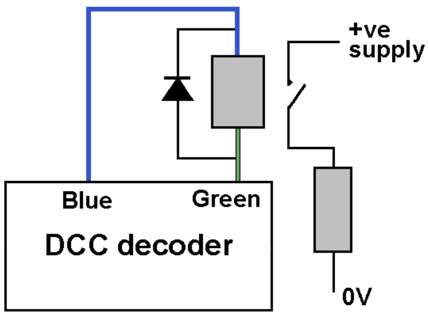
There is also a lot of information on smoke units at:

www.trainweb.org/girr/tips/tips3/smoke_tips.html

Power for additional devices

The voltage level from decoder function outputs are just a little less than the DCC voltage and this is sufficient for anything we want to switch. In fact, we will have to reduce the voltage for connecting servos and LEDs.
The bigger issue is the current limit from any one function output – often a 100mA maximum. Since the Seuthe Smoke Unit uses 149mA at between 11V and 16V, and the Seuthe Steam Generator needs 130mA at between 10V and16V, we cannot drive these directly from a decoder output.
Instead, we use the decoder output to operate a transistor switch and/or a relay, which in turn supplies power to the device (see the chapter on interfacing).

This circuit gives an idea of how to interface a Seuthe smoke unit. A 12V low current relay is connected between a decoder output and the blue wire which is always positive Suitable relays are widely available, such as those from Rapid Electronics (part number 60-2410 or 60-4662), or the BT 47W/6H relay from MERG.
Don't forget to fit a protection diode, such as a 1N4148 across the relay coil.



The decoder has its own circuitry that converts the incoming DCC into a DC supply. This must **not** be used to power any high-current devices.
Use the bridge rectifier, with the voltage regulator if necessary, described earlier in this chapter.

Chapter 10

Adding sound

Back to Contents Page 5

Adding sound

Many layouts are miniature masterpieces, yet still seem to lack something. That something is often the extra authenticity that comes from movement (other than trains), lighting and sound.
Mostly, a model railway is a visual experience, watching trains travelling through scenic layouts. We can add to the effect by animating lineside and scenic features using motors and servos. We can further improve the layout further with the careful use of scenic and train lighting.

This chapter looks at improving a layout by stimulating another of your senses – that of hearing. Noise is all around us, yet most layouts produce only the sound of electric motors and points operating. This can be corrected, as everything we hear in real life can be reproduced on your layout. This is a much wider area than may initially be thought and this chapter is really an introduction to the subject.

Layout sounds can be divided into two broad categories.

Train-related sounds
These are associated with the direct running of the railway and include:

- Hissing and chuffing
- Diesel engines
- Horns and whistles
- Wheels squealing round tight corners
- Station announcements

Some of these sounds are more easily created using DCC sound decoders, while others can be easily created for both DC and DCC layouts.

Scenic-related sounds
These sounds add to the general 'feel' of the layout and might include:

- Factories and mills
- Quarries and mines
- Traffic
- Water
- Farms
- Weather

All of these can be created for both DC and DCC layouts.

Sound Types

- General sound categories include:
- Man-made sounds (e.g. guns, machinery)
- Natural sounds (e.g. water, thunder, rain)
- Ambient/Environmental/Atmospheric sounds
  This involves the combination of various sounds to create the overall effect.
  For example, a rural scene might combine the sound of a brook with bird and/or animal sounds, while an urban scene might mix traffic noise with general population hubbub. Likewise, a fairground needs not only organ sounds but the sounds of children and laughter while a factory complex will mix a variety of mechanical sounds.
- Incidental (i.e. one-off) sounds, such as:
  Gunshots, foghorns, dogs barking, crossing bells, station announcements

Sound in perspective

Before we start, we must bear in mind the importance of 'getting it right'. The indiscriminate addition of sounds can easily produce a horrendous cacophony which seriously detracts from the experience.
Just because you can reproduce the sound of everything on a layout does not mean that they all have to be played, at the same volume, at the same time.
The coherence principle rests on the maxim that *'less is more'* – i.e. adding too many unnecessary sounds impedes the user's appreciation of the main content. This situation can be avoided by a few simple rules.
Layout sounds should be**:**

Appropriate

The sounds should add to the layout, not detract from it. A layout may offer a range of possible sources of sound. However, it is likely that balancing some of the most predominant sources would suffice. Even then, common sense should prevail. For example, it is unlikely that you will hear factory sounds on a day that church bells ring.
In the words of Joe Meek (1960's record producer)
"If it sounds right, it IS right!"

Atmospheric

The sounds chosen must reflect the nature of the layout. While layouts may have some elements in common (e.g. vehicles, people, stations, etc.), the dominant sounds must be those most sympathetic to the layout's character. Rural, urban and industrial layouts all have different characters, as do different periods in time. Better to choose the sounds that most characterise the layout than trying to include every possible sound. The sounds chosen should be both harmonious and proportionate.

Scaled

Your layout is a miniature of a much larger area. The baseboard may only be a few feet deep but that will represent quite a scaled distance. The addition of low relief building and detailed backdrops add another layer of perceived depth. The smaller the scale you model in, the greater the scale distance between the front of the baseboard and the rear details.
It follows, therefore, that sounds from the rear of the layout should be quieter than those closer to the front. Otherwise, the overall perception is jarred.
The sounds should provide the impression of depth to a layout. This supposes different volume levels for different distances (see later).
Overall though, we are discussing *background* sounds, so a subdued sound level should apply even to sounds from the front of the layout.

Switchable

Above all, layout sound should be under the control of the operator, who can decide what is played and when. Lots of noise might be perfect for the visits from children and grandchildren but be most unwelcome at exhibitions or during late at night/early morning sessions.
Firstly, then, is the ability to disable all sound at the flick of a switch or the click of a mouse. You may also wish to make individual sounds switchable (e.g. disabling diesel sounds and enabling steam sounds).
After that, there are decisions regarding timing of sounds.
For example, running water is expected to make a continuous sound, while incidental

sounds such as bird or animal sounds would only occur at irregular intervals.
Other sound effects, such as whistles and wheel squeals, would only be triggered by events such as a loco entering a specific section of track.

This chapter looks at the following areas:

- Deciding which sounds are appropriate and which reproduction system to adopt.
- Finding and using layout audio clips.
- Sound players.
- Speakers and their placement.
- Triggering of sounds.

Sound options

There are different methods available for adding sound to your layout.
There two different categories;

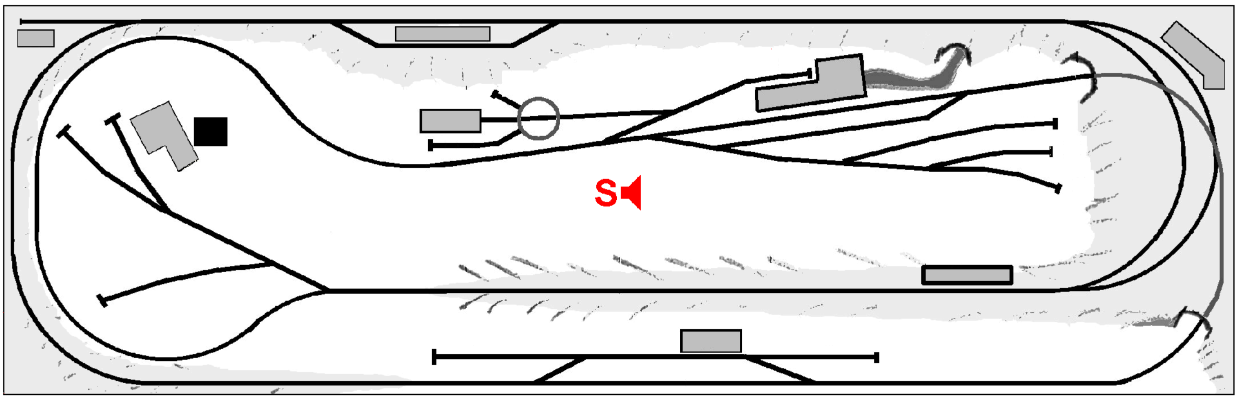
- Trackside sound, on the baseboard
- Sound modules in locos

Trackside layout sound

A number of different approaches have been adopted, both for commercial and homebrew sound systems. The main approaches are shown below, on a typical layout.

Centralised

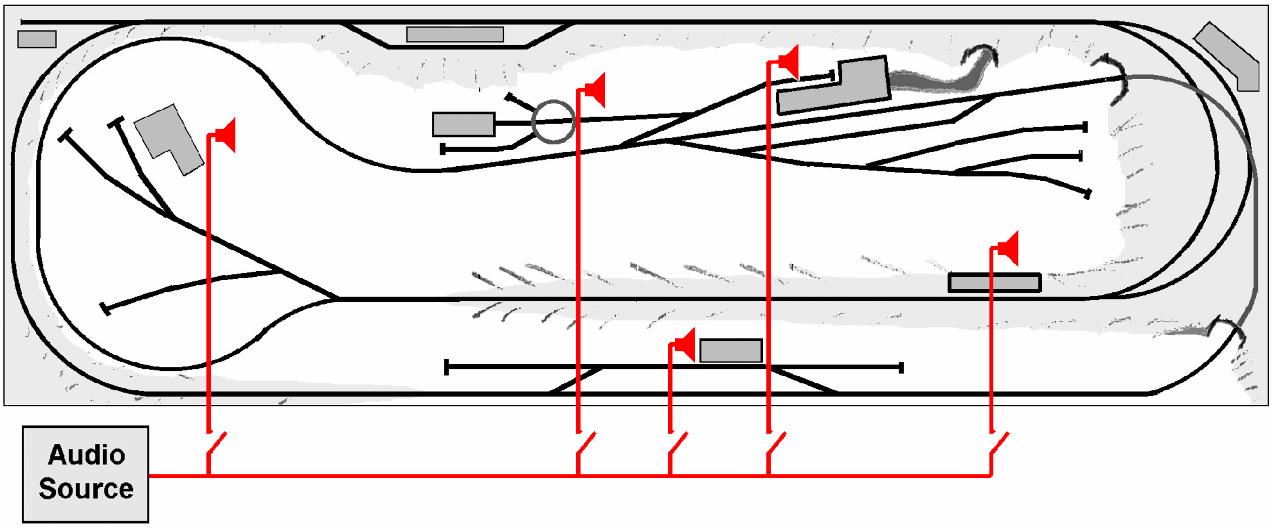
As the name suggests, the layout uses a single audio player and a single loudspeaker placed in the middle of the layout. This has the benefit of simplicity and cheapness (although some expensive commercial systems also use this method).



The downside to this method is that it does not deceive the listener, whose ears detect sounds in stereo. With this system, all sounds are detected as being in the middle, even although the church, factory, etc., may be situated at the extremities of the layout. While this method might work for small layouts, the difference between expectations and reality worsens as the layout becomes larger. It is not very effective in large layouts as the sound does not appear to come from specific objects or buildings.

Distributed

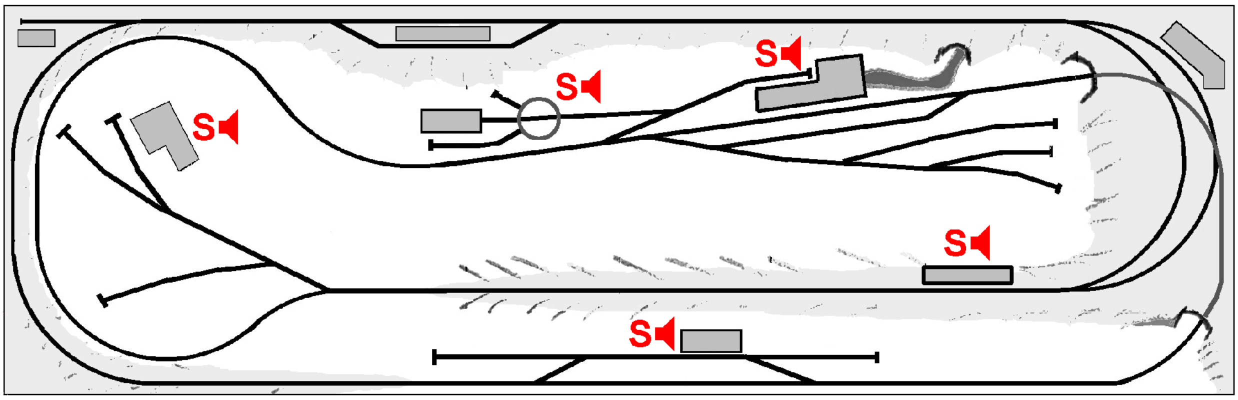
With this system, there is still a single source of all sounds but there are individual loudspeakers situated at each sound location and these are switched on and off to make the sound come from specific areas of the layout. In this way, the sounds are heard from their correct positions on the layout.



The downside in this method is the need to direct the audio from the player to the correct loudspeaker at the correct time, using either manual or electronic switching.
Also, only a single sound can be produced at any one time. This reduces the richness of the overall effect. It also prevents continuous sounds (e.g. rivers) from being played.

Local

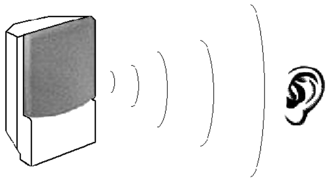
This system uses a separate sound player and loudspeaker for each location. This allows sounds to be heard from their expected locations and also allows multiple sounds to be played at the same time if necessary. So, a waterfall will have its own sound module, with another module producing sound for the sawmill, and so on. The speakers for these modules are located in, or next to, the supposed source of the original sound.



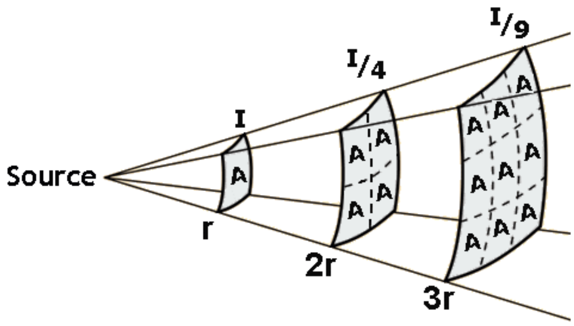
The modules can be left to play continuously, for background sounds such as with a waterfall or station hubbub. Alternatively, they can be wired to only play when triggered by train-on-track detectors. For example, the sound of a loco whistle/horn or wheel squeals could be produced when a train enters a particular section of track.
They could also be under operator control, switched on/off by MERG CBUS output modules (e.g. CANACC8) or controlled by switches on a control panel.
It can also save on all the extras switch wiring needed with distributed systems.
The downside is the extra expense of providing more players, although this not need be prohibitive.

Sound basics

Before moving on, it may be useful to recap on some basic sound concepts.
Sound is produced and carried by the rapid variation of the pressure of the air surrounding the object that is creating the sound.
Consider the loudspeaker inside your TV or HiFi. The speaker cone vibrates in sympathy with the audio content. As the cone vibrates, the air next to it is alternately compressed and rarefied. This constant change of air pressure sets up sound pressure waves, which travel through the air until they reach the listener's eardrum.

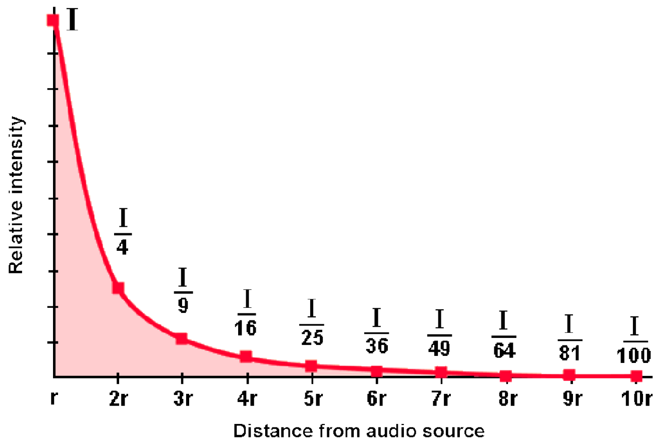


The air does not actually travel between the speaker and the ear. As the air next to the speaker cone moves back and forth, it passes on this movement to the air next to it - and so on. It is similar to the water in a pool when a stone is dropped in it. The waves move outwards but the water does not travel outwards. The knock-on effects of the variations diminish as the sound waves travel outwards, resulting in sound becoming quieter the further away the listener is from the source.



As this illustration shows, the audio source has a sound intensity I at distance r from the sound source. If we look at twice the distance (i.e. 2r), the area at r2 has to influence four times the area as that at distance r. Similarly, at distance 3r, it has to influence an area that is nine times larger. This effect is known as the “Inverse Square Law”.

This chart shows just how quickly sounds diminish with distance.



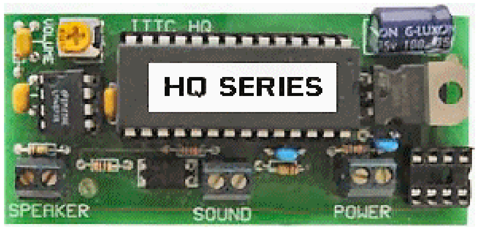
In model railways, this is sometimes considered as 'scale sound'. Although a layout may be physically small, it represents a much larger area (depending on the scale – e.g. 76 times greater in OO and 220 times greater in Z).
Clearly, then, sounds at the rear of a layout should not be heard at the same amplitude as those at the front of the layout. Also, if you are standing at one end of a long layout, you should hardly hear sounds from the other end of the layout.

In addition, we have to remember that the dominant impression of a layout should be the movement of the trains, not the sound from various features. Even the loudest sound should not detract from the layout.
In the words of Jim 'Enginears' Wells (of Fantasonics)

"your audience should feel like they want to lean in to hear the sound better"

Commercial sound systems

There are quite a few commercial systems that are specifically designed to produce sounds for model railway layouts. They are generally well specified but expensive.



This image shows a module from the HQ series of modules produced by ITTC (available from Bromsgrove Models). They are designed for fitting in local situations, with each module producing a different sound. The sounds vary from US loco sounds to lineside sounds such as a construction site, a sawmill, thunder, church bells, and much more.
The module is connected to power, a speaker and a connection for external switching.

This image shows the *'Master Blaster'* from Model Sounds. It is a combined sound module and loudspeaker in the one case.
It has a built-in battery, making it fully portable – it could be installed in a lineside building or fitted inside a coach or wagon (probably from 'O' gauge upwards).
It runs for three hours on a charge and can be recharged from a computer's USB port or from a battery charger.
It uses a MicroSD memory card to store the sounds and the different cards can be swapped in to get different sounds (purchased separately).
For details see:

http://www.modelsounds.co.uk/

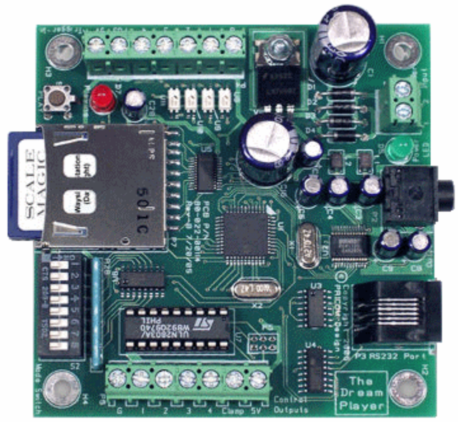
An example of a centralised sound module is the "Steam Whistle and Two Tone Horn" model EXH/TTW from Express Models.

http://www.expressmodels.co.uk/acatalog/Sound_Modules.html

It stores the sound of a whistle and a two tone horn and the sounds are played by pressing a push button (although this could be replaced by relay contacts, a reed relay, etc.).

This image shows another centralised sound system – the Model Rectifier Corporation Symphony 77.
The system is powered by an AC adapter and comprises two loudspeakers and a handheld controller
The speakers do not provide stereo sound. Instead, one speaker one set of sounds (e.g. steam chuff/diesel rumble, brakes, crossing gate, etc.) and the other speakers produces a different set of sounds (e.g. diesel horn/steam whistle, air release, uncoupling, etc.).
In all, over 25 different sounds can be produced by the operator choosing options on the controller.
Despite the variety of sounds, the sounds cannot be located next to all the objects that are supposed to be making the sounds.
Also, there is no provision for automatically triggering a sound.
It is available from US suppliers, Amazon or eBay.

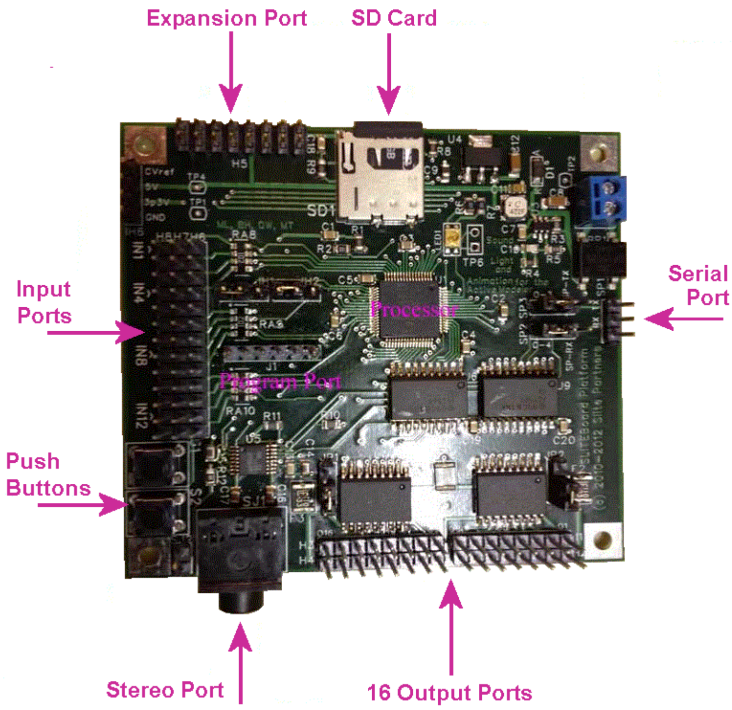
This is an example of a centralised sound module with additional facilities. It is the Pricom Dream Player.
At the top of the board is a row of inputs and there is a row of outputs along the bottom.
This allows a very flexible set of options. For example, an input can be triggered that plays a sound – and also changes the state of one of the outputs. For example, a change on one input could play a warning alarm while crossing gates close.
The sounds are stored in SD memory cards (see left of image) and therefore are changeable.



Although very flexible, its big downside that it has only one audio output (see right of image); this makes it less suitable to cover a large area of a layout.
For details see:

http://www.pricom.com/Trains/DreamPlayer.shtml
http://www.dccsupplies.com/shop/product_info.php?products_id=2088

An even more versatile, and more expensive, sound module is the RLW 4400 SLITE Animation Board from Republic Locomotive Works.
It has two audio outputs that can be used independently or used in stereo. The sounds are stored in a micro SD memory card.
It also has 12 input channels and 16 output channels and can be programmed by a computer through the serial port to carry out sequences of operations.
The program sequence is stored in the SD card.
The board has its own microprocessor that can run through a pre-programmed set of sounds and output switching.



The module provides facilities such as PWM (pulse width modulation) for dimming LEDs or setting the speed of motors, A/D convertors to read analogue changes such as light levels, and more. It is not currently marketed in the UK.

The top-of-the-price-range trackside sound system is the Surroundtrax DSP-80.
It is a distributed sound system, as it places six speakers around the layout.
Each channel can produce sound independently of the others.
Although it is a trackside system, it aims to provide the sounds of locos as they move round the layout.

It requires additional track block sensors, one for each speaker. As a loco enters a track block, the main module is informed and sound is directed to that speaker. Cleverly, as the

loco moves from one block to another, the sound in their allocated speakers cross-fade (the volume declines in one speaker while increasing in the other). This provides a smooth stereo-type sound transition.
Unfortunately, six speakers may not be sufficient for larger layouts.
It uses DCC technology, but without fitting sound decoders in each loco. This may be an advantage for those with large numbers of locos, or limited skills in fitting decoders.
On the other hand, it does not provide the large variety of options that are available in loco sound decoders. Also, it is limited to loco sounds only, with no lineside or other effects.
It is a very expensive system and further details are at:

http://www.soundtraxx.com/surround/

In the UK, it is sold by DCC Supplies

http://www.dccsupplies.com/shop/product_info.php?cPath=24_157_676&products_id=4780

Limitations of commercial options

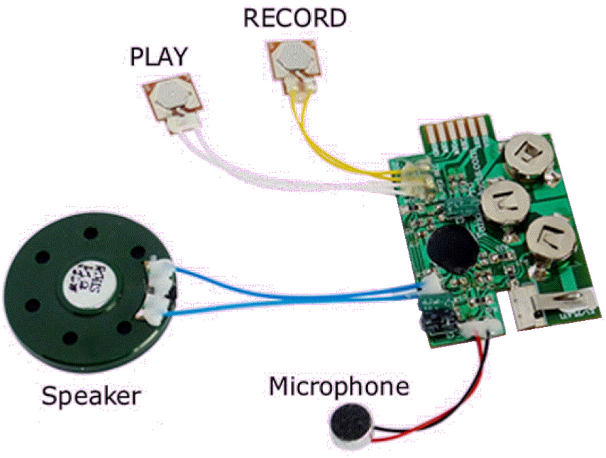
Most commercial modules only play built-in sounds with no provision for the user to alter or add their own. The Trax DHM-12 and the Express Models EXX/TTW each have two preset sounds, while the Quasar SG01M module provides 4 preset sounds. The HQ series modules produce only a single preset sound.
The Master Blaster module uses memory cards, each storing a different sound. These can be purchased separately, although only one can be installed and used at a time.
Others, such as the Symphony 77, provide more sounds but have to be manually operated; there is no provision for automatic triggering (e.g. from track detectors, relays, CBUS, etc.). The Symphony 77 has 25 unalterable sounds.
The top of the range SoundTrax DS-P80 only works with Digitrax block detectors and transponder equipped locos.

DIY sound systems

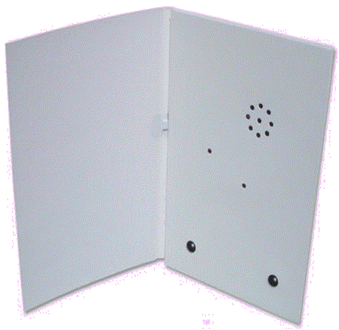
Despite their considerable prices, commercial modules have practical limitations.
However, there are a number of devices that are easily obtainable, are considerably cheaper and can provide more flexible and user-alterable sound players.
Here are five different devices, of varying complexities, all requiring little or no alteration.

Option 1 - Talking cards

Let's begin with the simplest and cheapest sound player that can be used straight out of the box. Although not of the greatest audio quality it is available for around £2 from eBay.
It is intended for incorporating into greetings cards, so that personal recorded messages can be included (*"Merry Xmas Gran, from little Timmy in Australia"*)
It is complete and ready to use, as it has its own built in batteries, microphone, speaker and operating pushbuttons.



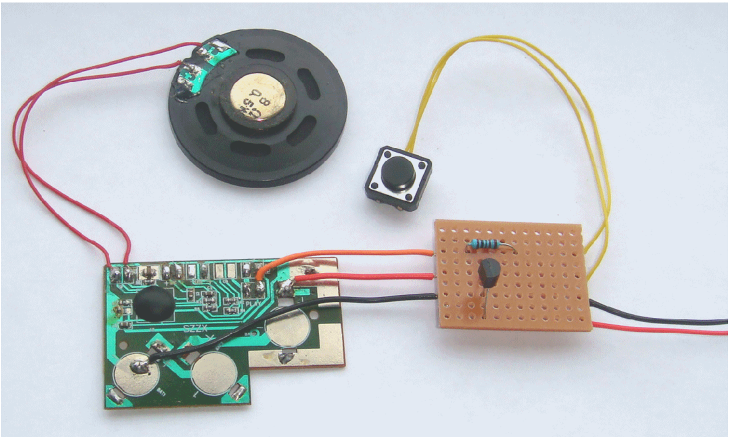
It may be supplied bare as shown in this photograph or be enclosed in a plastic case.
The boards are available with slightly differing layouts e.g. some have two batteries and some have three.



It is extremely simple to use. Just press the record button and hold the mike close to the source of the audio. This may be a sound clip being played through a computer's loudspeaker or be the spoken word (e.g. for a station announcement). The recording limit is 10 seconds.
To play back the recording, simply hold down the play button.
Although cheap and simple, this module can be surprisingly effective.

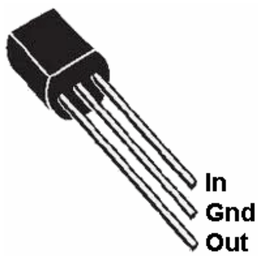
Possible modifications

- A likely modification is to remove the on board batteries and replace them with a 78L05 5V regulator so that the module can run from a standard 12V supply.



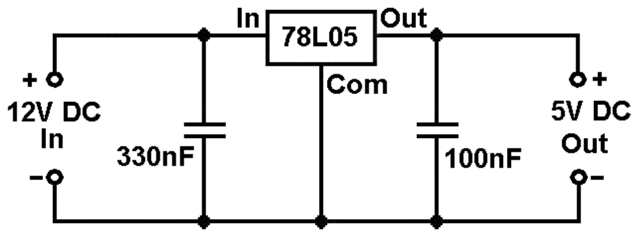
- I also unsolder the record button after making the recording, to prevent it being accidentally pressed later. It can always be re-connected later if you want to alter the recording.
- It is likely that you will want to operate the module from a distance and this would normally require two wires for the power and two wires for the play switch. This can be reduced to just three connecting wires, as shown in the image. In my module, one of the two 'play' connections has to be taken to +5V to play the sound (the orange wire in the image). This wire can be taken to any device that can provide a switchable +5V (e.g. a switch, relay contacts, output from atrain-on-track detector or a MERG CBUS CANACC8). I included a 1K resistor in series with the orange wire to limit the switching current.
- If you have an old loudspeaker lying around, you will probably find that it is a good replacement for the tiny speaker that comes with the card.
- It should also be possible to remove the microphone and record directly from the computer's 'Audio Out' socket. This is likely to improve the audio quality but I did not pursue the circuit for this.

Wiring a regulator



All five modules operate from either 3.3V or 5V supplies, so these notes on converting to use with a 12V supply apply to all options. Since they all consume relatively little current, they can use the 100mA style regulators (the ones that look like transistors) instead of the larger regulators (the ones with metal tabs). They have three connecting leads, as shown in the illustration. The 12V is wired to the regulator's 'In' pin and the regulated voltage of 3.3V or 5V is available from the 'Out' Pin. The third pin is shown as 'Gnd', also referred to as the COM (common) pin. This is wired to the 0V line of the incoming voltage and also acts as the 0V of the regulated output.The 78L33 is used for a +3.3V output and a 78L05 is used for a +5V output. Both have the same size and shape and identified by the printing on them.

Capacitors with values of 330nF and 100nf are fitted across the input and output pins, as shown in the diagram. They are used to smooth the incoming supply and to prevent any possible internal oscillation.
These values are non-critical and some circuits omit them altogether. However, given their small cost, they are best fitted.



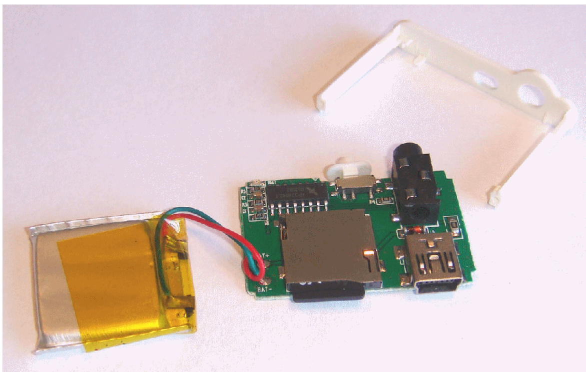
Option 2 – MP3 player

Many cheap MP3 players are flooding the market and can be purchased on eBay for around £2. Although designed for playing music on the move, they make great little sound players for a layout.
They store the sounds on a microSD memory card that is inserted into a slot in the player. These cards can be bought separately for around £2 and can store long sound samples. The sounds can be of higher quality and this player can handle stereo sounds,
The player has a built-in battery that can be recharged from any USB socket (often the one on your computer). The same USB lead is used to copy MP3 sound files from the computer into the player's memory,
The player is intended to be used with headphones so an external audio amplifier needs to be connected to the player's audio out socket.
The steps for this module are:

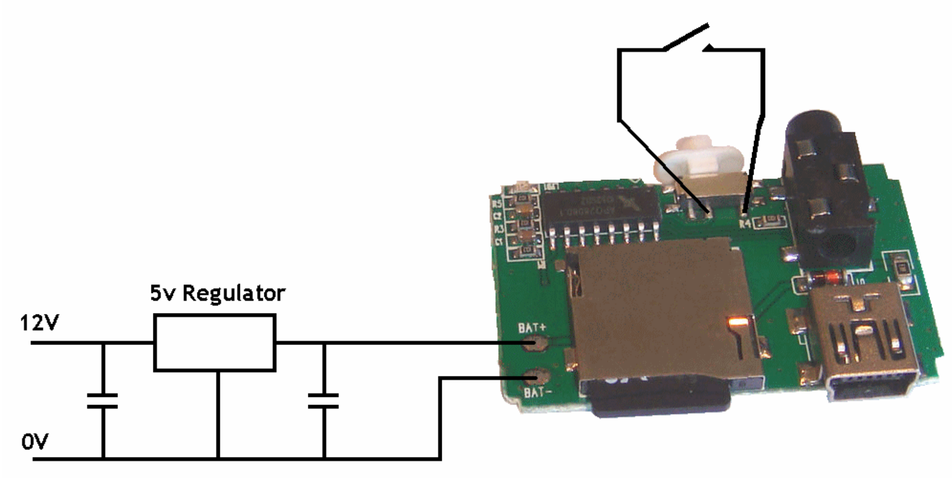
- Find, or create, a suitable sound file in MP3 format.
- Insert the microSD card into the player.
- Connect the player's output to an amplifier.
- Connect the supplied USB cable between the player and the computer's USB socket.
- Copy the sound file into the memory card.
- Remove the cable.
- Operate the power switch to hear the sound being played.

If you are happy to use the internal battery, then the player can be used as is, without any modifications.

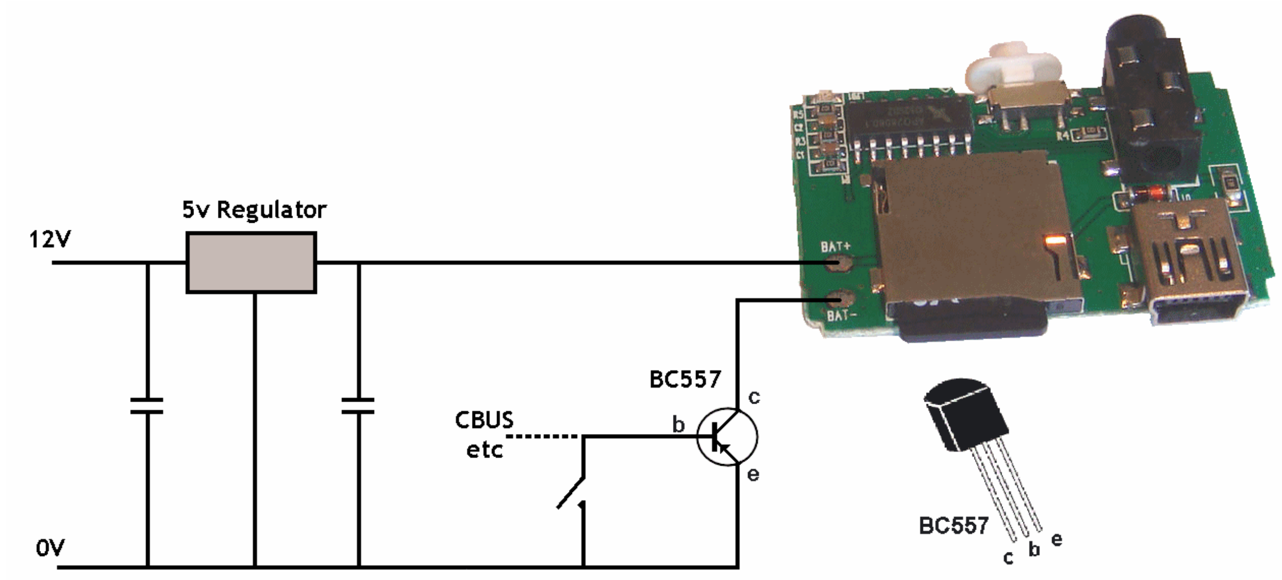
Possible modifications



- Having to regularly recharge the module's internal battery can be a nuisance. Fortunately, it can be easily replaced with a +5V regulator, allowing the player to be powered from a standard 12V supply. Two small screws secure the player's outer casing. Then, the internal battery can be unsoldered. A 5V regulator circuit can then be wired to the player's power connections as shown below.
- You can wire your own switch in place of the one fitted to the player, to allow you to operate the player from a distance (see the image below).

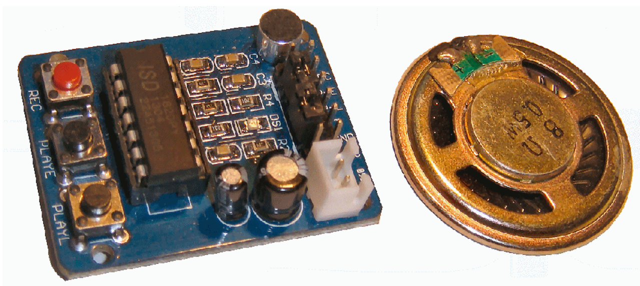


- You can also keep the switch in the ON position and switch the power to the player. The circuit below shows a transistor, type BC557, being used to act like a switch. When its base connection is taken to 0V, current is allowed to flow between the emitter and collector connections and the MP3 player is powered. The transistor's base can be taken to 0V with a switch, as shown in the diagram. Alternatively, the switch can be replaced with any other device whose output can switch between +5V



and 0V (e.g. relay contacts, the output from a CANACC8, the output from a track occupancy detector, etc.).

