Electronic Design Project 2

John H. Davies

2013 August 14

Preamble

This course covers a number of diverse topics. Some of these are practical and you will need the skills for projects in future years. Others simply didn't fit into other courses and were collected here! These are the main subjects.

1. Use of the **Cadence OrCAD PCB Designer** suite to draw, simulate and lay out designs on printed circuit boards (laboratories only).

2. **Analogue-to-digital** and **digital-to-analogue converters** (lectures and test, semester 1).

3. **Power supplies** and **passive components** (lectures and test, semester 2).

4. The **design project** itself: research, design, construction and test (preliminary laboratory, design assignment and final laboratories).

There is no complicated mathematics in this course. I try to keep the topics practical but you can't design or use either a power supply or an analogue-to-digital converter without some idea of how it operates.

The material overlaps considerably with other courses because of changes in the curriculum in the past few years. In particular, power supplies and related topics are also covered in Power Electronics 2. Both courses cover a broad range of material and we hope that the repetition will be helpful.

A glance will show you that this handout is not a collection of lecture slides. The notes contain far more material than the lectures. Instead of trying to cover everything in class I shall concentrate on the important concepts, leaving the details for you to study. This leads to the obvious question: *What should I learn for the tests?* Past tests are on moodle and you will see that the questions are very similar to the examples in each chapter.

This handout is more like a textbook, partly because no single book covers the course at the right level. Bonnie Baker's book [1] is good but a lot of the material is too advanced. You should also become familiar with *The Art of Electronics* [4], another wonderful book. See the chapter *Further reading* on page 143 for other books and application notes that might be useful. These items are referenced in the text by numbers in square brackets, such as [1] and [4] above.

Please take care of this handout because a lot of the topics are important for future projects.

Prerequisites

These are the main prerequisites from first year. We shall draw on all this material in the project.

- **Basic behaviour of the standard components** – resistors, capacitors and inductors, Ohm's law and so on (Electronic Engineering 1X). This includes their behaviour in *time*, not just in frequency. Impedance is useful only for sine waves (or signals that can easily be constructed from them, as you will learn in mathematics this year) but many of the currents and voltages that we consider are nothing like sine waves.

- ***General* relation between current and charge** – not just $Q = IT$.

- **Basic circuit analysis** – Kirchoff's laws, Thévenin's theorem, nodal analysis and the like.

- **Time-dependence of *RC* circuits** – the way in which a capacitor charges and discharges through a resistor (Electronic Engineering 1Y). This will be revised in Electrical Circuits 2. Inductors are major components in many power supplies and we'll need to look at their behaviour in time as well.
- **Operational amplifiers** – we need to use some unfamiliar circuits but they can all be analysed using the principles that you were taught in Electronic Engineering 1Y. Please do yourselves a favour and forget the rubbish that you were taught in Higher Physics: It won't work. (It's not your teachers' fault but the SQA.)
- **Operation of a bipolar diode and transistor** – outline only, no details (Electronic Engineering 1Y). You will learn a great deal more in Electronic Devices 2 and Analogue Electronics 2.
- **Microcontrollers** – used in the final project. The material from Electronic Engineering 1Y will be carried further in Embedded Processors 2. The microcontroller will be programmed in the C language, taught in Introductory Programming EE1.

Look back at your notes from last year or a standard textbook if you have forgotten any basic theory. You will need it for your other courses and I may well quiz you about these topics during the course.

Formal description

The university's formal description of the course is contained in the course specification, which can be found in the course catalogue. A link is provided from moodle. Here are the most important sections.

Minimum Requirement for Award of Credits

- Attendance at all tests, gaining a nonzero mark
- Completion of laboratories on printed circuit board design
- Attendance at all sessions for the project, making a worthwhile contribution to the team's work
- Timely submission of project reports and an acceptable laboratory record book

The only unusual item here is the third one. I will not tolerate 'passengers' on the project because it is not fair to the other students in the team.

Course Aims

This course addresses many of the issues that arise in the design of a real piece of electronic equipment. These include:

- provision of a power supply
- interface between analogue signals and digital components (such as microcontrollers)

- layout of a circuit on a printed circuit board
- interpretation of a manufacturer's data sheet
- selection and use of simple components, with heatsink if required

These issues are brought together in a project, carried out in a small team, which also draws on material taught in other electronics courses.

Intended Learning Outcomes of Course

By the end of this course students will be able to:

Printed circuit board design

- draw circuits using schematic capture, with about 20 components
- simulate a one-transistor amplifier and compare the results with analytical estimates
- lay out a printed circuit board, using local design rules, for single and double-sided boards, using manual and automatic routing

Power supplies and passive components

- Explain operation of traditional off-line power supply: transistor, rectifier and smoothing capacitor
- Calculate ratings of all components required
- Design shunt regulator using Zener diode and resistor
- Describe operation of linear regulator and state its key specifications
- List basic types of switching regulator and describe when they are appropriate
- Explain themal dissipation, calculate limits on operation of devices and specify suitable heatsink
- Describe basic passive components (resistor, capacitor, inductor), their different practica types, and choose an appropriate component for an application
- Describe construction of a basic printed circuit board (PCB) and different packages for modern components

