Advanced Electronics

Chris Carr
Space & Atmospheric Physics Group
Room 6M65 Huxley
c.m.carr@imperial.ac.uk

February 2023

All the passive components we studied in the *Basic Electronics* course were **linear** devices where the amplitude of the device's response is always proportional to the the amplitude of the applied stimulus. Ohm's law is a simple example of this. Further, when considering AC signals, we saw that the linearity principle also applies to capacitors and inductors. In this course, we will investigate the properties and uses of semiconductor devices such as diodes and transistors, which are inherently **non-linear** in their behaviour.

Contents

Lecture 1 The Semiconductor Junction Diode

1.1 Semiconductors

In a metal such as copper the atomic structure of the lattice is sufficiently densely-packed that the loosely-bound valence electrons are shared between atoms and hence are able to wander quite freely through the material. In an insulator, all the valence electrons are quite tightly-bound. Semiconductors are simply materials which, at room temperature, have valence electrons which are rather less tightly-bound than insulators, resulting in a resistivity which is intermediate between conductors and insulators (table 1.1). Heating a semiconductor results in more electrons becoming

available for conduction. Consequently, unlike a metal, the resistivity of a semiconductor *decreases* with temperature.

Element	**Atomic number**	**Room temperature resistivity** (Ωm)
Aluminium	13	$\rho = 2.8 \times 10^{-8}$
Silicon	14	$\rho = 2.3 \times 10^{3}$
Sulphur	16	$\rho = 2 \times 10^{15}$

Table 1.1: Resistivities: metallic conductor; semi-conductor; non-metallic insulator.

Intrinsic semiconductor

Silicon and germanium are both tetravalent semiconductors with four valence electrons in their outer shell. Both form crystals with four covalent bonds such that each valence electron is shared with a neighbouring atom. At low-temperatures, these materials are good insulators, however as the temperature rises, some of the valence electrons have sufficient energy to break free from their bond and are able to move about within the lattice, leaving behind a **vacancy** known as a **hole**. As the material is overall neutral, the hole represents excess positive charge and so if a wandering electron encounters a hole it is likely to be captured. At any given temperature, there will be an equilibrium between the generation of holes and the capture of electrons such that we would observe both electrons and holes moving randomly through the material. Conceptually, we can think of both electrons and holes as charge carriers, and indeed if an electric field is applied across the material then the electrons will tend to move towards the more positive side and as a consequence the holes will appear to move towards the negative side. However, the availability of these natural charge-carriers is tiny, the electrical conductivity[1] of such an **intrinsic semiconductor** is extremely low at room temperature, and it is necessary to artificially enhance the conductivity in order to make Silicon and Germanium useful for electronics.

1.2 Doping

Adding very small amounts of pentavalent or trivalent atoms such as antimony or phosphorous (5 valence electrons) or indium (3 valence electrons) has a dramatic effect on the conductivity. For example, in the case that antimony is added to germanium then an antimony atom occupies a location instead of a germanium atom and its very loosely-bound fifth valence electron is not needed for bonding and becomes readily available for conduction.

n-type semiconductor

Doping germanium with just 1 atom in 10^6 of antimony results in an increase in electrical conductivity by a factor about 10^4 and such a material is called an **n-type** semiconductor in which the **majority carriers** are negative electrons and the (much rarer, intrinsic) positive holes are **minority carriers**. There are far more conduction electrons than holes, though the material remains electrically neutral due to the excess positive charge on each antimony atom (this should not be considered a 'hole' since it is immobile and not inclined to capture electrons).

[1] the inverse of resistivity, $\sigma = 1/\rho$.

The n-type semiconductor is a reasonably good conductor with large numbers of negative-charged electrons as the majority carriers and very small numbers of holes as the minority carriers.

p-type semiconductor

Doping germanium with trivalent indium results in a p-type semiconductor. Each indium atom only requires three adjacent germanium atoms to form covalent bonds and has a tendency to **accept** electrons from neighbouring germanium atoms to fill the apparent vacancy. Consequently, the majority carriers in a **p-type** semiconductor are positive holes, and electrons are the minority carriers. Like the n-type semiconductor, the material is overall neutral.

The p-type semiconductor is a reasonably good conductor with positive-charged holes as the majority carriers and electrons as the minority carriers.

1.3 The p-n Junction

It is possible to manufacture a semiconductor crystal with one half n-type and the other half p-type. The boundary is called the **p-n junction**. Conduction electrons have thermal energy and are highly-mobile. At the junction, electrons from the n-type material diffuse into the p-type material and fill some of the holes there, leaving an excess of positive holes on the n-type side of the boundary. Conversely, we could think of this as positive holes from the p-type material diffusing into the n-type material. In this narrow (about $1\,\mu m$) **depletion layer** there is immediately an electric field and hence a potential difference which prevents further diffusion and establishes an equilibrium. This potential difference, typically a fraction of a volt, is called the **contact potential**.

The depletion layer is effectively devoid of charge-carriers, and the contact potential represents a barrier which would need to be overcome in order for further charge-carriers to cross the junction.

The sense of the contact potential means that the p-n junction behaves differently when we apply an external voltage across the semiconductor. Furthermore, any metal wires connecting to the semiconductor can only conduct electrons. Consequently, a current can only flow through the junction when holes flow one way and electrons flow the other way *simultaneously*.

Reverse bias

If we apply a voltage across the semiconductor in the same sense as the contact potential then this reinforces the depletion layer. Recall that the p-type material has a small number of electron *minority carriers*, and that the n-type material has a small number of holes as minority carriers. The contact potential is favourable for the minority carriers, but they are sparse, so only very small current can flow. In this situation, the junction is said to be **reverse-biased**. A reverse bias voltage of about $10\,\mathrm{V}$ will result in a very very small current $\ll 1\,\mathrm{pA}$.

Forward bias

If we apply a voltage in the opposite sense to the contact potential then we have to consider two cases:

1. If the applied voltage is less than the contact potential then the majority charge carriers on both side of the junction tend to want to cross the junction, but are prevented from doing so by the potential barrier of the depletion layer. Statistically, there will be a few charge-carriers with enough energy to jump the barrier, but in practice only a very small current will flow.
2. If the applied voltage is greater than the contact potential, then the majority carriers on both sides are able to flow across the junction and a current can flow easily. In practice, the current rises exponentially as the applied voltage increases.

1.4 The Junction Diode

The practical p-n diode as used in electronic circuits is a 2-terminal **non-linear device** made from a junction of p-type and n-type material. Silicon is most commonly used, but germanium diodes are useful for special applications. The circuit symbol is shown in figure 1.1. The diode allows current to flow one way: conventional current will flow in the direction of the arrow ('**forward biased**') but not against ('**reverse biased**').

Figure 1.1: Diode Symbol

For historical reasons, the terminals are called '**anode**' and '**cathode**', but the main thing to remember is that the diode is not an ideal one-way conductor since we need to ensure that terminal *a* is more positive than terminal *b* by about 0.6-0.7 V in order for current to flow. This value, known as the **forward voltage drop**, is an approximate 'rule-of-thumb' number for a particular diode type. This is sufficient to analyse the behaviour of diodes in practical circuits, however a more accurate description of the diode's voltage-current characteristic is given by the **diode equation**.

1.5 Diode Equation

The behaviour of a real diode is quite well modelled by the equation

$$I = I_s \left(e^{V_{ac}/nV_T} - 1 \right) \tag{1.1}$$

I is the current flowing in the a to c sense and $V_T = k_B T/e$ is the **thermal voltage** equal to about 26 mV at room temperature. n is device-dependent factor which is about 1 for a simple silicon diode. I_s is the **reverse saturation current** which will flow if V_{ac} is negative, but it is tiny - of the order tens of fA. Figure 1.2 illustrates the behaviour.

Clearly, the diode doesn't obey Ohm's law. Between about 0.6 and 0.7 V the diode goes from non-conduction to passing almost any current (up to some physical limit of the device). Typical diodes can be used for a few hundreds of milliamps and specialist diodes are available for high-current applications.

> Diodes are almost always used with some resistance in series to limit the current. Consequently we can assume that there is always a **forward voltage drop** of about 0.7 V associated with a **forward-biased** diode.

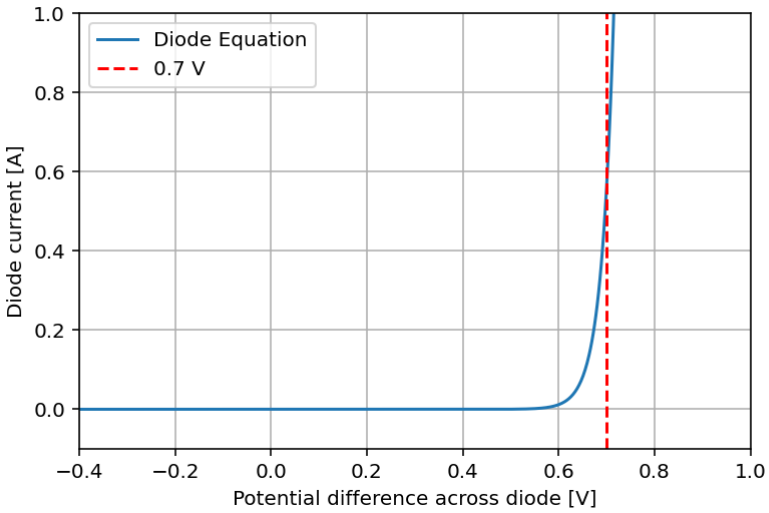


Figure 1.2: Voltage-current characteristic for the diode model (silicon diode at room temperature).

Fluid Analogy

The diode acts like a one-way valve for current. In the fluid analogy, the valve needs a certain pressure of water to open it, but once open, the water flows freely through.

1.6 Diode Model for Circuit Analysis

If we solve equation 1.1 for a diode current $I = 100\,\text{mA}$ we get a diode voltage $V_{ac} = 0.66\,\text{V}$ (at an operating temperature about 300 K). This voltage is shown with a dashed-line on figure 1.2. Since practical diodes are useful for currents in the range to about 1 A, the forward voltage drop will always be in the range $0.6 \ldots 0.7\,\text{V}$. To simplify circuit analysis, we can simply assume that the forward voltage drop is a constant of 0.7 V *whenever current flows*. This is of course an approximation, but it simplifies circuit analysis. For more detailed analysis, we could calculate the correct value, which is a function of temperature and current, but also the type of diode. For example, a germanium diode could have $V_{ac} \approx 0.3\,\text{V}$.

> For this course, we will assume use of silicon diodes with a forward voltage drop
>
> $$V_{ac} = 0.7\,\text{V}$$
>
> unless stated otherwise, e.g. in lab where you might encounter some specialist diodes.

Consequently, the general analysis of diode circuits is greatly simplified:

1. For the case that $V_{ac} \leq 0.7\,\text{V}$, no current will flow;
2. For the case that current does flow, V_{ac} remains fixed at 0.7 V, and the current which flows is controlled by other components in the circuit (typically resistances).

1.7 Power

If we connect a voltage source directly across a diode then, assuming the applied voltage is greater than about 0.7 V, an arbitrarily high current will flow and a real diode will fail. Recall that power is *always* voltage times current: though the diode is **not** an ohmic resistance, it dissipates some power as the current passes through the potential difference (the forward voltage drop). At 1 A, 0.7 W is dissipated, which is quite a lot of power to dissipate in a small p-n junction, and consequently typical lab diodes are only 'rated' for currents of a hundred mA or so. As mentioned previously, there are special diodes for high-current applications; these are adapted to dissipate more power. Nevertheless, as a consequence, you will almost always see a diode in series with a resistor. The resistor is there to limit the current drawn from the source. For example, if you use a light-emitting diode (LED) in a circuit, then it is important to place a resistance (typically 1 kΩ or so) in series to avoid overheating the LED.

1.8 Rectification

An AC voltage such as $v = V_0 \sin(\omega t)$ is typically associated with an oscillating current, i.e. the sense of the current will alternate between positive and negative within the circuit. We can exploit the one-way characteristic of the diode to make a **rectified** version of the signal. Rectification generally means to make a signal positive-only, i.e. any negative part of the signal is removed.