Option 3 - The ISD1820



This device is readily available on eBay for just over £3. It has a built-in microphone, built-in amplifier and is supplied along with a loudspeaker. The basic version is just 33mm x 42mm.
It needs a power supply of between 3V and 5V, so the 78L05 5V regulator circuit described earlier would allow it to be run from a standard 12V supply.
The module has three buttons. Holding down the red button allows a recording to be made using the microphone. The recording limit is 10 seconds.

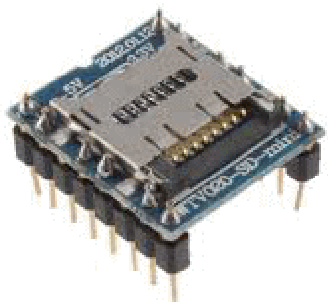
The module also has two black buttons, marked as *'single-pass'* and *'jog',* or *'PLAYE'* and *'PLAYL'.* PLAYE stands for edge-activated and even the shortest press of the 'single-pass' button results in the entire recording being played back. PLAYL stands for level-activated and the 'jog' button only plays back the recording for as long as the button is held down.

Possible modifications

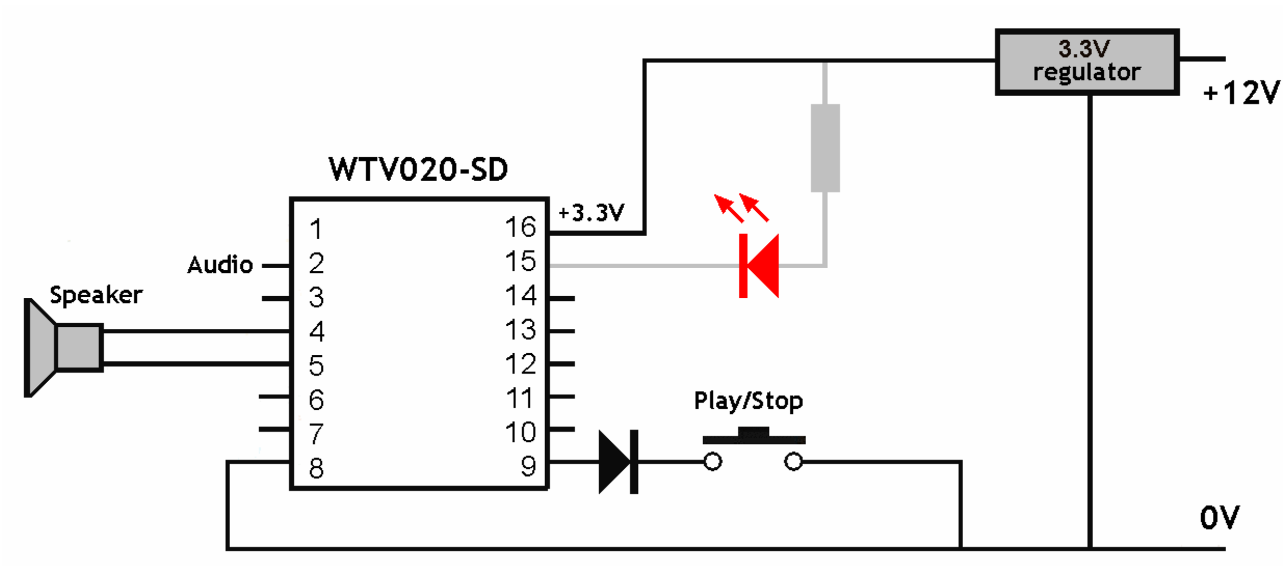
- The black playback buttons are handy for testing the module but it is likely that you will want to operate the module from some external triggering device. Fortunately, there is a row of pins on the board that replicate the job of the buttons. So, for example, taking the REC pin high is the same as pressing the REC button. Raising the PLAYE pin to 5v is the same as pressing the 'single pass' or PLAYE button. Similarly, raising the PLAYL pin to 5v is the same as pressing the 'jog' or PLAYL button. So the PLAYE or PLAYL pins can be wired to switches, relay contacts, the output from a CANACC8, the output from a track occupancy detector, etc.
- The microphone that is on board the module (shown in the top right corner of the module in the photograph) produces reasonable results for the price but there is scope for replacing the mike with a better quality model or with a connection to the audio output from a computer.

Option 4 - The WTV020-SD

This is another sound player that plays sounds stored on a microSD memory card. It supports up to a 1GB max microSD card so can play long sound files.
It is readily available on eBay for just over £3.
It has a built-in amplifier but needs a loudspeaker. Alternatively, it can be connected to an external amplifier.



It is about the size of a postage stamp and requires few other components, as can be seen from this diagram.



It is a 16 pin module, although most pins are not used in our circuit..
The LED and its resistor are optional. The LED flashes while a sound is playing and would not normally be needed except perhaps for test purposes. The module would normally be mounted under the baseboard or inside a building, so the LED would not be visible.

The module works from a power supply of between 2.5V and 3.6V, so a 78L33 3.3V regulator circuit described earlier would allow it to be run from a standard 12V supply.

Preparing the audio files

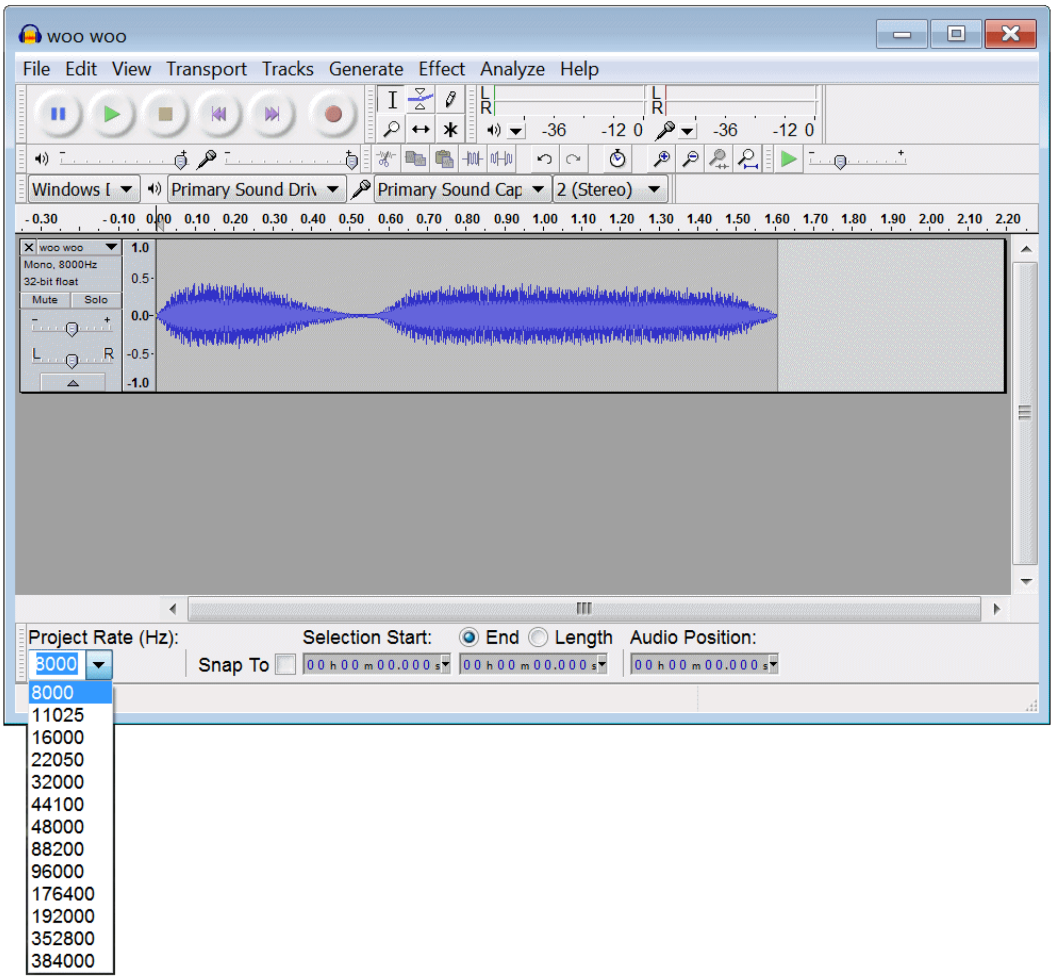
This and the next module require the preparation of the audio clips prior to them being installed in the module.
With the WTV020, the audio clips have to be stored in the memory card in ad4 format (4-bit adaptive PCM), with a sampling rate of 32KHz. Files should be saved as 0000.ad4, 0001.ad4, etc.
This is easier than it sounds, as there is free software that converts audio files into the required format. The process uses two software packages that are free and easily available.
After obtaining, or creating, your required audio file, carry out these two steps.

Step 1 - Convert the file to a sampling rate of 32khz

The most poplar software for this is the free open source application called 'Audacity', which is downloadable from

www.audacity.sourceforge.net

When you use the program to open an audio file, the audio track appears in the main window. The current sampling rate of the file is displayed in bottom-right corner. If it needs to be altered (i.e. it is not currently 32000) then 32000 can be selected from the drop down box. The example shows that the file has an 8KHz sampling rate.

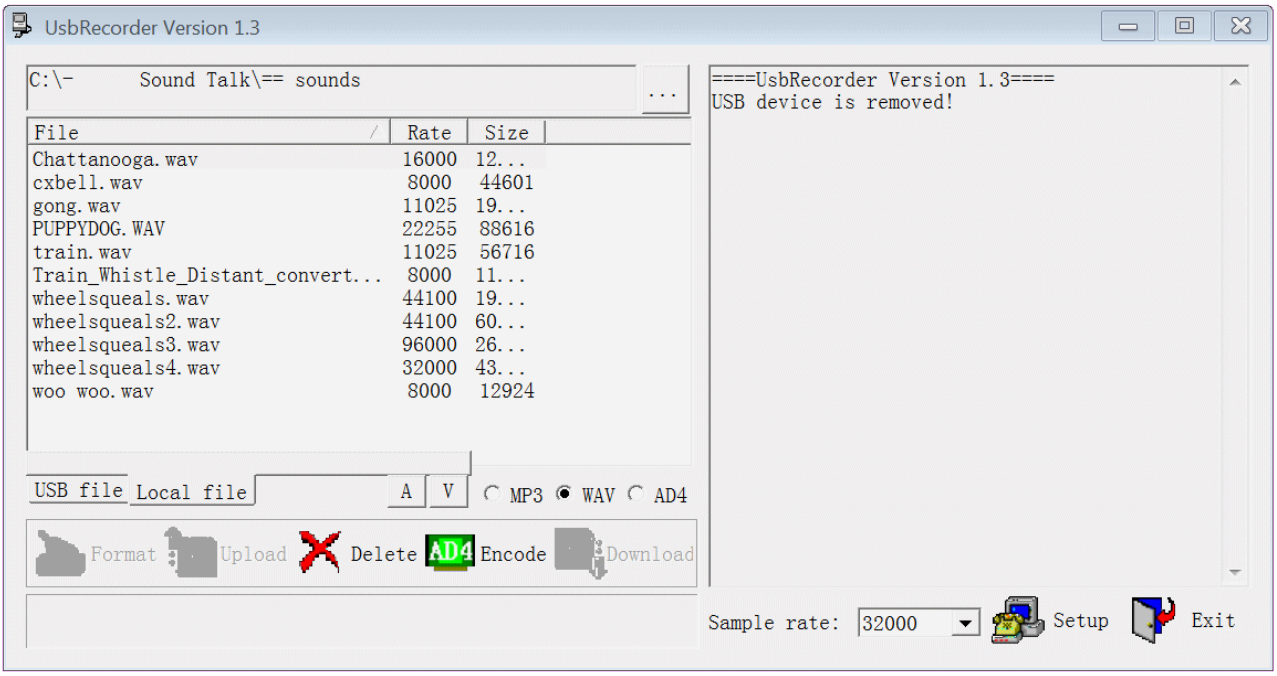


After selecting 32000, the File/Export menu option is chosen to save the file in the new sampling rate.
Although it takes a little effort to learn this technique, it is well worth it, as you can also use this software to edit the audio (convert from MP3 to WAV, mix two or more audio files, remove noise, add echo, introduce fading in/out, etc.).

Step 2 – Convert the file to ad4 format

The conversion software is called USBRecorder and is downloadable from
www.waytronic.com/download.asp

The software has a single window that looks like this:



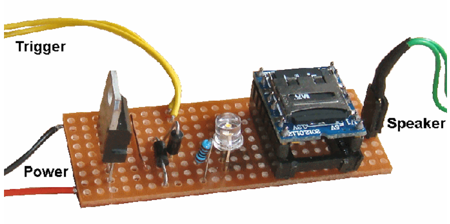
The left-hand pane lets you select a folder to view. From here, you select the file you want to convert.
The example shows that the wheelsqueal4.wav file is already using a 32KHz sampling rate, so an ad4 file version can be created by simply clicking on the 'Encode' button. The newly encoded file would be called whealsqueal4.ad4 and can then be renamed 0000.ad4.

Installing the audio clip

Now that you have the audio file stored in your computer in the correct format, you need a way to copy the file into your memory card. If you already own a digital camera, MP3 player or phone that uses a microSD card, then you are probably familiar with copying files between the device and your PC (probably using the computer's USB port). You simply, take out your existing microSD card from the device and insert the card you will use with the sound module. After copying the file to the memory card, you replace the original card in the device.
An alternative is to use a memory card reader (about £1 on eBay) to copy the file into the microSD memory card.

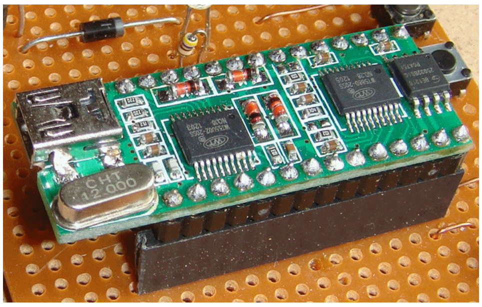
So, the steps are:

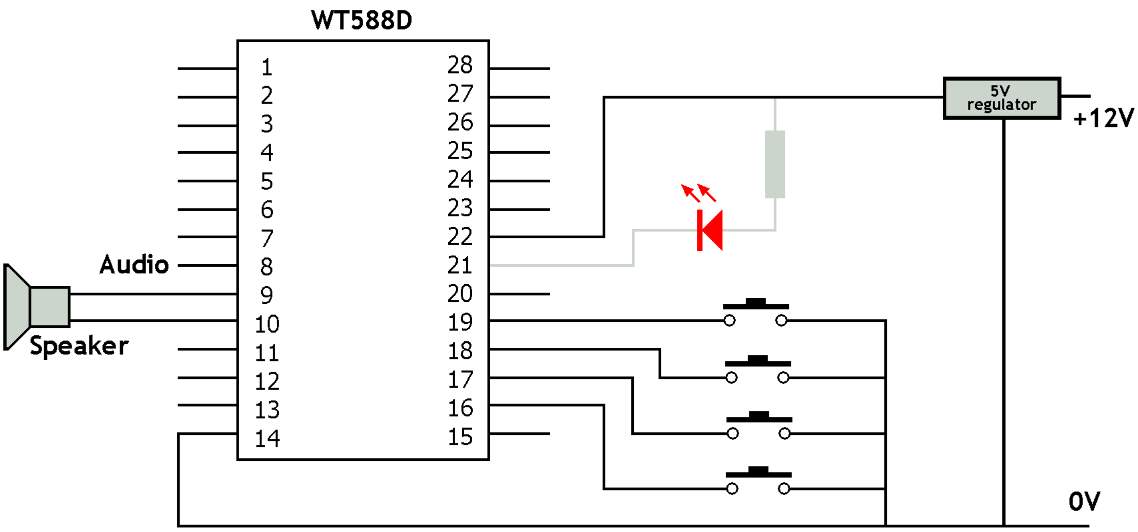
- Prepare the audio file as described above
- Copy the audio file onto the microSD memory card.
- Insert the microSD card into the WTV020 player.
- Connect a loudspeaker to the player's pins 4 and 5, or connect to an external amplifier using pins 2 and 8.
- Connect to a 3.3V power supply (see the diagram).
- Press to play, press to stop.



Option 5 - The WT588

This is the most expensive module (around £6 from eBay) but also the most versatile. Multiple configurations are possible, with the example showing four different inputs, allowing it to play four different audio tracks.
It has its own built in memory (no need for a memory card) and built in amplifier (although you have to supply your own speaker).
The image shows one of these modules mounted on a piece of stripboard. In the top left of the board, above the crystal, is a miniature USB socket (a Mini-B socket). This is used to transfer the audio clips from your computer into the module.
The pin numbering starts from the pin nearest the crystal. Pins 14 and 22 connect to the 0V and 5V respectively. Pins 9 and 10 can be connected to an 8 ohm/0.5W speaker. Alternatively, pins 8 and 14 can be used to connect the module to an external amplifier.





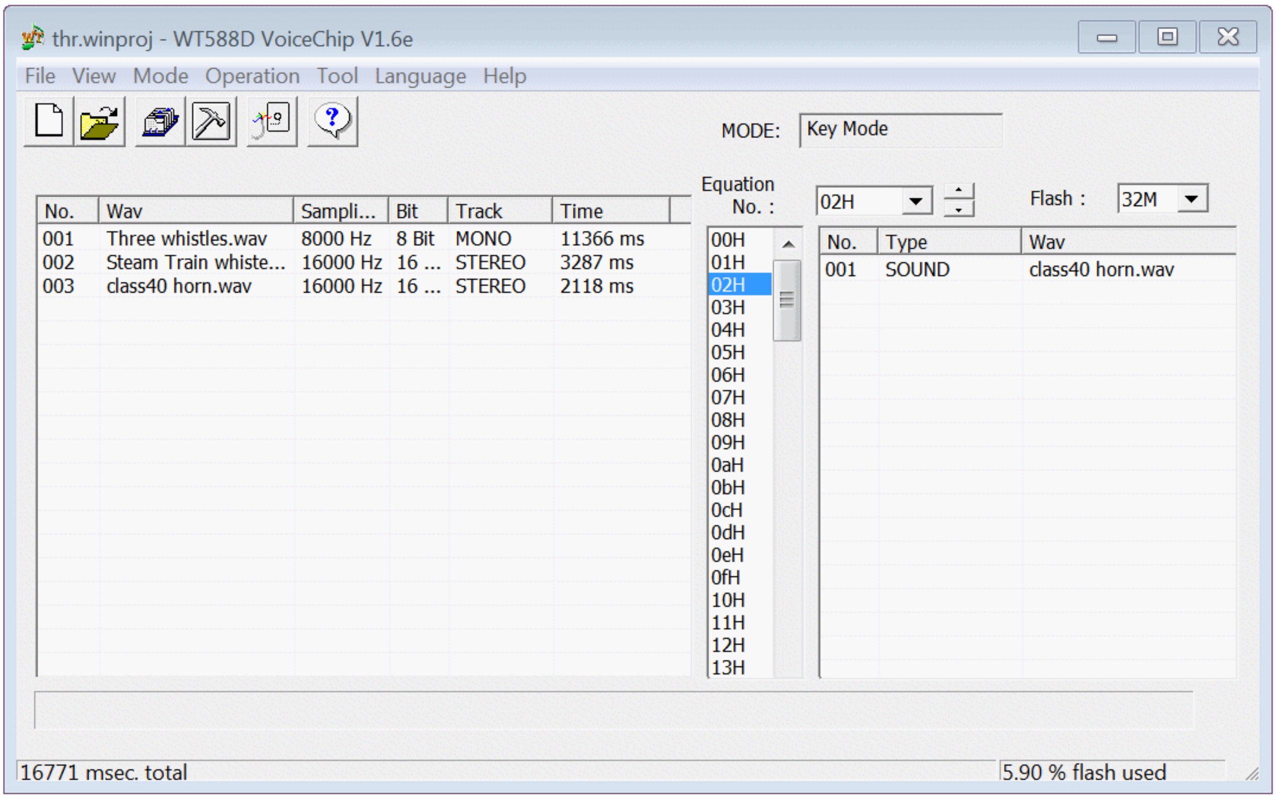
The WTV020 module supports audio files in WAV format and sampling rates of either 6KHz, 8KHz, 10KHz, 12KHz, 14KHz, 16KHz, 18KHz or 20KHz.

Linking sounds to switches

The module can store multiple sounds and so has to be told which sound(s) to play when a particular key is pressed.
This is achieved with yet another piece of free software and it can be downloaded from:

http://www.waytronic.com/download.asp

From this menu, choose 'WTD588d VoiceChip Setup'
When run, you may wish to change the default language to English.
The application has a single screen, as shown in the illustration.



When first run, both main panels will be empty.
Right-clicking in the left panel allows you to load your audio files.
It will refuse to load an improperly formatted audio file, displaying an appropriate error message (e.g. *"Sampling rate can not be greater than 22000"* or *"Inappropriate sampling rate").*
The narrow middle column corresponds to the possible key presses.
In the example, one sound is attached to a keypress. If required, multiple audio files can be attached to a single keypress and they will be played in succession. The program also lets you insert pauses between the playing of a set audio clips, when initiated from a single keypress. Imagine a station making four different announcements, depending on which train arrives (maybe using RFID or CBUS?)

DIY Comparisons

DIY sound players are of varying complexity, quality, length of audio, price, etc. This table gives a broad comparison of the different types.
As always, it is a matter of choice and ability – but think of the possibilities!

	Greeting card	MP3 player	ISD1820	WTV020	WT588
Play time	10 secs	Long	20 secs	Long	Long
Audio source	Mike	microSD	Mike	microSD	Memory
Type	Mono	Stereo	Mono	Mono	Mono
Looping	N	Y	Some models	Y	N
Amp	Y	N	Y	Y	N
Speaker	Y	N	Y	N	N

Sound sources

Some commercial sound systems are supplied complete with their own embedded sound(s). Often, their audio content cannot be altered, although some allow additional sounds to be purchased. Apart from the expense, the available sounds can be limiting.
Others sound players, including DIY versions, need to have suitable sound files installed in them. In these cases, you have to find your own sound files, usually from the Internet.
The most likely options are for commercial (but royalty-free sites) and public domain and free download sites.

Here are some commercial "pay per sound" sites:

www.pond5.com - royalty-free files
www.sounddogs.com
www.pricom.com/Trains/Downloads-DreamSounds.shtml
www.soundsnap.com

Here are some free download sites:

www.grsites.com/archive/sounds
www.audiomicro.com/free-sound-effects/free-transport-and-vehicles/free-train
http://trainweb.org/mdamtrak199/trainsounds.html
http://20th-century-fox-sound-effects.freedownloadmp3.net/steam-trains-trolleys-etc/
www.naturesoundsfor.me

Making your own

For those who have the time, the resources and the skills, home-recorded sound files from real world locations can provide unique layout sounds. Visiting specific locations (e.g. station announcements at a particular station, sounds from a local factory, etc.) can bring even greater authenticity to a layout.
Even very basic recording techniques can produce useful results, while additional recording and editing skills can further enhance the files. This could include adding echo, compression, file conversion, volume adjustment, noise elimination, panning, cropping, looping, fading up/down.
This is really beyond the scope of this book but a useful website on this subject is:

www.qsisolutions.com/products/techinfo/record-sounds.html

Once you have your audio files, downloaded or recorded, you have to to decide whether the sound is played as a one-off (e.g. a train whistle) or is played as a continuous loop (e.g. running water or city background hubub).

If you are creating continuous loops, leave breaks in the sound, or have variations in volume; this relieves the listener from a constant barrage of sound.

Alterations to a sound file require the use of a computer and audio editing software.
There are many such applications available and the most commonly used free software is called 'Audacity' and it can be downloaded for free from:

http://audacity.sourceforge.net/

It has an impressive list of features for a piece of free software.
The site even provides manuals and tutorials on how to use the software.
Check out:

http://audacity.sourceforge.net/manual-1.2/
http://manual.audacityteam.org/o/

For general discussion on matters relating to sound on your layout, you can join a well-used Yahoo Group:

http://groups.yahoo.com/group/LayoutSound/

Synthetic sound

For best results, use the actual recordings of the sounds you want to play on your layout. This is specially true for complex sounds comprising multiple sources (e.g. street scenes, football crowds and so on).

Fortunately, some of the more common railway sounds are less complex. That means that we can create an electronic approximation to the real sound – the result sound like the real thing. Such sounds are the hiss of steam, whistles and horns.

For these, you will find a number of circuits available on the Internet. For example:

www.ehow.com/how_12099618_build-electronic-steam-engine-sound-circuit.html
http://members.shaw.ca/roma/train.html
www.circuitstoday.com/dual-tone-train-horn-circuit

Audio players

An audio system comprises three parts:

- The audio file to be played
- An audio player
- A loudspeaker

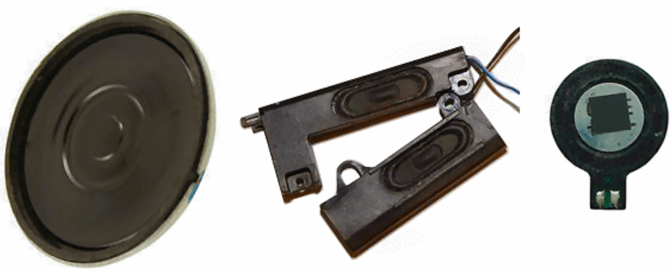
The player is an important part of the system, in terms of audio reproduction (e.g. quality, mono/stereo), facilities (e.g. looping, maximum length of audio clip supported) and method of operation (e.g. handheld controller, switches).

Most, but not all, players have built-in amplifiers. Some use internal batteries, some use power packs and some get their power from another device (e.g. phone or PC).

Types of speaker

The task of the loudspeaker is to translate audio signals into sound waves. Therefore, larger speakers are able to influence more of the surrounding air than smaller speakers. Of course, large speakers are not always practical on layouts. Fortunately, we don't need high volume levels on our layouts so small speakers are adequate in most situations unless we are going for high-quality reproduction. You may also consider using a larger speaker for reproducing the lower frequency rumbles.

Bear in mind that even a poorish quality clip will sound better on a better speaker, while a high quality clip will suffer when played on a cheap speaker. That said, experience shows that perfectly satisfactory results can be achieved with the many small speakers purchased from eBay. These range from speakers a few inches in diameter to the tiny speakers that are sold as spares for games consoles, mobile phones, laptops and toys.



The picture shows a 38mm speaker, a pair of replacement speakers for a Dell laptop and a replacement speaker for NDS Lite games machine.

For the budget conscious, you can also recover speakers from discarded radios, televisions, etc.

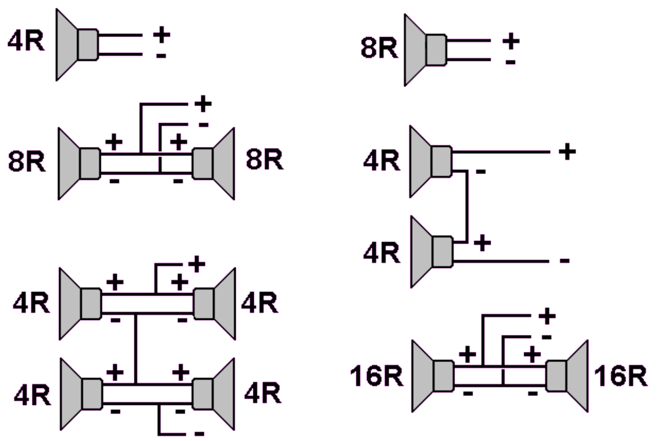
Speakers for mounting inside trains are discussed later.

Apart from size, there are two main considerations when buying or reusing a speaker – its impedance and its rated wattage.
The speaker coil commonly has an impedance of 4 ohms or 8 ohms. Impedance is the 'resistance' of a device when used in non-DC circuits; in these cases the inherent inductance and capacitance of the speaker coil has an additional inhibiting effect.

Speakers also have a wattage rating. This is not a measure of the power it normally consumes; it is a measure of the maximum audio power it can handle without producing distortion to the audio or damage to itself.
The maximum transfer of power to the speaker occurs when the speaker impedance matches the output impedance of the audio module/amplifier. Using a lower impedance speaker than that specified will usually result in poor audio reproduction and often destruction of the speaker. You can use a speaker with a higher impedance than specified, This will cause no damage but is likely to result in a reduced audio level.

Wiring speakers

Connecting a single speaker to an amplifier or sound module is straightforward, with only two wires to connect them. There may be times when you want to connect extra speakers to a single module (e.g. to spread the sound of thunder or traffic over a large area). Or, you may have some speakers whose impedance do not match the requirements of the module.
This illustration shows various ways to connect speakers.
The images on the left show two 8 ohm speakers wired in parallel to present a 4ohm impedance to the amplifier. If required, you could even wire two sets of 4 ohm speakers as shown. The upper pair and lower pair are both wired in parallel, resulting in 2 ohms each. The pairs are then wired in series, to present a final 4 ohm impedance to the module.
Similarly, the images on the right show combinations for 8 ohm systems.

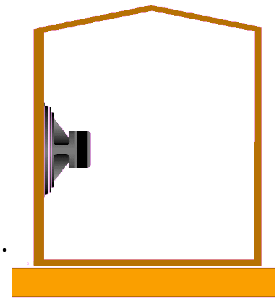


You must not end up with a final impedance that is less than the specified load.

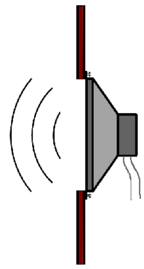
Polarity

The diagrams above show that many, but not all, speakers have positive and negative markings on them, or red and black leads.
A speaker should be connected with its positive lead to the amplifier's positive output terminal and negative lead to the negative terminal.
In practice, the ear is not sensitive to the absolute phase of an audio output. Connecting a conventional speaker in reverse has no impact on how we hear the sound.
However, things change when we connect two or more speakers to the same audio output. If one speaker is wired differently from another, they are said to be *'out of phase'*. This means that at any one given moment, one speaker will be 'pushing' the air while the other is 'pulling' the air. If you were positioned mid-way between the two speakers. The sound would be completely cancelled out in theory. In practice, the effect is less less drastic although noticeable, particularly at lower frequencies.

Mounting speakers



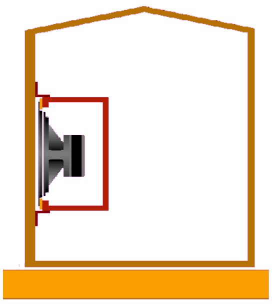
You may want to consider these suggestions before placing speakers on your layout.
A loudspeaker works by using a cone shape to vibrate the air. So, you have to ensure that there is no impediment to this free movement.
Firstly, don't just screw the speaker to a flat surface, such as the wall of a building or the underside of the baseboard.
This illustration shows the building being used as the speaker enclosure.
This method will weaken and muffle the sound. We also want to avoid unwanted vibrations – only the speaker should be producing the sounds.

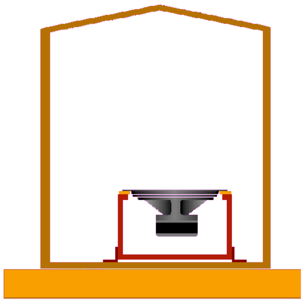


Instead, make a hole in the baseboard/building wall and mount the loudspeaker as shown in this illustration. This way, the material supporting the speaker acts as a baffle board. A baffle board has two purposes:

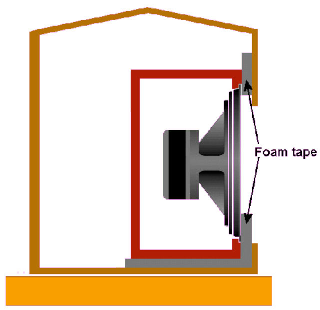
- It increases the volume and clarity. The air is no longer trapped between the cone and the surface it is mounted on and the pressure variations are not absorbed by the material.
- It prevents unwanted interference from the air vibrating at the rear of the cone. Remember that when the cone is 'pushing' air at the front it is 'pulling' air at the rear. We don't want the out-of-phase sound waves from the rear interfering with the wanted sound waves, partially cancelling or distorting them or arriving slightly later to produce unintended echo effects.

The hole in the board should be as large as possible, to allow the maximum effect of the speaker on the air, and should be no less than two thirds of the speaker size.
It is best not to screw or bolt the speaker directly to the material, as this can lead to the problems associated with sound transference, such as absorption and unwanted resonance.
The speaker can be insulated from the baffle material using doubled-sided foam tape, or a glue gun or silicone sealant.



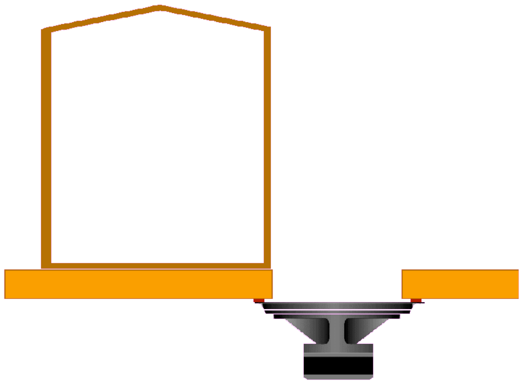


These illustrations shows the speaker mounted in its own enclosure then secured to the side or floor of the building.
Although these may be the only options in some circumstances, they are not the ideal mounting method, as the air is confined in the building and sound is only heard by transference through the building walls and roof.



A better method is shown in this illustration. The speaker is mounted on the rear wall of the building, out of sight. The foam mounting minimises sound transference to the building structure and the hole in the rear wall ensures free movement of air.
This is preferable to having speakers pointed directly at the listeners. The sound is being reflected off other buildings or bouncing back off the backscene towards the listener, a situation that often exists in real life. It's a bit like when you first hear a fire engine or the chimes of an ice cream van in the distance. You can hear the sound but can't initially work out where it is coming from. You may identify the general direction but can only pinpoint an exact source when it is closer to you.

An alternative, where the situation allows, is to mount speakers under the baseboard as shown in this illustration. You can position the speaker hole behind objects such as trees, bushes, walls, hills and buildings. It is best to cover the hole with a grill or piece of cloth, to prevent junk from falling into it. If the hole has to be in a less hidden position, it can be covered with a material that allows air to pass through it – such as lichen.



Lastly, it helps to use the largest size of speakers possible, subject to the physical restrictions on a layout. This improves the fidelity of the sound, particularly for low-frequency sounds such as thunder, machine rumble and diesel engines.

Sound switching

A variety of methods are available, depending on the effect that is required.
In addition, some sound modules only play as long as an input is held low/high, while others will play an entire sound clip when an input pulse is detected.

Permanent sounds

Some sound modules can be set up permanently plays sounds on a loop, perfect for running water, country sounds, etc. These modules run continuously when connected to power.

Random sounds

The sound module can be operated from a circuit that sends out randomly spaced pulses to start the module playing. This is useful for the sound of birds, dogs, saws, car horns, etc. Pulsing circuits are easily found on the Internet and the modules can run unattended.

Timed sounds

The sound module is operated from a timer circuit that runs on its own. This can be useful for any 'scale time' sounds such as a town clock, church bells, etc.

Triggered sounds

This is probably the most useful approach. The sound modules are operated by an event on the layout. For example, a Hector detector located at a tight bend could detect a train and initiate a 'wheel screeching' sound. Similarly, a DC or DCC train-on-track detector could initiate a station announcement when a train enters a station. If you don't want certain trains (e.g. goods trains) to initiate the announcement, then you could use a reed relay under the track as the detector and only fit magnets to the train that should trigger announcements.

Synchronised sounds

This approach allows a trigger to synchronise sound to other events on the layout. So, for example, a signal to a Servo4 board to operate a crossing barrier could also start the sound of crossing bells. Similarly, the flash from a welder (LED) could be accompanied by a 'frazzling' sound, or a flash of lightning could be followed by a roll of thunder.

Manual operation

The simplest approach is to have a manually-operated switch that powers or initiates a sound. Some users may wish to have control over every individual sound, while others may find it an unnecessary intrusion into the running of their layouts. Even automatic sound systems may benefit from the occasional manually-operated sound. For example, kids may like to press a button that starts all the fairground attractions light/moving while the sound of fairground organ is played.

Of course, you may also consider having a single master switch that disables all audio on the layout (very useful for those late night sessions).

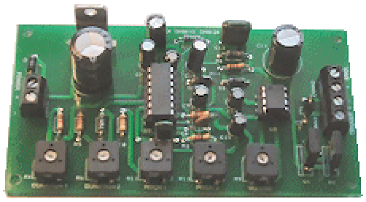
On board train sounds

The previous pages covered trackside sound but it is increasing common to also add sound modules inside the loco.

There are two types of on board sound module:

- Self-contained units that work on DC or DCC.
- DCC decoders that specially designed for handling sound

Self contained modules



This image shows the Trax Controls DHM-12 and DHM-24 Two-Tone Diesel Horn Modules, working on 12v and 24v respectively. Its size is 110mm x 60mm and it is intended for fitting inside O gauge and larger locos.

The row of small preset controls allow the pitch and duration of each tone to adjusted separately.

For onboard use, it can be triggered by having reed relay switch mounted under the loco; the switch is operated by a magnet set in the track.

For lineside use, it can be triggered by a pushbutton or by a reed relay embedded in the track that is operated by a magnet mounted under a passing loco.

The SWM-1 module is the Steam Whistle equivalent.