Analogue-to-digital and digital-to-analogue converters

- State mathematical expressions for an analogue-to-digital converter (ADC) and a digital-to-analgoue converter (DAC)
- Distinguish between resolution and accuracy with application to ADCs and DACs
- Describe common types of ADC and DAC (flash, pipeline, successive-approximation (SAR), sigma-delta and integrating ADCs; string, current-steering and sigma-delta DACs)

- Explain principleof operation of a SAR ADC and describe its input characteristics
- State Nyquist criterion to avoid aliasing and need for anti-aliasing filter
- Deduce key parameters of an ADC from its data sheet

Project

- Devise approach to address given requirements
- Perform appropriate preparatory experiments
- Design signal conditioning and specify ADC
- Design complete system, including power supply, decoupling capacitors etc
- Lay out printed circuit board
- Populate printed circuit board, test for continuity, rework as necessary, including surface-mount devices
- Write software for microcontroller with appropriate structure and documentation
- Test and debug complete, mixed-signal system
- Contribute to writing of user's manual and team report
- Work effectively as a member of a team of 3 or 4 students
- Keep an individual laboratory book

This course addresses many of the more general *graduate attributes* set out by the University.

Summative Assessment Methods

- 50% – Two class tests, each 1 hour (numerous past papers and solutions are provided on moodle)
- 10% – Laboratory on PCB design
- 40% – Project (practical work, team reports and individual laboratory record book)

There are two special conditions for the assessment.

- It is not possible to offer reassessment of the project because it is carried out in teams.
- To receive a grade D in this course, students must achieve at least a grade E in every component of assessment listed. The result will be capped at E1 otherwise.

The second condition is imposed by our Professional Engineering Institution, the Institution of Engineering and Technology, who accredit the programmes in electronics and computing science. They are concerned that nobody should be able to 'pass' the course overall, while failing a significant part of it. To get a grade D overall (typically 40

These special conditions affect only about one student per year and are of no concern to anybody who takes the course seriously.

Learning and Teaching Methods

The list gives the number of contact hours and (estimated notional learning hours – the total time that you are expected to devote to this course).

- Lectures: 10 (40)
- Tutorials: 2 (10)
- Laboratory work: 10 (10)
- Project work: 20 (30)
- Examinations: 2 (10)

Contents

Part I

Data conversion

1

Introduction to data conversion

A very large number of electronic systems have the overall structure shown in figure 1.1. These are the main functional blocks.

1. Analogue input comes from a **sensor** (temperature, microphone, antenna, . . .).
2. This analogue signal is converted to a digital value by an **analogue-to-digital converter** (A-to-D, ADC, A/D).
3. The signal is **processed** digitally. It may also be stored or communicated to another system.
4. The processed value is returned to an analogue signal by a **digital-to-analogue converter** (D-to-A, DAC, D/A).
5. Finally, a **transducer** or **actuator** (control valve, speaker, . . .) is driven with the analogue output.

A more specific example is shown in figure 1.2 on the facing page. This is a digital weighing machine, which was the project for this course long ago. In this case the output is only to a

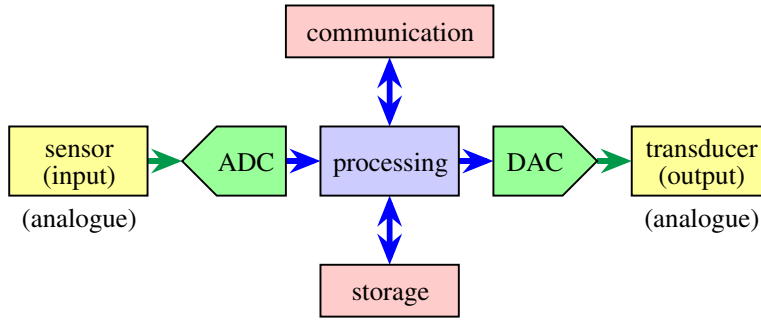


Figure 1.1 A generalized system with analogue input from a sensor, digital processing, storage and communication, and analogue output to a transducer.

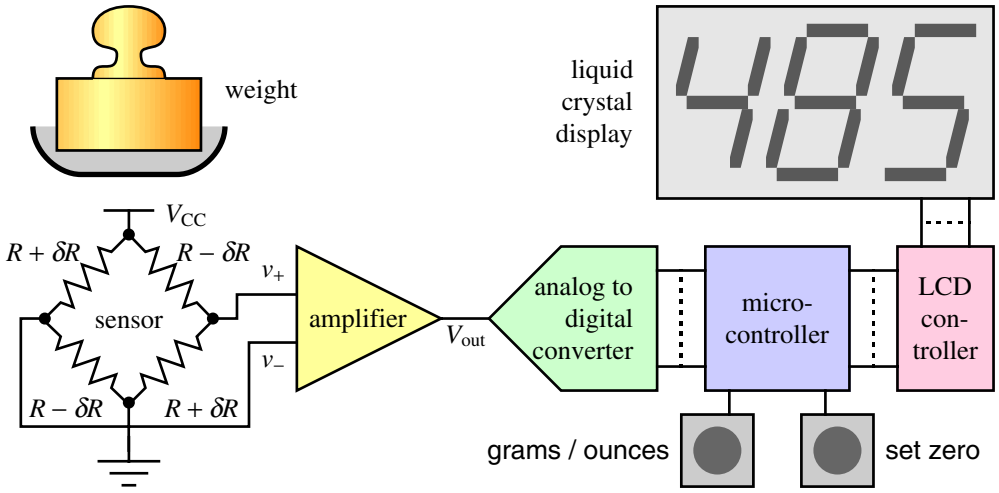


Figure 1.2 Block diagram of a digital weighing machine.

display although there would probably be a digital interface to a computer in a commercial system. It shows a bit more detail of the analogue input, which will be covered in chapter 6.