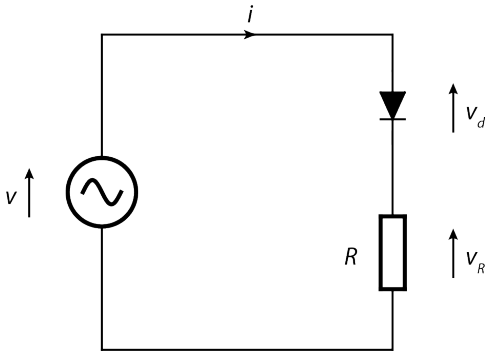


Figure 1.3: Half-wave rectifier circuit

For the case that the AC voltage is applied to a series combination of a diode and a resistor R (figure 1.3) then current will only flow when $v > 0.7\,\text{V}$ in which case $v_d = 0.7\,\text{V}$ and, by Kirchhoff's voltage law, $v_R = v - 0.7$. Consequently

$$v_R = \begin{cases} 0, & v \leq 0.7 \\ v - 0.7, & v > 0.7 \end{cases} \tag{1.2}$$

Note that the current $i = v_R/R$.

For a source amplitude $V_0 = 10\,\text{V}$, figure 1.4 shows, in the bottom trace, the voltage across the resistor. The **positive going** half of the input signal is preserved (minus one 'diode drop'), and the **negative going** part of the signal is removed. Consequently, this is known as a **half-wave rectified** signal.

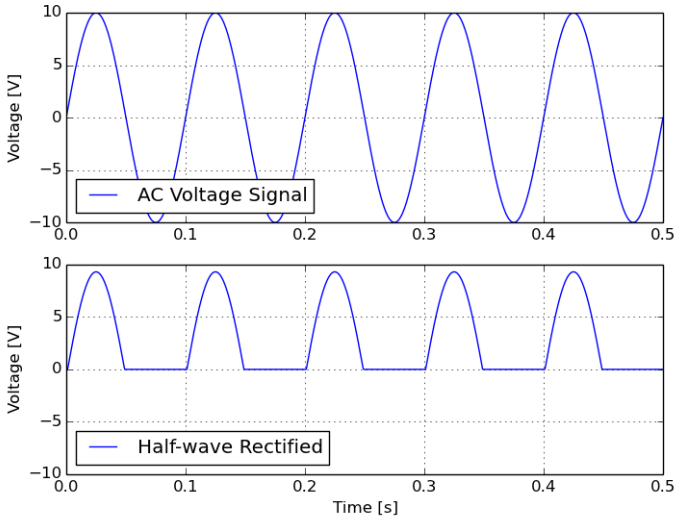


Figure 1.4: Half-wave rectified signal

1.9 Bridge Rectifier

We might like to make a **full-wave rectified** signal (mathematically the absolute value of the signal, ignoring the diode drop); figure 1.5 illustrates this and figure 1.6 gives the rectifier circuit. As the current flows alternately clockwise and anti-clockwise from the source, it can be seen that the current only flows one-way through the resistor, hence we only ever get a positive-sense voltage across the resistor.

Note that the output is only approximately $|v|$ since there is a reduction in the output due to the forward voltage drops across the diodes (for a positive output, there must be a current flowing through 2 diodes). Furthermore, there is a period of time each half-cycle when the output is zero as none of the diodes have sufficient voltage across them to conduct. This effect becomes more important for small signals (amplitude V_0 of the order a volt or so).

Further Reading

Sears and Zemansky's **University Physics** (*Young and Freedman*) is a good reference text for a more in-depth introduction to semiconductor physics. The relevant chapters are:

Chapter 42.5: Free-electron Model for Metals

Chapter 42.6: Semiconductors

Chapter 42.7: Semiconductor Devices

The Art of Electronics (*Horowitz and Hill*)

There is a good description of the rectifier circuit in Horowitz and Hill sections 1.25 and 1.26.

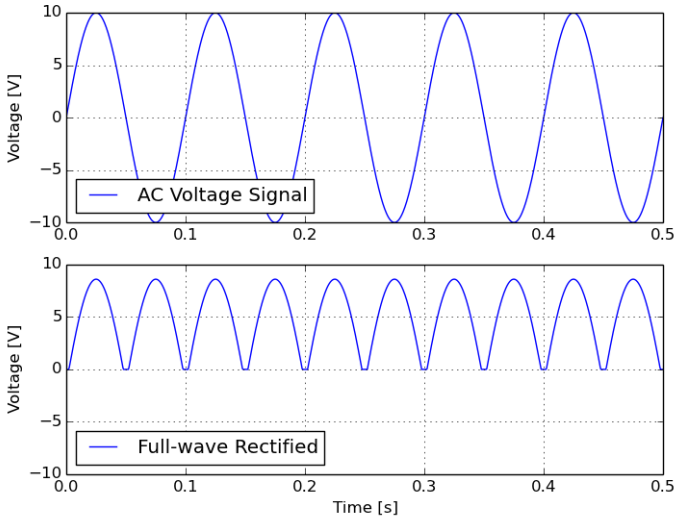


Figure 1.5: Full-wave Rectified Signal

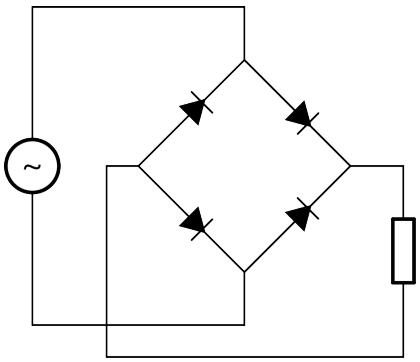


Figure 1.6: Bridge Rectifier Circuit

Lecture 2 The Bipolar Junction Transistor

2.1 Bipolar Junction Transistor

The npn transistor is a very thin layer of p-type semiconductor sandwiched in-between two thicker layers of n-type semiconductor. Each layer has an electrical connection and the two n-type layers are known as the collector (C) and emitter (E). The thin p-type layer is known as the base (B). Consequently, the npn transistor has two junctions: n-p and p-n.

2.2 Transistor Action

In typical use, the collector is made more positive than the emitter, however with this alone no current will flow as the base-collector n-p junction appears reverse biased. However, making the base more positive than the emitter by, say, 0.8 V makes the base-emitter p-n junction forward biased and conventional current will flow from base to emitter. At this point it is more instructive to think about electrons flowing the other way, from emitter to the base layer. As electrons diffuse from emitter to base they have three options:

1. combine with holes in the base layer;
2. leave via the base connection resulting in a base current I_B;
3. diffuse completely across the base until, encountering the base-collector junction, they are swept-up by the electric field of the depletion region and carry through into the collector thus forming a collector current I_C.

In practice, about 99% of the electrons do option (3), and the majority of the rest do option (2). This is only possible because the p-type layer is very very thin (less than the natural diffusion scale for electrons injected into the base region from the collector). If the base were thick then the npn transistor would just look like a pair of 'back-to-back' diodes and this effect would not occur.
The process, whereby electrons go straight through from emitter to collector, is known as **transistor action**. With no base current flowing, there is no possibility for electrons to be injected into the base and hence transistor action can not occur: the transistor is 'off'. However with a small base current flowing then the contact potential of that junction has been overcome, electrons are injected, transistor action can occur, and the transistor is 'on'.

> A key point is that the number of electrons going straight through to the collector is very much greater than the number going to the base terminal. Consequently, a small current in the base controls a much larger current in the collector. This provides the possibility for amplification.

2.3 npn Transistor Equations

Due to the opposite polarities of the two junctions, this type of transistor is known as a **bipolar junction transistor**. The practical npn BJT, as used in electronic circuits, is a **three-terminal** non-linear device with pins labelled base (B), collector (C) and emitter (E). Figure 2.1 shows that the base-collector and base-emitter connections look like diodes. Whilst it looks like no current can flow from collector to emitter, we have to remember that this is possible due to *transistor action*, which will occur if we fulfil two conditions:

Condition 1 The base-emitter junction must be forward biased such that a (typically small) current I_B flows into the base.

Condition 2 The collector must be more positive than the emitter such that a (typically large) current flows into the collector.

With these two conditions met, just two equations approximate the behaviour of the transistor:

$$I_C = h_{FE} I_B \tag{2.1}$$

where h_{FE} is the transistor's **current gain** (typically about 100) and

$$I_E = I_B + I_C \tag{2.2}$$

which is required due to Kirchhoff's current law (charge is conserved).

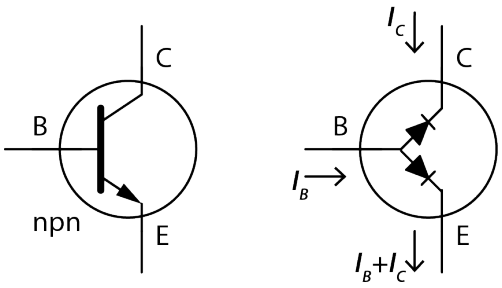


Figure 2.1: npn Transistor Symbol and 'Equivalent Circuit'

Fluid Analogy

The transistor acts like a valve for current flowing from collector to emitter, but the amount of current is controlled by the base. In the fluid analogy, we can imagine that the collector represents a large tank of water at some pressure. The amount of water which is allowed to flow through to the emitter is controlled by injecting just a small amount of water into the base.

2.4 The Transistor as an Amplifier

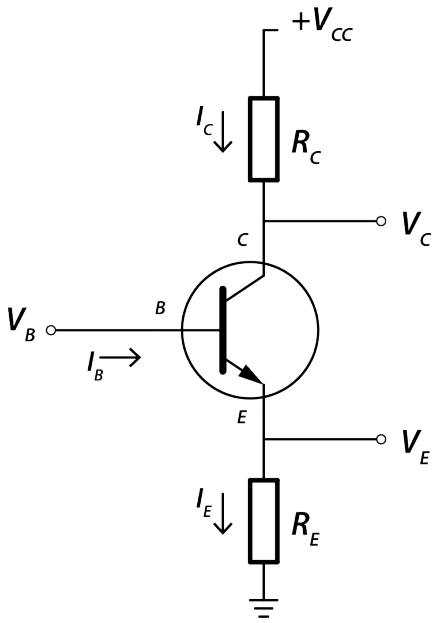


Figure 2.2: npn Transistor Circuit

Figure 2.2 illustrates how the transistor can be used to perform practical operations such as **buffering** and **voltage amplification**. Assuming that we have met both of the conditions described on page

10 we can write

$$V_E = V_B - 0.7 \tag{2.3}$$

where 0.7 volts is the typical **forward voltage drop** of the base-emitter junction. We can also conclude that $I_C \approx I_E$ since h_{FE} is large and hence conclude that a small current I_B controls a much larger current flowing through the resistors and the transistor. From these relations we can determine both the collector and emitter voltages; that is to say we can predict the output at both these points based on a knowledge of I_B.

When using the transistor as an *amplifier*, we will always provide the circuit with an input signal V_B connected to the base. We have two options for the output signal:

Emitter Since $I_E \gg I_B$, at the emitter, the transistor provides current amplification.

Collector By taking the output at the collector V_C, we can get voltage amplification.

The first of these options will be used in section 2.5 to build a device called an **emitter follower** or **buffer**. The second will be used in section 3.1 to build a voltage amplifier known as as **common emitter amplifier**.

Power Supply

$+V_{CC}$ here represents a power-supply of some kind. Typically this might be a voltage source of 10 V or so. Since it is an ideal source, it supplies whatever current I_C is required and its voltage doesn't change. This illustrates a key point about the transistor: it requires an 'external' power supply in order to function. The transistor does not generate current, it simply controls how much flows from the supply (the fluid analogy is helpful here).

2.5 The Emitter Follower

The circuit of figure 2.2 shows two possible outputs. When we take the output V_E at the emitter then the circuit is the **emitter follower** and equation 2.3 applies. Frequently R_C is omitted and the collector is connected directly to V_{CC} (figure 2.3). Note that the collector is then at a fixed voltage and the emitter voltage is defined by equation 2.3, so there is a potential difference V_{CE} across the transistor which depends on how much current flows. When the transistor is 'off' (no current flowing) then V_E is at ground potential. At the other extreme, maximum current flows from the emitter then V_E can be no higher than V_{CC}, and is typically a few tenths of a volt below this. The useful output range of the circuit is limited to $0 < V_E < V_{CC}$ and it's important to note that the output can never be either (a) negative or (b) greater than the power supply voltage.

Voltage Gain

In general **voltage gain** is output/input voltage, however from equation 2.3 we can see that this is non-linear and it is more useful to define the **small signal gain** dV_E/dV_B which is

$$gain = \frac{dV_E}{dV_B} = 1 \tag{2.4}$$

i.e. the emitter follower circuit has **unity gain** for variations in the base voltage. Consider the graph of V_E versus V_B: the gradient is 1 so long as V_B is sufficiently large to ensure current flows through the transistor so for example if we apply an input

$$V_B = 5 + 0.1 \sin \omega t$$

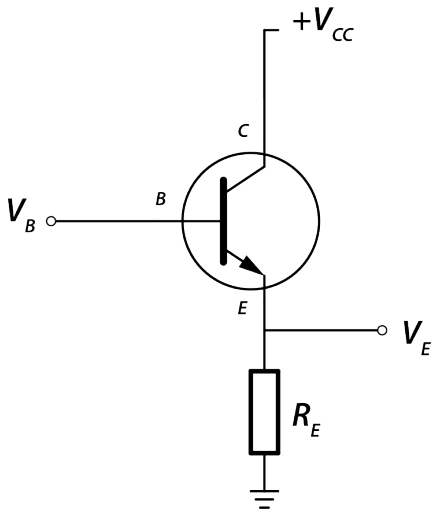


Figure 2.3: Emitter Follower Circuit

then we will have an output

$$V_E = 4.3 + 0.1 \sin \omega t$$

and we can see that the amplitude of the time-varying part of the signal is the same at the output as the input.