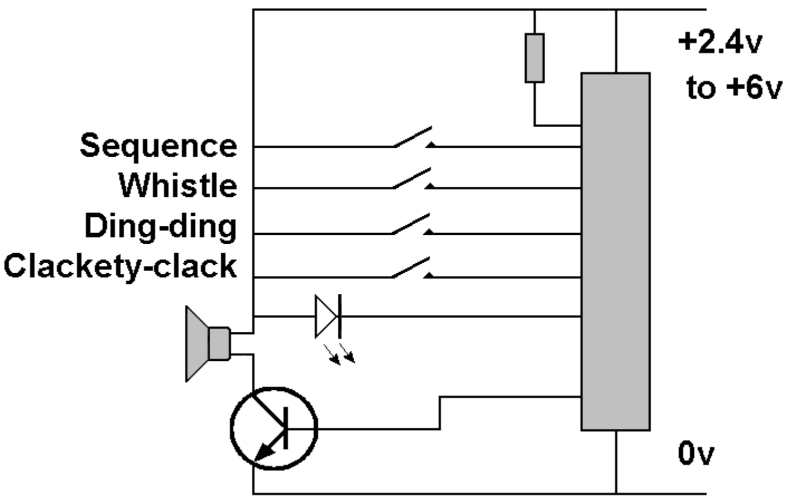
Trax Controls also produce steam and diesel sound modules that reproduce the sound of the locos in motion. The sound content and volume alters as the voltage to the track increases. In the steam module, the sound varies from a hiss while stationary to a chuff repetition that increases with increasing track voltage.

For more details, see:

www.traxcontrols.com/

Quasar Electronics produce the SG01M module.

At 30mm x 16mm, it can be fitted into smaller gauge locos, maybe in the tender or a wagon.



Like the DHM-12 above, it needs a separate power supply.

It produces three different sounds and these are triggered by fitting reed relay switches under the train.

Magnets can be fitted round the track, wherever a sound should be triggered.

Careful positioning of the reed relays and magnets is necessary, if you wish to use all three sounds.

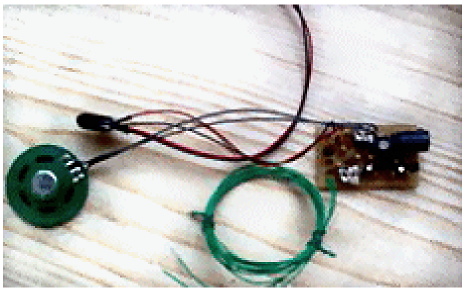
It is very simply wired, as shown in this illustration.

It uses a tiny 1 microamp standby current and under a milliamp when producing sounds. Being so economical, the battery could be left installed for long periods.

For more details, see:

www.quasarelectronics.com/sg01m-four-train-sounds-generator.htm

Mr Rails produce the MistyMK2 sound module, as shown in this image.



There are two versions, for OO and O scale. They use the same 50mm x 29mm board, with the O gauge version using a larger speaker.
It uses a 9v battery which has to be fitted somewhere on the train.
Like the DHM-12 module, it senses the voltage level coming from the track and produces a chuffing rate that matches it. It has a control for adjusting the voltage/chuffing rate for the particular loco it is fitted to.
For more details, see:

http://www.mrrails.com/html/misty.html

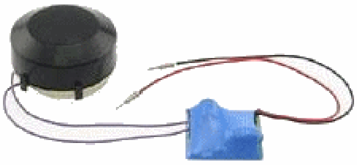
DCC sound

If your layout is DC only, then the above sound modules are the types you have to use to achieve on board sound.
These boards can also be used with DCC systems but the modules designed solely for DCC working, although more expensive, are considerably more versatile and configurable.

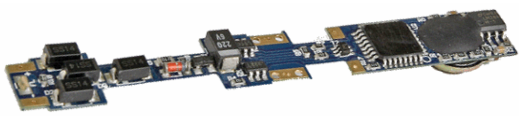
At its simplest, a DCC sound decoder is a small board with sophisticated electronics that has two wires that are connected to the loco's power pickups, and two wires that connect to a small loudspeaker.
At the other end of the range, some decoders can handle 8 simultaneous sound channels, multiple speakers, etc.
Because it is DCC, the decoder's sounds can be triggered by lineside events (e.g. track occupancy detectors), manually from the operator's handheld controller, or through a computer program.
See the chapter on DCC for details on how the system works.

Sound decoders are available in a variety of formats.

- Some are sold pre-fitted in locos, so the user has nothing to do but learn how to use it.
- Some handle sound only and a separate loco decoder is needed for the motor's speed and direction control. The image shows an MRC sound decoder, available in steam and diesel versions. The user has to fit and wire the decoder, or have it professionally done. Since it is a standalone unit, it need not be fitted in the loco along with the loco decoder; it could be fitted in the loco's tender or first coach.

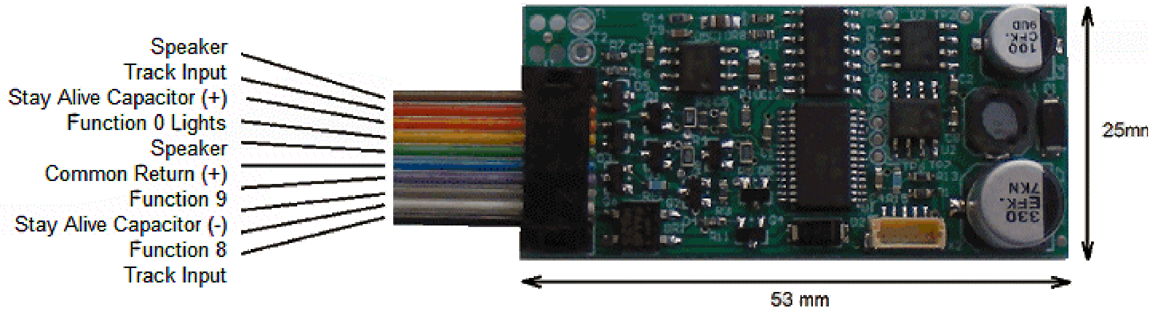


- Some combine handling sound and controlling the loco's motor (Digitrax, Zimo, etc.).
- Some are designed as a drop-in to particular loco chassis, as in this MRC module for a drop-in to a Kato chassis.

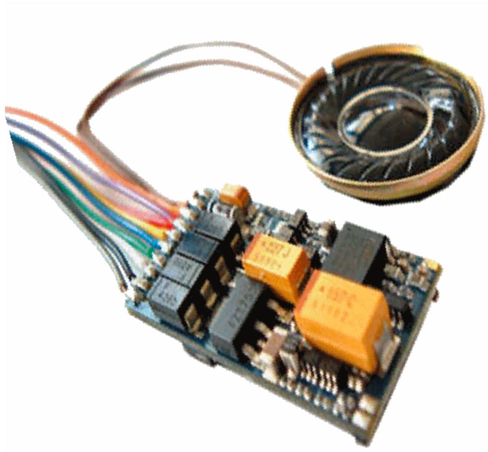


- Some are designed as plug-ins to loco decoders (e.g. the Digitrax Soundbug plugs into the Digitrax DH165 series decoders).

This is the MyLocoSound DCC Steam Decoder



Apart from handling sound, it handles a couple of function outputs. It also has option to fit a stay-alive capacitor. This is a storage capacitor that powers the sound module during any short interruptions in power from the track (e.g. dirty track, navigating frogs, etc.). It is used to prevent the sound from dropping out during these interruptions.
For more details see:
www.mylocosound.com



This image is the LokSound 4.0 from ESU
It can store up to 276 secs of sound and can play eight different sounds simultaneously through its eight audio channels.
You can purchase an (expensive) add-on, the LokProgrammer, that allows you to change the sounds stored in the decoder.
Apart from handling sound, the loco motor and various function outputs, it has the ability to drive four servos. This allows additional possibilities such as a rescue crane, a guard waving a flag or a lamp, and so on.

Function	Soundslot	CV	Range	Value
Sound on/off	1	259	0 – 128	99
Sound on/off	2	267	0 – 128	99
Whistle (playable)	3	275	0 – 128	128
Bell	4	283	0 – 128	128
Coal shovelling / Oil burner (AUX2 Firebox)	5	291	0 – 128	128
Air pump	6	299	0 – 128	128
Station announcement #1	7	307	0 – 128	128
Coupler sound	8	315	0 – 128	128
Cylinder blow	9	323	0 – 128	128
Conductor whistle	10	331	0 – 128	128
Sanding valve	11	339	0 – 128	128
Safety valve	12	347	0 – 128	128
Brake Set / Brake Release (automatic)	13	355	0 – 128	128
Drain valve	14	363	0 – 128	128
Curve squeal	15	371	0 – 128	128
Short whistle	16	379	0 – 128	128
Rail joints	17	387	0 – 128	128

It provides for the addition of a “PowerPack”, a stay-alive of 2 secs.

As with all DCC decoders, values are stored in configuration variables (CVs).

This table shows part of the CV list for steam. There is another table for diesel and one for electric locos.
For more details, see:

www.esu.eu/en/products/loksound/loksound-v40/

Speakers for locos

Large, quality speakers can give a rich sound that spans the frequency range from high-pitched whistles to the deep grumbling of diesels.

However, unless you are working in larger gauges, speaker size is a considerable factor given the limited amount of space inside most locos.

Smaller speakers suffer from two main performance limitations:

Frequency range

Their small size means that they never produce full deep bass sounds – although some models are better than others.

Audio output

They produce limited audio power output without distortion – although is that really a problem? Their output is more than sufficient in most situations.

What's available?

Speakers are available from major model railway suppliers such as ESU, SoundTraxx, Zimo and TCS.

The image shows a tiny *'Sugar Cube'* speaker.

Speakers are also available on eBay at much lower prices, although you would be best to buy just one try it before buying in any quantity.

Many commercial speakers for model railways are variants on the types that are used in laptops, tablets, MP3 players and cellphones.

Here are just some example speakers:

Manufacturer	Dimensions	Impedance	Wattage	Frequency range
ESU Sugar Cube	12x14mm	8 Ohms	1~2W	
ESU 50331	20mm	4 Ohms	1~2W	
Zimo MicroCube	12x8x8mm	8 Ohms	1W	
Zimo Sugar Cube	15x11x9mm	8 Ohms	1W	300Hz-20KHz
Zimo SugarLump	15x11x12mm	8 Ohms	1W	“More bass”
Zimo LS20R	20mm	8 Ohms	0.5W	
SoundTraxx 810053	20.5mm	8 Ohms	0.2W	550Hz-7KHz
SoundTraxx 810089	15mm	8 Ohms	0.2W	540Hz-20KHz
SoundTraxx Mega Bass	28x28x11mm	8 Ohms	2W	310Hz-1KHz
TCS WOW	20mm, 28mm	8 Ohms	0.8W	

The above table details some of the smaller speakers, with varying power handling capacities. There also have differences in the frequency ranges they can handle (it is difficult to find frequency range data from some manufacturers).

The SoundTraxx Mega Bass speaker is intended for reproducing mainly low frequency sounds, with its larger 39.5x19.5mm speaker covering the wider range of 550Hz-12KHz.

Larger gauges

The range for larger gauges is greater, with circular, oval and rectangular shapes, larger power handling – and increased size.

For example, the Zimo LSFRS5 is a large 50mm 8 Ohm speaker but can handle 5W at 150Hz -20KHz. Note that to fully utilise larger speakers (e.g. greater than 3W) you will probably need a large-scale decoder.

Mounting speakers

This is always a problem, particularly in smaller gauges.
While speakers are more easily fitted in coaches or wagons, this loses some of the authenticity of the sound (sounds come from locos not rolling stock). However, in extreme cases, this may be an only option.
From OO upwards, life is a little easier, as there is often room to squeeze in a speaker, e.g. inside the cab of a diesel or inside the boiler of a steam loco.
Some of the smaller speakers, e..g. the Sugar Cubes, are already mounted in their own enclosures to maximum sound output.
Some notes on mounting Sugar Cube speakers can be found at:

http://www.sbs4dcc.com/tutorialstipstricks/sugarcubespeakernotes.html

Fitting two speakers , where space is available, will result in increased volume.

Loco sounds

There is a huge range of authentic recorded loco sounds available in DCC sound decoders.
For example, check out the YouChoos website:

http://www.youchoos.co.uk/Index-Shop.asp?L1=Sounds

This company provides sound decoders with steam sounds from GWR, LMS, LNER, SECR, LSWR, SR and BR standard sources. Its diesel selections include shunters, locos, and DMUs. It also has a selection of electric loco sounds.
Many other DCC sound decoder manufacturers offer similar products.
If you still can't find what you want, Howes Models offer a service that replaces the existing sound files on ESU LOKSOUND decoders with our own sound recordings .

Chapter 11

Track occupancy detectors

Track occupancy detectors

Detecting the presence of a train has been a useful facility on many model railways, although perhaps still under-used.
This chapter looks at different methods of detecting track occupancy and compares their relative merits. The chapter covers nine different methods and MERG currently sells components or kits for six of these (using reed switches, Hall Effect switches, current detection kits, infra-red and RFID modules).
Others technologies exist, utilising radio transceivers or even GPS systems.
Track occupancy detectors are often called *'TOTIs'* – train-on-track-indicators.

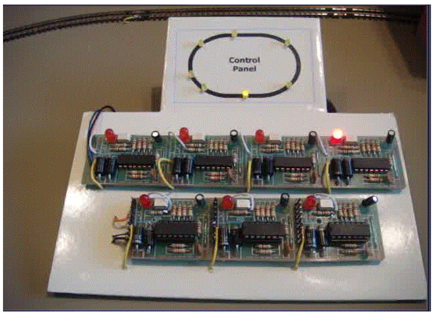
Track occupancy detectors uses

Hobbyists will have different uses for detectors, including a mix of some of those below.

Improving look/feel

- A train enters a siding and the yard light comes on.
- Track side animations.
- On-screen timetable announcements.
- On-screen train descriptions.
- Sound effects (station announcements, whistles, bells).

Improving functionality



- Feedback to mimic panel (the image shows part of an exhibition layout with TOTI modules lighting LEDS as a loco travels round a loop).
- Automatic crossing gates and barriers.
- Automatic signalling.
- Reverse loops.
- Collision prevention.

Improving control

- Detecting train lengths.
- Detecting direction of travel.
- Detecting a wagon that becomes detached from a train.
- Selective switching of points.

This list is by no means exhaustive. You can no doubt add many more possible uses, as the only limit is our imagination.
While many of these features can be implemented by all users, some require the use of computers.

Levels of detection

Not all detectors are interchangeable, as they work and operate differently.
Firstly, there is a distinction between *'spot detection'* and *'stretch detection'*.

Spot detection	The detector's output is only changed when a train passes the detector. The output reverts to its original state when the train leaves the detectable area (although sometime you can add a delay before switching back). This method, therefore is momentary and relies on movement past the detector. This method is useful for triggering

	events (sounds, CBUS notifications, etc.)
Stretch detection	The detector's output is changed when a train sits in any part of a track section. The output only reverts back to its original state once the train (usually the loco or a specially adapted coach) vacates the entire track section. This method, therefore, does not rely on movement but on occupancy. This method is useful for stationary information (knowing a hidden siding is occupied, maintaining a signal at green while the train passes, etc.)

Another distinction is between *'Occupancy'* and *'Identification'*.

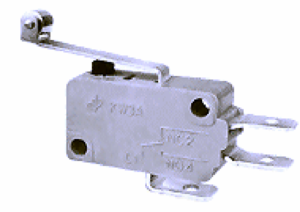
Occupancy	The operator, or computer, knows that a track section is occupied by 'something'. This is not as bad as it sounds and this method is still very useful and is the most widely used.
Identification	The operator, or computer, knows not only that a track section is occupied; the details of the train are made available (allowing for station announcements, different train whistles, selective action for point operations, etc.)

Starting at the beginning

The chapter looks at the different methods, starting with the simplest and ending with the most complex. Hopefully, readers at all levels will get something out of it.

So, lets start with a detector that even the newest beginner can tackle.

Microswitches

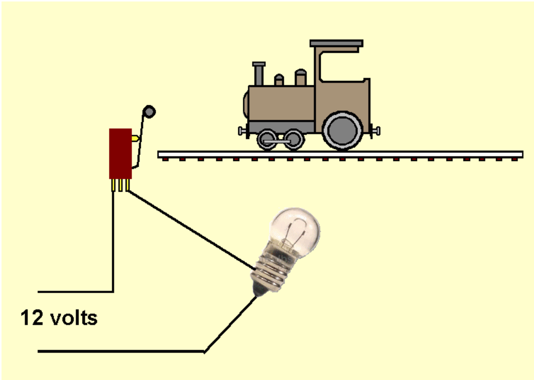


Microswitches are available in a range of sizes, current handling abilities and switch contacts arrangements.
As the image shows, all have a plunger that operates the switch.
This is often supplemented by a lever that operates the plunger.
As it provides greater leverage, it needs less effort to operate the switch.
Some levers are a simple strip of metal, while other have rollers (see the photo) to reduce any friction against the lever.
Some microswitches are simple ON/OFF devices and have two tags that are shorted when the lever is pressed. This simple action, known as single-pole single-throw (SPST) is similar to a push-button - the switch is designed to release again when the pressure is removed from the lever.
The microswitch shown in the image is a single-pole double throw type (SPDT). It has three tags with the lower tag shorting to one of the upper tags when the lever is pressed and shorting instead to the other tag when the lever is released.

An alternative is the 'leaf switch'. It is much smaller and need very little operating force. It is designed for low current switching and can be use for indicating point positions, when mounted close to the point's tie bar.
Its main weakness is that is open to the elements.

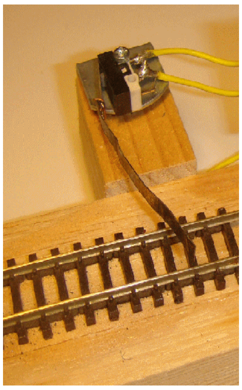
Basic use

We all know that if we take a power supply and wire it across a bulb, the bulb lights. We also know that we can wire a switch in series with one of the wires, so that we can switch the lamp on and off.
Let's start from there.



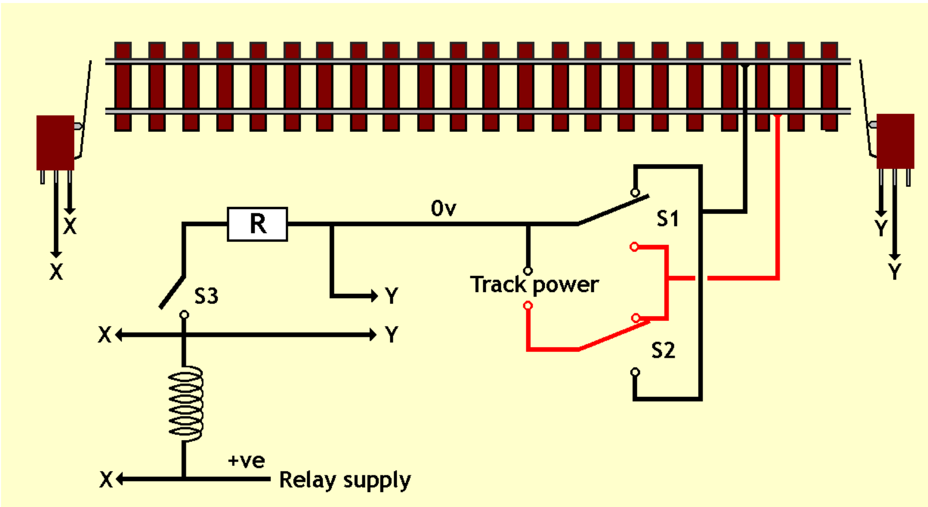
The illustration shows the simplest form of train detector possible.
The lever of the microswitch protrudes from between the buffers of a siding. When the loco reaches the end of the siding, the lever is pressed and the lamp lights. So simple, so basic, yet so handy when you have hidden sidings.
The next step might be to replace the filament lamp with an LED and dropper resistor.

If we now use two microswitches and add in a relay and a resistor, we have a cheap alternative to a rolling road, or even a very basic DC shuttle system.



Microswitches are positioned at each end of a section of track. The image shows that I added an extension to the microswitch operating arm, using a piece of tin taken from a beans can.
The relay is a 3-pole double-throw type. Two change-over switches are used to switch the polarity of the controller power to the track, while the other is used a single switch to the relay. The relay is available from RadioSpares.

The circuit relies on the *'hysteresis effect'*. This means that the relay takes a certain current to move the switch assembly but, once activated, requires a lesser current to hold the relay switched. In my own circuit, my 12v relay needs 15mA to operates the relay contacts – but they will stay operated until the current drop below 5mA.



The value of the resistor depends on the relay's operating voltage and its internal resistance. In my case, a 220 ohm resistor was used.

The sequence of events is:

- When power is applied, the relay is not operated and the power is switched as shown in the illustration.
- The loco moves rightwards until it presses against and switches microswitch Y.
- This creates a path from the relay's +ve supply, through the relay, through microswitch Y to 0V. The relay operates.
- This results in two activities. The relay switches the two changeover contacts S1 and S2, resulting in the loco power to the track being reversed. It also switches S3, which connects the relay to 0V through resistor R.
- When the loco starts moving leftwards, microswitch Y is released but the relay remains held operated as there is sufficient current being supplied through resistor R to hold in the relay.
- When the loco reaches the end of the track, it operates microswitch X. This shorts across the relay, making it drop out (there is no short to 0V due to the presence of resistor R).
- The relay releases all the switches and the wiring reverts back to that shown in the illustration. The process starts all over again.

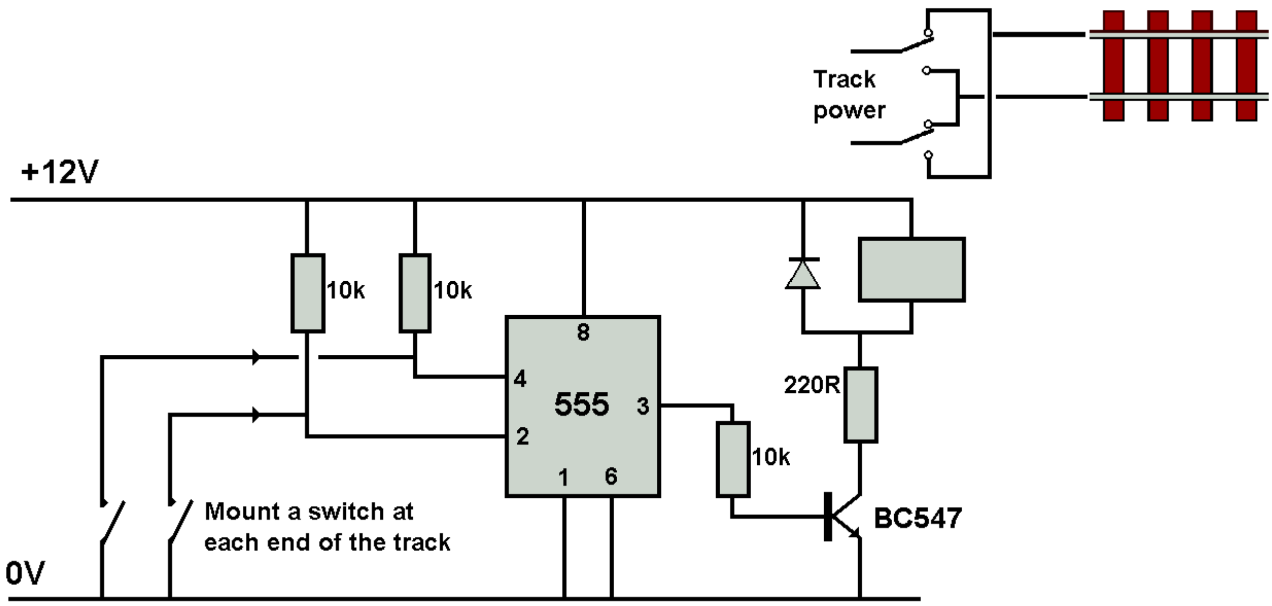
Place a loco on the track and connect the power to watch the loco traverse up and down the track at a speed determined by the controller. This allows it to warm up locos, instead of using an expensive rolling road.
It can even act as basic shuttle, although there is no pause at each end before changing direction.
Again, this is not the most sophisticated project but, given the small component count, it makes a good starter project.
Pros:

- Cheap, easily obtained, simple. Good for beginner's projects.
- Easily fitted to an existing layout.
- No modifications to track/power/etc. are required.
- Unaffected by light
- No modifications to locos or rolling stock.

Cons:

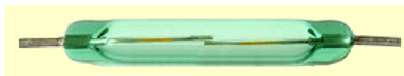
- Limited use, due the way the detector can only be placed at the ends of track sections (although experiments have been made with the microswitch being under the baseboard with the operating lever being operated by trains passing over them).

Here is an alternative approach that uses a 555 chip in *'bistable'* mode. That means that it happily stays in either one of two possible states – i.e. with its output pin 3 or low.
A microswitch, or leaf switch, is mounted at each end of the track and one of these will bring pin 2 down to 0V when the loco pushes it. The other microswitch brings pin4 low when the loco pushes it. When pin 2 goes low, it makes pin3 goes high. Pin 4 going low makes pin 3 go low. The output from the chip is taken to a transistor that switches the relay on and off, as the loco reaches on end of the track then the other.



Reed switches

This was one of the most popular early detectors used for model railways.
Its is covered in the MERG Technical Bulletin A7/1.
Two strips of metal are held in a sealed glass tube.
When a magnet is positioned nearby, one of the strips is attracted by the magnetic field and bends; in so doing it makes contact with the other strip.



They are in common use in burglar alarm systems and a range of industrial applications such as in wind turbines, mines, and as fluid level switches. Their airtight situation makes the contacts immune from external problems such as dampness. Also, there is no danger of sparking hazards in potentially combustible areas. This means that they are made in large quantities and easily available from a range of suppliers.

The image shows part on exhibition layout. A reed switch is embedded under the track of the siding. A loco pulls a wagon that is fitted with a magnet. When the wagon passes over the reed switch, the switch operates and lights the yard lamp (on the right); it also sends back a pulse to a PIC that controls the overall sequence. The PIC makes the train stop and activates a series of servos that opens a door, rotates a crane, moves a lorry, etc.

Pros:

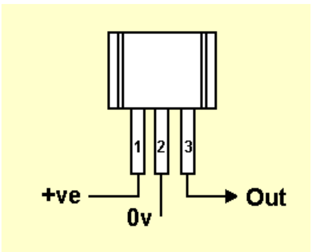
- Cheap, easily available.
- Easily fitted.
- No modifications required to locos or rolling stock.
- No modifications to track/power/etc. are required.
- Unaffected by light.
- Good for garden layouts, due to their waterproof construction.
- Can have conditional trains, with some fitted with magnets and others not. So, for example, points may only move to let goods trains enter a shunting yard; while the express train (not fitted with a magnet) cannot activate the point.

Cons:

- Each piece of rolling stock requires its own magnet.
- Only produces a momentary pulse output, so will probably require additional circuitry (such as latches or delays) in most circumstances – unless the pulse is used to send event information back to computers.

Hall effect switches

Hall Effect switches are widely used in the automobile industry and most cars have many such switches (transmission sensors, ignition sensors, gearstick position sensors, etc.). More details are available from Technical Bulletin A7/2, on the MERG website.
Like reed switches, Hall Effect devices are switched by the presence of a magnetic field. Unlike reed devices, they need power to operate.



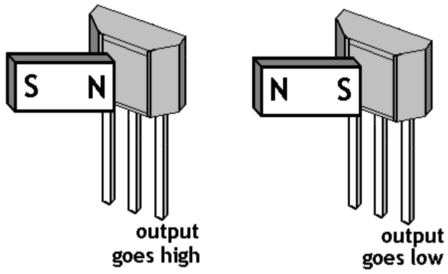
They are simple, 3-pin devices, with two of the pins connecting to a power supply (typically between 4.5V and 24V). The third pin is the output pin that changes state when the device detects the presence of a changing magnetic flux.
With output currents of 10mA upwards, their outputs can be directly connected to LEDs, via the usual voltage drop resistors. These devices are open collector, which means that they need a load between the collector and the positive supply.
These switches are available from the MERG Kit Locker.

An interesting feature of Hall Effect switches is that they can differentiate between a magnet's North and South Poles, unlike reed switches. The output changes when a North pole passes the flat side of the device, or when a South pole passes the chamfered side. So, a North pole passing the chamfered side creates no change in output. Similarly, nothing happens when a South pole passes the flat side.

Latching types

A variation on these switches is the 'latching' switch, available for about 40p each from eBay. It is still a 3-pin device but it behaves like a toggle switch (a bistable).
It needs to use both the North and South poles of a magnet to switch states. As the illustration shows, a North pole makes the output go high. However, when the magnet is removed, the output remains high (unlike the normal Hall Effect switch which reverts back to its previous state).

When a South pole now passes the switch, its output goes low – and stays low until switched again by a passing North pole.



The MERG Technical Bulletin A7/4 suggests a use for a latching switch involving putting magnets at both end of a train, The front magnet would be mounted with its North face pointing towards the Hall Effect switch, while the rear magnet mounted to present its South face. As the train passes the switch, its output goes high and stays high until the end of the train passes. Paul King cleverly shows how a normal Hall Effect switch can act like a latching type, at the expense of an extra magnet being placed behind the switch. Although requiring an extra magnet for each switch, it has the additional benefit of increasing the range over which the switch operates. Of course, this extra benefit is only available when used in latching mode.

Pros:

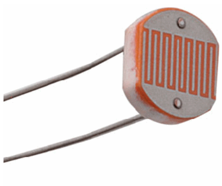
- No modifications to track/power/etc. are required.
- No modifications required to locos or rolling stock.
- Unaffected by light.
- Small component count.
- Clean ONs/OFFs (they use internal Schmitt triggers).
- Can be used as a latching sensor

Cons:

- Possible problems in positioning magnets and detector switches.
- Needs a powerful magnet, as the detector has only a small sensing distance (around 5mm).
- Only a momentary pulse output, unless using latching type.

LDRs

Light dependent resistors (LDRs), as you might expect from the name, are devices whose resistance changes with the amount of light that falls on them.
They are two-wire devices and can be wired either way round.

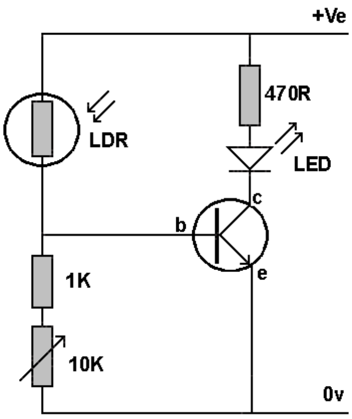


In their 'normal' state (i.e. in darkness) they have a very high value of resistance, usually measured in Millions of Ohms. As the amount of light shining on them increases, their resistance progressively lowers to a low point measured in hundred of Ohms.
This output is different from the other detectors which are either fully ON or fully OFF. The output resistance can be at any value within its range, depending on the amount of light shining on it. This characteristic could be used to improve the lighting on a layout. If the layout is operated in a subdued light, the lights inside the buildings could be dimmed a little to prevent an unrealistic glare. If the layout, however, was then operated in a brightly-lit room, the lights could be brought up to full illumination.
In most cases, though, LDRs are used to detect the presence of objects passing along the track.

LDRs are usually mounted under the track with the light sensitive area located between the sleepers.

The normal ambient light in the railway room results in the LDR having a relatively low resistance.
As the loco/wagon/coach passes over that section of track, it lowers the amount of light that reaches the LDR and the LDR resistance increases. This change of resistance is used to sense the event and trigger some activity.

The diagram shows a typical circuit diagram for boosting an LDR's output. The 1k resistor and the 10k variable resistor are used to tweak the light level at which the LED will illuminate. These can be replaced by a single resistor once the required value for a particular LDR and light surroundings is determined.
Mostly any NPN transistor will do in this circuit, including the BC547B that is available from the MERG Kit Locker.
MERG Technical Bulletin T33_25 describes another interface using a LM358 dual op-amp IC.I



If you Google for *'LDR circuit'* you will find a wide range of different circuits using different transistors and ICs.
Pros:
- Cheap.
- Works for locos or wagons/coaches.
- No modifications required to locos or rolling stock.
- No modifications to track/power/etc. are required.

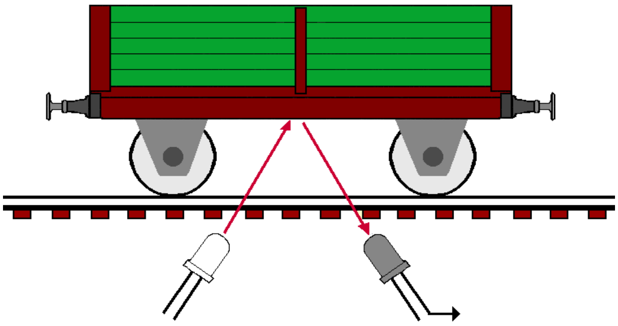
Cons:
- Sensitive to ambient light changes (e.g. at exhibitions).
- May require careful shielding to minimise unwanted switching from peripheral light changes (e.g. as you stretch across your layout).
- Only a momentary pulse output.
- Not a clean switch between on and off, unless using additional circuitry.

Infrared detectors

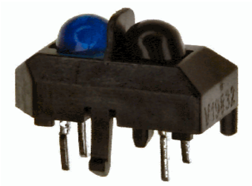
Infrared (IR) waves sit between microwaves and visible light in the electromagnetic spectrum.
Humans radiate some infrared energy and this is utilised by the motion detectors that are used in burglar alarms and automatic security lights. Almost every member will have a TV that control with remote that uses infrared..
In the modelling world, if we use a source of infrared emissions in conjunction with an infrared detector, we can detect the presence of locos and/or rolling stock as they pass by.

There are two methods of detection:
- The IR is bounced off the rolling stock.
- An IR beam link is broken when rolling stock passes through it.

In the first configuration, the emitter radiates the IR, which reflects off the undercarriage of any passing rolling stock. This reflected IR is detected by the IR diode, which allows current to flow and generates a voltage pulse.

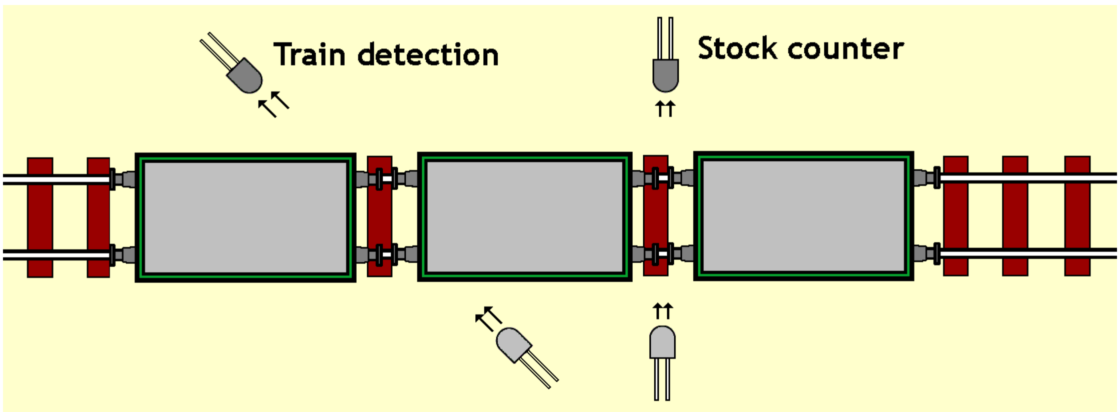


The illustration shows the radiation from an infrared transmitter bouncing off the undercarriage of a wagon and reflecting back to an infrared receiver.



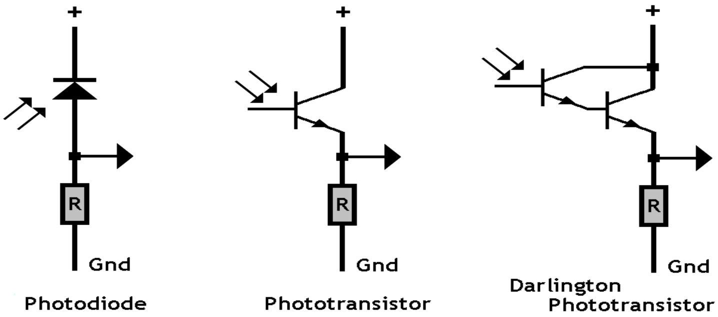
The emitter and the detector diode are available as separate components. They can also be bought with both enclosed in a single case and these are ideal for mounting under the track.

In the second configuration, we use a separate IR emitting diode and IR receiver diode.
This allows us to mount the two devices above the baseboard. They are spaced apart, such that the emitter points directly at the detector across a railway track.
The illustration shows two possible mounting methods.
On the right of the illustration, the two devices are placed directly opposite each other across the track. In the example, there are three items passing the detector. Since there are gaps between each item, it would result in the beam being broken three times and three output pulses being generated. Mounted this way, it could act as a counter of how many items were in a particular train.



However, most users would prefer an output that stays set for the duration of the passing of the entire train, no matter how many items were attached.
The left hand of the illustration shows the two devices set at a diagonal across the track.
The gaps between items of rolling stock are no longer detected. This results in the output being set when the first item passes and only being reset after the last item passes.

IR detector types

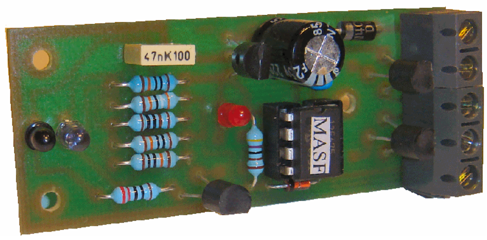


The three types shown in the illustration provide different levels of sensitivity. The basic photodiode can be improved by using a phototransistor, which acts like a photodiode and amplifying transistor in one device. The Darlington phototransistor provide even greater levels of sensitivity with the extra high gain provided by the two transistors.