The complete system would have operated on analogue signals in the past. For example, television used to be an analogue signal and was stored in this form on VHS tapes. The same was true of sound. Control systems were also entirely analogue (the department owned an analogue computer when I arrived in 1986). A controller for the temperature of an industrial process would have used an analogue sensor, processed the signal using op-amps with feedback networks, and produced an analogue output to drive the heater. The circuit with the op-amps would be designed specifically to achieve the desired function and the components would need to be changed if a different control law was needed. Controllers used to be made with terminals on the back so that the critical components could be exchanged easily.

Now the trend is to do as much as possible with digital signals. Television is again a good example. Analogue transmission will cease in the next few years and all broadcasts will be digital. Programmes can now be recorded in digital form on hard disks or DVDs. This has long been the case for audio, where CDs were introduced about thirty years ago. Here the amplifiers used to be analogue systems but even that is changing, and many audio power amplifiers are now 'class D' systems, which use various forms of pulse modulation (such as pulse width modulation or PWM). The conversion back to an analogue signal takes place at the last possible point, in the speaker itself.

The same trend can be seen on the input side of systems as well. For example, it is convenient for manufacturers of mobile phones if they can be adapted with minimal effort to the various systems in use around the world. The radio receivers therefore do as little as possible in the analogue domain before they convert the signal to digital form. It is relatively easy to reprogramme a digital signal processor to work in the USA rather than the UK, for instance.

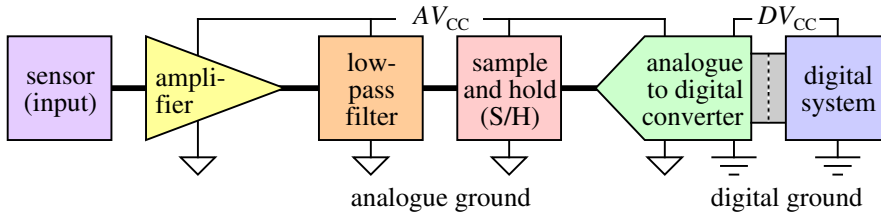


Figure 1.3 Block diagram of a data acquisition system, including the functions needed to condition the signal for the converter itself.

This all makes it seem as though analogue electronics is being pushed to the periphery of many systems (apart from the power supply, although even they use an increasing level of digital control). While this is partly true, it requires very high performance from the analogue circuitry that remains. Major electronics companies therefore promote their analogue products vigorously. For example, Texas Instruments is probably best known for digital signal processors (DSPs) nowadays but the title of its home web page is currently *Analog Technologies, Semiconductors, Digital Signal Processing*. Analogue is unavoidable! Often the interface between a sensor and the ADC is the most difficult part of a system to design. It can be very tricky to eliminate noise and present a clean signal of the correct magnitude to the analogue to digital converter. You will have to handle all this in future projects.

A great deal more than just an ADC is needed to turn a signal from the analogue voltage of a sensor to a digital output for further processing. Some of this was shown in figure 1.2 and figure 1.3 shows an expanded version of a typical system. Not every block may be needed and the order may vary slightly. Here is an outline of the function of each block.

1. The input comes from a **sensor** of some sort. The output is often a voltage but sometimes a current or change in resistance, capacitance or inductance – an *enormous* variety of sensors is available.

2. Many sensors produce very small outputs, perhaps only μV, and an **amplifier** may be necessary to raise them to a suitable level for the ADC.

3. A **low-pass filter** is almost always needed for two reasons.

   - **To remove noise from the signal**. Typically the signal has a low frequency, in which case any high frequencies present are unwanted noise and can be suppressed with a low-pass filter. Some types of noise have a well-defined frequency, such as the mains at 50 Hz and harmonics. A notch filter can be used to eliminate these. Some types of ADC operate in a way that removes particular frequencies intrinsically.
   - **To avoid aliasing**. If the signal is sampled at frequency f_s, then frequencies greater than $\frac{1}{2} f_s$ must normally be eliminated from the input before it is sampled. This will be explained in section 4.1.

4. The input to the ADC should be held constant while it is being converted to ensure that the sample refers to a well-defined time. The **sample-and-hold** (S/H) circuit does this

under control from the ADC. Many types of ADC act as their own S/H circuit and do not need an external one.