Input Resistance

It is useful to know the **input resistance** we would measure 'looking into' the base terminal. For a normal linear device this would be the input voltage divided by the input current but figure 1.2 implies that this will not be a constant value. However, assuming we have already brought the base-emitter junction into conduction we can ask the question: how much more current dI_B would we get if we increase the base voltage by a small amount dV_B?. Then the **small signal input resistance** is dV_B/dI_B. This is equivalent to taking the slope of the V-I curve at some point of the curve of our choosing which we will later describe as the **quiescent point**. In lectures we will show that

$$R_{in} = \frac{dV_B}{dI_B} = (1 + h_{FE})R_E \tag{2.5}$$

and since R_E will typically be 1 kΩ or so the input resistance will be high, which is generally good since it means that the base can be connected to a signal source such as a signal generator without **loading** the source; i.e. not much current will be drawn. We will look this in more detail during labs.

Output Resistance

We can also define the **small signal output resistance**, equivalent to the resistance 'looking into' the emitter and find

$$R_{out} = \frac{R_s}{1 + h_{FE}} \tag{2.6}$$

where R_s is the **source resistance**. For example, if we were to apply the signal to the base from a signal generator in the lab then $R_s = 600\,\Omega$.

Buffer

The significance of equations 2.5 and 2.6 is that the emitter follower is a good **buffer** circuit. We can look at this 2 ways:

1. at the output, the relatively high $600\,\Omega$ output resistance of a signal generator is transformed into about $6\,\Omega$
2. at the input, a relatively low load resistance $R_E = 1\,\text{k}\Omega$ is transformed into about $100\,\text{k}\Omega$

Either way, the result is that the buffer allows a signal generator to drive a signal onto a load resistor R_E with negligible loss of amplitude, even if $R_E < R_s$. This is the most common and useful application for the emitter follower circuit.

Power Gain

The circuit has a voltage gain of 1 but it does provide **current gain** in that the output current I_E is much greater than the input current I_B. This means that the circuit provides **power gain**. Note that we have the constraint that the output can only be a positive voltage, but we will soon see how to overcome this limitation and use the circuit for true AC signals.

Further Reading

The Art of Electronics (*Horowitz and Hill*) has a good introduction to the transistor

Chapter 2: Transistors

Sections 2.01 and 2.03 are most useful.

Lecture 3 Transistor Amplifiers and Switches

Equations 2.1 and 2.2 give

$$I_C \approx I_E$$

so a similar (large) current flows in R_C and R_E and this current is controlled by the much smaller current flowing into the base. A change in voltage at the emitter must also mean that there is a change in the voltage at the collector, and this will depend on the value of the resistors. We can consider two cases:

Case 1: $R_C = R_E$**.** Any change at V_E must result in a change of the same magnitude at V_C.

Case 2: $R_C > R_E$**.** Any change at V_E must result in a *bigger* change of at V_C.

The second case means that we can build a **voltage amplifier** since we can control V_E by changing V_B (equation 2.3)

3.1 Common Emitter Amplifier with Voltage Gain

When we take the output of the circuit as the voltage at the collector then the circuit is called the **common emitter amplifier**. The term *common emitter* refers to the fact that the emitter current goes to ground (through R_E, though sometimes this is omitted). The circuit has the capacity for **voltage gain**.

Consider a small change in the base voltage dV_B which will produce a small change in the collector voltage dV_C. Then we can define the **small signal voltage gain** which we will show in lectures is

$$gain = \frac{dV_C}{dV_B} = \frac{-R_C}{R_E} \tag{3.1}$$

We can set the gain of the circuit by changing the ratio of the resistor values. The gain is always negative, that is to say the sense of the signal is inverted. This makes sense, since an increase in the transistor current results in an *increase* in V_E and a *decrease* in V_C (the potential difference across *both* resistors increases, so the collector voltage must go down).

3.2 Operating Limits

It can not be stated too often that the transistor only operates when the base-emitter junction is conducting, however we must also be mindful that the output voltage has to stay within the range 0 to $+V_{CC}$. We will look at this in more detail in simulation labs, however by way of an example consider a design with $gain = -4$, $R_E = 1\,\text{k}\Omega$ and $+V_{CC} = 10\,\text{V}$ (It would be helpful to sketch this or simulate with LTSpice).

1. For $V_B < 0.7\,\text{V}$ there is no I_B and hence no I_C and we would describe V_C as being **tied** to 10 V. In this case, the transistor is '**off**'.
2. For $V_B > 2.7\,\text{V}$ the transistor is 'fully on' and $V_C \approx V_E$. Equation 2.1 no longer applies since increasing I_B does not result in any further increase in I_C. In this case, the transistor is '**saturated**'. This region of operation is useful when using the transistor as an ON/OFF switch, as we will see in section 3.5, but useless for amplification.
3. For $0.7 < V_B < 2.7\,\text{V}$ we have the useable amplifying range of the device. A change δV_B in the base voltage will produce a change $-4\delta V_C$ at the collector. Equation 2.1 applies and we can use the transistor as an amplifier. This is the transistor's '**linear**' operating region.

Case 3 is quite subtle and is key to the understanding of the transistor amplifier and the concept of small-signal gain. A key point to take from this is that we must choose a 'default' starting point for V_B ideally in the middle of this range, and then the signal to be amplified can be added to (or subtracted from) this. The process of setting up the circuit to achieve this is generally called **biasing** the transistor.

3.3 Biasing the base

For many applications e.g. audio, the signal we wish to amplify will be AC with an average value of zero and as we have seen, for the amplifier:

1. the input needs a DC offset such that the signal zero is at the **quiescent point** and
2. the output can only be positive

We can solve the first problem with a simple potential divider to **bias** the base-emitter junction into permanent conduction and then couple the signal on top of this using a capacitor. We would describe the signal as having been **AC coupled** into the amplifier. In order to set the size of the capacitor, we need to think of it as a high-pass filter and set the cut-off frequency appropriate for our signal. For example, for audio, a cut-off frequency of about 20 Hz or so would be fine.

The second problem can similarly be solved by AC coupling the output with another capacitor. Figure 3.1 shows a complete design with a gain of -8.

3.4 Design Procedure

It should be clear by now that designing a common-emitter amplifier from scratch is quite a tricky process which requires quite a detailed knowledge of the transistor behaviour and the characteristics of both the source signal and the load which the amplifier will drive. What follows is a summary of the design methodology which leads to the circuit of figure 3.1.

1. Choose the desired gain - with knowledge of the input signal's amplitude and the power supply voltage - and hence set the value of the collector resistor, using an emitter resistor of about 1 kΩ.
2. Choose the quiescent output voltage at the collector to be close to the middle of the transistor's output voltage range. The quiescent voltage will be the collector voltage when the input signal from the source is zero. In this case, the collector is chosen to be at about 6 V.
3. Calculate the quiescent collector current which results from this potential difference across R_C.
4. Assume $I_E = I_C$ and hence find the quiescent emitter voltage, here about 0.5 V.
5. The quiescent base voltage needs to be held at about 0.7 V above this by equation 2.3.
6. Calculate the base bias resistors which make a potential-divider circuit to fix the base voltage as per the previous step.
7. The input resistance now looks like a parallel combination of the two bias resistors and the base input resistance (equation 2.5). Use the desired cut-off frequency to calculate the input coupling capacitor ($\omega_c = 1/R_{in}C$), which forms a high-pass filter with R_{in}.

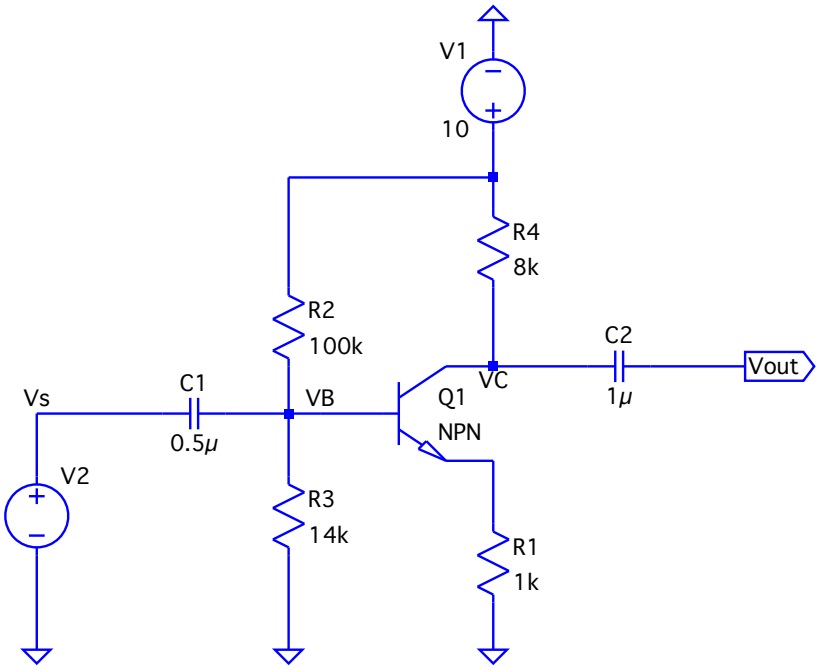


Figure 3.1: AC Coupled Common Emitter Amplifier

8. At the collector, the output resistance of the amplifier looks like a parallel combination of R_C and the resistance 'looking into' the collector, which is high (see section 3.4), so in practice this is just R_C. Set the output coupling capacitor using a similar cut-off frequency.

Output Resistance

For reasons we won't go into here, the resistance looking into the collector of a npn transistor is very high. We can think of it as being a 'sink' of current - like a current source but where the current goes in - and consequently it is possible to imagine that its apparent resistance is quite high. Consequently, as described in section 3.4, the effective output resistance of the common emitter amplifier is dominated by R_C which will be a few kΩ. This is really quite high - compare to the emitter follower - and this is one of the main drawbacks of the design. We can fix this, however, by connecting a second transistor in a follower configuration to reduce the overall output resistance of the design, and hence allow the circuit to successfully drive a low resistance load. We will look at this again during labs.

3.5 The Transistor as a Switch

We can use the current-gain capability of the transistor to control the current in a high-power device. Figure 3.2 shows an application known as the **transistor switch** where a low-current signal from a mechanical switch is used to control a high current into a bulb. Such a circuit could be used to control the headlights in a vehicle. The bulb draws 3 A from the supply (12 V battery). It is

undesirable to switch this with a mechanical switch for two reasons: firstly, the mechanical switch will generate sparks at this level of current, secondly, in case of a fault, a 3 A current inside the passenger compartment could lead to melted wires or worse. With this circuit, the high-current branch stays under the bonnet (battery, bulb and transistor) while only the low-current branch (mechanical switch) needs to be routed via the passenger compartment.

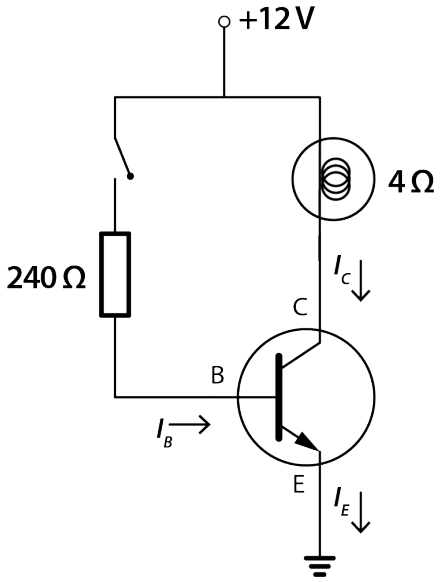


Figure 3.2: npn Transistor used to Switch a High-Power Load

Operating Principle

With the switch open, there is no base current and hence no collector current and the bulb is off. With the switch closed, the base-emitter junction looks like a diode and a current I_B flows.

$$I_B = \frac{12 - 0.7}{240} = 47\,\text{mA}$$

Equation 2.1 would lead us to think that 4.7 A flows through the transistor, but this would put the collector voltage below zero and is not allowed. Instead, the collector will drop to a voltage as close to zero as it can manage, resulting in the desired 3 A flowing through the bulb.