Mounting

Many IR phototransistors can be sensitive to light in the visible spectrum and it may be necessary to shield the detector to minimise unwanted effects from ambient room lighting. This may involve placing the detector in a lineside building or hiding it in a bush. Often, the detector is mounted inside a piece of tubing that is opaque (don't forget to seal or paint the back of the tube).
Line-of-sight mounting, particularly if the detector is shielded or at a distance, requires accurate alignment for reliable results.
Some users recommend using multiple LED emitters, to increase the usable range of the system.

Hector



The MERG 'Hector' (Kit 72) is an infrared detector kit. The two LEDs on the left of the board are the detector and the emitter. Although mounted on the board in this example, they can be mounted off-board and wired taken back to the board (e.g. for above-board operation).
Detected IR changes result in the indicator LED (the one in the middle of the board) illuminating. The board also provides a variety of logic level outputs that can directly drive colour signal lights or interface to CBUS/other logic circuits. It can detect a piece of white paper at around 2”.
Pros:

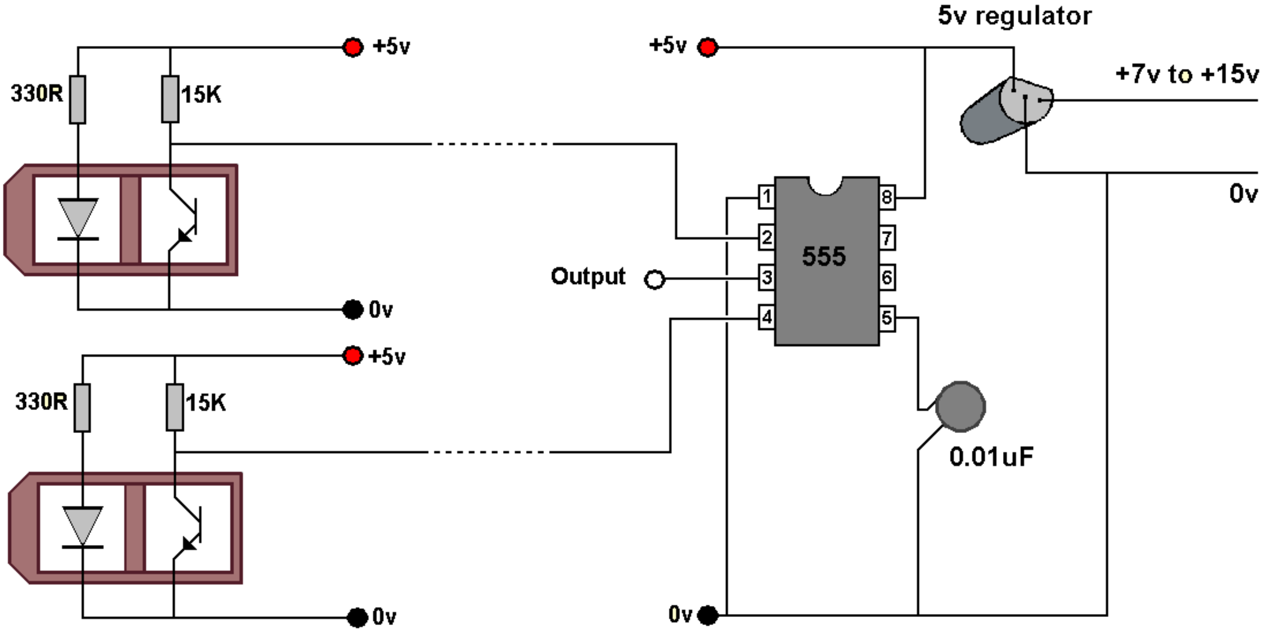
- Can work over a long distance.
- Can be mounted under track or trackside.
- Low power requirements.
- Can detect entire trains, or act as a stock counter.
- No modifications to track/power/etc. are required.
- No modifications required to locos or rolling stock.
- Output is digital (clean ON/OFFs).

Cons:

- Can be a little affected by ambient light.
- May need shielding.
- May need careful alignment.
- Relatively expensive.

Shuttle using infrared detectors

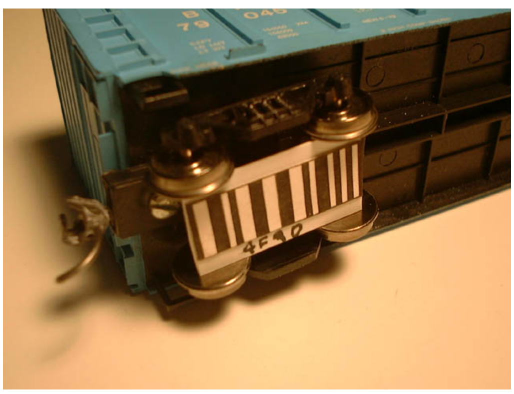
This circuit employs the same 555 bistable chip as used in the previous shuttle circuit that used microswitches as detectors.



The microswitches are replaced with two infrared emitters/detectors, one at each end of the track.
When the loco moves over one of the detectors, its output goes low and the output is wired back to the 555 chip where it triggers a change of output state.
The output can be wired directly to a relay if it is a low current type, or it can be used to switch a transistor on and off.

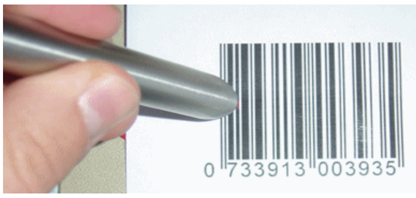
Barcode readers

We have all watched or used barcode scanners when our groceries are swiped across a supermarket checkout. The technology is widely used and bar code labels can be found on most goods.
At its simplest, a bar code is a series of black bars on a white background, the width and spacing of the bars representing different characters. The bar code reader bounces light off the label as it passes and detects the changes in in the amount of light that is reflected back.



In the model railway world, the reflected data is more than just an indication of occupancy; it can send back detailed information on every loco, coach or wagon that has a label fitted (see the photo for an example).
Free software is available that can print barcodes on to sticky labels.
However, this system is not widely used, mainly due to the complexities of integrating it into a model railway control system.
Commercial barcode readers have USB interfaces, requiring computer software to decode the barcodes and act on the data found.
Where multiple readers are to be used, as in most layouts, ways have to be found to

interface them all to a computer. While not insurmountable, it requires hardware and software expertise not widely available.
Alternative options include:



- Hack a USB interfaced scanner to get at the TTL output.
- Use a home-made infrared detector as discussed earlier.
- Buy a barcode scanner wand that already has a TTL output. The image shows one sold for under £10 on eBay.

The output from these readers still has to be decoded to extract the data. This could be achieved via the computer's software or by feeding the TTL output into a PIC whose output pins would feed other logic (or a CBUS CANACE8C).
Until this system is available into a reliable kit that can be implemented by most hobbyists, it remains mostly for those who relish a challenge.
Pros:

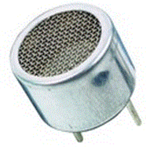
- No modifications to track/power/etc. are required.
- Returns detailed info.
- Can be used to take selective action (e.g. moving points, train whistles)
- Can be used for on-screen announcements.

Cons:

- Requires rolling stock to be labelled.
- Needs reader under every spot to be detected.
- Not easily integrated.
- Expensive option.
- Relies on keeping labels clean and unsmudged.
- The greater the detailed required, the longer the printed label. This may present problems with smaller gauges.

Ultrasonic detectors

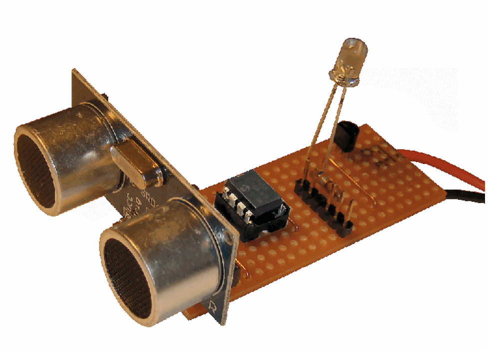
Ultrasonics are acoustic frequencies (i.e. they work by vibrating air) that are higher than those audible to the human ear.
An ultrasonic detection system can therefore be thought of as a loudspeaker and a microphone, tweaked to only operate within a small range of ultrasonic frequencies.
The products marketed by Parsonics range from operating frequencies of 14.6KHz to 420KHz and can operate at distances up to 150 feet.
Ultrasonic detectors are used in car parking sensors, burglar alarms, range finding, process control, etc.
The most commonly used systems operate at around 40KHz.



The units can purchased separately, as shown in the image, and they are marked as TX (transmitter) and RX (receiver). Other transducers can be used as either transmitters or receivers (known as 'transceivers').
Like infrared systems, they can be used facing each (e.g. beam breaking) or mounted to detect reflections.

Many circuits are available for providing the 40KHz transmission and for the detection of the ultrasonic waves. They all tend to have many components.

Another option is to use the HC-SR04 module, as shown in the image. This module can be picked up for around £2.50 on eBay and consists of a PCB with both transmitter and receiver modules mounted on it, along with much of the required circuitry already mounted. Apart from the power connections, it has only two other pins. One pin, when pulsed low by a PIC, sends out a short ultrasonic burst. The other pin changes state if it detects an echo being returned (i.e. something is within its reflective range). I can adjust the PIC to switch anywhere between 1" and 24".



If necessary, you can even check the time it took for the echo to return to calculate how far away an object is (as used by car parking sensors).

The problem with the cheaper ultrasonic receivers, apart from reduced range, is the wide angle of detection (some 15° in either direction). While this might be useful to see if there is any movement in an entire marshalling yard or group of hidden tracks/sidings, it limits its use to short distances in situations where tracks are closely laid (i.e. most layouts apart from larger gauges and garden layouts).

Pros:

- No modifications to track/power/etc. are required.
- Can operate over long distances.
- Unaffected by light.
- Low current consumption.

Cons:

- Needs additional circuitry (either lots of components or a PIC)
- Large; problems in hiding on layouts.
- Wide detection spread, particularly at longer distances.
- Not recommended for continued operation at high temperatures.
- Not recommended for continued operation at high humidity.
- Subject to vibration.

Current Detection systems

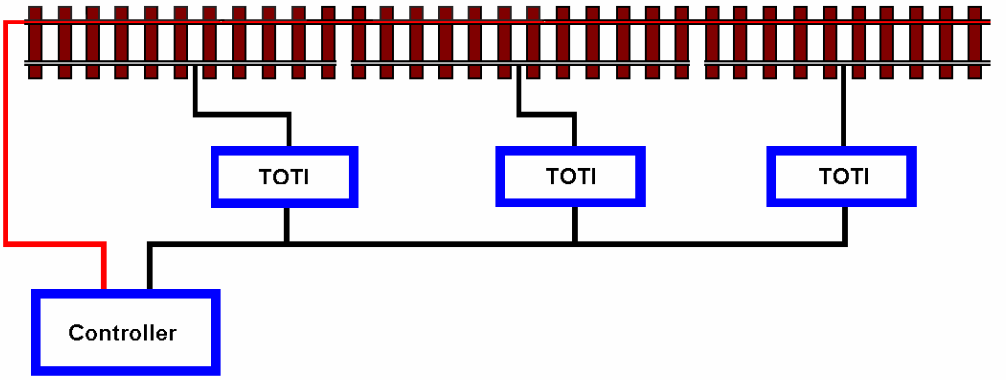
Previously discussed detection systems were *'spot detectors'* – they only triggered when a loco or wagon passed a specific point in the track. When the loco/wagon moved beyond that point, the detector reverted to its previous state (unless a delay was added to the output). This momentary or extended change in output is often sufficient to carry out a required task such as changing a signal or sending a pulse back to a computer or a PIC.

There are many occasions, however, when you want the output to stay changed as long as a section of track is occupied, only reverting back to its original state when the track is unoccupied.

Consider, for example, a hidden siding in a tunnel or in a building. The train may have passed beyond a detector but still be occupying the track. You can avoid running a train into an occupied hidden section by relying on your memory or your jottings. Better still is to have an indication that all operators can see, using lights on a control panel or on a computer screen.

On a layout, we could simply use street/yard lamps to indicate track occupancy. This same output can be sent back to computers for use in automation software.
To achieve this, the detector must activate when a train is sitting in any part of a track section, not just adjacent to a spot detector.

Current detectors work by sensing even a small current being drawn from a section of track, through a loco or adapted wagon/coach sitting on that section.
Since we need to detect occupancy on different sections of track, it follows that the layout has to incorporate track sections that are isolated from each other – the biggest drawback with current detectors.



The above diagram illustrates a controller feeding one of its outputs to a common rail, while the other output is fed to each individual track section by passing through a TOTI detector.

Note

> This is not the same as a block control system. In our example, all the sections are powered all the time.

Three types of current detectors are sold by MERG

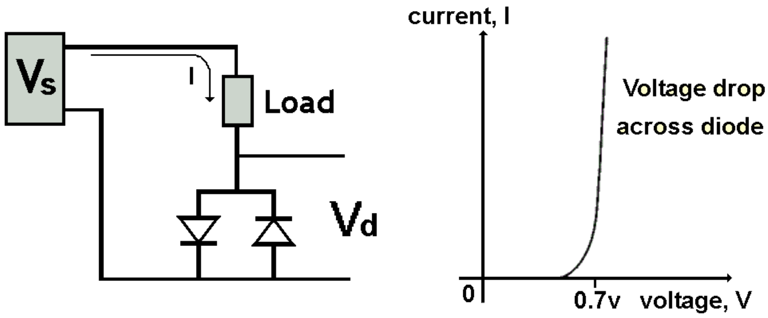
- TOTI-4 and TOTI-12 – diode drop based.
- DTC8 – transformer based.
- PMP7 – diode drop module for a single section.

How current detectors work

There are two methods that are widely used.

Diode Drop Detector

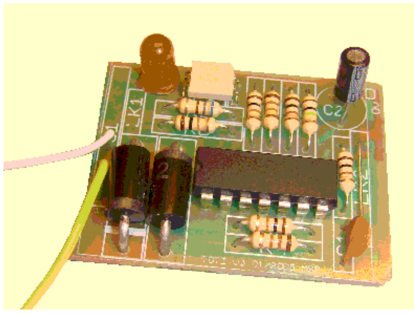
A diode, when a current is passed through it, develops a voltage across itself.   This voltage remains almost unchanged over a range of different current levels, from small to large.



As the illustration shows, a varying voltage source Vs (e.g. a loco controller)   can produce varying current draws through a load, while the voltage across the diode, Vd, remains

almost constant at 0.7V.
If we place two diodes in parallel, then a voltage drop of +0.7V or -0.7V can be detected across them when a load is present to complete the circuit. This positive or negative voltage depends on whether the controller is moving the loco forward or in reverse.

In practice, the load is usually a loco that is sitting in a particular track section. However, a load can be rolling stock fitted with resistors on their axles, as in the image.
The loco or rolling stock does not need to be moving to be detected; it could be be stationary in a siding and still be detected.
The current required to allow this detection is so small that it will not cause the loco to move or overheat, but is still sufficient to generate a 0.7V voltage across the diode combination.

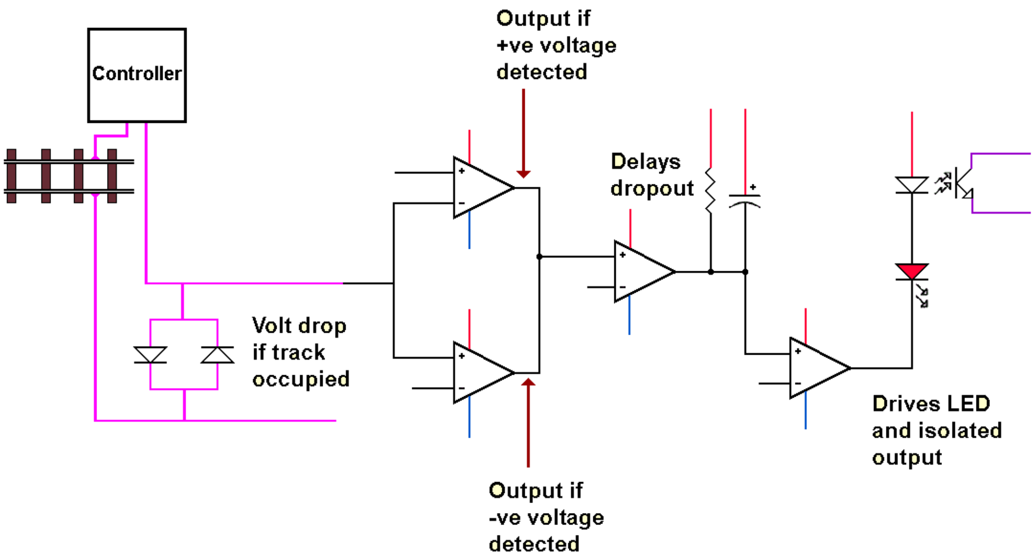


The block diagram below shows the four main stages that the MERG TOTI board (Kit 62) uses to process the output from the diode combination.

The lines shown in purple indicate the initial detection of current produced by a loco occupying the track section. The voltage developed across the diodes is passed to a chip that has four separate parts.

One part of the IC produces an output if a positive voltage across the diodes is detected, with another part producing an output if a negative voltage is detected.

The third part of the IC provides a delay before reverting back to a previous state, once a



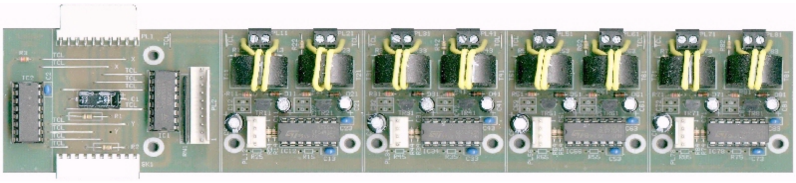
voltage drop is no longer detected, while the fourth part interfaces the module to the outside world.

The necessity to handle both positive-going and negative-going voltage means that the module needs a dual supply of +5V and -5V (available as MERG Kits 64 and 65).

The TOT-4 and TOTI-12 modules work with both DC and DCC systems.

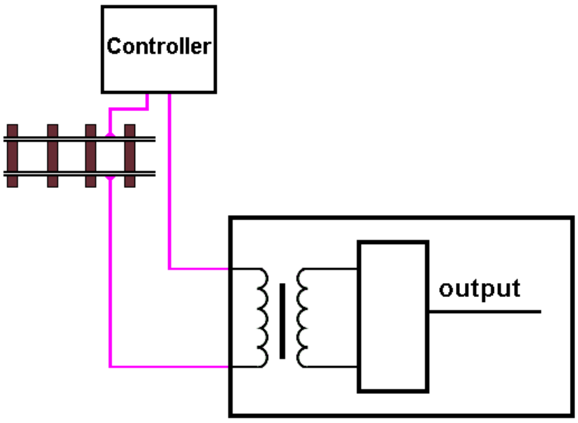
Transformer Block Detector

MERG's DTC8 module (Kit 56) also detects the flow of current but only works with DCC systems, unlike the TOTI-4 and TOTI-12 modules which work happily with either DC or DCC.



Again, the output from the controller passes through the detector module. This time, instead of passing through diodes, the current passes through a transformer's primary winding on the detector module. If the track is occupied, a current flows through the transformer secondary and is detected.
Unlike the diode drop version, the module's circuitry is isolated from the track circuitry and there is no voltage drop caused by its use.

Current detection systems require no additional circuitry and can interface directly to mimic panel LEDS, servos boards, CBUS modules, etc.



Pros:

- Detects any occupancy on an entire section of track.
- Does not rely on movement.
- Can detect rolling stock.
- Unaffected by light.

Cons:

- Need to create track sections.
- Requires modification to rolling stock.
- TOTI modules needs a split supply.
- TOTI modules slightly reduce the track voltage, due to the voltage drop across the diodes.

Summary of features

	Micro switches	Reed switches	Hall Effect	LDRs	Infrared	Barcode	Ultrasonic	Current detectors	RFID
Cheap	Y	Y	Y	Y	n	n	n	Y	n
Affected by light	n	n	n	Y	Y	Y	n	n	n
Must modify loco	n	Y	Y	n	n	Y	n	n	Y
Modify rolling stock	n	Y	Y	n	n	Y	n	Y	Y
Must modify track	n	n	n	n	n	n	n	Y	n
Computer 'essential'	n	n	n	n	n	Y	n	n	Y
Stretch detection	n	n	n	n	n	n	n	Y	n
Identification	n	n	n	n	n	Y	n	n	Y
Detect stationary train	n	n	n	n	n	n	n	Y	n
Good immunity	Y	Y	Y	Y	n	n	Y	n	n

Notes

- Some of these are characteristics are definite (e.g. either train modifications are required, or they are not).
- Others are relative to each other (e.g. an RFID reader can cost over £20, while a reed switch can cost 30p, but is £6.60 for a Hector kit cheap or expensive?).
- Others are indicative (a barcode reader may only be susceptible to ambient light under certain conditions).

The table is a quick guide only; refer to specific pages and product specifications for greater details.

Interfacing detectors

Many of the detectors simply produce a change of voltage when activated and may require additional circuitry to be usefully deployed.

Some examples of interfaces

Direct connection	Controls devices by direct wiring to them (e.g. switching a solenoid point)
Delay circuits	Introducing a pause before reverting to previous state (e.g. keeping crossing gates open while train passes)
Latched circuits	Short pulse switches an add-on's output to stay switched, until unswitched by another pulse (usually from a different detector).
Computer	Providing control of semi/fully automated layouts via serial or USB ports.
PIC	Controlling devices by direct wiring to them (e.g. operating multiple devices, LEDs servos, etc. - including delays, latching and conditional operations).
CBUS	Providing control of semi/fully automated layouts via a CAN_SB or CANRS module.

Chapter 12

R F I D

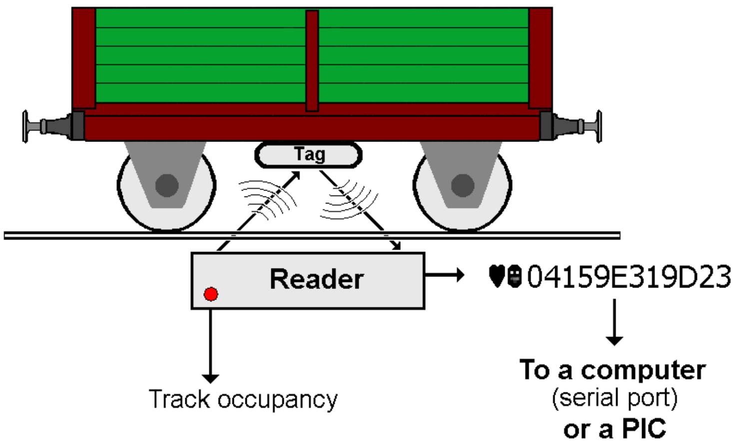
RFID

Radio-frequency identification is probably the most complex and most expensive of the various detection systems. It also offers the greatest possibilities, particularly when used for automation of layouts and other computer-controlled activities.
The system is covered in MERG Technical Bulletins A39/1 and A39/2.

> **Note**
> RFID is not a beginner's project, requiring some electronics and computing expertise.

How it works

Locos, coaches or wagons are fitted with 'tags'. Each tag is a little transmitter that has its own unique code (ID).
A RFID reader module (a little radio transmitter/receiver) is placed under, or to the side of, the track. When the tag passes close to the reader it transmits its ID.
The reader receives this ID and lights a LED to indicate successful reception.
In this way, it acts as momentary track occupancy detector.



More usefully, the module passes the ID out for processing

- via a serial port, for processing by a computer
- via a PIC, for local/special activities

The reader may consist of two parts. There is a module that is used to read and decode the tag's ID and there is a separate circuit that interfaces the reader to a computer. Both these modules, and the tags, are available from the MERG online Kit Locker.

How it can be used

When linked to the programming power of a computer, the RFID system can be used for everything from a simple block detector system to hugely complex automated system.

Below are the most common model railway applications as provided by RFID users:

Control operations

- Move points; e.g. trigger point routing for express/shunter.
- Detect wagons coming detached, if last wagon is tagged.

- Detect a loco's DCC address, in order to control it (e.g. assigning the engine to a throttle).
- Generate switch lists (handling instructions) for operators of fiddle yards, goods yards, etc.
- Determine an item's location by looking back to the last reader it passed.
- Calculate the length of the train, if all rolling stock is tagged.
- Know when train has passed (tag in last wagon/coach).

Maintain stock lists

- Spot if a new loco has same DCC address as one on the track.
- Compile sidings contents (hump yards).
- Generate a list of cars on the train.
- Scan a manifest upon coming in/out of a yard.
- Generate a list of discrepancies (cars not on the train that should be; cars on the train that shouldn't be).
- Ascertain the location of missing cars between operating sessions.
- Sign in and sign out locos at club layouts/exhibitions.

Trigger sound effects

- Station announcements
- The fiddle yard one speaks out the destination of the approaching trains.
- Whistles / bells.

Trigger screen announcements

- On-screen timetable updates. Readers can pass information onto the fiddle yard/goods yard operators as to which trains are approaching them.
- On-screen display of loco picture/info and train info that describes the train that has just passed over a detector.

Tags

The tags used in the MERG kits conform to the EM4102 specification, which is the low frequency band operating at 125KHz. Tags are available in a range of outlines, including key fobs, wristbands, stickers and cards.
The most commonly used for our purposes are either discs (available in 20mm and 30mm) or glass tags (available through MERG in 12mm and 34mm lengths).

These are known as 'passive' tags, which means that they do not need a power supply to operate (very handy – no need for track pickups or batteries).

As the tag passes over the reader, the little coil (look inside the glass tag) passes power from the electromagnetic field generated by the reader to the circuit inside the tag. When charged up, the tag transmits its ID.

Every tag that is manufactured has a different ID number. Don't worry, we won't run out of numbers any time soon, as there are 100 billion possible tag numbers!

Tag positioning

Since we are working with low power levels, the tag needs to be as close to the reader as practical.
The larger tags can harvest more power and therefore work over a greater range – which is OK when you are working with 00 upwards.   However, it takes a bit of ingenuity to fit a 30mm or 34mm tag in N gauge or smaller.
In most cases, the reader is mounted under the track (see later).
For smaller gauges, it is common to fit the tags as low as possible under the loco/ wagon/ coach This ensures that it is as close to the reader as possible.   It also minimises any losses caused by the presence of metal (e.g. weights placed in wagons).

A glass tag is normally positioned from end to end along the length of the loco or wagon.
Alternative positions include:

- Mounting the tag vertically, where undercarriage space is restricted.
- Mounting the tag across the buffers, with the reader standing vertical to the track.
- Mounting the tag on/under the roof of the coach, with the reader on an overhead mount.

The larger tags and discs, due to their increased range, are often placed inside a coach or wagon's body (particularly if there is no metal floor on its chassis).
Lastly, it should be noted that if two tags are located too close to each other, neither will be read by the reader.

Readers

A number of 125KHz RFID readers are available commercially. The two that are used and sold by MERG are made by ID-Innovations and are the:

ID-12

Smaller (26 x 25 mm x 7 mm - approx 1" square by 1/4" thick)
Smaller working read range (12cm+)
Uses 30mA at 5V (4.6V min, 5.4V max)

ID-20

Larger (40 x 40 mm x 9 mm - approx 1.6" square by 1/3" thick).
Greater working read range (16cm+)
Uses 50mA at 5V (4.6V min, 5.4V max)

The advertised working ranges are over-optimistic due to the usual obstructions and reflections in railway layouts.

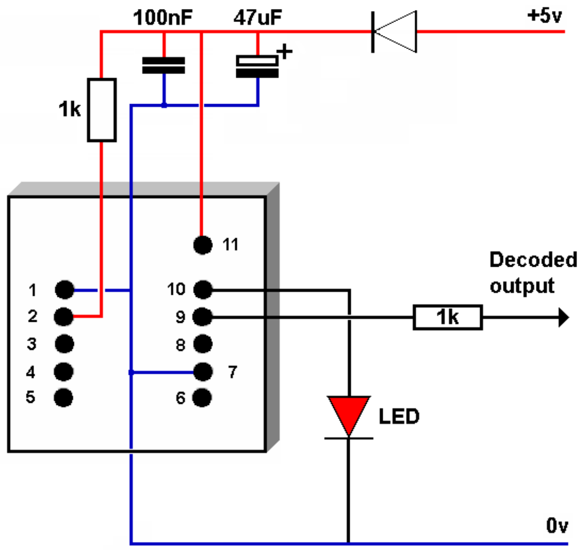
Both modules are cased in black plastic, with the connecting pins under the body.
The picture shows an ID-12 mounted on the MERG reader PCB (Kit 31).
Both these models have an internal aerial, with pins for connecting an external aerial.

Another model, the ID-2 has no internal aerial and is not used in MERG projects as external aerials tend to be bulky and therefore prone to picking up tags from adjacent tracks.
The ID-12 and ID-20 readers are both 11-pin devices and the circuit diagram shows the minimum connections used in the MERG reader PCB (viewed from underneath the reader).

When a tag is passed near the reader, the LED will light and the decoded ID will be sent out on pin 9. Pin 9 can also be used to sound a beeper, although the beeper should be driven by an added transistor.

To use the reader with a higher supply voltage, a 5V voltage regulator would have to be included in the circuit.

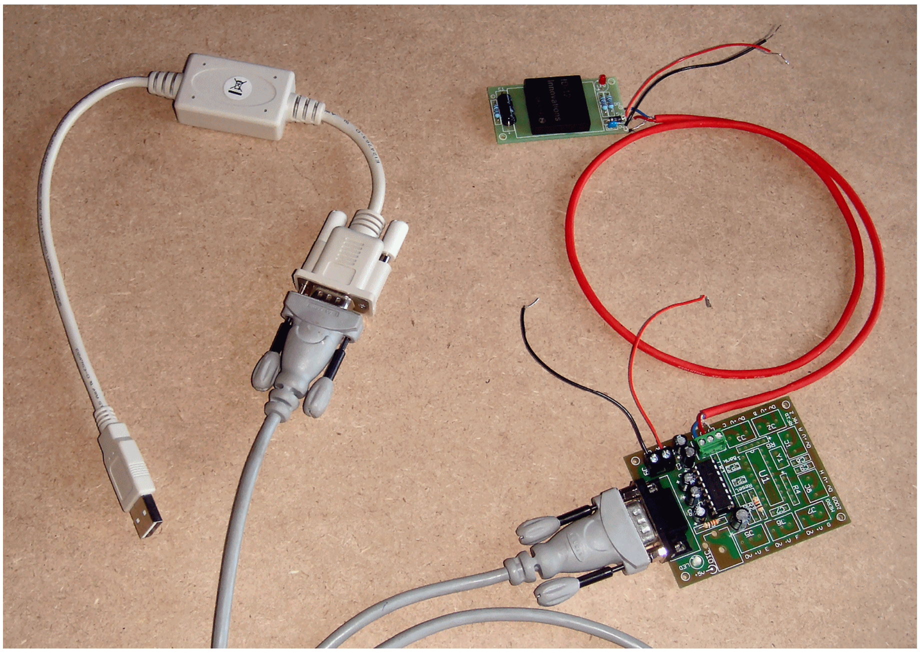
Note that no dropper resistor is required in series with the LED.
Pins 3 and 4 are used if an external aerial is to be connected.
Pin 8 is the same decoded output as on pin 9, only inverted.



Connection to your computer

The output from the above circuit can be fed directly into a PIC, with no other circuitry in-between.
Of course, you then have to know how to write PIC programs to read the data being fed in.
More common is that the output is fed into a computer, using its serial port or USB port.
MERG Kit 30 includes a module that converts TTL level output from pin 9 to the RS232 signal expected by a computer's serial port.

The picture shows a typical setup for a single reader.



The reader (Kit 31) feeds its ID data into the MERG serial output module. The output from the serial output module is connected to a serial cable. If your computer does not have a serial port, a serial to USB converter cable can be used, as shown in the picture. How the computer uses the processed ID is looked at later.

Reader positioning

As with tags, the positioning of readers has a marked effect on performance. There is a trade-off between sensitivity and selectivity. On the one hand, we want the best sensitivity possible. On the other hand, we don't want to pick up signals from tags on adjacent tracks. For example, using 34mm tags increases the effective range but also increases the possibility of triggering neighbouring readers that are insufficiently spaced apart.
Ideally, the reader should be mounted under the baseboard, out of sight.
Depending on the thickness/density of the baseboard material, this may not be possible.
Other alternatives include mounting it immediately under the track (removing some of the baseboard material), at the side of the track, or above the track in certain circumstances.

The picture shows an extreme case where the sleepers have been removed so that the reader can be as close to the rolling stock as possible.
User experience suggests that each reader should be allocated at least a 6" square area to itself – preferably more.
Other approaches include:

- Mounting readers slightly offset from the centre, with two adjacent tracks, to maximise the separation of the readers.
- Placing shielding materiel round the reader, to minimise unwanted pickup.

If you are using DDC on your layout, it would be advisable to minimise potential interference between DCC wiring and reader cabling. This could be aided by careful routing of the cables.
Also, as the reader connections are 3-wire (+5V, 0V and data), it is possible to use a screened single pair cable for the connections. 7/0.2mm screened wire can be used here.

Reading speed

Clearly, there is a trade off between the speed of the tag, the distance between the tag and the reader, the efficiency of the tag (usually its size) and the sensitivity of the reader used (e.g. ID-12 or ID-20).
The reader takes about 400mS to process a single read, so the reader has to be allowed that amount of time before having another tag passed over it. This determines the maximum reading speed for any given setup.

- If the tag is whisked past the reader too quickly, it may not be read properly.
- There is no such thing as reading tags 'too slowly'.
- Leaving a tag stationary over a reader, or moving it slowly across a reader, does not result in continuous readings. A tag has to leave the reading zone and re-enter again before it will be read again.
- If tags in rolling stock are not spaced sufficiently apart, both tags may be in the reading zone at the same time, preventing any being read successfully.
- Six inches to 8 inches has been suggested as a suitable distance between tags in different items of rolling stock. However, this might vary depending on the size of the tags and the efficiency of the reader. The 12mm tags used with the ID-12, for example, present a smaller reading zone than 34mm tags with an ID-20 reader.

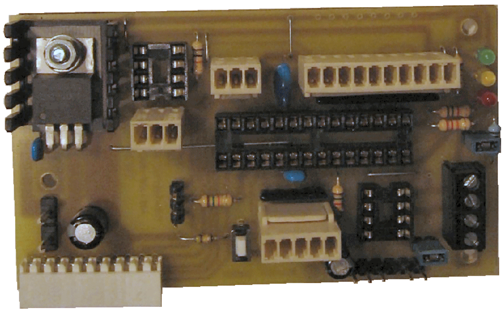
Concentrators

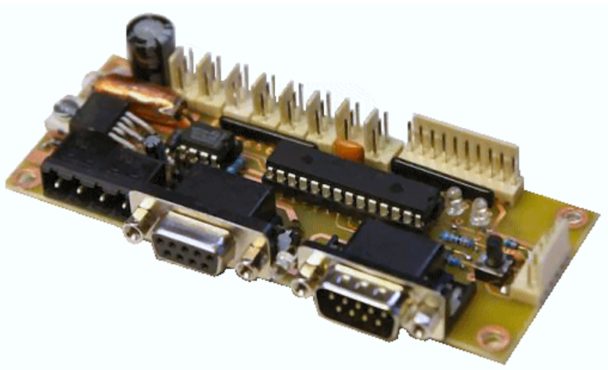
So far, we have looked at connecting a single reader. In practice, a layout will require multiple readers and we can't allocate a separate serial port for every reader.
The answer is to use a 'concentrator'. This module accepts the input from eight different readers and produces a single serial output.
The data from individual readers are still able to be uniquely identified and processed (see later).

Kit 33a is MERG's complete concentrator kit or Kit 33b can be purchased to upgrade an existing serial output board into a concentrator module.
The MERG concentrator has a serial output and connects to a computer via its serial port, or USB port is a serial-to-USB converter adaptor is used.

When you need more than eight readers, you can wire extra concentrators to other serial/USB ports on your computer. The MERG concentrator kit includes 'bank jumpers' that allow multiple concentrators to be used with each being individually recognised by the software.
Clearly, for large layouts, there will be many readers and concentrators, resulting in extra cables between the layout and the computer and added complications in the computer software that handles them. Also, of course, you will eventually run out of computer ports.
Another approach, if you are a CBUS user, is to feed the outputs from the concentrators into the single CBUS cable.

A possible such interface option is the CANRP164, which allows reader output to be fed into the CBUS system.
There is no kit or PCB in the Kit Locker for this module but Bob Vetterlein placed the documentation for this module in the 'Files' section in Yahoo.
See the can_rpc folder in the 'Files' area of the MERGCBUS Yahoo group.





Finally, in cooperation with the Rocrail team, Peter Giling has developed a range of hardware kits and PCBs that match the MERG CBUS specification. It includes an 8-channel interface between RFID reader modules and CBUS.
It is known as the CAN-GC4 mergCBUS 8 Channel RFID and more details are available at:
http://www.phgiling.net/

The cables that connect the readers to the concentrators can be of a considerable length, providing a little care is taken with routing the cables away from power and DCC cables.
A quoted maximum "guarantee" limit of 15m has been stretched in practice to two or three times that distance which is more than enough for even the largest of layouts.

Reader output

The final output from a reader is a stream of RS232 serial data in ASCII format and this can be read by any terminal program. The terminal program has to be set to read at 9600 Baud, 8 data bits, no parity and 1 stop bit to match the format of the reader's output.