5. Finally, the **analogue-to-digital converter** turns the analogue signal into a digital representation.

The blocks before the ADC will be covered in chapter 6. Note that the analogue and digital blocks have separate power and ground rails (supplies). This keepa the measurement free of digital noise. The ADC is the interface between the analogue and digital worlds and may therefore need both power supplies and grounds.

2

General features of analogue-to-digital converters

A computer processes digital data, which has a well-defined value that depends on the bits used to represent it. However, most data in the real world is analogue, with a continuous variation between some limits. The number of levels that can be detected depends only on the resolution of the measuring instrument. Conversion between the two forms is done by *digital-to-analogue* (DAC, D-to-A, D/A) and *analogue-to-digital* (ADC, A-to-D, A/D) converters. These use a variety of methods that differ in resolution, speed, accuracy and price. DACs are simpler but we'll start with ADCs because they are much more common.

We should consider some important general features of analogue-to-digital conversion before we look at the operation of particular types of converter.

- The input voltage is a *continuous* quantity, which means that it can take any value (within a practical range), but the output is an *integer* – a digital value with a given number of bits, which can take only a finite range of values. Information is therefore lost on sampling.
- Similarly, the input voltage is a continuous function of time but the output is a discrete sequence. Samples must be taken at a sufficiently high rate to get a faithful representation of the input.
- All ADCs compare their input against a reference voltage. The output value is a *ratio* of the input to this reference, not an absolute value.

Thus the input $v(t)$ to an ADC is a continuous quantity that varies continuously in time, while the output $N_{\mathrm{ADC}}[n]$ is a list of integers. The process of converting $v(t)$ to $N_{\mathrm{ADC}}[n]$ is called *sampling* and can be analysed mathematically to determine how much information is lost in amplitude and time. I shall do only a little of the theory but the basic facts are essential.

2.1 Input–output characteristic of an ideal ADC

The behaviour of an electronic system is expressed mathematically by its *transfer characteristic* or *transfer function*, the function that gives its output in terms of its input. For an ADC the input

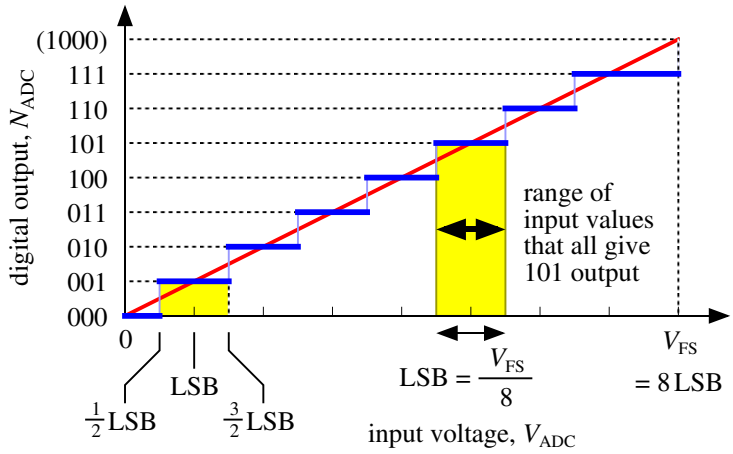


Figure 2.1 Relation between analogue input and digital output (transfer function) for an ideal 3-bit converter. This can produce 8 digital values and therefore has LSB $= V_{FS}/8$. The straight line shows the ideal transfer function without quantization.

is a voltage, V_{ADC}, and the output is a (binary) number, N_{ADC}. Two parameters of the ADC are needed before we can write down the transfer function.

- **Number of bits in the output**, N. An output of N bits can represent 2^N distinct values.
- **Full-scale voltage of input**, V_{FS}. The ADC is designed for work for input voltages from zero to V_{FS}. The full-scale voltage is typically the same as the reference voltage V_{ref}. (Some ADCs have different input ranges, as I'll mention later.)

It seems reasonable that $V_{ADC} = 0$ should give $N_{ADC} = 0$, and that $V_{ADC} = V_{FS}$ should give $N_{ADC} = 2^N$ for the maximum value. A first stab at the transfer function is then the simple, linear relation

$$N_{ADC} = 2^N \frac{V_{ADC}}{V_{FS}} \qquad \text{(conceptual)} \tag{2.1}$$

Unfortunately this cannot be correct because the input voltage V_{ADC} is a continuous quantity, so the right-hand side can take any value between 0 and 2^N. In contrast, the left-hand side is an integer. A realistic expression is therefore

$$N_{ADC} = \text{nint}\left(2^N \frac{V_{ADC}}{V_{FS}}\right) \tag{2.2}$$

where the nint() function gives the nearest integer to its argument. This is the transfer function for an ideal ADC.

Another complication is associated with the output values. If these have N bits, they can represent 2^N values, which go from 0 to $2^N - 1$. It is not possible to represent the number 2^N with N bits. For example, $4 = 2^2 = 0\text{b}100$, which just needs 3 bits. The largest number

that can be represented with 2 bits is 3 = 0b11. The prefix '0b' means a binary number and is available in some programming environments but is not standard C.