Saturation

This is an example of a transistor driven into the **saturation** region of operation, which is outside the conditions described on page 10. Specifically, we have violated condition 2 which requires a potential difference V_{CE} in order for equation 2.1 to be valid. In saturation, we are putting sufficient current through the base in order to ensure that the transistor looks like a nice low-resistance path to ground for the collector current (i.e. like a switch in the 'on' position'). Strictly speaking, we would only need 30 mA base current to get a 3 A collector current, but here we are using more current into the base just to ensure that the transistor is fully 'on'. It is important to note the

distinction between this rather 'brute-force' use of the transistor in the saturation region and the more subtle and tricky circuit designs required for amplifiers where equation 2.1 applies.

3.6 Power Dissipation in the Transistor

It was stated in section 3.5 that the collector goes very close to zero volts, however there are practical limitations and, depending on the transistor type, there will always be a small voltage drop V_{CE} even when the transistor has been driven hard into the saturation region by a large base current. This may be about 0.1 to 0.2 V. Since current flows through a potential difference there will be power dissipated in the transistor which may be $0.2 \times 3 = 0.6$ W. This is tiny compared to the 35 W dissipated in the bulb, but non-negligible for a transistor and we'd need to ensure that we choose a special device with a sufficiently high **power and current rating**.

Further Reading

The Art of Electronics (*Horowitz and Hill*)

Chapter 2: Transistors

The transistor switch is covered in section 2.02

Lecture 4 The Operational Amplifier

In lectures 2 and 3 we saw how hard it is to design a transistor circuit to amplify an AC voltage signal by a small factor such as 4. If we need the circuit to work with DC voltages, the design gets even harder. Nevertheless, using mainly transistors, it is fully possible to build all of the electronic circuits you're ever likely to need in a lab environment. *Horowitz and Hill* (The Art of Electronics) will show you how to do this. However, transistor circuits tend to be quite tricky to get right in practice. The **Operational Amplifier**, however, can take the pain and inconsistency out of discrete transistor circuits.

The op-amp actually pre-dates the transistor; the first op-amps used vacuum valves instead. These were used in early analogue computers where the designers needed simple circuits to perform mathematical *operations* (hence the name) such as addition, subtraction, integration, differentiation, as well as simple amplification. The op-amp provides repeatability and consistency, that is to say we can take one op-amp out of a circuit (maybe because we have broken it) and replace it with another and the circuit will behave exactly the same. This is often not the case with transistors as the value of h_{FE} can vary quite a bit from one part to the next. The key advantage of the op-amp is that we don't need to know what is going on inside the device. Even simple op-amps contain about 25 transistors and their designs have been refined over decades such that the device's behaviour can be abstracted into a simple set of rules. We can then use the op-amp to build really quite complex circuits which have a good chance of working first time. There are also a huge number of different op-amps available on the market with many different kinds of specialist properties. It is for these reasons that scientists and engineers tend to prefer the op-amp when constructing circuits. While there are still times when only a specialist discrete-transistor circuit will do, the op-amp is the 'work-horse' component of most circuit designs, and can be found inside almost every piece of equipment from a 'scope to a laptop.

4.1 Differential Amplifier

The operational amplifier is a **Differential Amplifier** which means that it amplifies the *difference* in the electrical potential between two inputs. This is illustrated schematically in figure 4.1. V_{in} is

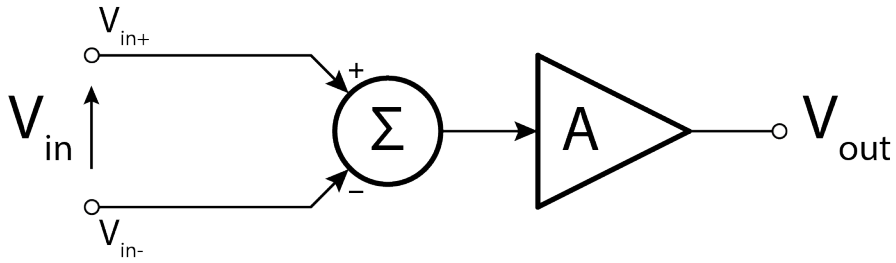


Figure 4.1: Differential Amplifier

the *potential difference* between the two inputs, i.e.

$$V_{in} = V_{in+} - V_{in-}$$

The amplifier has a gain A which can, in general, be both complex and frequency-dependent. Recalling AC circuit theory, we could write

$$\tilde{v}_{out} = \tilde{A}(\omega)\tilde{v}_{in}$$

This interpretation of *gain* is exactly the same as the gain of an RC filter as seen in the *Basic Electronics* course. The output voltage will, in general, have a phase-change with respect to the input given by $\arg(\tilde{A}(\omega))$. The amplitude of the output will be related to the amplitude of the input by a factor $|\tilde{A}(\omega)|$.

We will later see that the op-amp does indeed behave very much like a **'low-pass filter with voltage amplification'**. However, to help simplify the analysis of op-amp circuits, it will be sufficient to assume that we can replace $\tilde{A}(\omega)$ with a constant A_0 which is the magnitude of the gain at DC, i.e $A_0 = |\tilde{A}(\omega = 0)|$. This is a reasonable approximation at low frequencies, in which case

$$V_{out} = A_0 V_{in}$$

4.2 Op-Amp Characteristics

The gain A_0 is known as the **Open-Loop Gain**. Physically, the op-amp is usually supplied as an 8-pin **Integrated Circuit** or IC (figure 4.2).

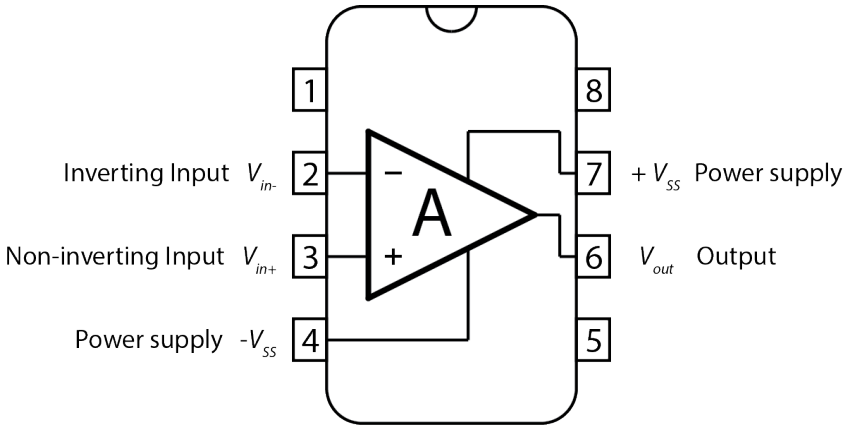


Figure 4.2: Op-amp connections ('pin-out') e.g. TL081 from *Texas Instruments*

An op-amp requires a minimum of 5 connections:

- The **Non-inverting Input** often called V_{in+} and labeled '+'
- The **Inverting Input** often called V_{in-} and labeled '-'
- An output V_{out}
- 2 power-supply connections $\pm V_{ss}$ often +15V and -15V

The device is governed by three key relations:

$$V_{out} = A_0(V_{in+} - V_{in-}) \tag{4.1}$$

On the manufacturer's data-sheet, A_0 is commonly described as 'large signal voltage gain', or similar, and it will usually be large number: 10^5 or more.

Because it is an active device, requiring power-supply connections (like the transistor) the output cannot exceed the power supply 'rails' (voltages) so

$$-V_{ss} \leq V_{out} \leq +V_{ss} \tag{4.2}$$

If we do try to drive the op-amp output to levels greater than the power supply rails then we will see **Clipping** where the signal is sharply trimmed-back to the $\pm V_{ss}$ levels. In practice, many op-amps will 'clip' a few 100 mV below V_{ss}, so it's a good idea to design op-amp circuits which don't require the full-range of $\pm V_{ss}$ on the output. Also, while the op-amp is a *fast* device the output cannot change instantaneously and we find there is a maximum **Slew Rate**

$$\left| \frac{d}{dt} V_{out} \right|_{\max} = S \tag{4.3}$$

Where the slew-rate S is typically of the order $1\,\text{volt}/\mu\text{s}$.

Figure 4.3 gives a 'model' representation of the op-amp. The potential difference between the two inputs is $\delta v = V_{in+} - V_{in-}$ and the output is $V_{out} = A_0 \delta v$. The input resistance is high ($> 10^6\,\Omega$) and output resistance is low (a few tens of Ω). These facts, taken together with equations 4.1 through to 4.3, give a quite realistic representation of op-amp behaviour, which is summarised in table 4.1.

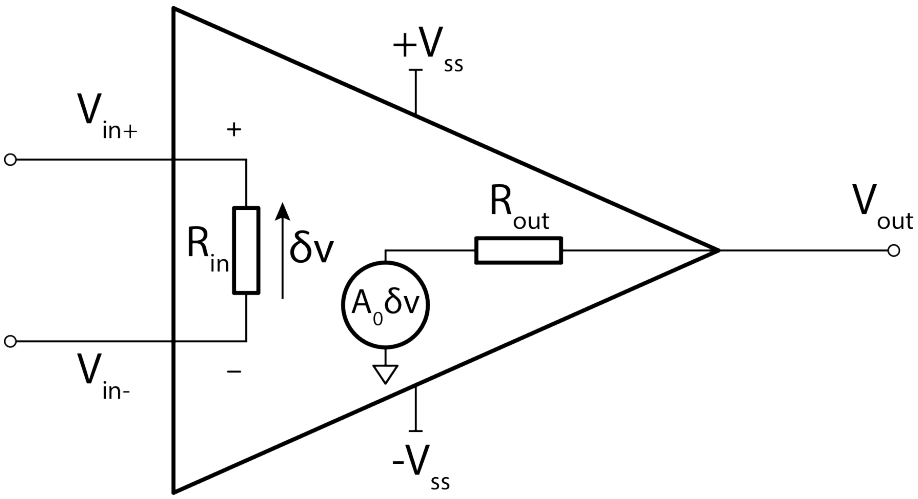


Figure 4.3: Model Representation of the Operational-Amplifier

4.3 The Comparator

The simplest op-amp circuit is the comparator (figure 4.4). This is very effective for comparing two voltage levels and deciding which is greater. Due to the very high gain A_0 and slew-rate S, the op-amp will respond almost instantaneously to any potential difference between the inputs $\delta v = V_A - V_B$

Table 4.1: Op-amp summary characteristics

Parameter	**Typical value**
Power Supply $\pm V_{SS}$	$\pm 15\,\mathrm{V}$
Differential inputs	V_{in+}, V_{in-}
Input resistance R_{in}	$> 10^6\,\Omega$
Open loop gain A_0	10^5
Output V_{out}	$A_0(V_{in+} - V_{in-})$
Output range	$-V_{SS} < V_{out} < +V_{SS}$
Slew rate S	$1\ \mathrm{V}/\mu\mathrm{s}$
Output resistance R_{out}	$\sim 20\,\Omega$

greater than a few 100's μV or so, by going very quickly to one of the two power supply voltages. This is fully consistent with equations 4.1, 4.2 and 4.3. The behaviour is as follows:

$$V_{out} = \begin{cases} +V_{ss} & A > B \\ -V_{ss} & A < B \end{cases} \tag{4.4}$$

We can do something useful with this such as controlling a switch. For example if the two voltage inputs represent temperatures in an experiment we can do

```
IF
    temperature_A > temperature_B
THEN
    switch_on_heater
```

The output could, for example, control the on/off status of a transistor switch. The comparator relies on the op-amp being always in **saturation** so that the output is either $\pm V_{ss}$. While this is useful, we normally want the output to have some proportional relationship with the input in order to use the op-amp as an amplifier or other linear circuit element. The comparator is the *only* design we'll look at which uses the op-amp in saturation; in all other circuits we will want to keep $-V_{ss} < V_{out} < +V_{ss}$ at all times.

Note that in figure 4.4 the power-supply voltages $\pm V_{ss}$ are not included in the diagram. This is normal practice, to avoid clutter on diagrams, but it's important to remember to wire these connections when building a circuit for real.