Older versions of Windows provided an application called 'HyperTerminal'.
This handy utility was dropped in more recent versions such as Vista or Windows 7, although you can install HyperTerminal on these newer versions if you have a valid CD of Windows XP. There are full details available on the internet if you Google it.
Instead, I use a free application called PuTTY.exe. You can download it at www.putty.org

The output format

Each time a tag is read, the reader module sends out 16 ASCII characters out on the RS232 connection. In fact, the actual ID is made up of only 40 bits (ONs and OFFs) but is packaged in a way that allows easy transmission and reception. Apart from the first and last byte, the contents of all the other bytes have an effect on the screen content.

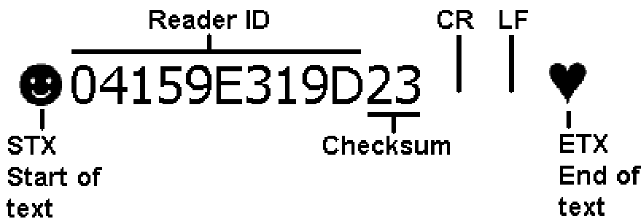
STX	Reader ID	Checksum	CR	LF	ETX
1 byte	5 bytes (40 bits)	2 bytes (16 bits)	1 byte	1 byte	1 byte
'smiley face'	10 characters	2 characters	Carriage return	Line feed	'heart'
Start of text (02h)	The ID shown as ASCII text	Check shown as ASCII text	Go back to start of line (0Dh)	Go to new line (0Ah)	End of Text (03h)

The STX byte informs the terminal program that a string of data is ready to be processed.
The 5 bytes of the reader ID are converted into 10 readable ASCII text characters.

Note

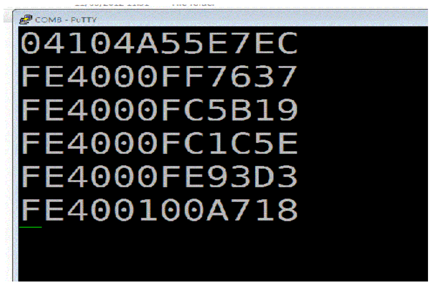
All the ASCII characters are in the range 0 to 9 or A to F. Why? Because these are the values for hex notation. It makes for simple conversion from bit patterns to hex characters. The 40 bits of the ID are split into ten groups of four bits (four OFFs and/or ONs). So, for example, if a group was 0011, then that represents the value 3 and the ASCII character '3' is transmitted. If the a group was 1011, that represents B in hex and the ASCII character 'B' is transmitted.

The illustration shows the construction of a typical ID.



The STX is displayed as a 'smiley' face, while the ETX is displayed as a heart shape.
In practice, using HyperTerminal, the effect of the CR and LF result in the ETX character appearing at the start of the next line, as shown in the next illustration.

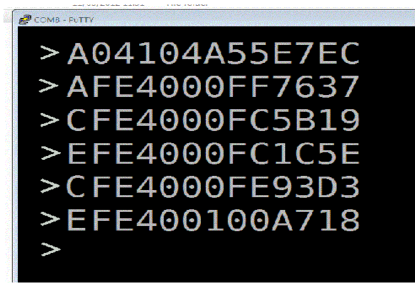
The checksum is generated by adding together all the values of the reader ID and sending this value as the checksum bytes. The terminal program also creates a checksum from the receiver ID. If both checksums are identical (it is called the XOR checksum method) then the terminal program knows that the ID has been received correctly, with no alteration due to interference, etc.
The CR and LF bytes are used to ensure that the next ID read appears on the following line on the monitor screen.
Finally, an ETX byte is sent to let the terminal program know that this is the end of that transmission.



This illustration shows the output as seen on the PuTTY screen. Note that it ignores the STX, CR, LF and ETX characters and only displays the 10 ID characters and the 2 checksum characters.

Concentrator output

The above illustrations and screenshots show the output from a single reader.
Where multiple readers are in use, the output has to indicate not only the tag ID but the reader that is sending the ID.
This is achieved by amendments to the ID string.
A concentrator can handle 8 readers and it recognise input 1 as 'A', input 2 as 'B', up to input 8 as 'H'
The STX character is then replaced by the appropriate letter.
So, for example, if tag FE4000FF7637 was read from input 4, the beginning of the string is changed from STX FE4000FF7637 to DFE4000FF7637.



The ETX character at the end of the string is replaced by the > character.
Now every character has an impact on the screen display; all the characters are printable, with the exception of the carriage return and line feed.

Software

Many existing layout control applications already support RFID input. These include RocRail, JMRI, Tcc and SSI. Users can use existing scripts, or create their own scripts, to handle a variety of tasks, including visual display updates, point switching, etc..
MERG members have also produced applications that are available for download from the MERG website. Phil Herbert describes his Train Announcer program in the Winter 2010 Journal.

Mark Riddoch has written two programs.
One uses tag inputs to display different pages from a PowerPoint slide show (great for impressing visitors at exhibitions). The other is “an Excel application based around the concept of providing a timetable sequence to a fiddle yard operator. The RFID readers are used to feed information to the operator as to the train movements.”
Although the most expensive and complex of systems, RFID offers the greatest scope for layout control; scan previous Journals and the MERG website for plenty of discussion and suggestions.

Chapter 13

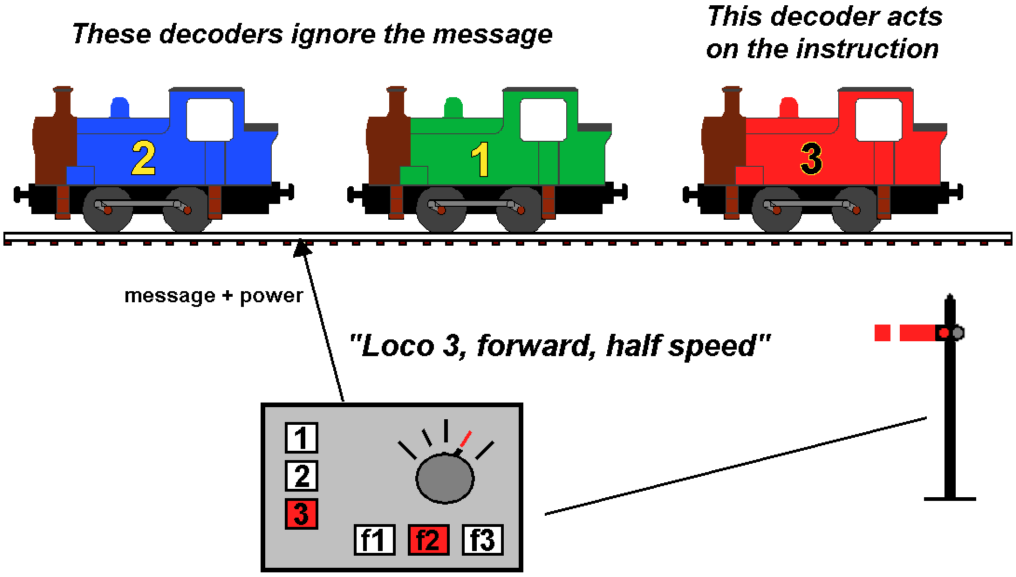
Digital Command Control

DCC - Digital Command Control

Chapter 2 on Loco Controllers covered the basics of DCC.
Let's start with a quick recap before we look deeper.

How does DCC work

Unlike DC, which varies the voltage to the track to control motor speed, DCC has the track powered at full voltage all of the time. Altering the voltage to the electric motors takes place within each individual loco. This is achieved by fitting a little circuit board, called a *'DCC decoder'*, between the motor connections and the loco's wheel pickups.
A DC controller only sends out the power to the track. A DCC controller, called a *'Command Station',* also sends out the power. Additionally, it sends out commands to the various decoders that are connected to the track Usually the decoders are fitted to the locos but others can be fitted at the trackside, to control points, lights, signals, etc.



The illustration shows three locos on the same section of track, being operated from a single command station. They could be actually anywhere on the layout, since all the track is regarded as a single piece.
Each loco is fitted with its own decoder, each with a unique identifying address (1, 2 and 3 in the example). Decoders only respond to messages that are sent to its unique address, ignoring all other messages. When a message is received that matches its address, the instruction from the command station is acted upon. The message could be to go faster, go slower, stop, reverse, etc.
In the example, the operator wants loco 3 to drive forward at half speed, so he/she selects loco 3 on the command station and turns the speed control to half way forward. The command station converts those movements of the speed control into an electronic message that is sent out along the track. Although the message is received by all the loco decoders, only the decoder with address 3 acts on the message.

A similar process takes place when the operator wants to move a point or accessory. The accessory decoder for that device also has its own unique address.
In the example, the command station has a set of function buttons and pressing function 2 operates a semaphore signal.

With accessory decoders, the message is not picked up from the loco's wheels and pickups. The decoder is permanently wired across the track near where it is to be used. Although the decoder will be receiving messages via the track, it may require additional power for running the accessories it is attached to.

Its also worth repeating some of the benefits of using DCC

- Simplified wiring: no complex block/section control with switches and/or relays; fewer wires, fewer faults.
- Simplified working: multiple locos on same track section; consists.
- Improved running: PWM; fewer problems with voltage drop.
- Motor tuning: set the min/max voltage, acceleration/deceleration rate .
- More than loco motors: point, signals, lights, sound.
- Computer control easier: control locos; don't have to worry about block switching.

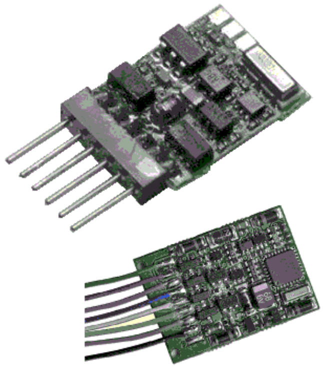
Comparison of DC and DCC systems

	With DC	**With DCC**
Track voltage	Varies from 0V to max (12V?).	Always at max voltage.
Power availability	Only on sections that are switched.	All track powered at all times.
Multiple trains on same track at the same time	No	Yes
Need to split track into blocks	Yes, for anything but the simplest layout.	No, unless you want a track occupancy system, or a very large layout.
Multiple operators	Each operator needs own power controller.	Entire layout controlled by a single command station. Each operator has their own CAB controller that links to the command station.
Run consists	Only with some difficulty.	Yes
PWM running	Depends on controller used.	Yes
Changes to loco	None.	Need to fit encoder in each loco.
Voltage drops	Can be a problem.	Less of a problem.
Handles bemf	Depends on the controller used.	Depends on the decoder used.
Motor tuning	No	Yes
Points and accessories	Need separate wiring.	Can use DCC accessory decoders.
Computer control	Generally complicated.	Easier.
Locos run straight from the box	Yes	Needs to be given a unique number.

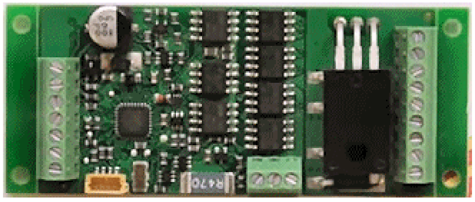
For a working DCC system, as we have seen, we need a command station to send out the commands and a decoder that will act on these commands.

So, let's start by looking at decoders in some more detail.

DCC decoders



Locos draw different amounts of current depending on their size. An N gauge motor may only consume a quarter of an Amp, while large scale motors may consume a number of Amps.
Accordingly, DCC decoders are produced with different current handling capabilities, at different prices.
For example, the Lenz Silver Mini decoder only handles 0.5A, while the Lenz Gold Maxi decoder handles 3A.
Remember to choose a decoder that can handle the loco's stall current, not just its running current.
As expected, the decoders that handle larger currents are also bigger in size. The Lenz Silver Mini is only 1mm x 9mm. The Lenz Gold Maxi, shown in the bottom image, is 70mm x 28mm.

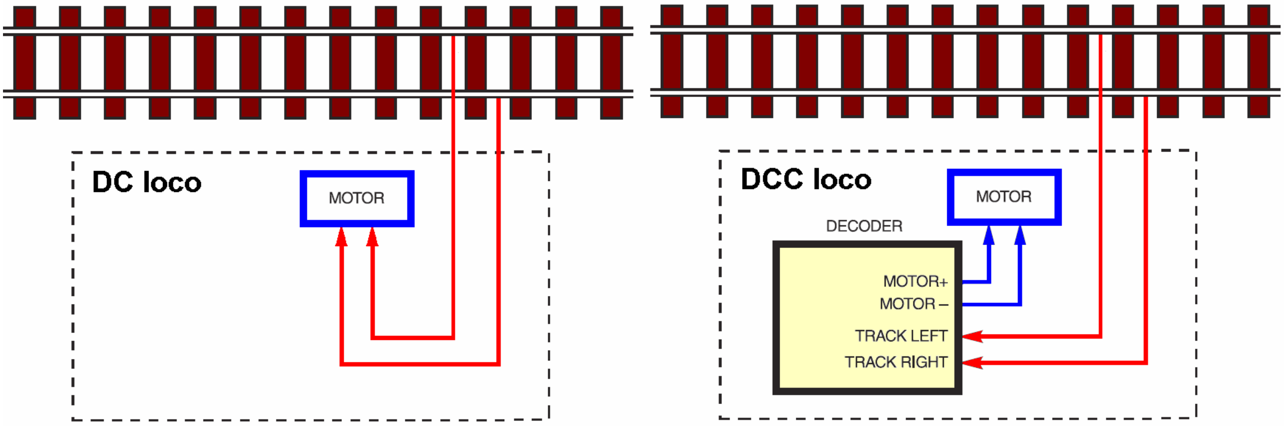


A basic loco decoder is able to control the speed and direction of the motor. Most also have function outputs – connections to switch lights on headlights or lights in the cabin.

Wiring decoders to locos

You can buy locos with decoders already installed in them. There is nothing inside the loco that you need to touch, although you might want to tweak the decoder to maximise performance (this is carried out externally - see later).
DC locos can be converted to DCC working by installing a decoder.
If you are wary about tackling the conversion, or have a difficult installation (e.g. a loco with a non-isolated frame), you can use a fitting service to do the job for you. Of course, this adds more expense to the conversion.

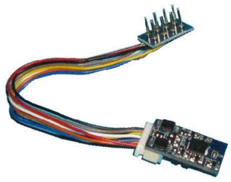
These illustrations show what is involved in upgrading a DC loco to DCC.



The left hand drawing shows a DC loco with the power from the track being wired directly to the motor. The right hand drawing shows the track power only reaching the motor through the electronics on board the decoder. This is the electronics that controls the motor speed and direction.

There are three types of decoder connection, and you can see two of the images at the top of the previous page.

- Wired – You have to solder the wires yourself.
- Plug – The end of the decoder has connection pins that plug into a socket.
- Harness – Like the plug version, only with a wire extension (as in this image).



Installing Wired Decoders

If you can solder, you can fit your own decoders. There are only a few things to watch out for:

Space

Finding space inside a loco body can be a problem, especially for smaller gauges. While there is often sufficient room inside a large diesel, an N gauge 0-4-0 poses a bigger challenge. Fortunately, smaller gauge locos use motors that draw less current and can use the smaller decoders. In extreme cases, the decoder can be fitted inside the loco's coal tender or a following coach/wagon.

Isolation

Some, specially older, locos had a split chassis. The left and right hand sides of the chassis were isolated from each other, but were used to carry current from the wheels to the motor.
Ensuring isolation from chassis. This direct connection between wheel/chassis/motor tag has to be interrupted to allow the decoder to be wired in. A quick test with a multimeter will determine whether your loco uses a split chassis and requires alterations to isolate the motor from the chassis.

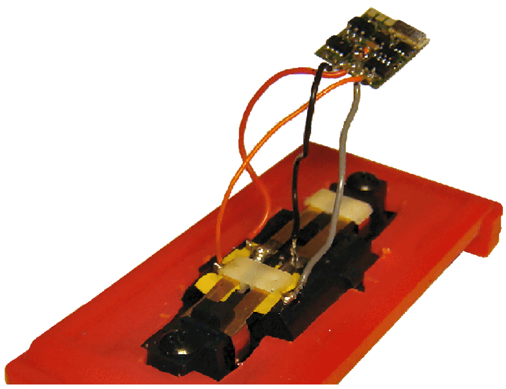
Following colour codes

The wires that emerge from a decoder follow an agreed code, as laid out in the NMRA Recommended Practice RP9.1.1.
This means that whichever decoder you buy, you should be confident that the main connections are identical with other decoders.

The four main wires, available on all decoders, are:

Red	Goes to the loco's power pickup from the right hand track rail, when in forward travel.
Black	Goes to the loco's power pickup from the left hand track rail, when in forward travel.
Orange	Goes to the motor's positive brush.
Grey	Goes to the motor's negative brush.

The image shows a decoder wired to a Kato chassis. In some loco's, the motor is fitted in a chassis in such a way that the orange and grey leads need to be reversed, otherwise the loco would run backwards. Alternatively, the decoder's CV can be altered, as discussed later.

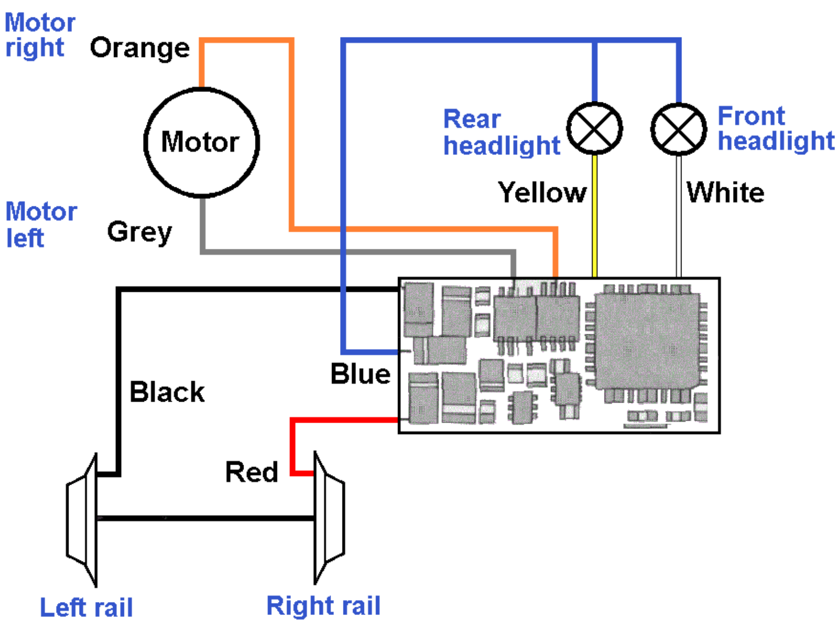


> **Note**
> Take great care to avoid the red and black wires shorting to the orange or grey wires, as this will almost certainly blow the decoder's main chip.

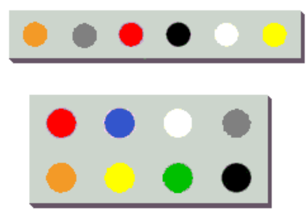
The other decoder wires may include:

White	Goes to the forward headlight.
Yellow	Goes to the rear headlight.
Blue	This is always positive and is the common return for the lights and any other functions. Not provided on a 6-pin socket; available on 8-pin sockets.
Green	Provides an extra function output (if the decoder supports it). Not provided on a 6-pin socket; available on 8-pin sockets.

Here is the wiring diagram for a decoder that provides basic motor and headlight functions.



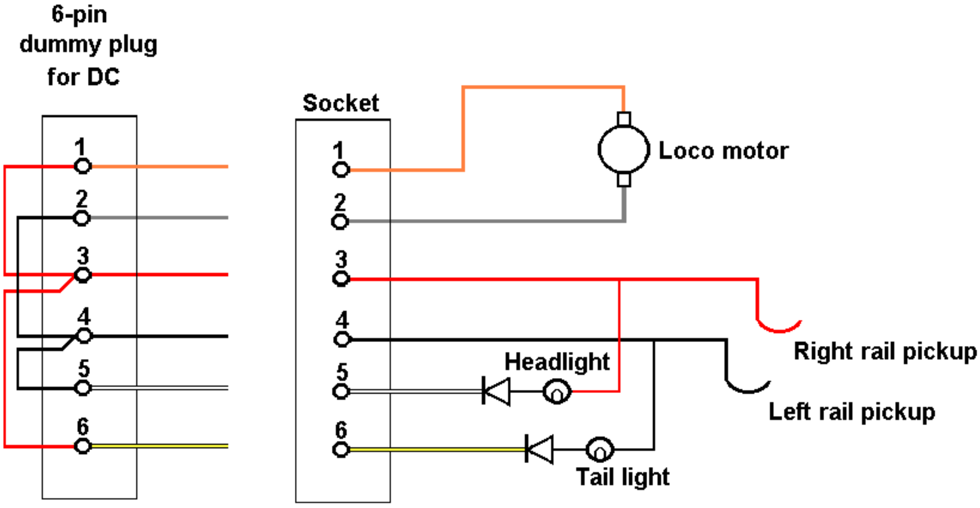
Plug-and-play decoders



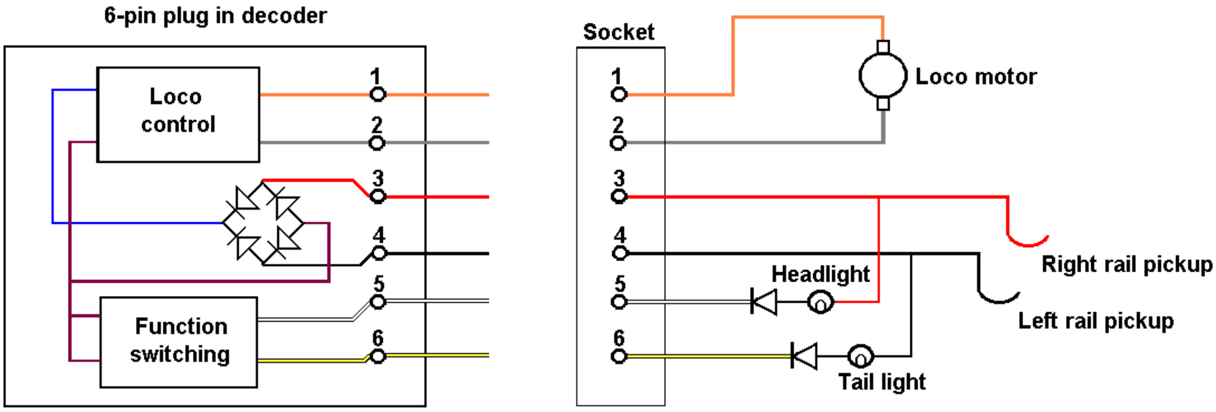
The decoders that have pins or harnesses are used in *'DCC Ready'* locos.
A DCC ready loco is one that has the wires from the track pickups, the motor and the lights all connected to a socket. The illustrations show the two most popular socket types – the 6-pin and the 8-pin.
The 6-pin socket is designed for N gauge motors and is rated at 500mA continuous (750mA peak).
The 8-pin socket is rated at 1.5A continuous (3A peak).

Not all decoders match this specification, so check before buying.

In DC mode, a 'dummy plug' sits in the socket, as shown in the following diagram.



This allows the loco to function as a normal DC loco. If you trace the wires, you will see that the motor and lights are simply wired in parallel across the track pickups.
For conversion to DCC running, the dummy plug is removed and the decoder's pins or harness is plugged directly into the socket.
The loco is now a DCC loco, with no soldering involved.
You will have noticed that the 6-pin socket has no blue or green connections.
Since the blue wire is used as a common return for the headlights, this method is not available. Instead, the lights are taken to the track pickups. That's OK for DC, but for DCC they only work on half wave. This reduces the light's brightness which some overcome by reducing the value of the LED's series resistor.
This illustration shows how a 6-pin decoder interfaces with the loco wiring.



The DCC power from the track pickups are connected to a bridge rectifier in the decoder, which provides a DC voltage for the decoder's circuitry. The decoder shown has circuitry for controlling the loco speed and direction, along with additional circuitry to switch headlights.
An 8-pin version would use its blue connection to carry a positive voltage across from the decoder into the socket, where it would act as a common return for the two headlights.

Note

DCC sockets and dummy plugs are available as separate items for purchase. You can buy a socket and wire it yourself, to create a DCC Ready loco. That way, you can use the loco on a DCC layout or a DC layout by simply swapping the decoder and the dummy plug.

Back to Contents Page 5

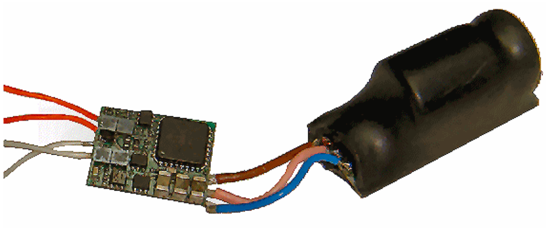
Stay-alives

You can never guarantee that a loco will always get the power it needs from the track, without any losses whatsoever. Problems caused by dirty track, dirty wheels, dirty pickups, insulfrogs, short wheelbases, etc. can lead to power glitches. In extreme cases, the loco just comes to a halt. Temporary power dropouts can lead to the loco lurching or a sound decoder's sound dropping out or starting to repeat from the beginning.
Gradually, more decoders provide for the additional of an on-board power store. This stores power during normal running and powers the decoder and loco during short track power glitches. A common demonstration shows a loco running over a length of track that is covered with an insulating material. Check out:

www.tcsdcc.com/public_html/Customer_Content/Web_Tools/Videos/Main.html

to see an N gauge loco fitted with a TCS stay-alive running over an 18"section of unpowered track.
A heavy OO gauge 16-wheeled diesel, with its large wheels and multiple power pickups can span across dirty track sections or a point frog. On the other hand, an light N gauge 0-4-0 with its much shorter wheelbase, smaller wheels and fewer pickups would benefit the most from fitting stay-alives.

Commercial versions



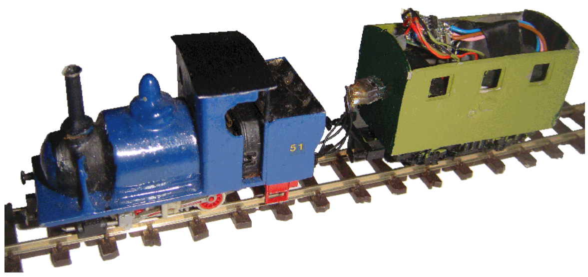
In some cases, the stay-alives are sold as optional extras, while others are sold as a matched pair.
The image shows a Lenz Gold Mini fitted with its optional stay alive, known as a 'Power 1'. An effective, although expensive, combination.
Alternatives are the TCS KAT24 mentioned above and the DCC concepts *"DCC-S4SAP 4 Function 8-Pin S Series Decoder With Stay Alive"* available from eHattons.com. Bear in mind that commercial stay-alives have different storage abilities, with some providing much shorter support; check before you buy.
Zimo decoders allow you to buy and add your own capacitors, rather than having users tied to branded storage products.

DIY

If you have the experience, the skills and the confidence, you can make and fit your own stay-alives and save money. The website at:

www.members.optusnet.com.au/mainnorth/alive.htm

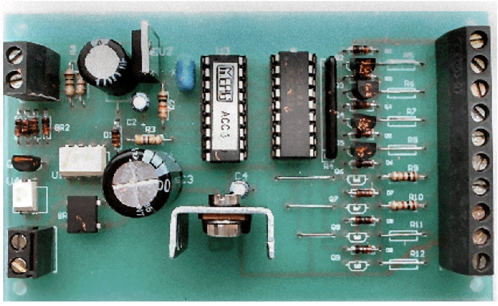
provides details on how to tackle a range of different manufacturers' decoders.
A Google search for *"dcc decoder hacking"* also provides details on this.
Remember, this will almost certainly invalidate your decoder warranty.



There is a bit of a predicament here; the smaller locos are the ones to most benefit from stay-alives, yet they are the ones with the least space to house them.
While it is usually possible to install a decoder inside a small loco's body, the stay-alives are much larger.
The image shows a decoder and stay-alive installed inside a coach, with wire links between the coach and the loco.
This works but means that the coach must always be attached to the loco.

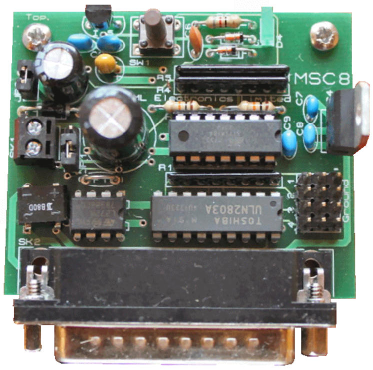
Accessory decoders

Accessory decoders also connect across the track and get their instructions from the command station (e.g. put a light on, switch a point, move a semaphore arm). Some accessory decoder outputs can handle lights or Tortoise motors directly, using the power derived from the DCC on the track. For more power-hungry devices (e.g. solenoid motors), the decoder would need a separate power supply.



This image shows the MERG Steady State Accessory Decoder kit that has 8 outputs. These outputs can control low power devices such as LEDs and Tortoise-type point motors (when used in pairs).

Examples of commercial versions are the DC10 and DC20 from CML Electronics, the Lenz LS100, the Zimo MX82E, the Digitrax DS64, the ESU SwitchPilot 4.



Other accessory decoders are designed to control servos. For example, the ESU SwitchPilot Servo module and the Team Digital SMC4 module both control up to 4 servos.

The LDT TT-DEC-B kit is specifically designed for controlling a turntable.

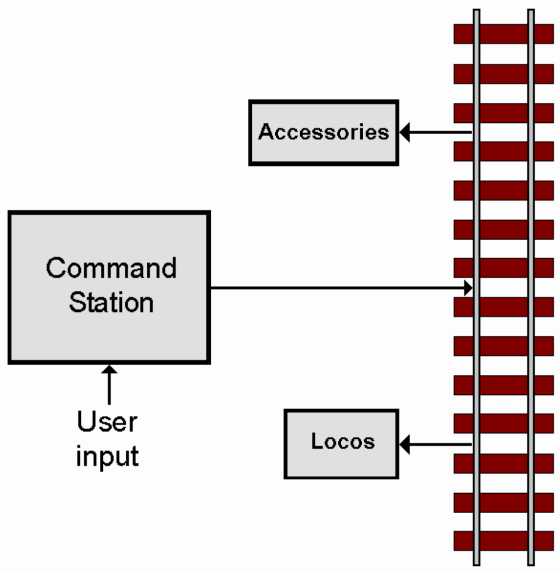
CML sell the MSC Scenery Animator, shown in this image. It can control a mix of devices including motors, servos and lights. It also supports creating sequences and 'flicker' effects.

All the accessory decoders have one thing in common- they all are controlled by the messages issued from the DCC command station.

Wiring for DCC

Chapter 3 explained that DC layouts split the track into multiple sections/blocks using lots of cuts in the track. Each section has to be individually powered using switches (or sometimes relays).

The chapter also explained that for a one-operator DCC layout running a few trains, there was no need for track sections. The layout could be reasonably large and complex and still be run with a simple wiring scheme. That scheme involved running power round the entire layout, to every bit of track, allowing locos to go anywhere on layout without any section switching.



This illustration shows how a simple DCC layout operates. The whole layout is controlled by a single command station. Its output is permanently wired to every piece of track on the layout. The locos get their power and operating instructions from the track.

Commercial command stations

Command stations are also called DCC controllers, DCC control units and DCC throttles by various manufacturers.
They are available from many manufacturers including Digitrax, ESU, Hornby, Lenz, NCE, Viessmann and Zimo. These offer a range of facilities and cover a wide price range

At the lower end of the price range is the Hornby Digital Select, shown in this image. It is an entry level command station and comes with a separate power supply that provides 1A to the track. Locos are selected using the keypad and speed is controlled using the central knob. The direction of the loco is selected by pressing one of the arrow keys on the keypad. It provides basic functions such as setting a locos' unique address, its acceleration and deceleration. It can control up to 10 locos simultaneously, although the 1A maximum clearly limits the number (without extra power being added).
A similar product is the Digitrax Zephyr Starter Set, with a maximum output of 3A.

At the other end of the price range is the ESU ECOS 50200. It can control up to 16384 locos simultaneously, as well as 2048 points. It has a 7" colour touch screen display and the operator can use it like a mimic panel.. The user's layout is displayed on the 800x 480 pixel screen and points can be changed by touching them with a stylus.
Without any additions, it only provides 4A, which is not a lot at the price being asked.
The Viessmann Commander uses a similar approach with the user's layout displayed on a 7" 800x480 pixel colour touch screen, and 1.5A continuous current without a booster. The The ESU supports operating shuttles, while the Commander supports full automation if required.

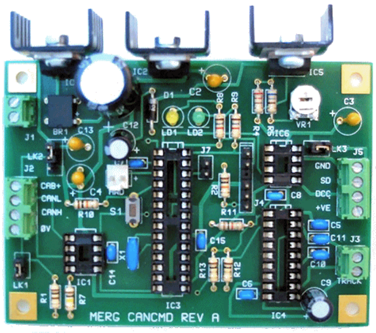
Both of the above expect the user to operate from wherever the command station is sited.

For those who prefer the flexibility of moving around a layout while operating, manufacturers provide handheld controllers. These may plug into an existing command station or be matched with a less glamorous box containing the power unit.
The image shows an example of this – the Lenz Start Set 100, which has an output of 5A.
Other similar products include the modestly-named Digitrax Super Empire Builder and the Super Chief, both working at 5A. The NCE Power Pro 10A Large Scale Set, as the name suggests, is intended for O Gauge and above.

For even more flexibility, some command stations get round the restrictions of having a handheld tied to the layout by a cable – by providing wireless communication between the handset and the power unit. The operator is free to move anywhere round the layout without having to find a socket to plug in the hand controller. Example are the Digitrax SCFXD8 8A system which uses infra-red communication (perhaps not the best choice for a link) and the soon to be released Zimo MX 12A system based on a radio link.

MERG command station

MERG has two kits that provide a cheaper alternative to the commercial offerings.



The image on the right shows the CANCMD kit. This command station module provides up to 1A but can be connected to a booster for extra power. It connects to MERG's CBUS system (see the next chapter), which allows it to be controlled either by a suitable handheld controller or by a computer. It provides the power and the instruction messages to the track.

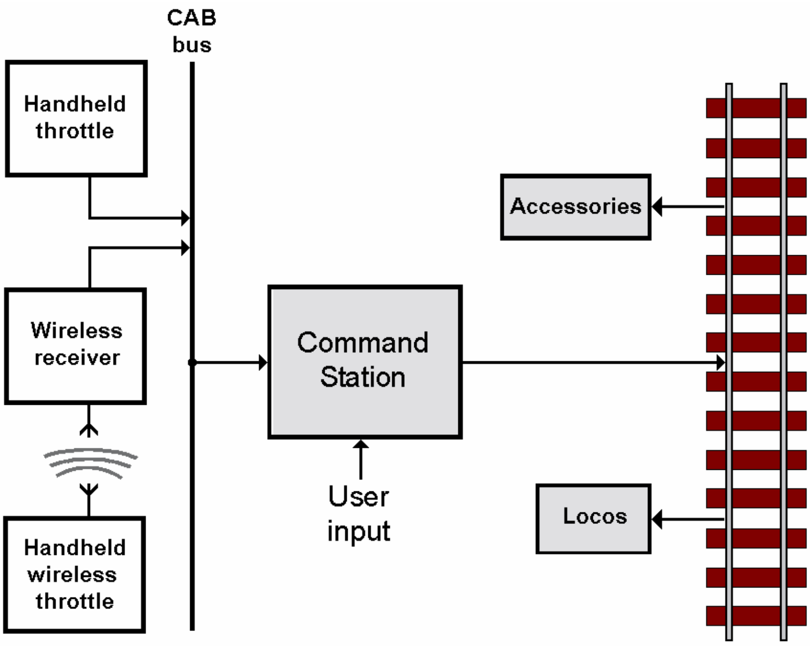
This image shows the CANCAB kit, a handheld controller that connects to the CANCMD. It can control up to 32 locos simultaneously, along with points and accessories.

Locos are selected using the keypad and speed is controlled using the thumb-operated knob on the top-left of the case. The direction of the loco is selected by pressing a button and LEDs indicate the current loco direction.

Pressing the STOP button once, stops the loco that is currently being controlled. Pressing the button a second time stops all locos on the layout.

Multiple operators

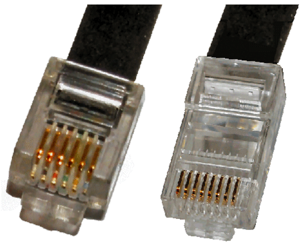
There may be circumstances when you want more than one person to be operating a layout. Examples are club layouts, large home layouts, or the grandkids coming round. Fortunately, many command stations allow you to connect more than one hand controller. The illustration shows three operators sharing a command station. One is using the command station controls, while the others use a wired throttle and a wireless throttle.



Although there are multiple operators, there is still only a single DCC command station which controls the entire layout. The operators press their buttons, etc. and these actions are sent as messages to the command station. The command station translates these messages into the instructions that are sent to the track.

The connection between the handheld controls and the command station is known as the *'CAB Bus'* or *'Throttle Bus'*. The way the bus is wired is often specific to a particular commercial DCC system, as there is no NMRA standard for CAB buses. Most manufacturer's modules are not interchangeable (e.g. CabBus/NCE, LocoNet/DigitTrax). Some use Expressnet (e.g. Lenz, Atlas, ZTC, Hornby, Roco – although they call theirs RocoNet).

To allow the maximum flexibility of operation, it is common to wire the CAB Bus round the edges of a layout, with sockets at strategic points. The operator can then plug a handheld controller into the nearest socket, while moving round the layout.