The transfer function is often expressed in terms of another voltage, called LSB for *least significant bit*. Unfortunately the name has a quite different meaning in digital systems. In analogue systems, LSB is the change in input voltage required to change the output by exactly one bit. Looking at equation (2.1) or (2.2) shows that

$$\mathrm{LSB} = \frac{V_{\mathrm{FS}}}{2^N}. \tag{2.3}$$

This is the range of input voltages divided by the number of possible output values. It is a convenient quantity for analysing the behaviour of ADCs and DACs. The ideal transfer function can be rewritten simply in terms of LSB as

$$N_{\mathrm{ADC}} = \mathrm{nint}\left(\frac{V_{\mathrm{ADC}}}{\mathrm{LSB}}\right), \tag{2.4}$$

provided that the output does not go outside its limits of 0 and $2^N - 1$.

As an example, figure 2.1 on the preceding page shows the transfer characteristic of an ideal 3-bit converter. This has 8 possible digital output values and therefore LSB = $\frac{1}{8}V_{\mathrm{FS}}$. The straight line on the plot from bottom left to top right is the conceptual transfer function in equation (2.1), which allows the output to take any value. In reality the output must be an integer, which turns the straight line into the staircase shown in the plot. This means that a range of analogue inputs give the same digital output, a feature called *quantization*. Consider the behaviour of the output as the input is raised from zero.

- An input of 0 naturally gives 000 output.
- Raising the input at first gives no change in output because of quantization.
- The output jumps from 000 to 001 when the input passes through $\frac{1}{2}$LSB = $\frac{1}{16}V_{\mathrm{FS}}$, because higher values are closer to 1 LSB than to 0.
- The output remains at 001 until the input rises above $\frac{3}{2}$LSB, when the output jumps to the next value of 010. Thus a range of input values from $\frac{1}{2}$LSB to $\frac{3}{2}$LSB gives the same output of 001, a spread of 1 LSB.
- This continues until we reach the maximum value 111 of the output, which appears when the input passes through $\frac{13}{2}$LSB. It should continue until the input rises to $\frac{15}{2}$LSB, at which point the output ought to jump to the next value. However, it can't because there are no more values. Thus the output must remain at 111 until the input reaches its full-scale value of 8 LSB.

The problem at the ends of the range is that a range of inputs of only $\frac{1}{2}$LSB gives zero output, so an interval of $\frac{3}{2}$LSB is needed for maximum output to fill the full range of 8 LSB. It would be more convenient if a 3-bit converter could give 9 output values, 0–8, rather than only 0–7. DACs have similar behaviour, as we shall see in section 8.2.

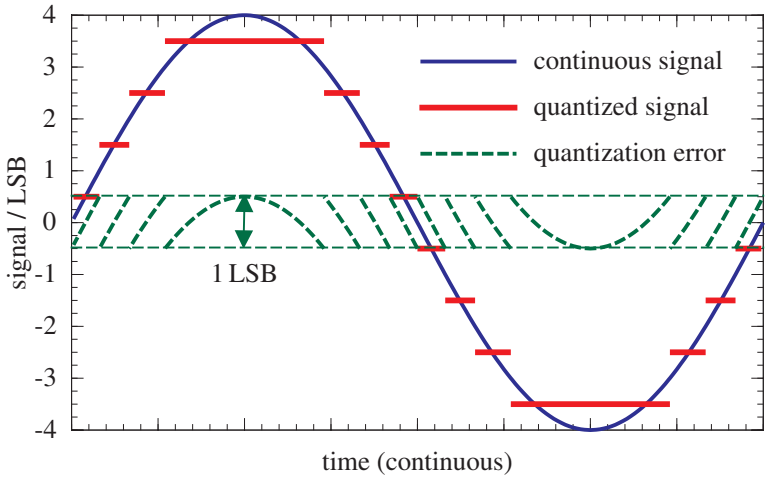


Figure 2.2 Quantization error in continuous time for a sine wave of peak amplitude 4 LSB and a 3-bit converter. The intervals at the ends are symmetric for simplicity.

Figure 2.2 shows a sine wave, before and after passing through a 3-bit ADC. This shows clearly the effect of quantization – the damage done to the signal by converting it from analogue to digital. The intervals at the ends are symmetric for simplicity. The sine wave lies between ± 4 LSB to use the full range of inputs. The quantization error lies between $\pm\frac{1}{2}$LSB. It has a distinct pattern for this simple sine wave but looks random for a less predictable signal and is often called quantization *noise*.

2.2 Resolution, precision and accuracy

These must be among the technical terms that cause most misunderstanding! (Well, there are other candidates, such as synchronous and asynchronous...) The distinction is vital for data converters.

- **Resolution** or **precision** tells you the number of distinct output values that a measurement can provide. This is 2^N for an N-bit ADC.

  Alternatively, it can be specified as the change in input that corresponds to the minimum change of 1 bit in the output: the smallest change in input voltage that can be resolved. This is just LSB.

- **Accuracy** tells you how close a measurement is to its 'correct' value – the value that would be produced by an ideal system.

To give a trivial example, a 4-digit voltmeter that gives a reading of 1.234 V is more precise than a 3-digit meter, which reads 1.32 V. On the other hand, the second meter is more accurate if the true voltage is 1.321 V. A practical difference is that it is fairly easy to get high resolution but much more difficult (and expensive) to get high accuracy.