4.4 The Voltage-Follower or Buffer

The **Voltage Follower** circuit is made by looping the output voltage V_{out} back to the inverting input. Note that figure 4.5 shows the inverting and non-inverting inputs at swapped locations compared to figure 4.4. There is no convention as to which way up we show the op-amp so it's always important to check the labelling of the inputs. Since A_0 is large, the *only* feasible solution to equation 4.1

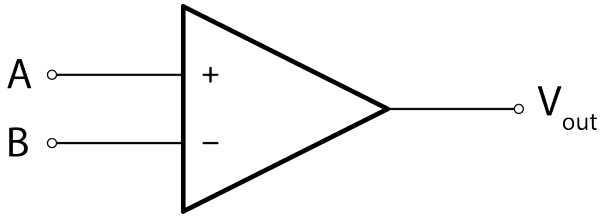


Figure 4.4: Op-Amp Comparator

requires $V_{in+} - V_{in-} \approx 0$ so the follower is governed by

$$V_{out} \approx V_{in}$$

The practical application of a circuit where $V_{out} = V_{in}$ comes when we consider that the input resistance of the circuit is very high, so it draws negligible current, while its output resistance is very low. This is useful where we have a source which is 'weak' and easily loaded (typically having a high source resistance). We can **Buffer** the source using a voltage-follower. The op-amp is then able to drive any current needed for a low-resistance load without pulling much current out of the source. This avoids the problem of **loading** the source.

The buffer circuit is the op-amp equivalent of the emitter-follower transistor buffer circuit seen in section 2.5. The op-amp buffer is superior since

- the input resistance is much higher;
- the output resistance is much lower;
- it works for both positive and negative input voltages;
- it works for DC voltages;
- for AC voltages, there's no need to AC-couple the input and output.

The only drawback is that the op-amp's output current is typically limited to 40 mA or so, which means that the op-amp can't directly drive a large voltage signal onto a low-resistance load. We can solve this problem, as we'll see in labs, using a clever arrangement of transistors on the output.

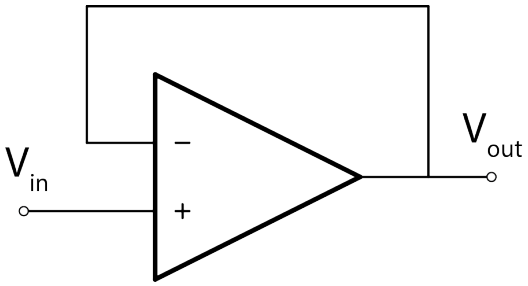


Figure 4.5: Voltage-Follower or Buffer circuit

Further Reading

The Art of Electronics (*Horowitz and Hill*)

Chapter 3: Feedback and Operational Amplifiers

Note however that Horowitz and Hill *start* with a discussion of feedback, which will be the subject of our *next lecture* (Lecture 5).

Lecture 5 Negative Feedback

The buffer is an example of a circuit using **feedback** to reduce the overall gain of the op-amp circuit.

5.1 Closed-loop amplifiers

When we connect the output (or some fraction of it) back to the input of an op-amp then we have created a feedback loop. In the case of the voltage-follower we have returned *all* of the output back to the input, which creates a feedback amplifier with a *gain* of 1. In practice, all useful op-amp circuits (except the comparator) use feedback, and when the output is connected back to the inverting input then this is described as **negative feedback** or **closed-loop**. All our subsequent op-amp circuits will use negative feedback. Figure 5.1 illustrates *closed-loop* negative feedback schematically; compare this with the schematic of the *open-loop* amplifier of figure 4.1.

> In the general case, some fraction of the output β (where $0 < \beta < 1$) is fed back to the input in such a way that the output tends to *reduce* the voltage at the input. This is the definition of negative feedback.

In lectures we will show that the gain of the overall circuit becomes

$$gain = \frac{V_{out}}{V_{in}} = \frac{A_0}{1 + \beta A_0} \tag{5.1}$$

Equation 5.1 gives the **closed-loop gain**, which we must distinguish from the open-loop gain A_0. The buffer is a circuit where $\beta = 1$ (all output goes back to the inverting input). Open-loop gain A_0 is large, and closed-loop gain is ≈ 1.

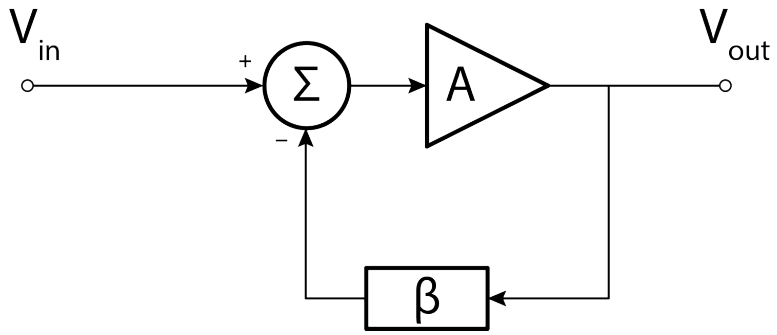


Figure 5.1: Feedback Amplifier Model

Figure 5.2 illustrates how an op-amp is connected for negative feedback. The op-amp implements everything within the dashed-box, and everything else is done with external connections. β is easily implemented with a resistive *potential divider*.

Note that, in the case we assume that the open-loop gain is very large, i.e. $A_0 \to \infty$ then the closed-loop gain

$$gain \to \frac{1}{\beta} \tag{5.2}$$

The importance of this is that:

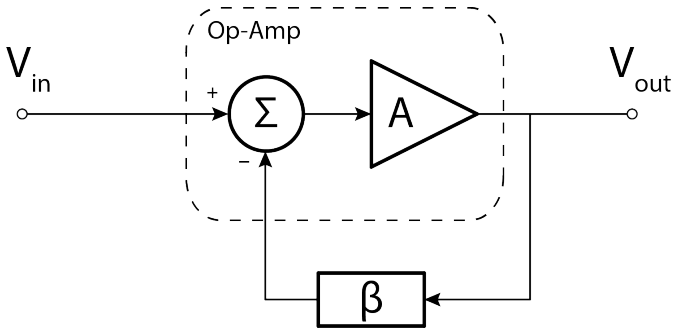


Figure 5.2: Negative Feedback with the Operational Amplifier

When using an op-amp with very high open-loop gain, the behaviour of the circuit depends almost entirely on the feedback circuit and *not* on the particular op-amp used.

5.2 Ideal Amplifier

This observation leads to the concept of the **ideal operational amplifier** which has the characteristics of table 5.1, compared to a typical 'real-world' op-amp such as the LM741.

Parameter	Ideal Op-Amp	LM741
A_0	$\to \infty$	2×10^5
R_{in}	$\to \infty$	$2\,\mathrm{M}\Omega$
R_{out}	0	$\sim 20\,\Omega$
S	$\to \infty$	$0.5\,\mathrm{V}/\mu\mathrm{s}$

Table 5.1: Ideal amplifier vs. real amplifier

If we use the **ideal amplifier approximation** of table 5.1 then the analysis of circuits containing op-amps becomes *much* easier. We shall see that, for most practical purposes, the assumption of ideal behaviour gives results which are very close to reality, so for most every-day op-amp circuits, the ideal amplifier approximation is all we need. It's only when we start to use op-amps towards the extremes of their capabilities then we encounter problems. For example

1. at very high-gains;
2. at very high frequencies;
3. or to drive large currents.

So long as we avoid the three cases above, the assumption if ideality gives valid results. We will look at some of these exceptions in Lecture 6. However, for the rest of Lecture 5, we will assume our amplifiers are ideal. Note that in some texts you may see the ideal amplifier referred to as the **infinite gain approximation**.

5.3 Golden Rules

The assumption that the op-amp is 'ideal' leads to a couple of circuit rules which *greatly* simplify the analysis of op-amp circuits, meaning that we can forget about equations 4.1 through to 4.3, and rely instead on the two **golden rules**:

> **Rule 1** The output will act in such a way as to drive the potential difference between the inputs to zero.
>
> **Rule 2** The inputs draw no current.

Rule 1 follows as a direct consequence of the assumption $A_0 \to \infty$ and Rule 2 follows from $R_{in} \to \infty$.

As already mentioned, **ideal amplifier approximation** has a high level of validity and hence circuit analysis using the *golden rules* gives a good representation of the true behaviour.

With the assumption of the ideal amplifier, we can use the golden rules to analyse the behaviour of some common op-amp circuits.

5.4 Non-Inverting Amplifier

This circuit (figure 5.3) illustrates how a feedback fraction β of the output voltage is 'fed-back' to the inverting input. R_1 and R_2 form a potential-divider such that the feedback voltage V_{in-} is

$$V_{in-} = V_{out}\frac{R_2}{R_1 + R_2} \tag{5.3}$$

Using golden rule 1, $V_{in+} = V_{in-}$ so

$$gain = \frac{V_{out}}{V_{in}} = 1 + \frac{R_1}{R_2} \tag{5.4}$$

By the definition of β as '*the fraction of the output fed-back to the input*'

$$\beta = \frac{V_{in-}}{V_{out}} = \frac{R_2}{R_1 + R_2}$$

so it is clear that

$$gain = \frac{1}{\beta}$$

Note that the voltage-follower circuit of section 4.4 is the same as the non-inverting amplifier with $\beta = 1$.

5.5 Inverting Amplifier

This common circuit is shown in figure 5.4. The non-inverting input is grounded so golden rule 1 says that the inverting input sits at zero volts.

$$V_{in-} = 0$$

This is known as a **virtual ground**; it is not actually connected to ground, but the feedback action keeps this point in the circuit at zero volts, which allows us to make easy calculations of the currents in the two resistors.

$$i_1 = \frac{V_{out}}{R_1}$$

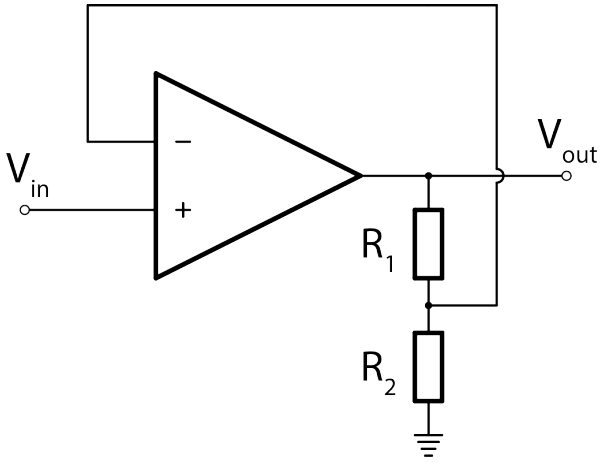


Figure 5.3: Non-Inverting Amplifier

$$i_2 = \frac{V_{in}}{R_2}$$

Rule 2 assumes that no current flows into the inverting input of the op-amp, which requires (Kirchhoff's Current Law)

$$i_1 + i_2 = 0$$

Giving

$$gain = \frac{V_{out}}{V_{in}} = -\frac{R_1}{R_2} \tag{5.5}$$

The output voltage is a scaled and inverted copy of the input. Sometimes we might want the inversion, and sometimes it doesn't matter. For the cases when it is a problem, use the non-inverting amplifier. For many AC applications, e.g. audio, the inversion of the signal is not so important.

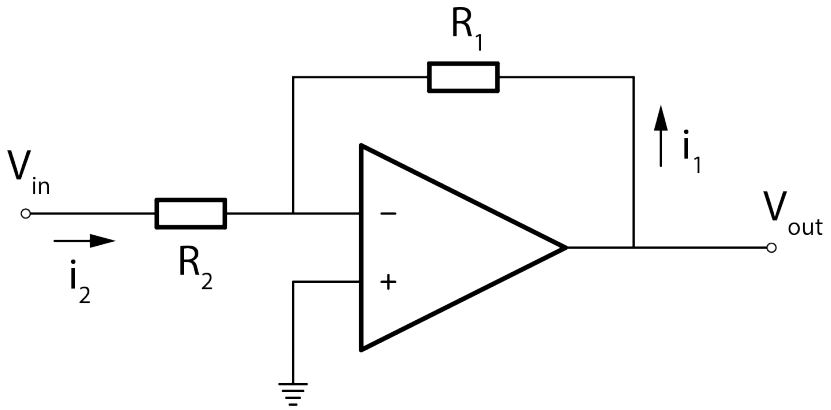


Figure 5.4: Inverting Amplifier

5.6 Mathematical Operations

Since our voltage signals typically represent physical quantities, e.g. as measured by sensors, it is useful to be able to perform mathematical operations such as addition, subtraction, integration and differentiation. The circuits to perform these functions are all variations on the basic *inverting amplifier*. The circuits are summarised in figure 5.5 and their analysis (left as an exercise) follows the same method as section 5.5.

Hint:
For the integrator and differentiator, recall the differential relation between current and voltage for the capacitor. For the summing amplifier, remember that the inverting input is a virtual ground. The analysis for the difference amplifier requires a bit more effort, but the same principles are used. Start by calculating the potential at the non-inverting input then use golden rule 1.