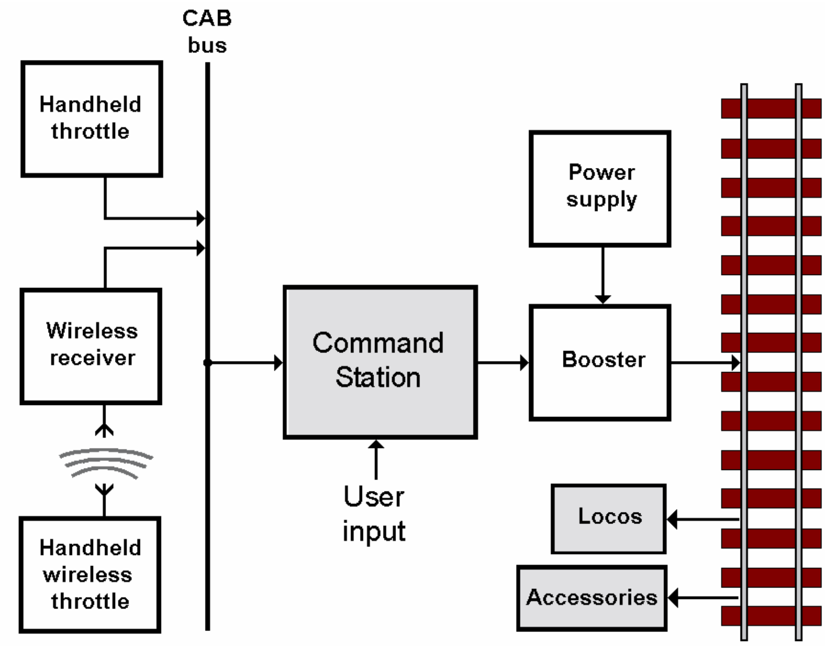


CAB cables commonly terminate in an RJ12/RJ25 plug or a RJ45 plug. The RJ25 plug, shown on the left looks like a plug you find on the end of a telephone, but it is different. The wires are straight-through (no inversions) and it has six contacts. These are found on Lenz, Hornby, NCE, XpressNet/Loconet Atlas, Roco, etc. handhelds. The RJ45 plug is the type used for Cat 5 Ethernet cabling. It uses duplicate connections for increased reliability. Used by MRC, Gaugemaster, etc, handsets.

Although not shown in the illustration, the CAB bus can also be used as a means of controlling a layout using a computer or a control panel.

Handling high currents

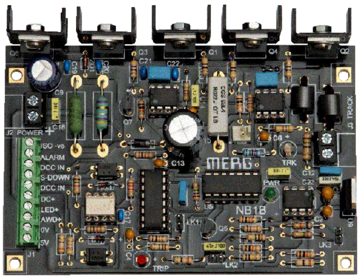
As your layout grows, you may find that the 1A or 3A command station you own is not sufficient to meet the current demands of your expanding fleet of locos and accessories. If you are working in larger gauges, or running multiple locos, you can expand the capabilities of your existing command station by adding a *'power booster'*, sometimes also known as a *'power station'*. Many command stations provide for this improvement.



The illustration shows that the booster provides the track power and requires a beefed-up power supply to provide the higher amperage. The booster also receives the messages from the command station and send them out on the track.

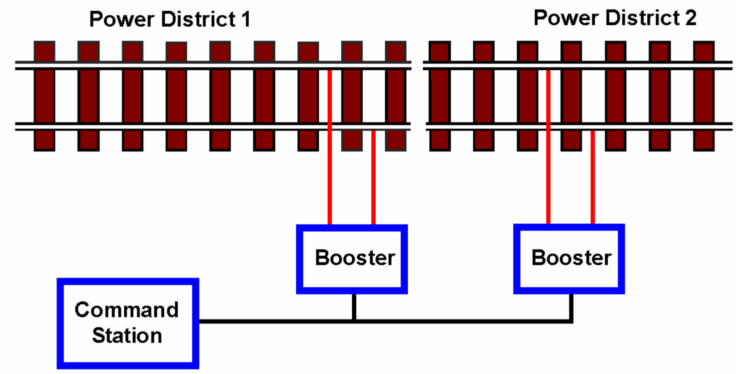
In all other respects, the setup runs as before.
Locos still move freely throughout the layout without any block or block switches.

A power booster unit is usually housed in a plain case, as can be seen in the earlier image of the Lenz Starter 111 Set. Digitrax sell 5A and 8A boosters, while NCE have a 10A version.
The image shows the MERG NB1B Booster, which can provide an output of 5A or 10A depending upon the components that are used. It connects to the MERG CANCMD command station.



Large scale DCC

When we look at *really* big layouts, the ones with many large power-hungry locos running at the same time, they need even more power than a 10A booster can handle.
The answer is to use extra boosters. Unfortunately, its not that easy, as multiple boosters can't be wired in parallel with each other.
The answer still lies in using multiple boosters – but with each booster providing power to a *part* of the layout.
For the first time, we have to break the layout up into independently powered blocks, known as *'power districts'*.

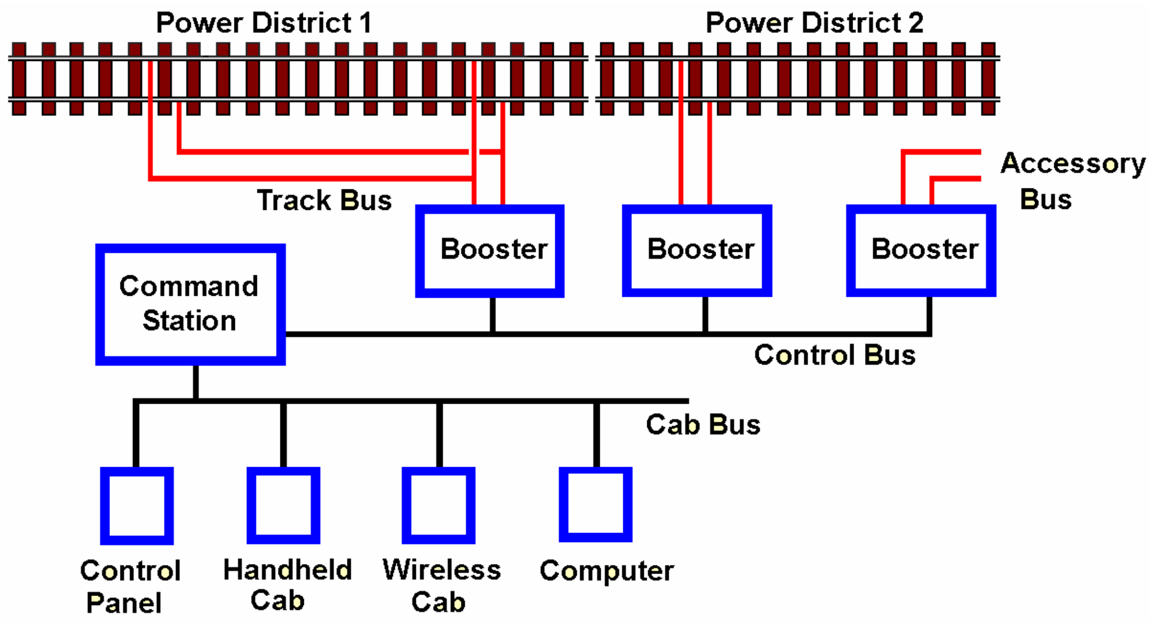


This is different from DC block working in a number of important respects;

- All pieces of track are still powered at all times.
- There is still a single command station that controls the entire layout. The instructions from the command station are fed to all boosters and from them to all power districts.
- Locos can still move freely between districts, without the need for switches or relays.

It makes sense to separate a layout into distinct work areas (e.g. goods depot, marshalling yard, mainline station, etc.) with an operator in charge of the loco operations within each work area. Since each work area can expect heavy traffic, it is allocated its own power booster. That way, power districts match the main activities of the layout and the command station and its wiring does not have to carry the total current of the entire layout. Power districts result in more complex wiring. On the other hand, having separate identifiable sections can help help in tracing layout faults.

Here is a possible configuration for a large layout.



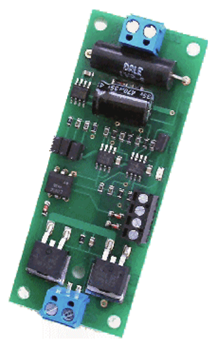
In this example, the layout's DCC accessories require too much power to connect them to the track. In such a case, a separate cable can be taken from a booster and fed directly to the accessory modules. This is known as an '*Accessory Bus*'.
Bear in mind that power districts are not required for most layouts – although there is still a case for creating independent track sections for feedback purposes (see the chapter on track occupancy detectors).

Handling shorts

Despite our best efforts, short circuits are a common occurrence on model railways, mostly caused by loco derailments. Short circuit protection is an absolute must to protect the layout and to avoid the serious risk of fire.
You may have wondered why we bothered to have multiple power boosters instead of one giant power booster. We could, of course, build a 50A booster, instead of using five 10A power districts. Now, consider what would happen if there was a short with a booster running at 20V. Power, you will know by now, is voltage times current. So the power that would be dissipated in event of a short would be:

$$P = I \times V$$
$$P = 50 \times 20 = 1kW$$

In other words, a short would produce the same heating effect as a single-bar electric fire!
No wonder, then, that command stations and boosters provide:



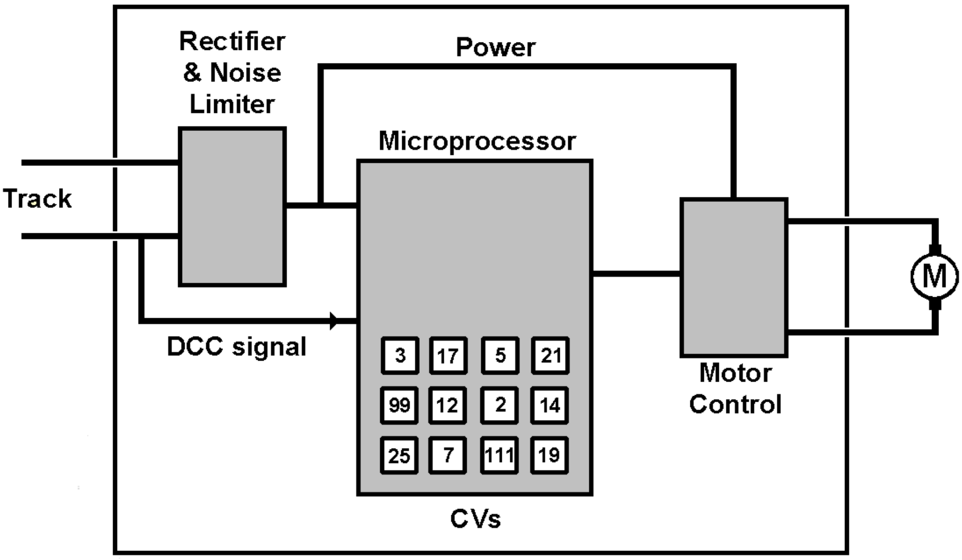
- Short circuit protection – closing down the power when an excessive current is detected.
- Thermal protection – closing down the power when the components inside the command station or booster are detected as overheating.

Most cutouts spring into effect after about 50mS. They then test every 0.5 secs to see whether the short is removed and power can be restored.
The image shows the NCE EB1 circuit breaker.
Chapter 5 looked in detail at DC and DCC overload protection.

Back to Contents Page 5

Setting up decoders

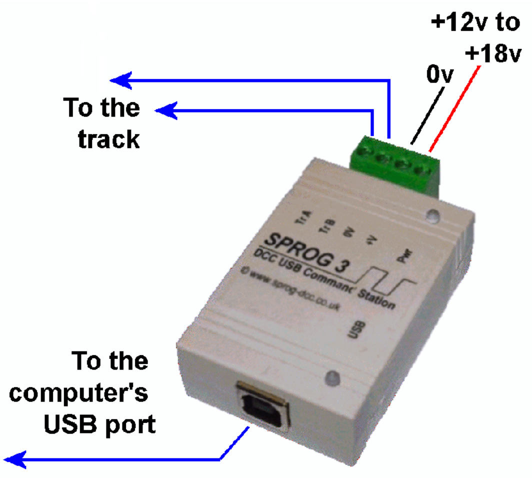
A decoder is supplied with some default settings. For example, the unique address is always set to 3. If you want to run other DCC locos on your layout then the decoders have to be given new addresses. This is called 'programming the decoder'.
Apart from addresses, decoders can store all kinds of details regarding the loco's performance – minimum and maximum motor voltage, acceleration, deceleration, lights, and much more. All these performance factors are stored inside the decoder and can be altered by the user. These values are known as CVs (Configuration Variables). This illustration shows some of the internal workings of a decoder.



The loco's motor, lights, functions, etc. are all controlled by a microprocessor on the decoder. The chip contains a number of separate memory locations, each one able to store details of the way that particular loco is configured.
They are called 'variables' because they are alterable and may well be different for different locos (each decoder would certainly have to store a different address). By altering the values stored in the CVs, you can tweak each loco for maximum performance.

Programming hardware

Most command stations can program the contents of a decoder, although it can often be a laborious process requiring multiple button presses and some knowledge of all the CV numbers and their options. Command stations may have two outputs, with a separate output for Service Mode programming (see next).



A popular alternative is to use the SPROG DCC decoder programmer, as shown in this image.
You connect a power supply and the track to the terminal block and use a USB cable to connect the unit to a computer.
For more details on the SPROG, check out:
www.sprog-dcc.co.uk

Wiring for programming

There are three ways that you can connect your programmer and decoder.

Ops Mode

This is often referred to POM (Programming on the Main).
The output of the command station or programmer is fed to the main layout track.
This means that a loco can be anywhere on the layout and still be programmed.
Even better, the alterations can be carried out while the loco is moving. So, you can check the effects of altering loco parameters (e.g. max speed, matching locos for consists, adjusting sound levels, etc.).
This mode works by specifically sending the new data to a specified loco address.
An example Ops Mode instruction could be considered as

"Loco 7, change your acceleration momentum to 109"

All the decoders in other locos should ignore the instruction and carry on as before.

> **Note**
> The programming hardware, the decoder and the software must all support Ops Mode for this to work. Older command stations (Lenz Compact, Atlas Commander, Digitrax UT4 Throttle) and older decoders may not support POM. Check before using POM. Unless you are absolutely sure, consider Service Mode programming.

A couple of other limitations:

- You cannot read and examine the existing stored CV values while in Ops Mode; you need to be in Service Mode.
- Some systems don't allow you change the unique address while in Ops Mode.

Service Mode

The output of the command station or programmer is fed to a length of *'programming track'* that is electrically isolated from the rest of the layout. This could be an unused siding, a length of track in the fiddle yard, or even a piece of track sitting on your workbench. The programming track must not connect to the layout track in any way.
An example Service Mode instruction could be considered as

"Change your acceleration momentum to 109".

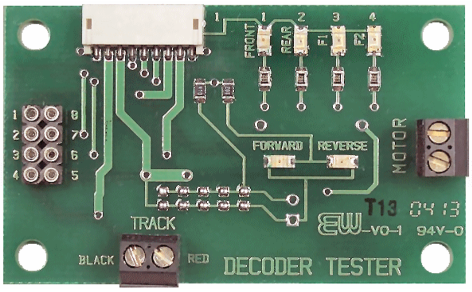
This is called *'broadcast programming'*, as the instruction is not specific to any one decoder address. Since no loco address is specified, all locos on a programming track would have their decoder CVs programmed at the same time.
Your command station may have separate outputs for the main layout and for a programming track.
Although this mode loses the benefit of altering CVs while the loco is in motion, it provides other benefits:

- You are guaranteed not to program other decoders by mistake.
- You can change the decoder's address.
- You can read existing CVs values stored in the decoder. You can only read one decoder at a time, so disconnect the sound decoder to read the loco decoder – and vice versa.
- You can reset a decoder back to its factory defaults (see manuals) if you get totally muddled or get unexpected results.

Test Rig

Some users prefer to use a 'test rig' for their decoders. This is a board with all the connections to a spare motor, lights, DCC power etc. The decoder is either plugged in or wired to these connections. It has similar benefits to working on a programming track. However, the decoder can be programmed and tested *before* fitting it into a loco. That way, you know everything works prior to the fitting.

The image shows a commercial tester, the NCE DTK Decoder Tester, available from Bromsgrove Models, Digitrains, etc. It has connections for both wired and plugin decoders, plus connection to the DCC power (via the track) and a connection for an external test motor. Once wired up, it can test for motor control including direction, plus four functions. The ESU 51900 is similar but has a motor mounted on the board.



If preferred, a homebuilt version is easily made up from a piece of stripboard, some LEDs with resistors, and various connectors.

Programming software

You can download a very useful software application called *'DecoderPro'* which makes the process of setting up a decoder much more user friendly. DecoderPro works with Windows, Linux, Macintosh (OS 8.6 through 9.2 and MacOS X), and OS/2

For a detailed tour and guide to DecoderPro, check out:

http://jmri.sourceforge.net/help/en/html/apps/DecoderPro/

or look at tutorial videos at:

http://model-railroad-hobbyist.com/mrht_decoderpro

The examples that follow are from DecoderPro.

Decoder standards

Before looking at programming a decoder, we should be aware of the standards expected from a decoder. The NMRA (National Model Railroad Association) presents itself as the lead body on helping to set standards with the model railway industry.

In fact, the NMRA produce both *'Standards'* and *'Recommended Practices'* in relation to many electrical and electronic issues – including DCC decoders.

Standards are mandatory, if a manufacturer want to claim NMRA compliance for their products. For example Standard S-9.1 describes the electrical standards for DCC (e.g. pulse amplitude, length, etc.) while S-9.2 describes how instruction packets are coded.

For a mine of information, check out:

www.nmra.org/standards/sandrp/consist.html

Recommended Practices cover decoder abilities but, as the title implies, not all its proposals are mandatory.

NMRA do have a 'Baseline' standard for compliance, covering issues such as basic speed and direction, the need to offer up to 127 unique addresses, etc. All manufacturers have to meet this baseline to be able to describe their products as meeting NMRA conformance.

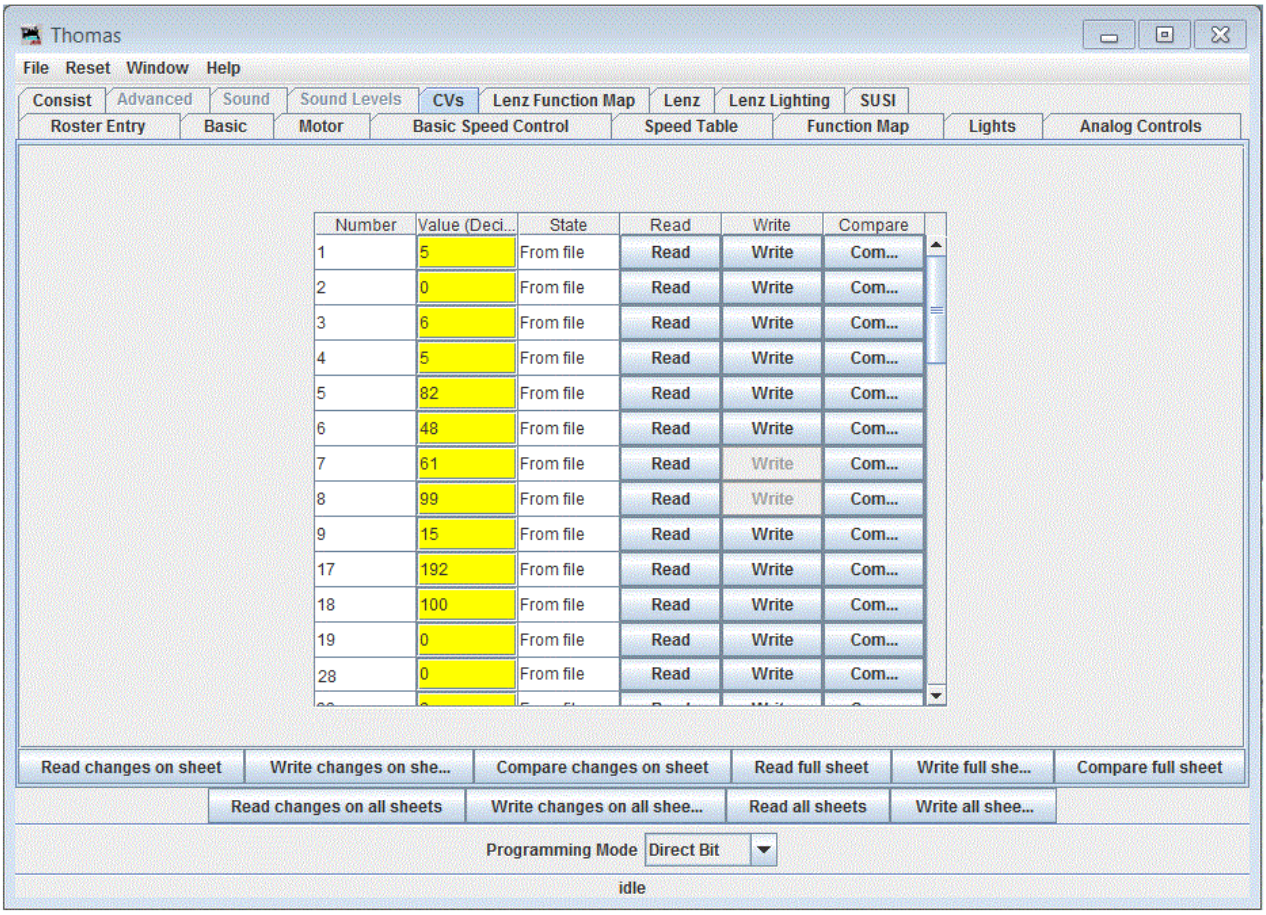
The 'Extended Packet Format' (RP-9.2.1) covers additional features such as offering more than 127 addresses, functions, accessories, consists, etc. Since these are not mandatory, the list of facilities offered in a decoder varies between manufacturers and individual models.

Decoder CVs

Here is a sobering thought - decoders CVs are so useful that there are hundreds of them!
The NMRA standard CV definitions go up to 512, including some for future expansion.
The Lenz Gold Mini, for example, stores 128 CV values.
Inside the decoder, the CV values are stored as a long table containing CV numbers and the values they are currently storing.

The hard way

This image shows only part of the long list of CVs, being viewed on a DecoderPro screen.



This screen can be used to alter the contents of any of the CVs as long as:

- you know what each CV number is used for
- you know the effect of any value you change.

Over a period of time, with regular usage, you might start to recognise the main CVs and their likely settings.   Fortunately, there is an easier way.

The easy way

This chapter does not include a tutorial on using DecoderPro, as there is a lot of material already available on that subject, including the links provided earlier.
However, a number of screenshots from that package are shown to demonstrate some of main CVs that hobbyists alter or tweak.

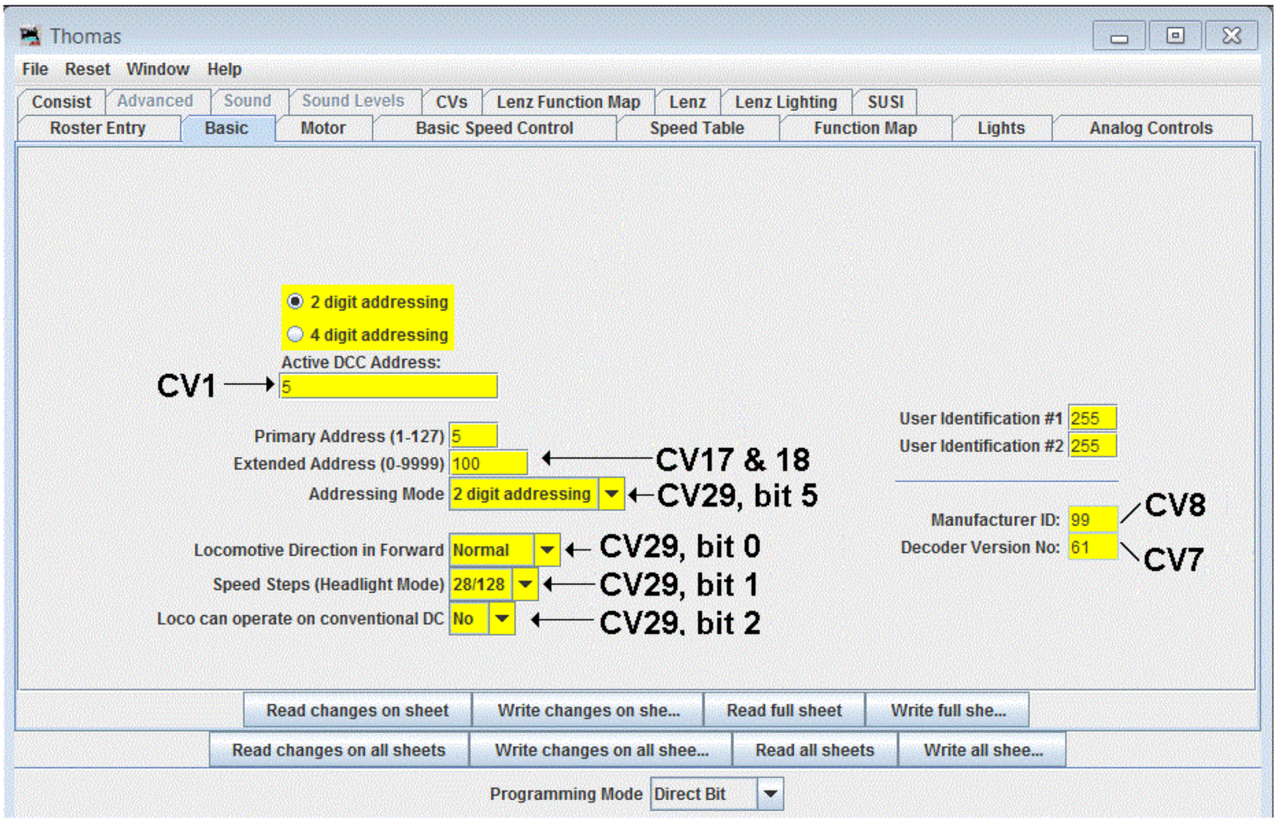
So, if there are hundreds of possible CVs, how do you know which ones are usable with your particular decoder?

Back to Contents Page 5

There are two ways to go:

- Look up the tables provided in the manual that come with the decoder. This will list the CVs supported by your decoder. The hard way.
- Let DecoderPro interrogate your decoder, find out what make and model it is, and only show you the CVs that you can view and alter. The easy way.

Here is a screenshot of a decoder, using the *'Basic'* tab.



You will notice that all the entries are in plain descriptive English; there is no mention of CV numbers. They are hidden from the user, so there is no need to know the CV that tweaks each parameter.
In fact, two of the CV locations (CV8 and CV7) store the make and model of the decoder. When DecoderPro reads these CVs, it knows what that particular decoder is capable of and only shows you its variables.
In the example, CV8's stored value is 99, which mean that it was made by Lenz, while the value of 51 in CV7 means that it is a 4th generation model that is capable of handling BEMF. But you don't need to know this, as Decoderpro figures it out for you.

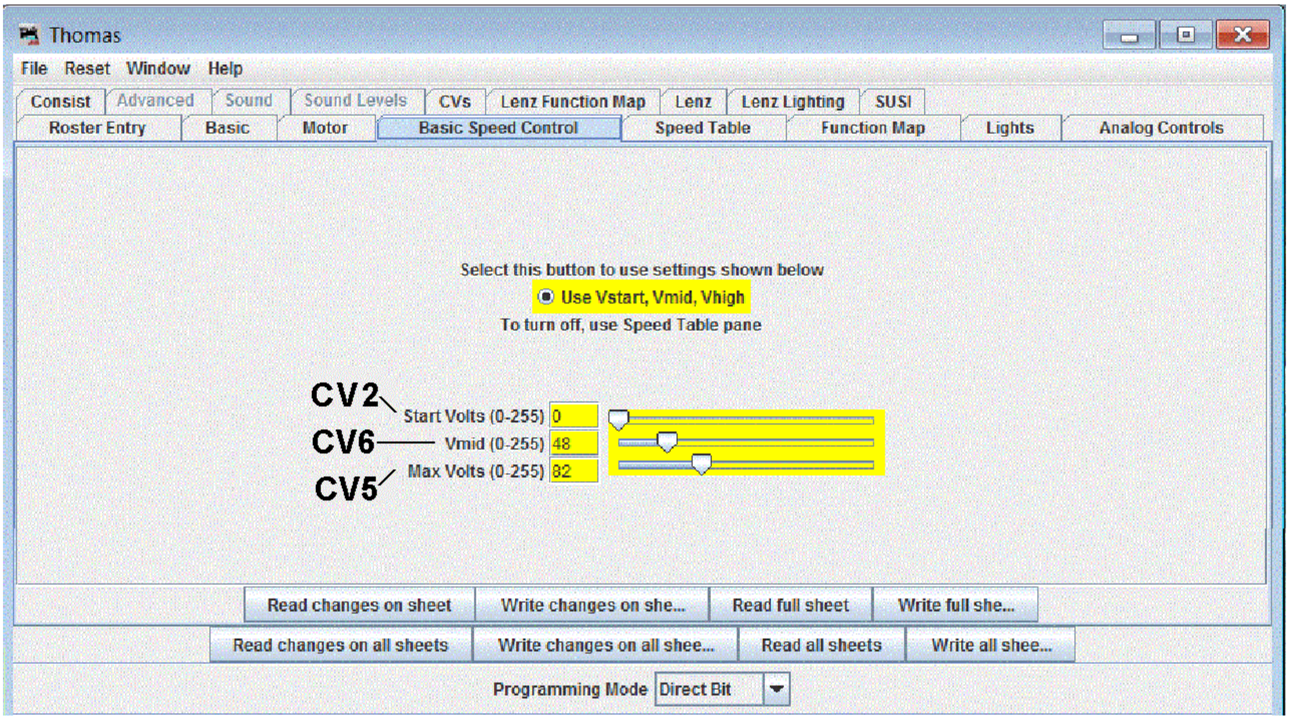
CV1 stores the current unique address of the decoder. This is the one that is factory set to 3 when purchased and needs to be altered when using multiple locos.

You may have noticed that some CVs have a numeric value, while others are simply storing *'On'* or *'Off'*. Each CV can store a byte of data, which means values between 0 and 255. That is a waste for on/off values, so some CVs split the byte up into 8 separate bits. Each bit can then store a 1 for On and a 0 for Off. Again, this is all invisible to you.
In the above example, CV1 CV7 and CV8 store values, while CV29 uses individual bits.

CV29, bit 0 is handy if you have wired your decoder's motor wires incorrectly and the loco goes in reverse when you want it to go forward. You can either dismantle the loco and rewire the motor – or alter bit 0 so that the error is corrected in the decoder's electronics.

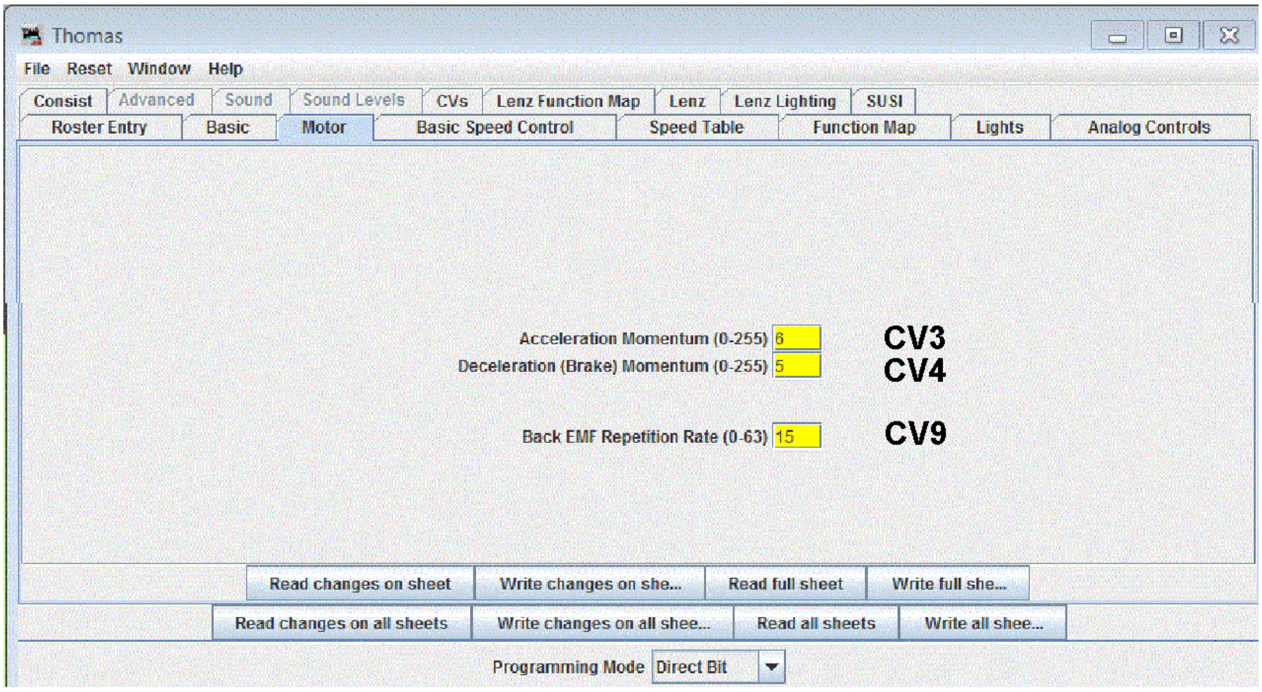
If you want the loco to be able to run on DC layouts as well as DCC tracks, you set CV29, bit 2 to Yes.

Here is DecoderPro's *'Basic Speed Control'* screen:



CV2 sets the loco's minimum working voltage, so it stores a byte of data. The motor is supplied with this voltage as soon as the operator advances the speed control to the very first notch. There is no need to turn the speed knob part way round, just to get the loco moving. **CV5** sets the loco's maximum voltage that will be supplied to the motor. So, for example, a shunter will be prevented from racing round the yard.

The screenshot of DecoderPro's *'Motor'* screen shows three options.

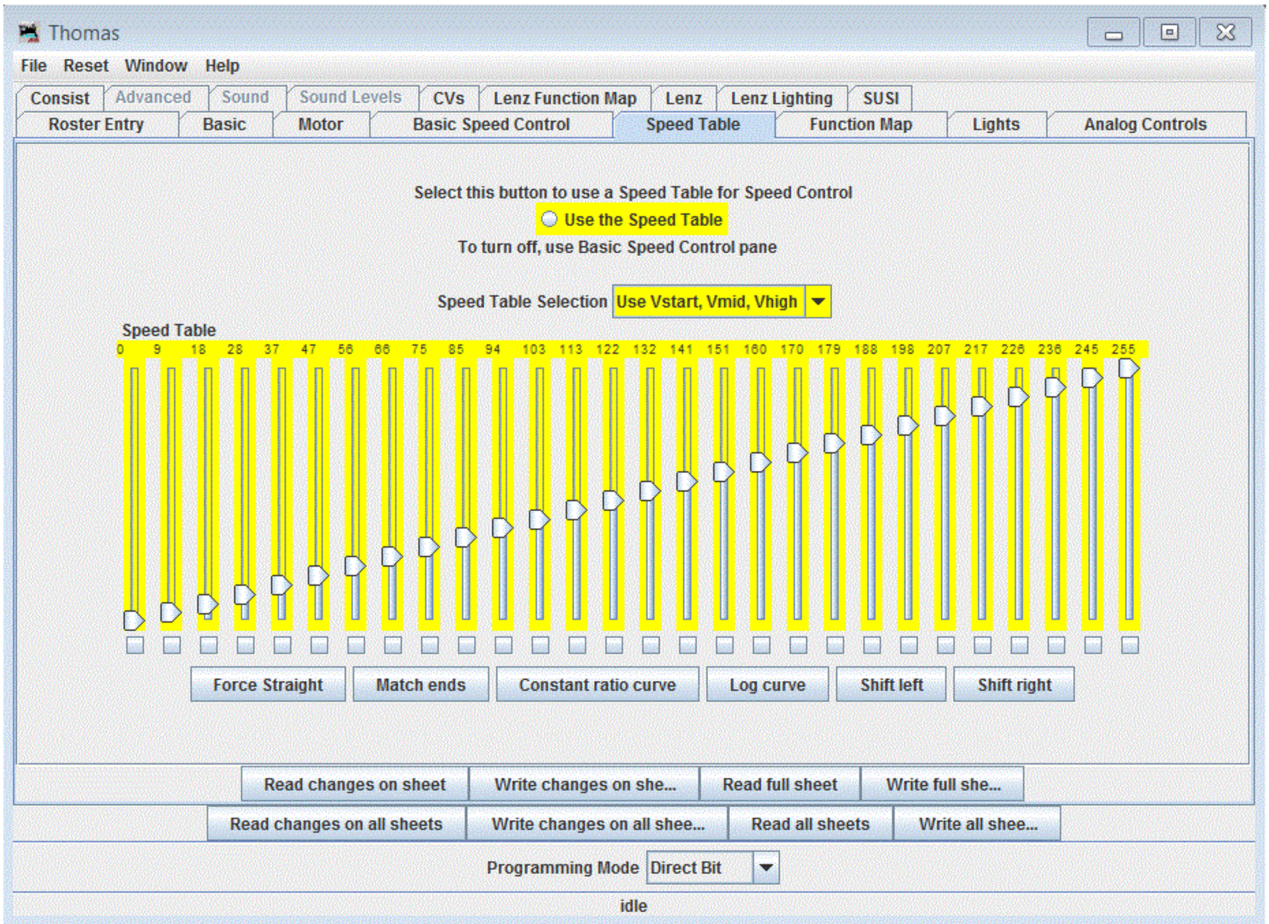


CV3 sets the acceleration momentum. If you suddenly jerk the speed control higher, the loco will not take off like a bullet. Instead, it will gradually accelerate until it reaches the speed set by the command station. The rate of acceleration is controlled by the value stored in CV3 and may be different for different trains. For example, a small 0-4-0 pulling a long rake of coal wagons would be set to a slower than normal rate of acceleration.