The resolution or precision can be quoted in different ways for an ADC, depending on whether you are looking at the ADC alone or its overall behavior in the system. Consider a 10-bit ADC, for instance.

- Its output is a binary value of 10 bits, which can represent $2^{10} = 1024$ distinct values. Thus its resolution is 10 bits or 1 part in 1024.
- We also need to know the range of input voltage to determine the resolution on the input, which is LSB. Suppose that the range is from 0 to a full-scale value of $V_{FS} = 3\,\text{V}$. Then a change of one bit in the output corresponds to a change of $\text{LSB} = (3\,\text{V})/1024 \approx 3\,\text{mV}$ on the input. We can therefore say that the ADC converts its input to a precision of 3 mV.

Accuracy is much harder to define and measure and the topic rapidly gets very technical; data sheets for ADCs are complicated. The simplest specification is the *total unadjusted error*, which is the largest difference between the actual transfer function and the ideal straight line. It is usually quoted in bits but sometimes in % or parts per million (ppm). An ideal ADC has a total unadjusted error of $\pm\frac{1}{2}$ LSB as in figure 2.1.

Larger errors often arise from other parts of the system, such as the voltage reference. You need to consider all of these to evaluate the overall accuracy. For instance, how accurate is the gain of the amplifier if you need one? Noise is another major problem in practice.

2.3 Summary

- The action of an ADC is most simply expressed in terms of the voltage $\text{LSB} = V_{FS}/2^N$.
- The ideal transfer function is then $N_{ADC} = \text{nint}(V_{ADC}/\text{LSB})$ within limits.
- Make sure that you know the difference between precision (resolution) and accuracy.
- Resolution (precision) can be expressed as the number of bits in the output or as LSB, the change in analogue input that corresponds to a change of a single bit in the output.
- The total unadjusted error is the simplest measure of accuracy for an ADC.
- A complete data acquisition system has many sources of error, not just the ADC.

2.4 Examples

Example 2.1 A 12-bit ADC has a full-scale range from 0.0–3.3 V. What is its resolution in voltage (LSB)?

Example 2.2 Calculate the digital output in hexadecimal from a 12-bit ADC with a range of 5 V when the input voltage is (i) 0.1 V (ii) 1 V (iii) 4 V, (iv) 5 V. Remember that the output must be an integer. [0x052]

Example 2.3 An 8-bit ADC has a full-scale range of 0–2.048 V. What ranges of input voltages would give hexadecimal outputs of (i) 0x00, (ii) 0x09, (iii) 0xAB and (iv) 0xFF? Quote your answers to the nearest millivolt. [0–4 mV]

Example 2.4 An ADC is required to convert a voltage between zero and 1.8 V with a resolution of 1 mV. Specify the converter.

Example 2.5 Why does an ideal ADC have a total unadjusted error of $\frac{1}{2}$ LSB rather than zero?

3

Basic types of analogue-to-digital converter

3.1 Introduction

These are the common types of ADC in current use.

- **Flash** – fast, low precision, high power
- **Pipeline** – essentially a sequence of low-resolution flash converters
- **Successive-approximation** (SAR) – most popular general-purpose ADC, widely found in microcontrollers and systems on chip (SoCs)
- **Sigma–delta** (Σ–Δ) – radically different approach, deferred to chapter 4
- **Integrating** – slow, high precision, low power (fading away?)

The list goes from fast, low precision devices to slow, high precision and their applications are summarized in figure 3.1 on the facing page. I'll go through them in turn, in more or less detail. Most converters take samples of the input at the same rate at which they produce output values and are sometimes called *Nyquist* converters (we'll see why in chapter 4). In contrast, sigma–delta converters sample the input at a *much* higher rate than they produce output values. We need to do a little more theory before looking at these so I shall put them off until chapter 4.

Most basic ADCs work in essentially the same way: the input is compared with a set of known voltages and the closest match is found. The difference is the way in which the known voltages are generated and whether this is a sequential or simultaneous process. The technology used to generate the voltages has changed over the years. Currents and resistors were used in the bipolar days, so $V = IR$. A problem with this is that it is difficult to make accurate resistors in integrated circuits. The need for a current also makes it difficult to reduce the power. Nowadays the voltages are generated by using charges on capacitors and $V = Q/C$. It is easy to fabricate accurate capacitors in MOS technology; in fact the gate of a MOSFET is essentially a capacitor. However, this can lead to awkward problems because the input to such an ADC looks like a capacitor rather than a resistor. I'll explain this more for SAR ADCs in section 3.6.