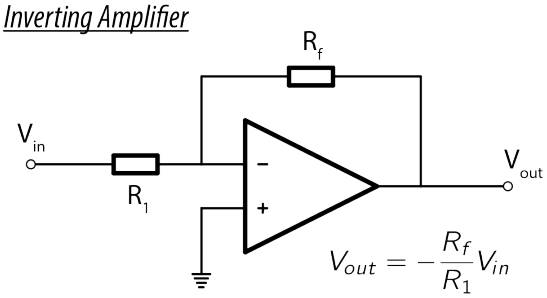
Further Reading

The Art of Electronics (*Horowitz and Hill*)

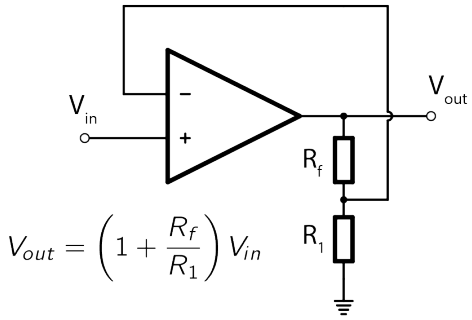
Chapter 3: Feedback and Operational Amplifiers

Sections 3.01-3.06, 3.09 (part), 3.18-3.19 (part)

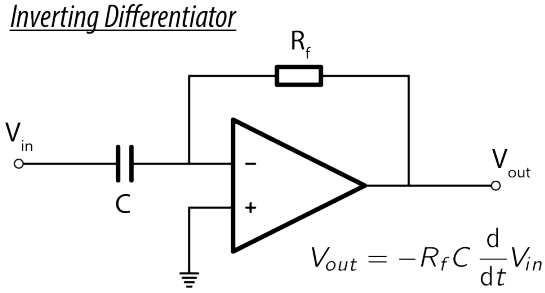
Inverting Amplifier



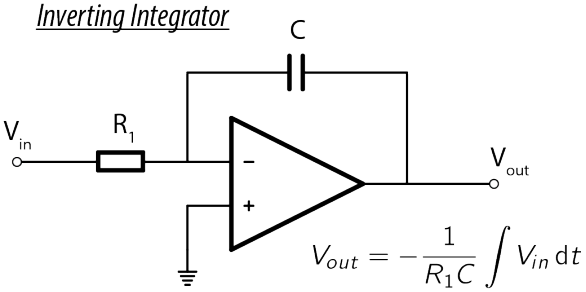
Non-inverting Amplifier



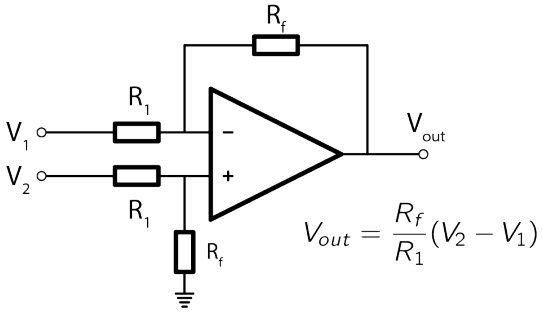
Inverting Differentiator



Inverting Integrator



Difference Amplifier



Inverting Summing Amplifier

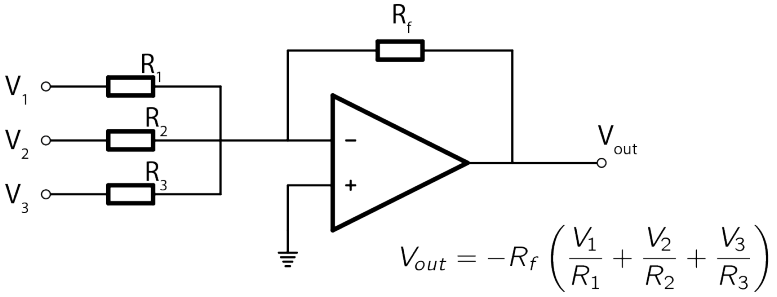


Figure 5.5: Common Op-Amp Circuits

Lecture 6 Real-World Op-Amps

All of the circuit analyses in Lecture 5 made the assumption of the ideal amplifier, however under some circumstances it is necessary to take into account 'real-world' non-ideal behaviour, and this also will help to illustrate how feedback really works. To do this, we will take another look at the quantities of table 5.1.

6.1 Finite Gain

For finite open-loop gain A_0 it is clear that there must be a potential difference between the inputs of the op-amp otherwise the output would always be zero, according to

$$V_{out} = A_0(V_{in+} - V_{in-})$$

The buffer circuit of figure 4.5 helps to illustrate this. Here:

$$V_{in} = V_{in+}$$

$$V_{out} = V_{in-}$$

So we can re-arrange to get

$$V_{out} = \frac{V_{in}A_0}{1 + A_0}$$

For an input $V_{in} = 1\,\text{V}$, the ideal amplifier has an output $V_{out} = 1\,\text{V}$ and the real op-amp will give $V_{out} = 0.99999\,\text{V}$, which is close enough for most purposes. Note, however, that when the circuit adds some closed-loop gain, e.g. figure 5.3, then the discrepancy between real and ideal behaviour grows. The assumption of ideal behaviour under closed-loop requires that, in equation 5.1, $\beta A_0 + 1 \approx \beta A_0$, that is to say

$$\beta A_0 \gg 1 \tag{6.1}$$

> The parameter βA_0 is called the **loop-gain** for the circuit and so long as $\beta A_0 \gg 1$ we can assume that the *gain* of the real circuit will, at low-frequencies, be similar to the prediction made using the golden rules.

6.2 Finite Input Resistance

In figure 4.3 the input resistance R_{in} is modelled as the resistance between the op-amp's inputs. If we applied a potential difference of 1 V between the inputs, we would expect an input current of about $0.5\,\mu\text{A}$ as a result. However, when the op-amp is used in closed-loop, then the action of feedback drives the potential difference between the inputs to the level of some μV, which means that the current drawn into the inputs is truly tiny, and the *apparent* input resistance is very much larger than R_{in}. We can define the apparent input resistance, closed-loop, as $R_{in,CL}$ then it can be shown that

$$R_{in,CL} = (1 + \beta A_0)R_{in}$$

> Closed-loop operation dramatically increases the apparent input resistance of the operational amplifier. Therefore, golden rule 2 is almost always a valid assumption, and we can conclude that the op-amp's input draws negligible current.

6.3 Finite Output Resistance

So far, we calculated output voltages 'open-circuit', i.e. assuming that the op-amp circuit's output isn't connected to any load. Since the op-amp drives no load, we can be confident that the output voltage will be as predicted. However, as we have seen previously, output resistance becomes a problem when a device is required to drive some current, resulting in a potential drop across the device's output resistance (see figure 4.3). However, as with input resistance, feedback improves the situation. Again, the buffer circuit is useful to understand how this works. Look again at figure 4.5. If we connect a resistor between V_{out} and ground, then the op-amp is required to drive some current, and we would expect V_{out} to drop. However, feedback works by always acting so as to drive the potential difference between the inputs towards zero, and consequently the op-amp's output voltage will rise to counteract the 'loading' of the amplifier's output. To be clear, this can only happen because the comparison point - V_{out} - is *after* the op-amp's output resistance.

This brings us to a key point in the understanding of a feedback loop as a self-regulating system which acts so as to maintain a constant output. In this case, the constant output parameter is the voltage. In essence, the op-amp will try to drive whatever current is necessary to maintain the desired voltage at it's output. An analogy which may be useful is the 'cruise control' on a car, which is setup to maintain a constant speed. If the car encounters a hill then the speed would tend to reduce, which the system detects by measuring the difference between the desired and actual speed, and so acts to increase the engine output to drive this difference back to zero. The op-amp feedback loop is doing exactly the same thing. The comparison-and-adjustment 'action' is done automatically and continuously through the differencing and amplification of the input voltages, as encoded in equation 4.1.

> As a result of this self-regulation on the output, the apparent output resistance is dramatically reduced when the op-amp is used closed loop, and consequently the op-amp's output appears to act as a very good *voltage source* of negligible internal resistance.

It can be shown that

$$R_{out,CL} = \frac{R_{out}}{1 + \beta A_0}$$

6.4 Frequency Response

There are a number of factors affecting the frequency response of the op-amp but it will be clear that the finite slew-rate means that the op-amp cannot amplify signals of arbitrarily high frequency. Figure 6.1 shows, with the solid-line, the natural open-loop gain (magnitude part) of the Bode plot for a typical op-amp with $A_0 = 10^5$ (or 100 dB). This is the curve we would measure if we applied a very small amplitude AC input to the open-loop op-amp, measured the corresponding output amplitude, and then repeated this for all frequencies along the horizontal axis. Note the following features:

- The general shape of the gain curve is very similar to the curve for the low-pass filter. This is in fact deliberate. The reasons for this are beyond the scope of this course, but the reduced gain at high frequencies guarantees that the op-amp will remain stable (feedback systems with high-gain can be prone to uncontrolled oscillation).
- The gain starts to 'roll-off' at the -3 dB frequency, here 10 Hz, so in this case we would say that this op-amp has an **open-loop bandwidth** of 10 Hz.

- At a frequency of 1 MHz, the gain has fallen to a value of 1 (0 dB).

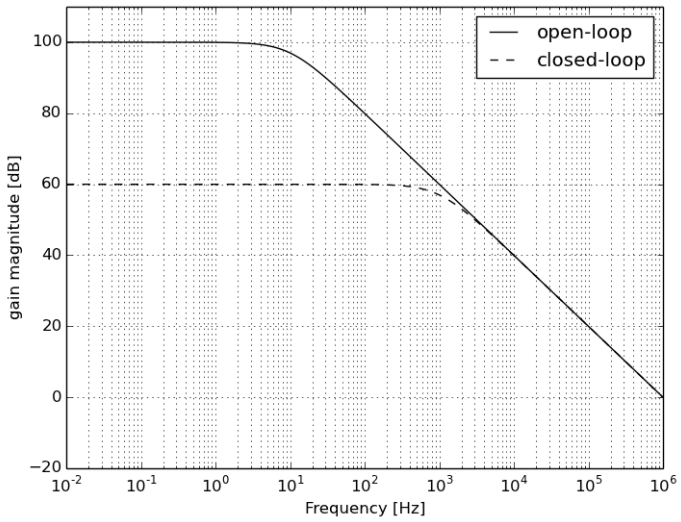


Figure 6.1: Frequency Response of an Op-Amp, Open and Closed Loop examples

The gain of the op-amp can be described by the function

$$\tilde{A}(f) = \frac{A_0}{1 + jf/f_0} \tag{6.2}$$

At low frequencies, and ultimately DC ($f \to 0$) then $\tilde{A} \to A_0$, which we have previously used for the gain of the device. Note the similarity of equation 6.2 to the equation for the gain of the low-pass RC filter. The op-amp has a DC gain A_0 and a cut-off frequency f_0. Taking the magnitude of $\tilde{A}$, using $A_0 = 10^5$ and $f_0 = 10\,\text{Hz}$, then expressing this using dB, results in the open-loop gain curve of figure 6.1.

6.5 Closed-Loop Gain

In section 5 we analysed feedback using the scheme of figure 5.1 assuming that the op-amp's gain was a constant. If we repeat the analysis using the true, complex and frequency dependent gain $\tilde{A}(f)$ then we find a complex and frequency dependent closed-loop gain $\tilde{G}(f)$ given by

$$\tilde{G}(f) = \frac{\tilde{A}(f)}{1 + \beta\tilde{A}(f)} \tag{6.3}$$

Substituting using equation 6.2 and re-arranging gives

$$\tilde{G}(f) = \frac{\left[\frac{A_0}{1+\beta A_0}\right]}{\left[1 + \frac{jf}{f_0(1+\beta A_0)}\right]}$$

We can define new constants A_c and f_c such that

$$\tilde{G}(f) = \frac{A_c}{1 + jf/f_c}$$

where

$$A_c = \frac{A_0}{1 + \beta A_0}$$
$$f_c = f_0(1 + \beta A_0)$$

and by similarity with equation 6.2 it is clear that A_c is the closed-loop gain at DC, while f_c is the closed-loop cut-off frequency. $G_{dB} = 20 \log_{10} |\tilde{G}(f)|$ is displayed as the dashed-line in figure 6.1 for $\beta = 10^{-3}$.