CV4 sets the deceleration rate. Trains jerking to a halt is not very realistic, so CV4 lets you set the rate at which the train comes to a halt. Even if you suddenly reduce the controller's speed to zero, the train will slowly brake to a halt. Again, the value set will depend on the loco is an express train or a shunter.

CV5 sets the BEMF repetition rate for maximum performance for that loco. See the chapter on Loco Controllers to remind you of how BEMF works. This option will not be present if your decoder does not support BEMF.

Here is a screenshot of DecoderPro's 'Speed Table' screen.



It is used for ultimate control over the speed of the loco for different setting of the command station's speed control. It is mostly used to match the performance of multiple locos, to allow smooth-running consists.

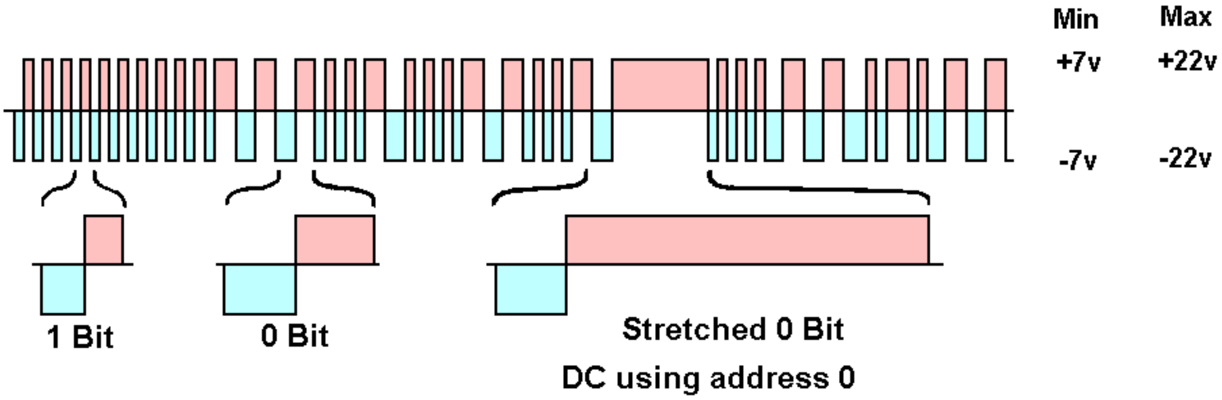
As you alter CV values, they are temporarily stored by the software. When you are finished altering values and setting, you write all the changes away to the decoder.

Remember, these are only samples to give you an idea of the process. Read the tutorials for more details.

The remainder of this chapter is for those who are curious about the inner workings of DCC. It is not required to use DCC, but gives an insight into the way messages are constructed and handled.

Inside a DCC packet

A DCC command station sends both power and decoder instructions to the track – both at the same time.
The power is delivered as a continuous repetition of pulses, as shown in the illustration.



First it swings positive then negative, in what is described as *'bipolar DC'*. This is passed through a bridge rectifier in the decoder to obtain the DC power that the decoder, motor and lights use.
The decoder's specification calls for a minimum swing of 7V in each direction and a maximum of 22V. Here is a table of common voltages.

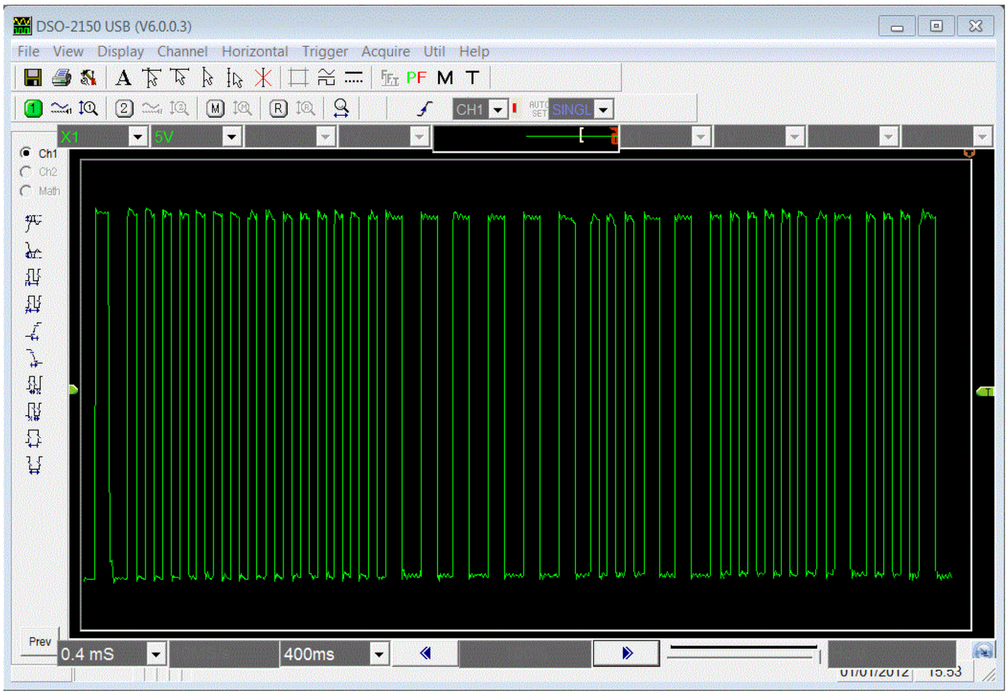
Common DCC track voltages, by gauge			
Z	N	OO/ HO/ O	Large scale
10V-12V	12V-14V	14V-16V	20V-22V

Since the voltage is constantly changing polarity, we can slightly alter the length of time it remains at at any one polarity. In other words, we can alter the pulse width to introduce digital information on the signal.
In the above illustration, we can see that a shorter pulse represents a '1' logic while a longer pulse represents a '0' logic.
A 1 bit last 58usecs (microseconds) positive and 58usecs negative – a total of 116usces.
A 0 bit last 100usecs positive and 100usecs negative – a total of 200usces – although this can be increased to 9900usecs under certain circumstances.
This is the approximate equivalent to a 7KHz signal.
The command station's instructions to change speed, direction, etc. are encoded into a stream of 1 and 0 bits to send out with the DCC signal.

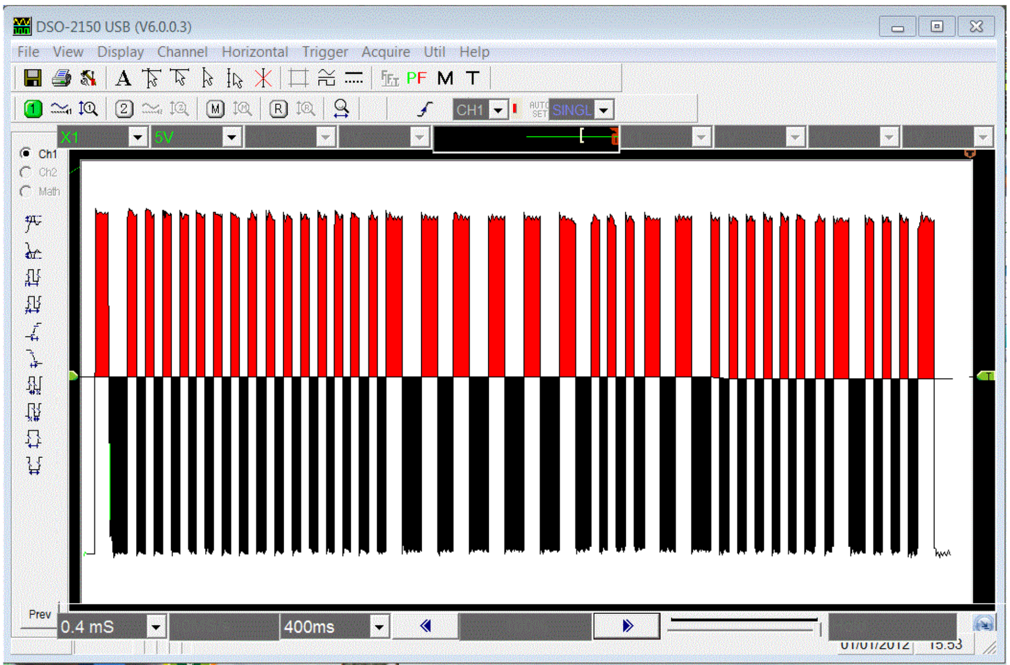
Note

It is possible to run a single DC loco on a DCC layout, although not part of the official DCC specification. You set the command station to an address of zero and moving the speed control results in the 0 bit being stretched so that it appears more like a DC signal.

This image is an example of an actual DCC signal, as captured by an oscilloscope and shown unedited.
This is the signal that is sent out of the command station and this is what is received and processed by the DCC decoder.

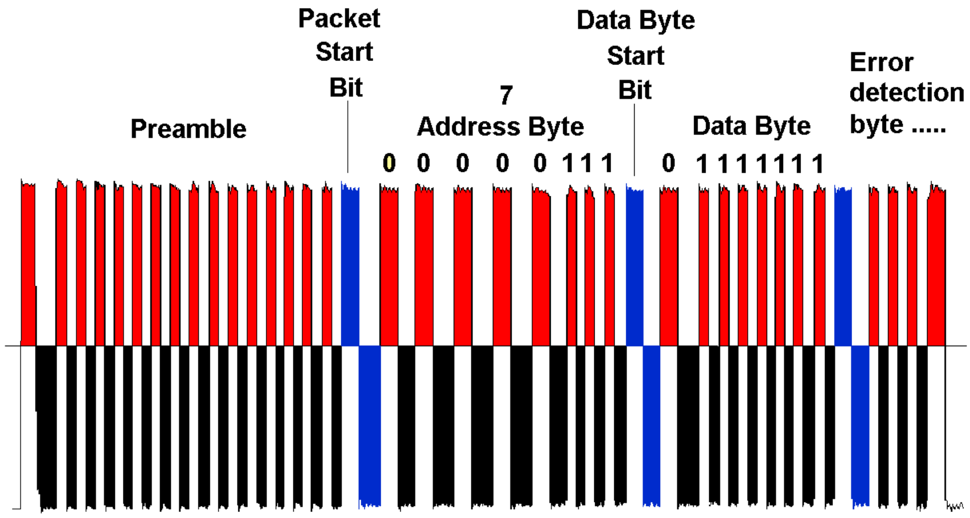


And here is the same image, edited so that the positive and negative transitions are more clearly identifiable.



Although it looks a bit of mess, it contains the DCC instructions, packaged in a way that does not get in the way of supplying the power to the decoder and motor.

This illustration shows how a baseline DCC packet is made up.



Both positive and negative transitions are used to provide power. From a packet point of view, the same information appears in both the positive and negative transitions. That is why the bottom section of the illustration is ignored.

The DCC packet is made up of several different parts:

The Preamble

This is a series of '1' bits that is used to alert decoders that a packet is about to be transmitted, allowing the decoder to synchronis its timings with the incoming signal. There is a degree of confusion on the number of bits required, with 10, 12 and 14 bits being stated from different sources. The answer lies in the NMRA Standard S-9.2:

> *"A digital decoder must not accept as valid, any preamble that has less than 10 complete one bits, or require for proper reception of a packet with more than 12 complete one bits. A command station must send a minimum of 14 full preamble bits."*

The Packet Start Bit

This tells the decoder that the preamble is ended and the message data is about to begin. If the first bit of the message happened to be a '1' bit, the decoder would not know when the preamble ended and the message began. To prevent this, the preamble is always followed by a '0' start bit.

The Address Byte

This byte contains the address of the decoder that is to act on the instruction. All others with different addresses will ignore the rest of the packet. If the address is one byte, it is for decoders in the 1 to 127 range. For address above 127, there will be two address bytes.

The Data Start Bit

This informs the decoder that what follows is the instruction data.

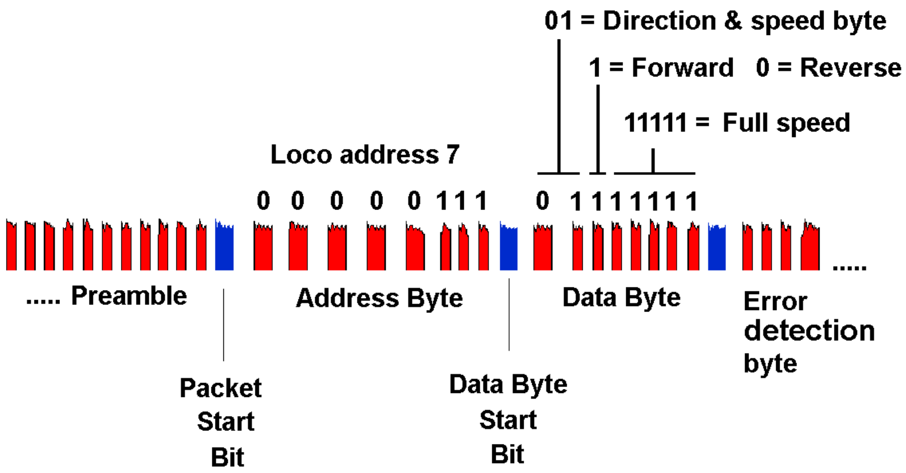
The Data Byte

For simple instructions, there will only be a single byte of data, with two bytes being used for more complex instructions.

The Error Detection Byte

There is the possibility of the incoming message being corrupted by electrical interference or interruption to the packet reception due to dirty track/pickups. To avoid acting on corrupted signals, the command station finishes the transmission of the packet with an error correction byte. This is a checksum on all the data sent in the packet. The decoder calculates its own checksum on the incoming data. If the two checksums are identical, the decoder knows that the packet is genuine; otherwise, it rejects the packet.

This illustration shows the actual message that was contained in the packet.



The Address Byte stored the eight bits 00000111. This is the binary number for 7, so the message was for the loco that is fitted with a decoder with the address of 7.
The Data Byte breaks up its 8 bits into three distinct groups of bits.
The first two bits stores 01 which indicates that what follows is loco direction and loco speed data. The next bit is a 1, indicating an instruction to go forward.
The remaining five bits are 11111, which indicates full speed (00000 being stop).

Handling errors

Not all the instruction packets arrive intact at the decoder end. Packets can get interrupted by dirt on the track or poor wheel pickups. Packets contents can also get distorted due to electrical interference. The error correction byte prevents incomplete or garbled messages from being acted on – but the packet is still lost.
In most circumstances, this is not a great problem, as the command station sends out the packets are repeatedly. So, if there was only one loco on the layout, all the packets would be directed at that decoder address. If there were three locos on the layout, each would receive their own packets on a time-sharing basis.
Given the pulse widths specified in the NMRA documents, a command station could issue up to 180 three-byte data packets per second. If there is only a single loco on the layout, using a baseline working, it is capable of receiving 180 instruction messages per second.
When there are 10 locos operating on a layout, they could expect to receive up to 18 instructions per second.
If the decoders are using extended packet formats, up to 110 five-byte packets can be sent every second.
Unless you are operating an extremely busy layout in a very harsh electrical environment, the loss of packets should not pose any problems.

Back to Contents Page 5

Handling congestion

Although packet losses don't usually prevent decoders from eventually receiving their own messages, the lost packets represent loss of 'bandwidth' – every interrupted or corrupted packet is one less useable time slot. The command station can only issue between about 100 and 180 messages per second and every lost packet reduces the number of valid messages on the system.

Let us consider a bad case of congestion.

A layout is running 50 locos, each with an additional sound decoder.

The 100 decoders have to share the 110 packets per second, which mean that they might expect an average of 1.1. packets per second. If a decoder loses one of its packets, it has to wait about 2 seconds before receiving a duplicate message. For a speeding express, that might mean the difference between stopping and crashing into buffers or points, overshooting a platform, etc. This can be minimised using CV11, which is a packet time-out variable. This can be set to bring the loco to a halt if no commands are received after a preset time. Such drastic occurrences are not likely to happen on most layouts but designers have to consider the worst case.

There are a number of ways to combat congestion, with the main ones being queuing and prioritisation.

Once a decoder receives a command, it acts on and continues to act on that same instruction until it is given a new command. In theory, therefore, a single command should suffice. In practice, as we have seen, periodic repetition of the command ensures that the packet arrives at the decoder.

At any one time, the command station is storing a list of all the packets that need to be sent out periodically. If there are too many additional incoming requests to the command station, there will not be sufficient time slots to send them all. The command station has to form a queue to allow all the packets to be sent eventually.

When there is a queue, the command station has to decide the order in which the packets will be sent out and this order could be constantly changing.

Decisions could include:

- Placing new packets ahead of older packets in the queue, since it is likely that the older packets have already been received by the decoder. This regards responding to changes in the system as more important than the repetition of older messages.
- Placing packets controlling sound and/or functions ahead of speed and direction messages. This recognises that the operator does not make rapid speed changes in any given time period (human movement being slow compared to the speed of the system).

If the system supports feedback, track occupancy sensors would be given a high priority, as they affect a loco's speed and position.

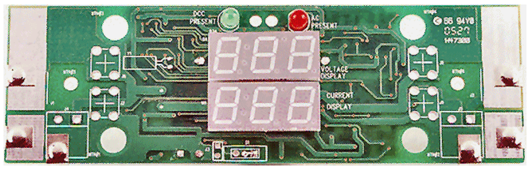
Advanced consisting, also called Decoder Assisted Consist, saves packet bandwidth by having a single command station packet controls all the locos in a consist.

With the use of congestion controls, several hundred locos can be run simultaneously – more than most operators could ever imagine

Measuring DCC voltage

Checking for the presence of a DCC signal is easy enough. You can use a car light bulb, a LED with a dropper resistor, or a multimeter set to the AC range. They will all tell you if DCC is available at the point you test. If you use the multimeter, it will also tell you if you are suffering from any voltage drop at any point, just by comparing the reading with that leaving the command station.

Getting an accurate measure of the DCC voltage is a little more complex than you might first imagine.



You can avoid any problems by using the RRampMeter by DCC Specialities, available from DCC supplies and Digitrains. Although expensive, it has a 2% accuracy.
Just connect the two left-hand terminals to the point to be tested, to get an accurate reading of the DCC voltage at that point. It is designed so that these terminals can be held across the track, making for a quick check on voltage presence and amplitude. It can be used as a hand-held tester or it can be permanently installed in a control panel. It can also read current, as described later.

Expensive models aside, most multimeters are unable to accurately read DCC voltages.

Multimeter on DC range

To measure DC, you just place a multimeter's leads across the controller's output terminals.
However, DCC is more like AC, since it's polarity constantly switches (hence its description of bipolar DC). So, every pulse of one polarity is followed by a pulse of the opposite polarity – cancelling each other out and producing no meter reading on DC.
The only exception to this is to set the command station to address 0. This places the output in DC compatibility mode and provides a reading, although not too accurate.

Multimeter on AC range

Most digital multimeters are designed to read voltages that are derived from the mains supply (i.e. either reading the mains voltage or the AC voltage output from a transformer). They expect AC readings at 50Hz (UK, Australia, etc.) or 60Hz (USA, Canada, etc.) and also expect the waveform to be a sine wave. The meter's electronics makes the necessary *'averaging'* processing prior to displaying the reading based on these expectations.
DCC though, operates at a much higher rate (approx 7KHz) and is not a sine wave. The meter cannot sample at a fast enough rate and this introduces readings that can be anywhere between 10% and 40% different from the true voltage. These readings are, at best, only a guide or can be used for comparison purposes.

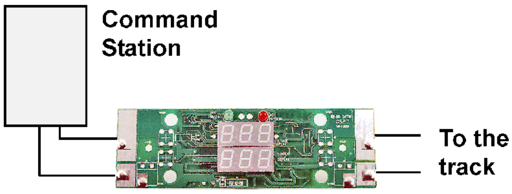
Using a true RMS meter

A meter than can accurately report voltages on waveforms that are not sinusoidal is known as a *'True RMS'* meter. A commercial example is the Fluke 115 for industry or the RRampMeter for model railways.
Unlike the cheaper average responding meters, true RMS meters can sample faster, process better and produce much more accurate readings.

If you know someone with a True RMS meter or a RRampMeter, you can compare the readings with your own meter, note the difference and use it in future meter readings.

Measuring DCC current

Like voltage, getting an accurate reading of DCC current requires using a True RMS meter. The illustration shows the RRampMeter connected between the command station's output and the track. Wired this way, it gives readings of both DCC voltage and current (assuming that there is a load on the track such as a loco to draw current). The meter can read up to 28V and 10A.



DIY projects

The RRampMeter, sells for over £70 at the time of writing, which is around the same starting price for industry's digital True-RMS meters.
If you are looking for a cheaper, homebuilt, alternative, try out the projects that are explained on these websites:

http://www.wiringfordcc.com/track.htm#a4

http://home.cogeco.ca/~rpaisley4/DCCvolts.html

http://www.awrr.com/dccacces.html

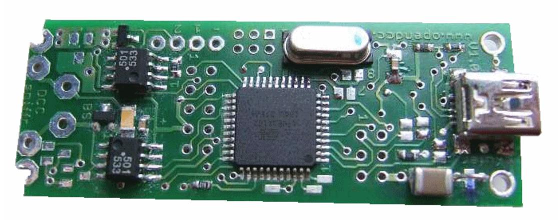
http://www.members.optusnet.com.au/nswmn2/DCC_Meters.htm#Ammeter

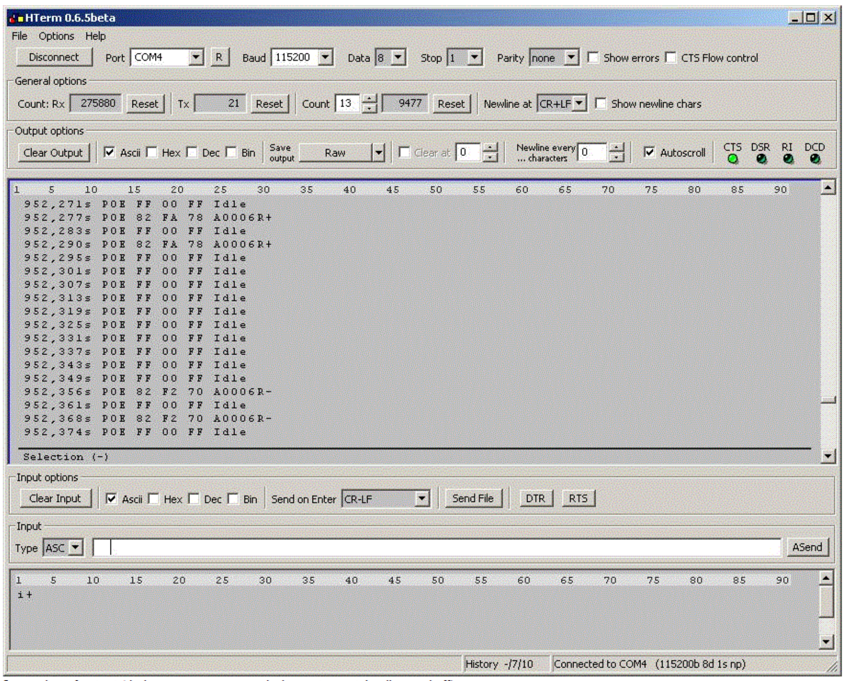
A DCC 'sniffer'

If you are adventurous and feel the need to examine the DCC packets, have a look at:

www.opendcc.de/elektronik/dcc_sniffer/dcc_sniffer_e.html

The module shown connects to the track and plugs into a computer's USB port.
The software description reads:
"This software allows for a complete tracking of all commands on the rails as well gives sophisticated trigger features to capture only special events. Local LEDs provide a quick overview of the status and the actual DCC commands."





Screenshot of output (during accessory command, shown: turnout 6 coil on and off)

Chapter 14

C B U S

An introduction to CBUS

Many excellent layouts would further benefit from improved efficiency and greater functionality if it could successfully tackle a couple of problems:

- The huge amount of wires and connectors used, particularly on layouts built using multiple boards.
- The lack of two-way communication, getting feedback from points, train detectors, automating some features, even adding computer control.

The CBUS 2-wire layout control system was developed by MERG members Mike Bolton and Gil Fuchs, to tackle these issues.

Lets start by looking at some the problems with traditional layout wiring.

The conventional way

These layouts have a central control panel that is used to activate points, switch lights etc. This involves running wires from each switch and pushbutton to each point, light or accessory. On a large layout, this results in masses of wires running under the baseboard. Where multiple boards are used, this collection of wires has to be carried over between baseboard sections, using multipole plugs and sockets.
The sheer number of wires and connections increases the complexity and fault liability of the layout.

The DCC way

The previous chapter on DCC showed how points and accessories can be operated by sending signals along with the DCC power. Accessory decoder modules connect across the track and get their instructions from the command station.
This cuts down on wiring, with the accessory decoders only requiring power from the layout's power bus when handling high-current devices.
In many respects, it is a 2-wire control system, although it is was really designed for one-way communication – from the command station to the decoders.
Although the NMRA recognises the need for bi-direction working (see RP-9.3.1 and RP-9.3.2), it is about the loco decoder sending information back to the command station. Helpful though this is, it does not cater for a range of other needs such as a point or track detector sending back information.
The other limitation is how the accessory decoders are controlled by the operator. While some operators are happy to use buttons on a handheld controller, many would still prefer to be able to operate the layout's points and accessories from a control panel.
They would also like feedback on track occupancy, etc. to be fed back to the control panel.
Another limitation of DCC is that all control passes through the command station and there is no provision for local autonomous activities using the DCC bus.

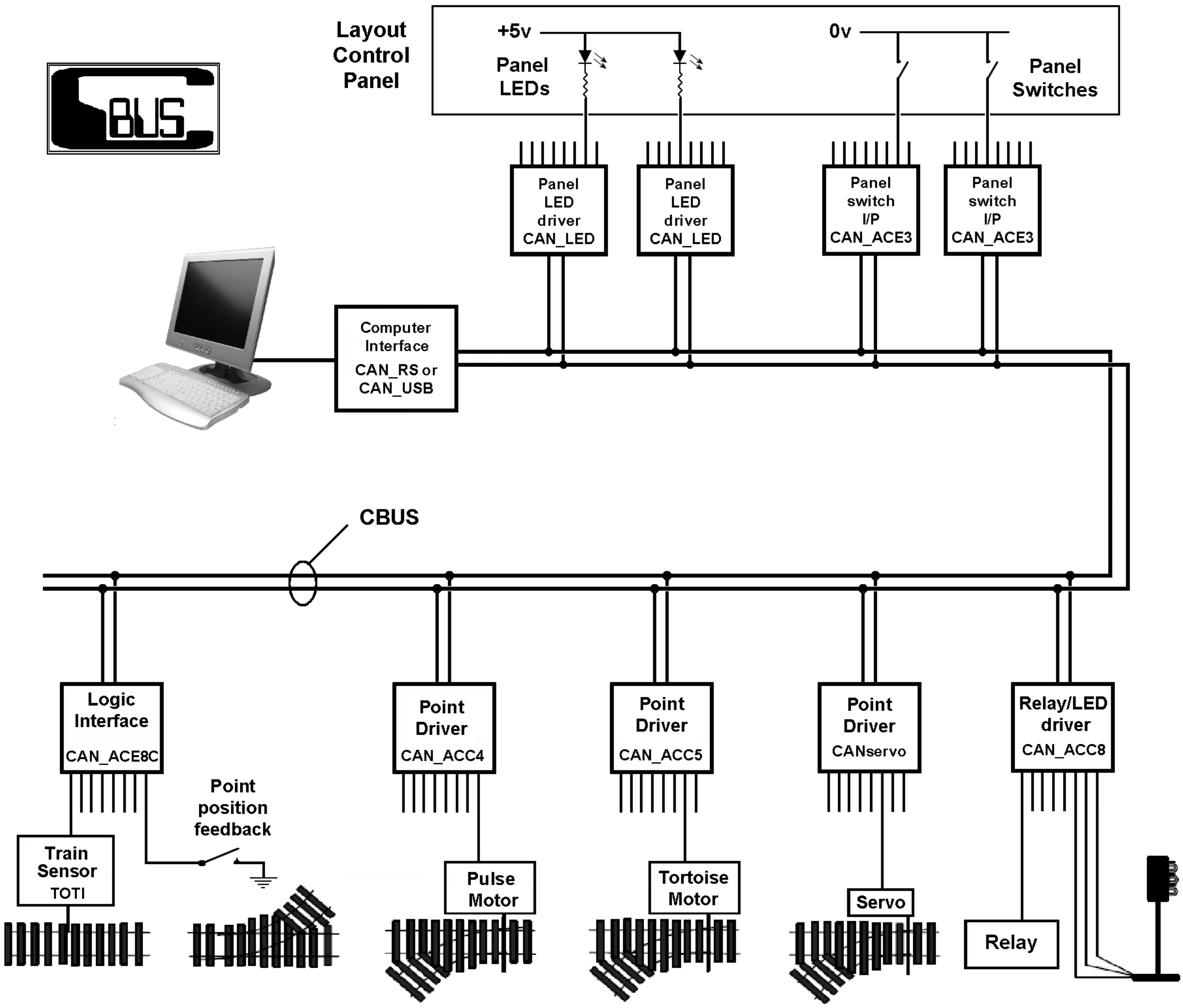
What we need, then, is an accessory system that combines the benefits of a 2-wire system with true 2-way communication, the ability to use mimic panels, the ability to have modules talk to each other over the two wires without using any central intelligence, and the ability to provide all this with or without the use of a computer.

This is what the CBUS system offers – plus more.

The MERG CBUS way

Imagine building a new layout consisting of four boards. With CBUS all you need do is run four wires the length of the layout – two for power and two for the control system. No matter how many switches, button, lights, points, track occupancy detectors, accessories, etc. you now add to the layout, you still only need those four wires. The accessory connections between boards are always just these four wires – not the scores of wires associated with conventional wiring.

Like DCC, all accessories are linked via messages. These messages can be transmitted by small modules that can be fitted anywhere on the layout. Indeed, you can have multiple control panels if you like. You could have a mimic panel for a main station, one for a goods yard, and so on. Or, if you prefer, you can control the entire layout from a single panel.
A message could be an instruction from a control panel switch to a point motor controller, telling it to switch the point. A message could be from a track occupancy detector to a light on a control panel, informing that a siding or loop is occupied.

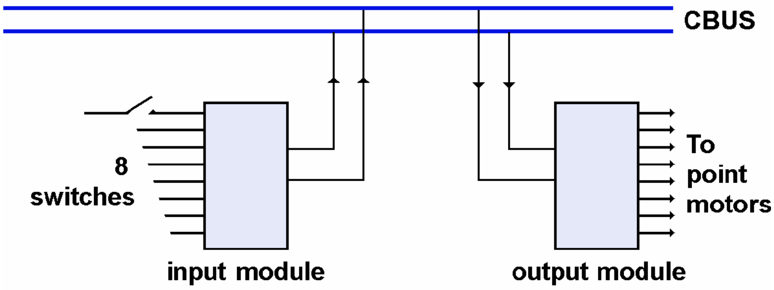


This diagram shows some of the range of options that are provided for by the CBUS system. Some of them give instructions, like the panel switches or the train sensors. Others act on these instructions, like the range of point motor controllers, lights, relays, etc.
All these modules, and more, are produced as kits by MERG and you can choose which

ones best meet your needs.
Lets look again at our imaginary four-board layout which is currently empty apart from a power bus and a two-wire control bus – both running the length of the layout.

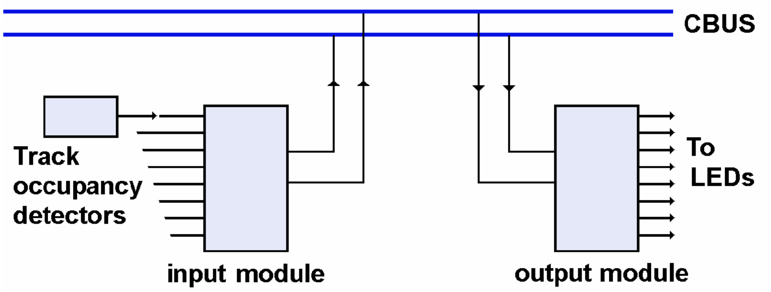
If we lay track for a goods yard on one of the boards, we can install a CBUS output module to control their points. The module is connected to the nearest point on the 2-wire bus, to its power supply and to the point motors. This is called a *'consumer'* module since it cannot create any messages. It can only act on incoming messages.

To control the points, we need an input module that can send the instructions to the points module. The input module is called a *'producer'* module since it creates messages when its switch inputs are altered. Like the output module, it connects to its power supply and to the nearest point on the CBUS. It uses a set of switches as inputs (ideal for a control panel). This illustration shows a basic minimum setup for a CBUS system.



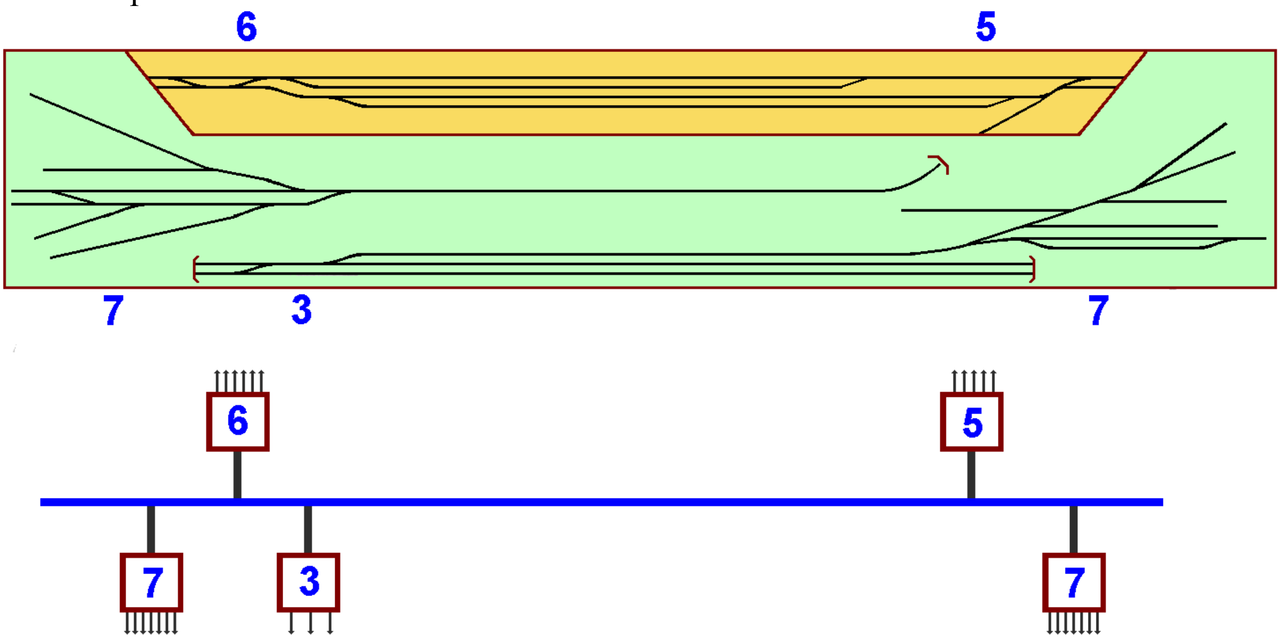
When you throw a switch, the input module recognises the change and sends a message on to the CBUS wires. It is picked up by the output module which translates it into a change of output on the corresponding pin. So, for example, throwing a switch on the input module's pin 3 might result in moving the point motor that is attached to pin 5 (or whatever pin you choose) of the output module .
Two relatively inexpensive modules allow 8 switches to control 8 points.
The modules could be at opposite end of our layout, since they can connect anywhere along the CBUS wiring.
As we extend our layout, we can add more switch modules and more point modules as before. We could have 50 or 100 switches on a control panel operating 50 or 100 points – and all using the same two CBUS wires!

Another benefit of CBUS is the ability to get feedback from the layout. This illustration shows an input module whose input pins are connected to track occupancy detectors.
When a train enters that section of track, an input pin's voltage is altered and this results in a message being sent out onto the bus. On the control panel, this messages is received by the output module which illuminates the corresponding LED on the mimic panel.



Example layout

Here is an example of implementing CBUS on a typical layout. The layout has 28 servo-operated points and is spread over several boards to allow long rural stretches of track.
The points are in four distinct groups. The fiddle yard has 6 points at one end and 5 at the other end. Each station has 7 points and there is a small group of 3 points on the main line.
The lower illustration shows how the points could be operated by five CBUS output modules (a typical output module having 8 outputs).
Since the points on the left side of the layout are fairly closely bunched, you could even use two 8-output modules at that end.



If we were to examine the underside of this layout's baseboard, we would only see the loco's power bus, layout power and the CBUS wiring. Four (or five) boards would be positioned close to the groups of points.
If we decided to add extra points or features to the middle section of the layout at a later date, we would only need to add another module to the CBUS.

The layout would also have a control panel attached to the same CBUS. It would contain the switches and the modules that send out the control instructions to the board's output modules. The panel can look and operate like a conventional 'hard wired' system.

Advantages of CBUS

The many advantages of CBUS include:

- A big reduction in wiring and connectors, with all the accessory modules sharing the same power bus and the same 2-wire cotrol bus.
- Fewer faults. With far fewer wires and connections (e.g. no multiplugs between boards), there is much less to go wrong.
- It supports feedback, You can get confirmation that points have moved, tracks are occupied, etc.
- It is easier to make changes and additions. Adding that extra railway station or goods yard requires no baseboard wiring changes. Just add the modules to the existing power and CBUS wires.
- It is suitable for automation / computer control.