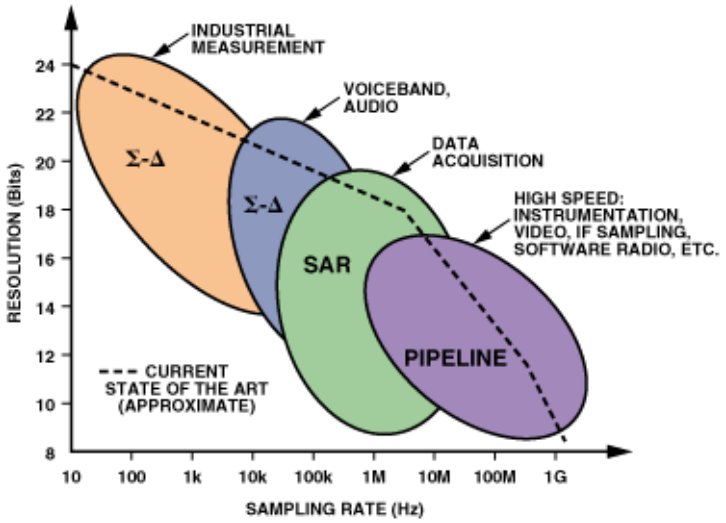


Figure 3.1 Resolution as a function of sampling rate for common types of ADC, showing their typical applications [from Texas Instruments catalogue]. Flash ADCs are omitted.

Before looking at ADCs themselves, it's amusing to look at their cost. Table 3.1 shows an example for a family of microcontrollers with different analogue inputs. The MSP430 is a simple, 16-bit microcontroller designed for low-power applications [2]. The cheapest option has only a comparator rather than a true ADC. A 10-bit ADC nearly doubles the price and the 16-bit ADC triples it. This table also shows an interesting feature of modern integrated circuits. Presumably about $1.00 of the $1.50 price of the MSP430F2003 goes on its ADC. This is two-thirds of the cost! The digital part of a small IC (by current standards) is almost free. Few standalone 16-bit sigma–delta ADCs cost less than $1.50 too. Crazy.

The cost is reflected in the area of the chip required for an ADC. Figure 3.2 on the following page is a low-resolution image of the die (the bare silicon chip) of an MSP430 microcontroller. This is a much larger device than those listed in table 3.1 and contains three 16-bit ADCs, seen on the left of the image. Each of these is larger than any other module on the chip except for the flash memory in the centre. Even the CPU is smaller than a single ADC. The cost of making a chip is roughly proportional to its area, and ADCs must be extensively tested on each chip, so I

Table 3.1 Price of different members of the Texas Instruments MSP430 family of microcontrollers, which depends on their analogue inputs. The devices are otherwise nearly identical.

Device	Analogue input	Price
MSP430F2001	comparator	$0.55
MSP430F2002	10-bit successive-approximation ADC	$0.99
MSP430F2003	16-bit sigma–delta ADC	$1.50

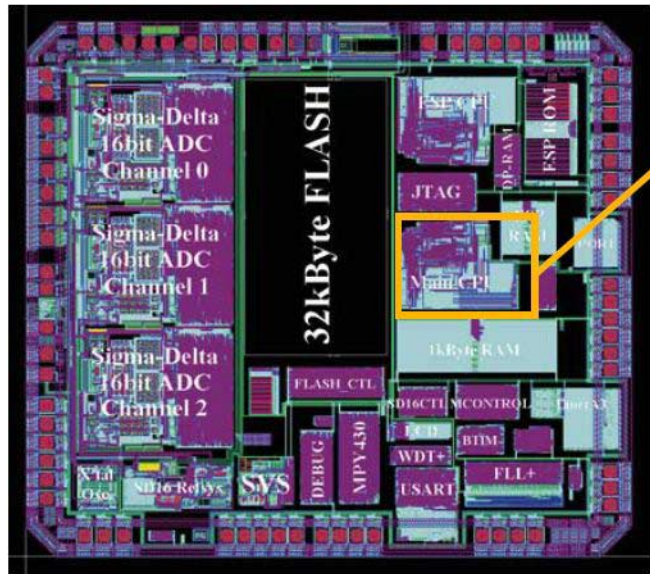


Figure 3.2 Image of the die (chip) for an MSP430 microcontroller with the main modules outlined. Each of the three large blocks on the left is an ADC. The CPU is outlined in yellow on the right.

suspect that the cost of this device is probably dominated by the ADCs. (It has three ADCs so that it can take three measurements simultaneously, which is required for measuring electrical power accurately.)

Exploring further, table 3.2 on the next page lists a selection of ADCs from the Texas Instruments *Amplifier and Data Converter Guide* (2009). I chose this catalogue because it includes prices, which is not always the case; unfortunately they do not make flash ADCs. The range of specifications is immense and this is reflected in the price. I have tried to choose a leading edge device (expensive, often called the bleeding edge), middle of the range and a cheap device.

3.2 Comparators

Before looking at any of the ADCs listed above, we'll take what looks like a diversion and explore the *analogue comparator*. This forms part of many types of converter and can sometimes be used as a substitute for a 'real' ADC. The obvious advantage is cost: a comparator is much simpler and therefore cheaper than a true ADC, as shown by table 3.1. Why pay for an ADC if a comparator will do the job?

A comparator is roughly like an op-amp used *without* negative feedback. The lack of feedback means that most of the usual rules about circuits with op-amps no not apply. For example, you *cannot* assume that $V_+ = V_-$ as for an op-amp with negative feedback. In fact real comparators use a little positive feedback rather than negative feedback to help their outputs to change quickly. The symbol and its output as a function of the inputs are shown in figure 3.3.

Table 3.2 Approximate price of a selection of ADCs from Texas Instruments (2009 Q1). The speed is measured in samples per second (sps).