6.6 Gain-Bandwidth Product

Using the -3 dB definition for bandwidth of the amplifier, the product of the DC gain and the bandwidth for the open-loop op-amp is $A_0 f_0$, and similarly when using feedback

$$A_c f_c = \frac{A_0}{1 + \beta A_0} \times f_0(1 + \beta A_0) = A_0 f_0 \tag{6.4}$$

This is independent of β, and so independent of any closed-loop gain we may choose, and just depends on the key characteristics of the op-amp. This is a parameter known as the **gain-bandwidth product** (units Hz), which is usually quoted on the data-sheet. Further, by writing equation 6.4 as $A_c f_c =$ constant, it's easy to see that we can increase the bandwidth of an op-amp circuit by reducing the closed-loop gain, or *vice-versa*.

Further Reading

The Art of Electronics (*Horowitz and Hill*)

Chapter 3: Feedback and Operational Amplifiers

Sections 3.11 and 3.12

Lecture 7 Digital Logic

Up to now we only encountered **analogue** signals which we separated into characteristically AC or DC quantities. In the physical sciences, a signal typically represents some physical quantity which is to be measured. The signal is generated by a **sensor** such as an accelerometer, which could be attached to an oscillating mass (generating an AC signal) or a temperature-sensor monitoring the status of some chemical reaction (typically thought of as a DC quantity, even though it is of course varying with time). In both cases, the signal is represented by a voltage which can be thought of as a continuous function of time. An example of such as sensor would be the LM35 temperature sensor, which produces an output voltage of 10 mV per degree centigrade[2]. We could use an op-amp amplifier to boost the output by a factor of 10, however we would need to keep in mind that any error in the gain of our amplifier will result in an error in our temperature measurement. With good design and careful checking, the error can be minimised and quantified, but it is clear that the more complicated the electronics, the more the possibility exists for error to creep in. One of the motivations for **digital electronics** is to convert an analogue voltage into a 'digital' number which can be manipulated, stored, reproduced and transmitted without - in principle - any further loss in quality. Digitally-stored measurements can be processed in ways impossible by purely analogue techniques, for example we can perform a Fast Fourier Transform (FFT) in order to discover the frequency-content of a dataset. The process of converting an analogue signal into a digital form is called **Analogue to Digital Conversion** (ADC) and often we would wish to be able to convert a digital quantity back into an analogue voltage (**Digital to Analogue Conversion**, DAC) in order to interface a computer to some experimental equipment.

Another motivation for digital electronics is the possibility to build **control logic** for making decisions, based on sensor inputs, for example, and for actions to be taken accordingly. Simple control logic exists in a home heating system, which controls the heating ON/OFF status using the input from a room thermostat. Digital electronics is not essential for this, but digital circuits are now the norm in almost all control applications from a toaster to an aircraft auto-pilot.

We encountered digital logic with the comparator circuit, which uses an open-loop op-amp to compare two voltages. We are only interested in the op-amp's positive or negative saturation outputs so it has one output state which represents $A > B$ and the other possible state is $A < B$.

7.1 Boolean Algebra

Equivalently, the comparator answers the question 'is $A > B$?' with a **True (T)** or **False (F)** answer. Logic with only true or false answers is **Boolean**. There are a limited set of operations which can be performed in Boolean algebra. The statement 'apples are fruit' is true while 'potatoes are fruit' is false, so 'apples are fruit' AND 'potatoes are fruit' is FALSE while 'apples are fruit' OR 'potatoes are fruit' is TRUE. Since potatoes are not fruit, we deduce that NOT 'potatoes are fruit' is TRUE. The last operation is exclusive-or which requires one and only-one assertion to be true, hence 'apples are fruit' XOR 'pears are fruit' is FALSE. By similar arguments, we can build up **truth-tables** which define all the possible inputs and outputs for these four operations. In table 7.1, each of the Boolean variables A, B represents a quantity which is itself either TRUE or FALSE. In some logic schemes, TRUE can be represented by the number '1' (one) and FALSE by the number '0' (zero). This is a **binary** number scheme, and we will look at the digital representation of binary numbers in lecture 8.

[2] Datasheet: http://www.ti.com/lit/ds/symlink/lm35.pdf

A	B	A AND B
F	F	F
F	T	F
T	F	F
T	T	T

A	B	A OR B
F	F	F
F	T	T
T	F	T
T	T	T

A	B	A XOR B
F	F	F
F	T	T
T	F	T
T	T	F

A	NOT A
F	T
T	F

Table 7.1: Truth Tables for Boolean Operations

7.2 Transistor-Transistor Logic Levels

In order to implement logic functions with electronic circuits it is necessary to define the voltage levels which represent TRUE and FALSE, or indeed '1' or '0', and also known as **logic high** and **logic low**. One popular 'family' of devices is known as 'Transistor-Transistor Logic' or **TTL** for short. These devices are made with bipolar junction transistors and use +5 volts to represent TRUE/1/high and zero to represent FALSE/zero/low. There are many different types of logic 'families', such as the 3.3 V CMOS low-power family, however we will use 5V TTL logic in lab. Logic devices in LTSpice are simply implemented as 'one volt logic', with the input threshold set at 0.5 V. Table 7.2 gives a comparison between TTL and LTSpice implementations.

Binary	Logical	Level	TTL Voltages	LTSpice
0	FALSE	Low	0 V (0..0.8 V)	0 V (<0.5 V)
1	TRUE	High	+5 V (2.4..5.5 V)	1 V (>0.5 V)

Table 7.2: TTL and LTSpice Logic Levels Compared

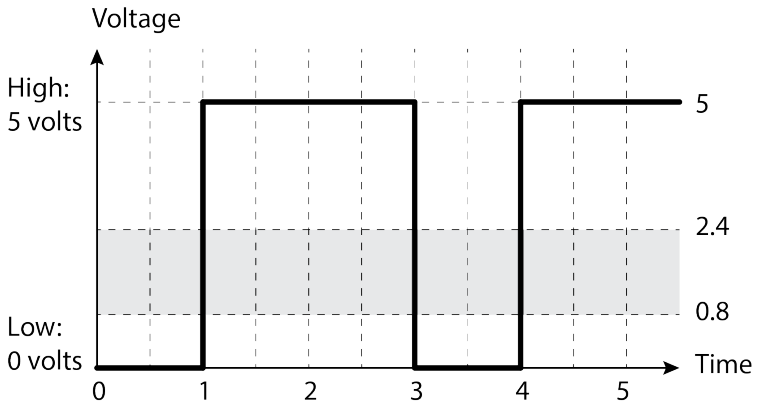


Figure 7.1: TTL Logic Levels Illustrated

Note that TTL recognises any voltage on an input between zero and 0.8 volts as representing FALSE or 'logic low' whereas any voltage in the range 2.4 to 5.5 volts represents 'logic high' (TRUE). Figure 7.1 illustrates this with a digital signal (thick black line) which - if present at the input of a TTL device - would be interpreted as logic low for time $0 < t < 1$ and logic high for time $1 < t < 3$. Note the 'no-man's land' between 0.8 and 2.4 volts. This avoids the possibility of uncertainty about the state of an input signal. The transition from 'low' to 'high' is rarely as sharp as shown in figure 7.1 and real signals will have finite rise-times, and also be subject to noise.

7.3 Resistor-Transistor Logic Circuits

The TTL logic functions described in section 7.2 are implemented with transistor designs which are not immediately intuitive so it will be useful to look at their predecessor known as **Resistor-Transistor Logic** or **RTL**. This was the standard for implementing digital logic in the early 1960's, but was superseded by TTL due to its reduced power-consumption. Nevertheless, RTL is worth studying since it illustrates how logic functions can be built from npn transistors and resistors. Figure 7.2 shows an inverter or NOT function implemented using RTL, and figure 7.3 shows the LTSpice simulation (horizontal axis is the input voltage A). When A is low then the transistor is off, there is no collector current so the output is high. Conversely when A is high then there is sufficient base current to saturate the transistor and hence the output is low. Note that this circuit is compatible with TTL logic-levels: any voltage $< 0.8\,\mathrm{V}$ on the input is interpreted as logic-low, and A$> 2.4\,\mathrm{V}$ is interpreted as logic-high.

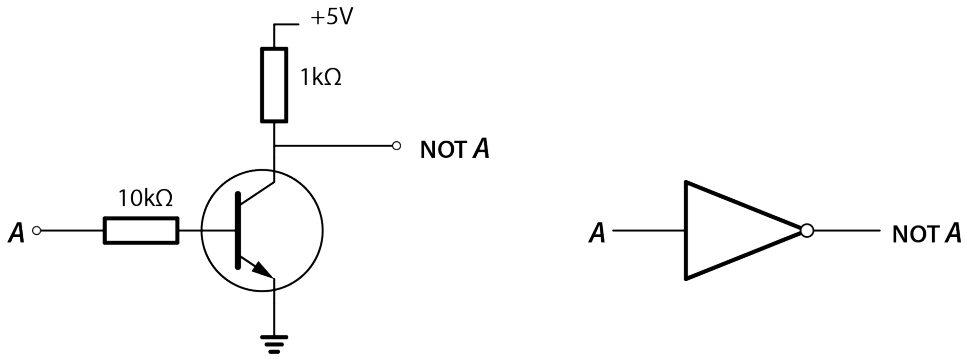


Figure 7.2: NOT function Implemented with Resistor-Transistor Logic. The figure on the right is the circuit symbol.

7.4 Logic Gates

As with the op-amp, rather than building logic functions with individual transistors, we use ready-made designs packaged as an integrated circuit. The devices are known as **logic gates** and their circuit symbols are shown in figure 7.4. The figure includes the NAND and NOR gates, which are the same as the AND and OR gates, but with their outputs inverted. This is indicated with a circle on the output. Consequently, NAND $\equiv$ NOT AND and NOR $\equiv$ NOT OR. The truth-tables for these gates are given in table 7.3, here using the binary representation, and the circuits are shown in figure 7.5. For the NAND, there are two transistors in series so both need to be saturated in order for the output to be low. Equivalently, if any input is low then the associated transistor turns off and the

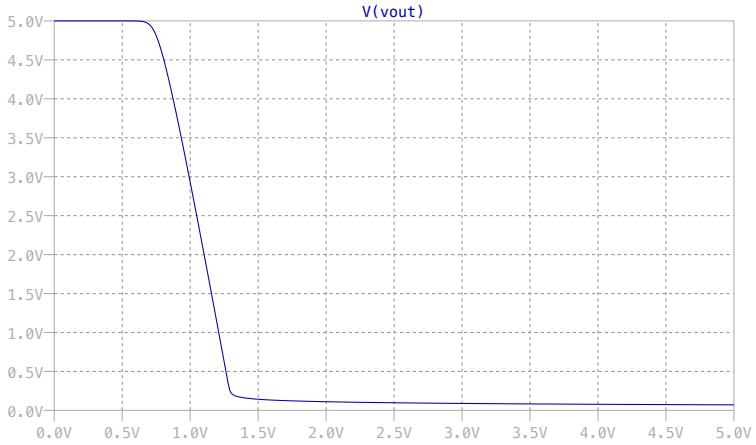


Figure 7.3: LTSpice Simulation of figure 7.2

output will be high. Figure 7.5 also shows a NOR function where the transistors are connected in parallel and hence either going into saturation causes the output to go low.

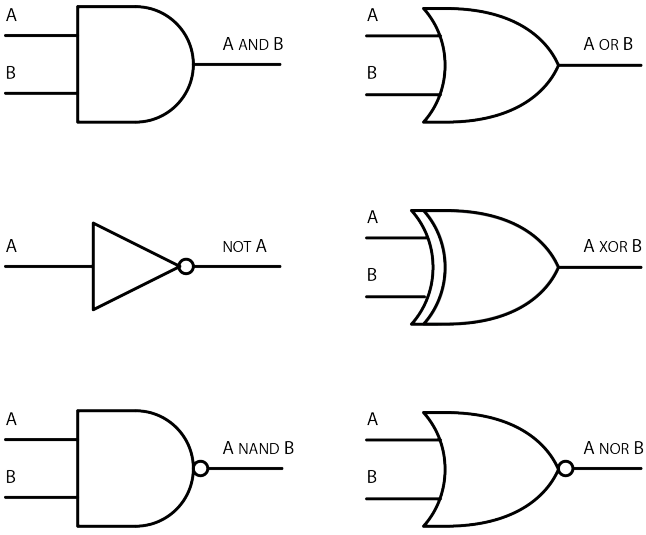


Figure 7.4: Circuit Symbols for the Common Logic Gates

Note that in digital electronics, **inversion** is synonymous with the NOT operation. We might write

that NOT A is equivalent to 'A inverted'. Sometimes, 'A inverted' is written with a *over-bar* i.e. NOT A = $\overline{A}$.