Different levels of use

While CBUS offers a comprehensive range of facilities, you can use as much or as little of it as you need. If you already have a fully functioning layout, you may want to use CBUS output modules to add additional features. If you are planning a new layout, you may decide to use a full implementation of CBUS across the entire layout. It is up to you.
Here are some suggestions for different levels of implementation:

- Output modules to switch on lights in buildings.
- Output modules to operate signals.
- Output modules to operate animations (e.g. cranes, level crossings).
- Feedback from track detectors to indicate occupation of hidden sidings.
- Control of point motors.
- Control of locos (CBUS includes modules to control DCC locos).
- Computer automation of layout.

The first four suggestions can be added to any existing layout without any changes to its existing wiring and control systems.

What is CBUS

CBUS is based on the widely used CAN Bus (Controller Area Network) system, adapted for use with model railways.
The CAN bus was developed for cars in the 80's and found its way into many industrial, medical and military pieces of equipment. Everything from a coffee machine to the Las Vegas monorail works using the CAN bus, not forgetting Otis elevators, fork lifts, radar stations, drones and CERN.
The CAN bus, with tens of millions of controllers sold each year, is a tried and tested system. It is robust, reliable and resistant to external interference. After all, you would expect nothing less from a chip that is controlling aircraft, cars, lifts and medical equipment. Due its widespread use, the main interface chip is easily available and is cheap.

CBUS is based on the industry standard CAN bus, using its interface chips and basic principles. In addition, it gives special consideration to how it can aid model railway hobbyists.

Benefits of CBUS

- Low cost.
- Not tied to any manufacturer.
- Works with any scale or gauge.
- Works with DC or DCC layouts (although loco control is DCC only).
- Works with both small home layouts and large club layouts.
- Operates on wiring lengths of up to 500 metres, enough for all but the most ambitious garden layouts.
- Provides manual, semi-automatic or automatic control of devices (see later).
- Can be used on new layouts or integrated into existing layouts.
- Available as a range of self –build kits, with all parts and construction notes supplied.
- Module construction only needs basic soldering skills.
- Easily set up. No programming skills required (additional modules are available for those who wish to have computer control of their layout).
- Lots of support from MERG members.

The CBUS cable

The pair of wires that carry the messages throughout the layout should be *'unshielded twisted pair'*. This simply means that the two wires should be twisted round each other to prevent unwanted interference. There is no need to buy shielded cables that have foil covering. In fact, you can make your own cable from any wires as long as they are loosely twisted together as shown in the image.

Since the wires only carry messages, there is no need to use heavy gauge wires. The speed of the system dictates that the maximum length of the bus is 500m, so it can handle very large layouts.
The bus has to be *'terminated'*. This involves fitting a resistor across the two wires. Without this, there is a danger that signals will reach the end of the cable and reflect back down the wire, interfering with the wanted signals. Depending on your age, you may remember seeing 'ghousting' on your TV picture, which is an identical phenomenon. Fitting a resistor allows for the signal to be absorbed and prevents reflected electromagnetic waves from distorting the messages. The resistor is also required for proper CAN arbitration to work (i.e. preventing two producer modules from sending messages at the same time). A 68 ohm resistor is fine for small cable runs; for most layouts, fit a 180 ohm resistor at each end of the bus.
The bus wires should run the length of the layout. It need not run in a straight line; it can snake along to go where it is needed. Stubs are allowed; these are short stretches of cable that run off from the main bus (e.g. to feed a board in a T-shaped layout). Stubs should not exceed 3m in length.

Connecting the modules

Adding a CBUS module to a layout consists of a few steps:

- Connect the module to its power supply.
- Connect the module to the CBUS cable.
- Connect the inputs (e.g. from track occupancy detectors) or outputs (e.g. to point controllers, lights, etc.)
- Configure the module the way you want it to operate (see later).

The connections can be soldered, or use connectors. Plugs and sockets can be used to allow the quick replacement of a faulty unit (handy at exhibitions).
Once connected to the CBUS network, modules can be used in various configuration:

- Between the layout's modules and a control panel.
- Within the layout (e.g. the arrival at a station brings on the station lights)
- Between baseboard panels (e.g. providing communication between multiple operators)
- Between the layout and a computer.

You can use one of these, or all of them, depending what you want to do.

Expandability

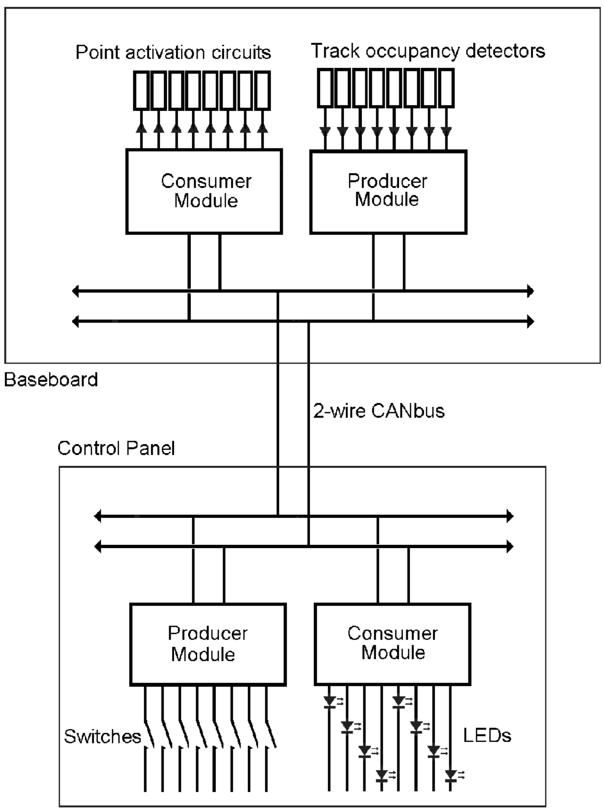
At its simplest, a layout using CBUS need only have one module that sends commands and one module that acts on the commands.
However, the CBUS was devised to allow for huge expansion potential.
You can fit an unlimited number of consumer modules, if they are used in SLiM mode (more later). Producer modules are limited to 110 nodes (the CBUS name for a producer module) on a single segment (an unbroken stretch of CBUS cable). Since some modules can handle 64 inputs and some can handle 64 outputs, a single segment can handle 3520 inputs and 3520 outputs. If that is still not enough, you can connect different segments

using a bridge module, until you reach a maximum of 65,536 nodes on the system! Depending on how they are configured, each node can produce up to 65,536 'events' (different instructions). So, the system can handle 4.29 billion different instructions!

How CBUS works

CBUS modules either send messages or receive messages (although there is one module that can both send and receive). Some modules only ever send messages on to the bus. They are called *'producers'* since they produce messages when activated. An example might be a module connected to track occupancy detectors. This small board may have inputs from, say, eight track occupancy detectors. The message sent by the module would depend upon which change is detected.
An even simpler example might be a bank of off/on switches used to control points, signals, trackside lights etc. The group of switches could be held in a frame or could be mounted on a mimic board/control panel. Throwing a switch produces a message that uniquely identifies which switch was thrown and whether it was switched on or off. In these examples, changing one of eight inputs to a module results in one of sixteen unique messages being transmitted on the bus (one message when an input goes high and a separate message when that input goes low).

Other modules are designed to only ever receive messages and to act upon them. These are called *'consumer'* modules and examples include point actuation and LEDs on mimic boards. So, for example, a message received by a module's might result in one of its output lines being switched on or off. These output lines can be used to switch eight different point motors. Another module can light up to 64 different LEDs depending upon the input messages received.
The user does not need to know how the message is constructed and how it works in order to use the system. The user simply throws a few setup switches and the module's electronics take care of everything.



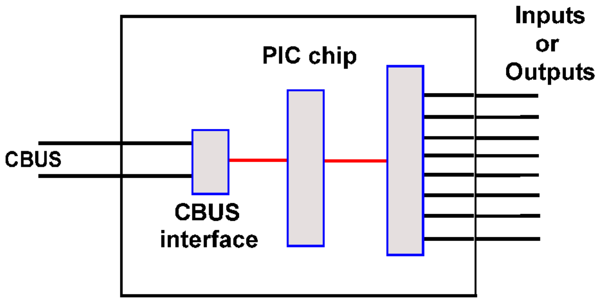
A small example

The diagram shows a simple setup, with a baseboard that has eight points and eight track occupancy detectors. The control panel has eight point switches and eight LEDs. Along with power connections, the CBUS links the control panel and baseboard and carries messages between them. The baseboard modules are connected to the user's point activation units and track occupancy detectors.
Assuming that the four modules have already been set up (see later) their operation is straightforward. So, if a train activates track detector number 5, the producer module sends out a unique message on the bus. The consumer module on the control panel

translates that message into switching on LED number 5. Similarly, throwing point switch 3 makes that producer module transmit a unique message over the bus that results in point number 3 being activated.

CBUS Modules

Most CBUS modules use the same basic arrangement shown in this illustration.
It has three main parts:



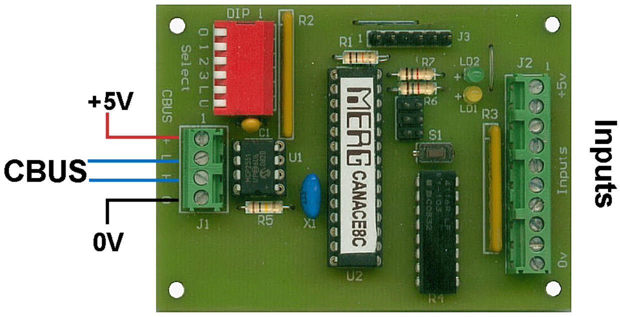
- The CBUS interface, known as a *'CAN transceiver'*. Every CBUS module needs one of these so that it can send to, or receive messages from, other modules connected to the bus.
- A PIC chip, which is like a small computer that stores a program and data. Its contents are programmable so that it has multiple possible uses, depending on what the module is used for. This is the 'brains' of the module.
- The Input/Output, which is the connection to the hardware devices. A producer module, for instance, might have switch inputs, while a consumer module may have outputs for controlling point motors.

MERG are constantly developing new modules, and providing new features to existing modules. Using PICs is a benefit, as the program built into it can be updated; there is no need to throw away a module when an improved version comes out.

Let us look at the range of CBUS modules that are available.

Eight input module

This is known as the CAN-ACE8C.
It has eight input connections. Each connection is normally held to +5V internally.
Bringing that connection down to 0V results in a message being sent out on the CBUS. Removing the 0V lets the connection revert back to +5V and this results in a different message being sent on to the CBUS. So, the module can produce two different messages for each connection – one when a connection is at +5V and one when it is at 0V.



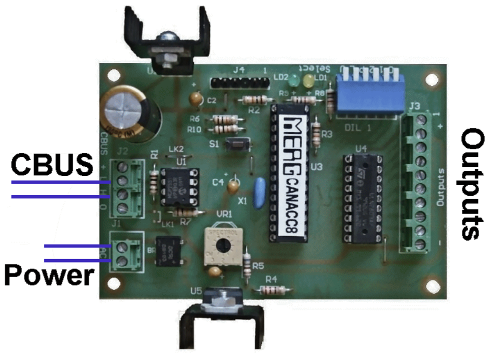
Any device that can switch between a high and 0V voltage can be used as an input. The most common inputs are toggle switches, microswitches and track occupancy detectors, although other electronic modules could also act as inputs.
If this module is mounted in a small control panel (e.g. to control a small layout or a small local goods yard in a larger layout), it would use switches to control points/lights on the layout. If the module is mounted under the baseboard, it could be used to send back information to a control panel, that a track section is occupied or to indicate the direction a point is facing.

Eight output module

The is known as the CAN-ACC8.

Apart from the connections for power and the CBUS cable, there are eight outputs that can be switched between 0v and anywhere from +5v and +15V. The little white square is actually a preset potentiometer that sets the voltage.

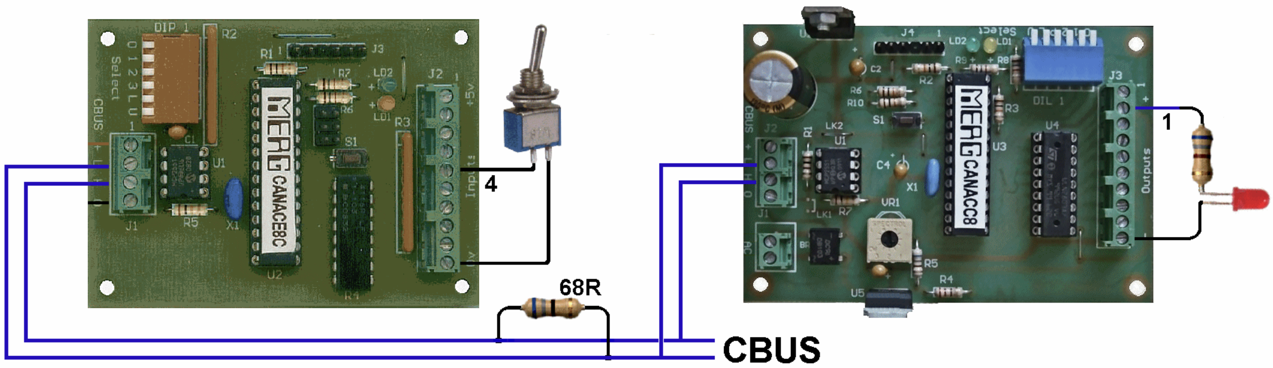
Being able to vary the output voltage means that it can switch 5V relays or 12V relays, LEDs, lamps and even small DC motors. Each output can handle a maximum current of 500MA, although the sum of all outputs must not exceed 1A. It is commonly used to operate points and animated accessories (crossing gates, semaphore signals, etc.) using servo motors. The outputs can operate relays to switch point frogs. Also used to remotely switch lights in buildings.



Simple example

This image shows how to connect a single input module (the CAN-ACE8C) to a single output module (the CAN-ACC8), so that throwing a switch controls the LED.

For simplicity, it does not show the power connections.



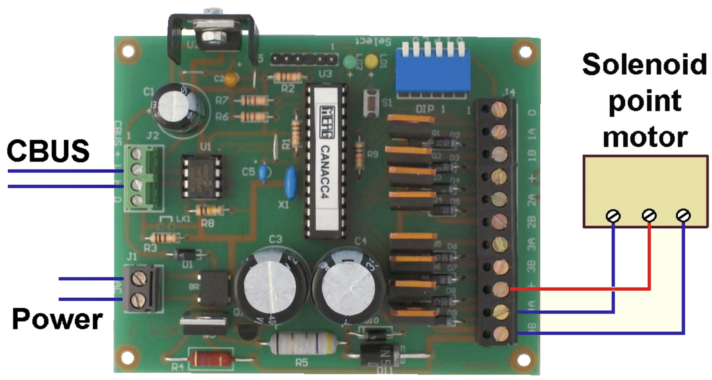
The points to note are:

- The CBUS has CANH and CANL wires, marked as H and L on the modules. The H connection on a module should be taken to the H wire of the bus; similarly, the L connection on all CBUS modules should be taken to the L wire of the CBUS.
- A terminating resistor of 68 ohms has been wired across the CBUS.
- The variable resistor has been rotated to produce a +5V output.
- The switch is connected to input 4 of the input module, while the LED is connected to output 1 of the output module. Input 1 does not have to control output 1 and so on. You can decide which output will be switched by any particular input.

Four solenoid module

This is known as the CAN-ACC4.

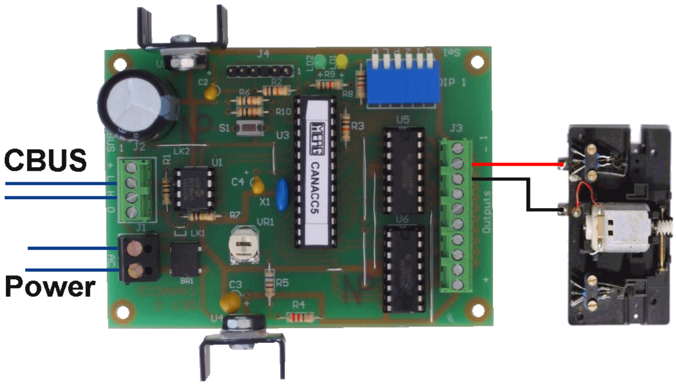
It is specifically designed to handle four solenoid-type point motors, such as the Hornby, Peco, Atlas or Seep type.



The module has a capacitive discharge unit (CDU) built in to the board and connects directly to the point motors as shown in the illustration.
Each point is allocated two messages – one to operate the point in one direction and another to operate the point in the opposite direction.

Four point motor module

This is known as the CAN-ACC5. It is specifically designed to handle four motorised point motors, such as the Tortoise, Hoffman or Conrad types. The illustration shows a Fulgurex point motor being connected – although not to scale. The module's outputs consist of four sets of two connections. Each point is allocated two messages – one to place a voltage across the two connection (e.g. as in the illustration) and another to reverse the polarity of the connections.

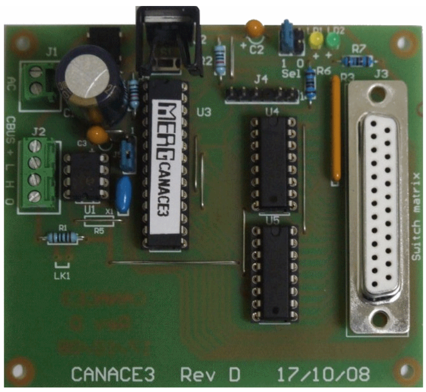


The recommended maximum current is 1 amp total for continuous operation.

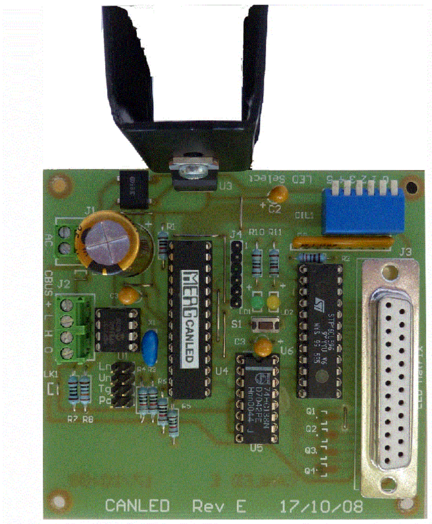
> **Note**
> The three output modules shown above can be controlled by using two unique messages for each output.
> If you wish, you can have several, or all, of the outputs controlled by the same pair of messages. This is useful for route switching, where one command can move a set of points and another command resets the route.

Mimic panel modules

The eight input and eight output modules are ideal for mounting under baseboards, to operate points and send back feedback. The benefit of the CBUS system is having small local modules where you need them rather than having wires trailing all over your layout.
They are not very useful, however, when used on a control panel or mimic panel, unless you have a very small layout or are using a number of small controls panels spread across the layout.
For large panels, there is a need for modules that can handle larger numbers of switches or LEDs.
Consider, for example, having a control panel with 50 switches and 50 LEDS. You would need 7 input modules and 7 output modules built into the panel- which would be very cumbersome.
So, MERG have a couple of modules that are specifically for use in control panels.



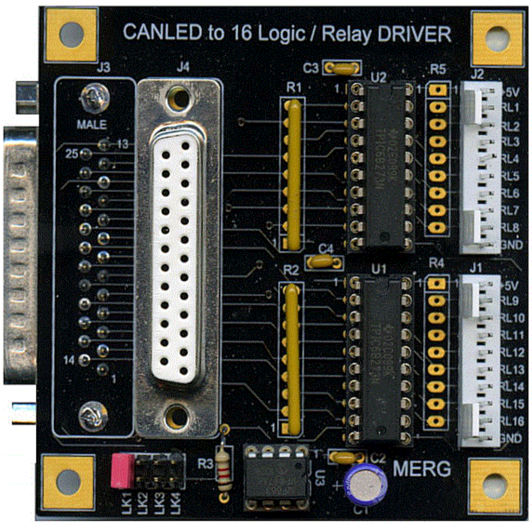
This module is known as the CAN-ACE3.
It is a switch interface that can handle 128 toggle switches or 64 pairs of push buttons.



The module on the left is known as the CAN-LED. It can handle 64 LEDs.
Using just one of each module in a control panel, you can control up to 64 points/accessories and indicate feedback from 64 devices.
The large sockets on the right of each module are provided for the connections to the switches/pushbuttons/LEDs.

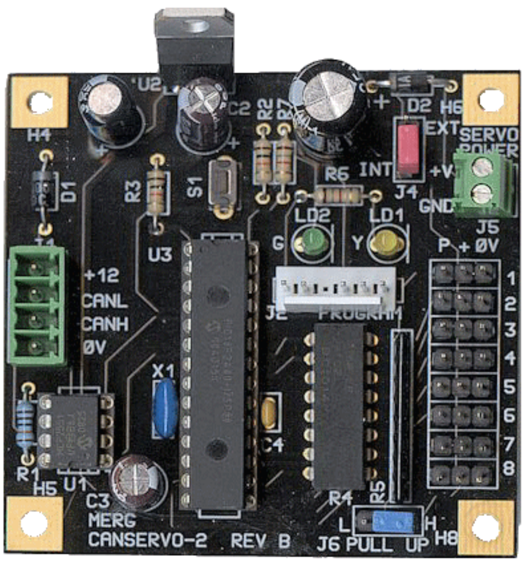
CAN-Relay 16

Although designed for control panels, the CAN-LED can be adapted for use on your layout, to operate relays.
If you have a need for a large number of relays on your layout, you can fit the *'CAN-LED to relay driver interface'*. As the name suggests, the CAN-LED board can be converted from operating 64 LEDs to operating 64 relays.
Each relay board can control up to 16 relays and you can attach four boards to one CAN-LED.
So you can, if you wish, control 64 relays over a single 2-wire connection.



CANSERVO8

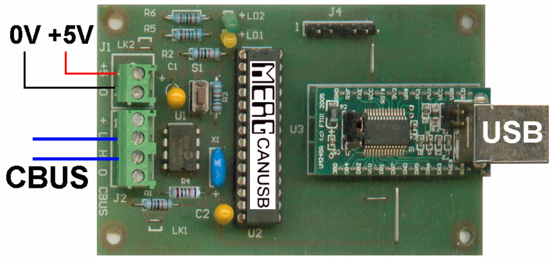
The CAN-ACC8 module can switch eight outputs on and off and these outputs can be taken to servo driver modules. MERG's Servo-4 modules handle four servos.
So, if your layout had multiple points in close location to each other (e.g. station throat, fiddle yard, goods, yard) you would need three modules to control those eight points.
For such dense use of servos, there is an alternative for CBUS users.
The image shows the CANSERVO8 module which combines a CBUS interface with the control of up to eight servos that plug into the connections shown on the right of the module.
There is manual method (switches) of setting up the CANSERVO8 module; each servo's settings have to be set up using computer software (see FLIM later).



Computer Interface

The CBUS system was designed to able to be set up and used without the need for a computer.

For those with computers and a bit of knowledge, MERG have a kit known as the CAN-USB.
It connects to the CBUS and plugs into a USB port on your computer.
With appropriate software, it can provide useful new facilities such as:



- Testing CBUS modules, to confirm whether they are functioning correctly. You can send messages on to the CBUS from the computer and see whether an output module responds as it should, Alternatively, you can check an input module by examining its messages on the computer screen.
- Controlling a layout. Using computer software (covered later), you can operate points, accessories and even trains from a computer screen, using your mouse.
- Setting up your modules, using the free configuration software for MERG members.

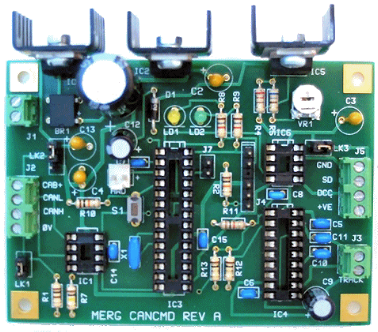
This saves fiddling with switches and

For those with older computers (ones with no USB sockets), MERG produce a kit for plugging into a computer's serial port. It is known as the CAN-RS and is useful for setting up modules, although a bit too slow for controlling larger layouts.

DCC Command Station

For many, the benefits of CBUS lie in controlling points and accessories, or getting feedback from track detectors, etc.

For DCC users, the CBUS system extends into controlling locos using a couple of modules.
The image on the right shows the CAN-CMD module which is a complete DCC command station.
It has connections for power, the CBUS and the layout track (either directly or through a power booster).



It can be controlled via the CBUS in either of two ways:

- Using a handheld controller.
- Using computer software

For those who like to walk round their layout while controlling their trains, MERG have a CAN_CAB kit.

The image shows a completed CAN-CAB module, with a cable to connect to the CBUS. Pressing the buttons, or turning the speed knob, results in instructions being sent to the command station which are translated into control messages for the layout's locos.

An alternative, for those so-minded, is controlling the command station via the computer. This would require a CAN-USB computer interface module to be fitted.
Suitable software is looked at later.

CBUS messages

So far, there has been a lot of talk about modules sending messages and receiving messages. We do not need to know exactly how a message is made up in order to use the system.
We do, however, need to consider a few general points:

At the sending (i.e. producer) end

- A message should be sent when something *changes* in an input module – e.g. a switch has been thrown or a button has been pressed. CBUS calls this an *'event'*
- The message should say *which* pin has changed – e.g. pin 4 in our example.
- The message should say *how* it has changed – e.g. it has been changed to a high or a low voltage.

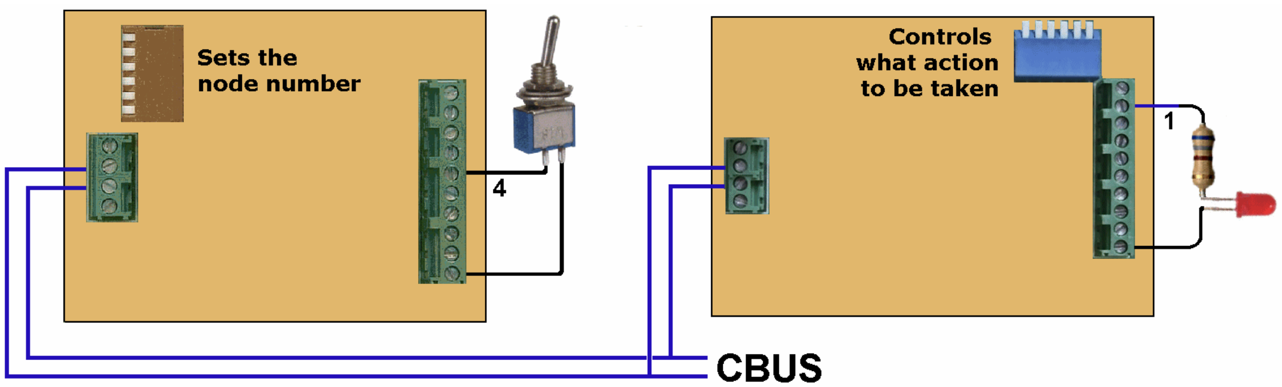
So, part of the message might say

"The fourth input connection has changed to low"

or

"The fourth input connection has changed to high"

So far, so good. But a system might have a number of producer modules and each one of them will have a input connection number 4. If they all sent out the same message, output modules would not know which one they were supposed to act upon.
The message, therefore, also has to contain information about *which module* is sending the message.
This is achieved by giving each producer module its own unique number, known as a *'node number'*.



"Node 7's fourth input connection has changed to low"

Consider the example in this diagram, if the producer module with the switch attached was given a node number of 7. Throwing switch 4 from high to low would produce a message with the following essential information:

"Node 7's fourth input connection has changed to low"

Every message now contains information on what has changed and what module sent the message.
The simplest way to allocate a module its unique number is by throwing the switches highlighted in the illustration.

On the bus

When an event occurs in a producer module, it sends out a message without knowing whether any consumer module will respond to it. It just sends the message anyway. This is

known as *'fire and forget'*. The producer module does not check for any acknowledgement that its message has been received by another module.
There are a number of possible situations.

- There may not even be any consumer modules attached to the bus. The messages are being sent but no-one is listening.
- There may be many consumer modules attached to bus. They will all receive the message but none may be set up to act on the message.
- A single module may receive the message and act upon it.
- Several modules may receive and act on the same message.

CBUS messaging is not equivalent to sending a text message to a particular friend, where that person's phone is the only one to receive the message. It is more like the Saturday night Lottery broadcast. The results of the draw are broadcast to all who want to listen but only a few will act on the message.

At the receiving (i.e. consumer) end

Since all producer modules have their distinctive node numbers, we can now distinguish between messages. We may know, for example, that all the messages from node 5 come from track occupancy detectors, while messages from nodes 4 and 6 come from switch panels, and so on.
In such a case, we could use an output module to drive panel lights when it receives messages from node 5, while using an output module for driving points motors when messages are received from nodes 4 and switching lights in buildings when messages are received from node 6.

The next step is to decide what each module should do, if anything, with an incoming message. You have a lot of flexibility in how you configure your output modules.
This is achieved using the bank of switches shown in the earlier illustration.
The options include:

- Ignoring a message. This will be the most common reaction, since most messages are of no interest to the majority of consumer modules.
- A consumer module's outputs all being dedicated to the same type of outputs – all LEDS or all point motors.
- A consumer module (e.g. the 8 outputs of the CAN-ACC8) being used various purposes – some connections switching lights, others switching relays or point motors.
- Different outputs of a consumer module being activated by the *same* message. This allows a single switch on a control panel to change a group of outputs – very useful for setting up routes and signal aspects.

Setting up modules

For a basic 8-input producer module, the only setup needed is allocating a node number to the module. At its simplest, you can throw a couple of switches and the job is done.
CBUS modules can be set up using switches or using a computer. Once set up, a module acts in exactly the same way, regardless of what was used to configure it.
It is just two different ways to achieve the same end. Modules set up by switches and modules set up by computer can be used happily on the same layout.

SLiM

The default mode is called *'SLiM'* (The Small Layout Model).
This method has the following advantages:

- It allows modules to be set up using switches – no computer is required.
- Modules can be configured and tested before fitting under the baseboard.

It also has a few disadvantages:

- The setup process can be fiddly and error-prone, unless you really concentrate.
- You have to carefully document what outputs are changed by which inputs, otherwise it will cause serious inconvenience if you have to make changes later.
- Replacing a faulty board is more time-consuming than with FLiM.
- Reconfiguring the module means fiddling with switches while perched under a baseboard, or removing the module for alteration.

The exact method of configuring SLiM CBUS modules is documented in each kit's instructions.

There are also excellent articles that cover practical implementations of CBUS.

For further reading on this, see

"Experiments with CBUS" by Howard Watkins
(MERG Journals for Spring, Summer, Autumn Winter 2008 and Spring 2009)
"A Starter CBUS Project" by Davy Dick
(MERG Journals for Summer, Autumn and Winter 2009).

FLiM

The method using a computer for configuring modules is called *'FLiM'* mode (The Full Layout Model). This method has the following advantages:

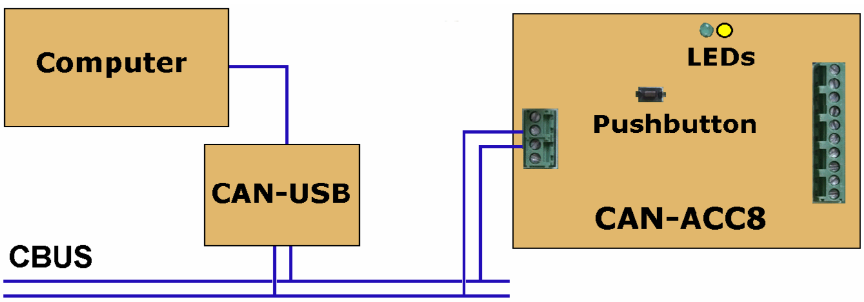
- You just fit and wire the modules, then all configuration takes place while sitting in front of your computer.
- The configuration software maintains a database of all the modules, what triggers them and what their outputs do. No need to keep paper records.
- Changes in your layout track plan or points routing are easily dealt with as the software updates all the changed relationships for you.
- It is quick and easy to replace a faulty module. Plug in the replacement module then download the previous module's settings into it from the database.
- Complicated configurations, such as the CAN-ACE3 and the CAN-LED are easier to accomplish using software.
- The CANSERVO8 module can only be configured via a computer. Future complex modules may also require computer configuration.

FLiM also has a few disadvantages:

- It needs a computer.
- It needs a CAN-USB module.
- The configuration process takes a bit of getting used to.

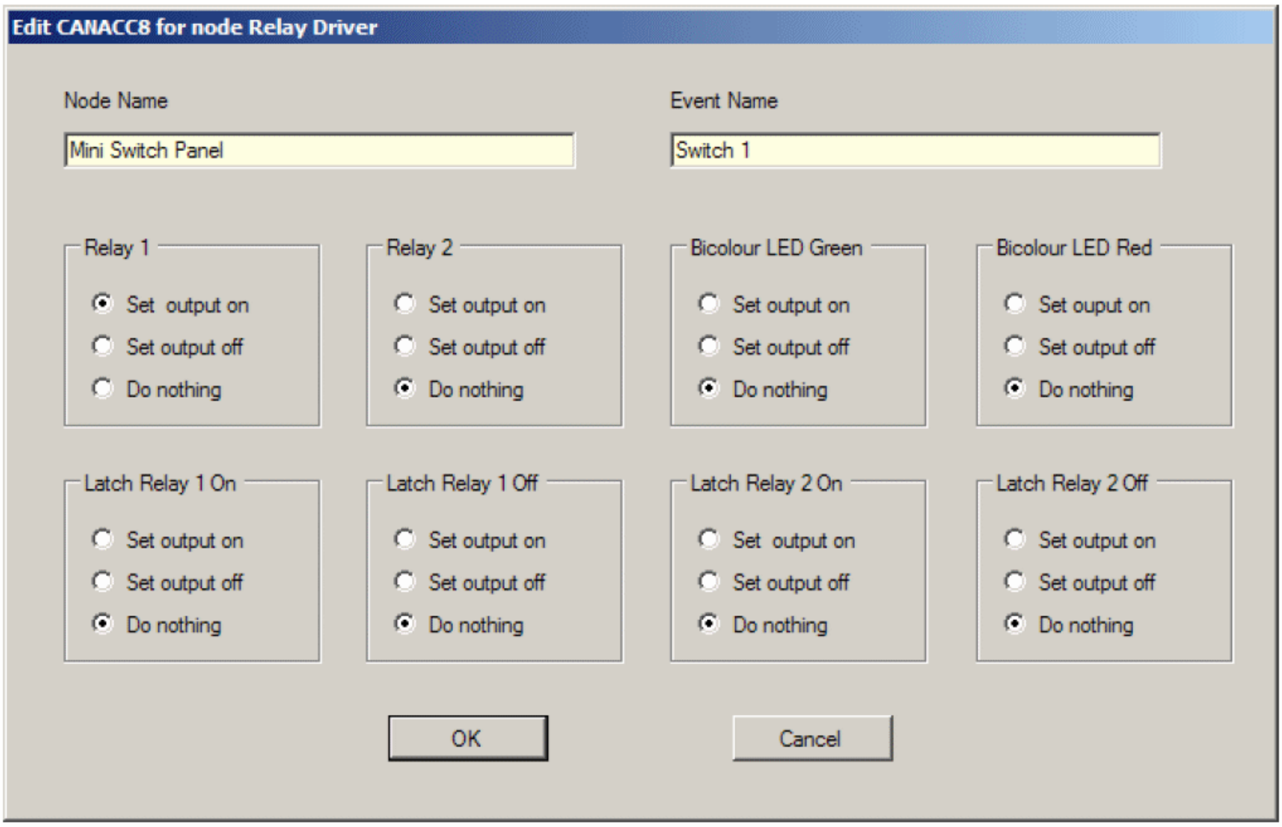
The FLiM setup

The illustration shows the setup for configuring modules.



A CBUS-to-computer interface module, the CAN-USB, is plugged into the USB port of a computer and connected to the CBUS cable.
The module to be configured is connected to the CBUS and to its power supply.
A consumer module can be switched between SLiM and FLiM modes by holding down the pushbutton for about 8 seconds, until the yellow LED illuminates (indicating FLiM mode) or the green LED illuminates (indicating SLiM mode).

The computer should have a software utility, called the FCU (FLiM Configuration Utility) installed on it. The utility creates and maintains a *'configuration file'* which stores the configuration details for all CBUS modules on the layout's bus.
Unlike SLiM, which only allocated node numbers to producer modules, every module is allocated a node number in FLiM mode. This is necessary so that the utility can identify individual consumer modules for configuration.
Here is a typical configuration screen for a CAN-ACC8, eight output module.
Each of its eight outputs has been given a handy text description (Relay 1, etc.) and the boxes shows what states the outputs can be switched to.



It shows that when the 'Switch 1' event in the producer node called 'Mini Switch Panel' is activated, the Relay 1 output is switched; all the other outputs are unaffected by this incoming message.

When many modules are connected to a layout, giving meaningful names to all the producer modules and their events, plus all the modules and their outputs, makes later identification and understanding much easier.
There is an excellent Guide, written by the utility's author Roger Healey, which explains the process in great detail.

You only use the computer during the setting up operation, or any subsequent additions or alterations. It does not need to be permanently connected, unless you decide to automate, or semi-automate your layout with other control software.

Notes

This chapter only provides an overview of the basics of CBUS.
CBUS is extensively covered in MERG's Technical Bulletins (there are 10 different Technical Bulletins in the G32 list).
There are also many practical articles in back copies of the MERG Journal.

The MERG CBUS system is constantly evolving with new boards being added that offer even more features and greater flexibility.
They begin as a member's idea, followed by a circuit diagram, followed by making a PCB available. If the demand is sufficient, it will appear as a full kit of parts.

So … watch this space!