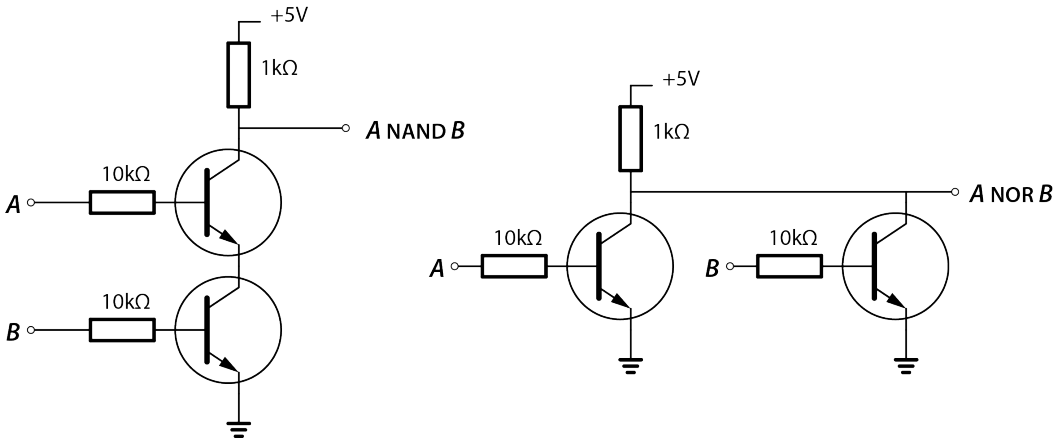


Figure 7.5: NAND and NOR Implemented with Resistor-Transistor Logic

Logic gates are **active** devices - they require a power supply and ground in order to function, as shown in figures 7.2 and 7.5. Each integrated-circuit, or 'chip' will have two pins reserved. One is for 'power' (usually +5 V) and the other is connected to ground.

A	B	A AND B
0	0	0
0	1	0
1	0	0
1	1	1

A	B	A OR B
0	0	0
0	1	1
1	0	1
1	1	1

A	B	A XOR B
0	0	0
0	1	1
1	0	1
1	1	0

A	NOT A
0	1
1	0

A	B	A NAND B
0	0	1
0	1	1
1	0	1
1	1	0

A	B	A NOR B
0	0	1
0	1	0
1	0	0
1	1	0

Table 7.3: Truth Tables for Logic Gates. '1' Represents logic TRUE or electrical level HIGH

7.5 De Morgan's Theorem

Figures 7.2 and 7.5 show that we can build NAND, NOR and NOT functions using very few transistors, and as a result the 'inverted' forms of these gates are more common than the AND function or an OR functions. As long as we have sufficient NOT gates to hand, we can always convert a NAND gate into an AND. Furthermore, it is also possible to convert between AND and OR, or *vice-versa*, using an important theorem in logic known as **De Morgan's Theorem**.

> De Morgan's theorem states that, if we consider AND and OR functions to be of alternate types, then inverting the output of one function results in the alternate function, but with inverted inputs.

Consequently, an OR gate with both inputs inverted behaves the same as a NAND gate.

Further Reading

The Art of Electronics (*Horowitz and Hill*)

Chapter 8: Digital Electronics

Sections 8.01, 8.02, 8.04, 8.05, 8.06, 8.07

Lecture 8 Digital Representation of Numbers

Computers can be programmed to represent numbers in any way we wish but fundamentally it makes sense to use the binary scheme at the 'hardware' level, as we have seen that digital electronics makes use of the two-valued logic high/low scheme.

8.1 Decimal (Base 10)

Recalling the decimal scheme we are used to it is useful to remember that the decimal number 163 is represented by three decimal digits representing 'hundreds', 'tens' and 'units' and

$$\mathrm{d}163 = 1 \times 10^2 + 6 \times 10^1 + 3 \times 10^0$$

where the 'd' indicates that the number following is decimal. If we wish to add d38 then we start from the **least significant digit** taking care to carry forward any **overflow**, e.g. $3 + 8 = 1$ in the 'units' column with 1 carried forward to the 'tens' column.

10^2	10^1	10^0	
1	6	3	
	3	8	+
2	0	1	
1	1		

Note that, in the decimal system, **shifting** all the digits left by one place results in multiplication by a factor 10. In the example given, we would either need to add another digit to the numbers, or indicate that the shift operation generates an overflow. This will always occur if the **most significant digit** is non-zero.

8.2 Binary (Base 2)

If our number scheme is limited to the **binary digits** '0' and '1' then we can use exactly the same principles, but we find we need more digits to represent the same numbers

2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0	
1	0	1	0	0	0	1	1	
0	0	1	0	0	1	1	0	+
1	1	0	0	1	0	0	1	
	1			1	1			

Subtraction follows the same principle. We can multiply by a factor of 2 by shifting left, and divide by a factor of two by shifting right.

8 binary-digits - or **bits** - are known as a **byte**, and one byte is sufficient to represent all numbers from zero to $2^8 - 1 = 255$. 16 bits - two bytes - are known as a **word**, and one word is sufficient to represent zero to 65535.

Negative Numbers

If we need to represent negative numbers then we can implement a convention where one bit - say the most-significant bit (MSB) - is reserved to identify the sign of the number. A convention known as **offset-binary** represents all positive numbers by placing a '1' in the MSB. For an 8-bit

representation, the number b10000000 (the b representing binary) is the offset-binary equivalent of zero. Zero is defined to be a positive number. +1 is b10000001. Similarly we can represent -1 by subtracting 1 from zero to get b01111111.

Offset-binary can be converted into another convention known as **Two's Complement** - a format popular with computer designers, by simply inverting the MSB (table 8.1).

Offset-Binary	Decimal	Two's Complement
11111111	+127	01111111
11111110	+126	01111110
11111101	+125	01111101
...	...	...
10000010	+2	00000010
10000001	+1	00000001
10000000	+0	00000000
01111111	-1	11111111
01111110	-2	11111110
...	...	...
00000010	-126	10000010
00000001	-127	10000001
00000000	-128	10000000

Table 8.1: Offset-Binary

8.3 Hexadecimal (Base 16)

The binary system is convenient for computers since it can be directly represented by digital logic, however it is quite hard for humans to read, and large numbers need to be represented by many digits. The **hexadecimal** system is best understood by splitting each byte into two 4-bit **nibbles**. Each nibble can represent a single digit from zero to d15, so if we wish to use a single character we need to extend our decimal characters 0...9 by adding the letters A...F. Thus

$$\text{d163} = \text{b1010 0011} = \text{hA3}$$

The binary nibble b1010 is decimal d10 which is represented as the single '**hex**' digit hA. The advantage of hex is that a hex number with n digits can always be represented by $4n$ binary digits, while no such direct correspondence exists between binary and decimal. Hex is almost exclusively useful for representing binary numbers in a convenient 'human readable' form.

8.4 Addition

The most fundamental operation to be performed on two digital numbers is addition. The addition of the two 8-bit binary numbers in the table on page 42 occurs 'bit-wise', which means that each pair of bits is added in a separate operation, taking into account any overflow from the addition-operation of the less-significant digit (to the right) and passing on any overflow to the more significant digit (to the left). This is referred to as 'carry-in' and 'carry-out' respectively. The addition of two single bits A and B to give a single bit sum S is performed by the operation $S = A \text{ XOR } B$. The essential addition operation can therefore be performed with a single XOR gate, however it is also necessary to take into account the carry-in and carry-out which will require additional logic gates.

Digital Logic for Addition

The starting point for designing a logic circuit to add numbers is to build the truth table (see table 8.2). From inspection of the truth table, the logic design can be constructed. There are several possible solutions, and one is shown in figure 8.1. For a complete circuit to add two 8-bit bytes, figure 8.1 would need to be duplicated 8-fold. We would then have two input values represented by bits $A_{0..7}$ and $B_{0..7}$ giving 8 result bits $S_{0..7}$. The carry-out (C_{out}) of each bit is connected to the carry-in (C_{in}) of the next more-significant bit. This process is illustrated in figure 8.2.

Overflow

The carry-out of the most-significant bit can be used to indicate an overflow, that is to say the summation has exceeded the largest number which can be represented by the 8-bit sum, which is 255. Note that there is no carry-in for the least-significant bits A_0 and B_0.

C_{in}	A	B	S	C_{out}
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	0	0	1	0
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

Table 8.2: Addition Truth Table

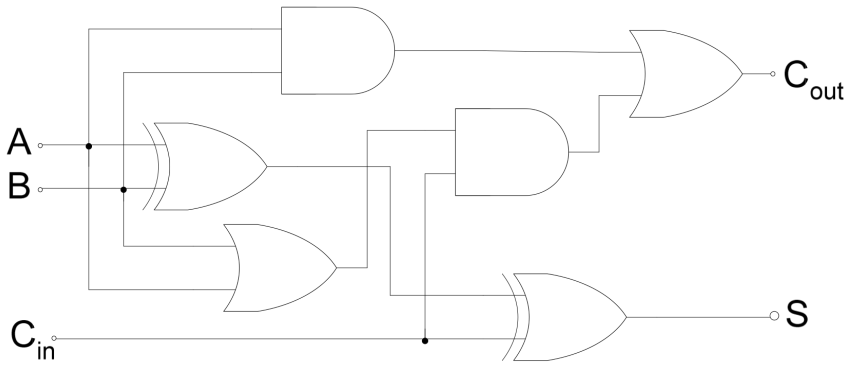


Figure 8.1: Single-bit Addition

8.5 Memory

One characteristic of all the logic circuits presented so far is that the outputs of the logic gates will react immediately to any change on the input. This is generally desirable - the adder circuit,

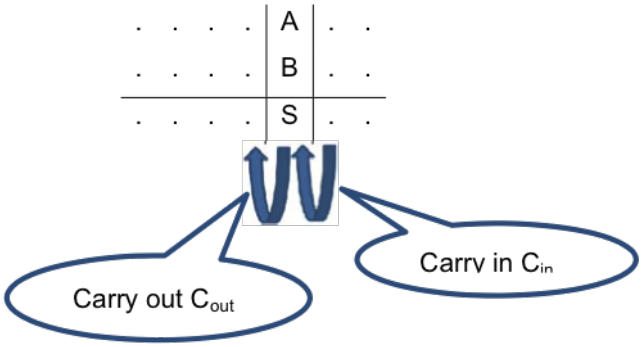


Figure 8.2: Carry in/out

implemented for 8-bits, performs an immediate and simultaneous addition of the two input numbers. If any of the input bits change, then the output always represents their sum. This parallel operation is an important factor in the speed of digital circuits. Frequently, however, we may wish to store a digital number for later processing. The key characteristic of a **memory** device is that the number it stores is changed only on command. The **D-type Flip-Flop** is a logic device with memory - also known as a **latch** - that is to say, it's behaviour depends on its past history. Figure 8.3 shows the pin arrangements and the truth table is given in Table 8.3

D-type Flip-Flop

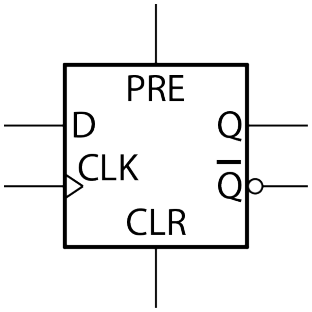


Figure 8.3: D-type Flip-Flop component schematic

There are four inputs: CLK (Clock), D (Data), PRE (Preset) and CLR (Clear). An 'X' in table 8.3 means that the input is irrelevant as the state of the flip-flop is controlled by other factors. The symbol ↑ indicates that the action happens when this input goes from Low to High. Preset is used to force the output Q to the high state and Clear is used to force Q low. $\overline{Q}$ is always an inverted copy of Q.

Edge-Triggered Clocking

From the truth-table for the Flip-Flop we can see that the output Q only changes to D when clock has a positive-going transition (↑). Q is unaffected by the ↓ transition.

CLK	D	PRE	CLR	Q	$\overline{Q}$
X	X	High	Low	High	Low
X	X	Low	High	Low	High
↑	High	Low	Low	High	Low
↑	Low	Low	Low	Low	High
X	X	High	High	Low	High

Table 8.3: Truth Table for the D-type Flip-Flop

The Flip-Flop stores one bit of information. The number stored is the value of D at the instant when CLK makes a low-to-high transition. Subsequently, the value on the output Q will remain unchanged even if the input D changes.

Registers

To store an 8-bit byte therefore requires 8 flip-flops in an arrangement known as a register. Typically, all the flip-flops would share a common CLK such that the register is loaded with a new 8-bit number with all bits loaded simultaneously. The 8 outputs $Q_{0..7}$ are a binary representation of the stored number.

Further Reading

The Art of Electronics (*Horowitz and Hill*)

Chapter 8: Digital Electronics

8.03, , 8.08, 8.09, 8.10, 8.16, 8.17