Device	Architecture	Resolution (bits)	Speed (sps)	Price
ADS5474	pipeline	14	400M	$161
ADS6122	pipeline	12	65M	$12
ADS8422	SAR	16	4M	$24
ADS7883	SAR	12	3M	$2.5
ADS7885	SAR	8	3M	$1
ADS1281	sigma–delta	31	4k	$30
ADS1248	sigma–delta	24	2k	$5
ADS1100	sigma–delta	16	128	$2

The 'crossed S' is often omitted, in which case the symbol is the same as an op-amp.

In words, the comparator tells you whether the difference in voltage between its inputs, $V_+ - V_-$, is positive or negative:

- the output goes to ground (or the negative supply), logical zero, if $V_+ < V_-$
- the output goes high (to V_{CC} here), logical one, if $V_+ > V_-$.

The transition occurs over a very small range of differences $V_+ - V_-$.

Comparators are used inside most types of ADC, as we shall see. In fact a one-bit ADC can be made from a comparator. Tie the input V_- to a reference at $\frac{1}{2}V_{FS}$, where V_{FS} is the full-scale voltage. The comparator now acts as a one-bit ADC for the input $V_{ADC} = V_+$ as follows:

- output = 0 if $V_{ADC} < \frac{1}{2}V_{FS}$
- output = 1 if $V_{ADC} > \frac{1}{2}V_{FS}$.

I'll say more about comparators and give examples of their use in section 6.5.

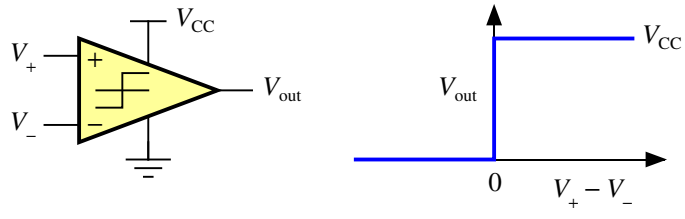


Figure 3.3 The symbol and behaviour of an analogue comparator.

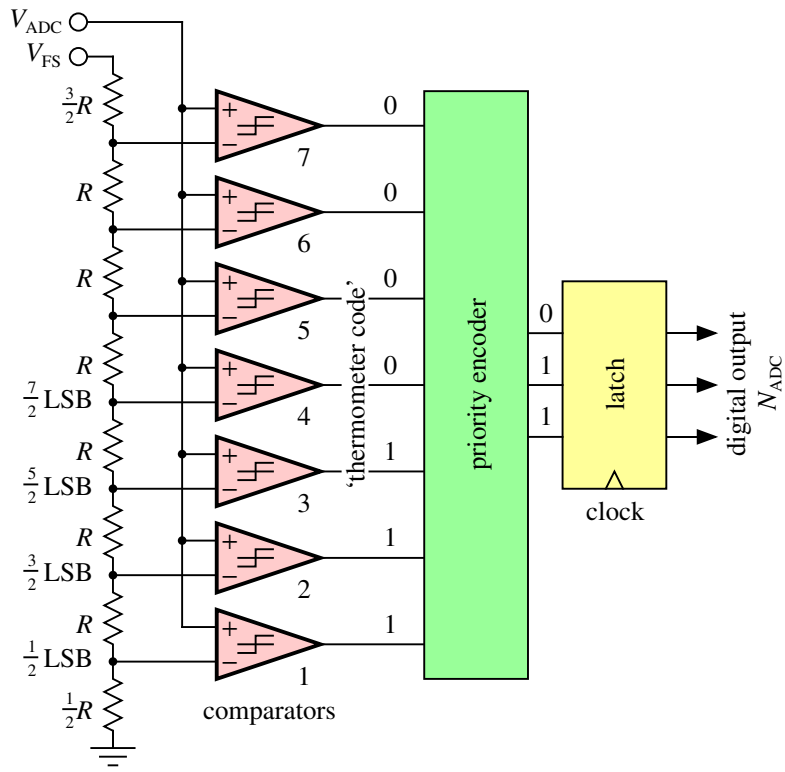


Figure 3.4 Structure of a 3-bit flash ADC. The input voltage V_{ADC} lies between $\frac{5}{2}$LSB and $\frac{7}{2}$LSB.

3.3 Flash converters

Flash or *parallel* converters are the fastest type of ADC at the cost of a large number of comparators: one for each of the 2^N possible output values except zero, not just one for each of the N bits. Their structure is straightforward, shown in figure 3.4 for a 3-bit ADC. This has $2^3 = 8$ possible outputs and therefore needs $8 - 1 = 7$ comparators. A chain of resistors, connected between V_{ref} and ground, generates a set of 7 voltages that define the transitions between one output value and the next. These voltages were shown in figure 2.1 on page 7 and lie at $\frac{1}{2}$LSB, $\frac{3}{2}$LSB and so on, separated by LSB $= \frac{1}{8}V_{ref}$, up to $\frac{13}{2}$LSB (but *not* $\frac{15}{2}$LSB!). Each voltage feeds the negative input of a comparator, whose positive terminal is connected to the input voltage V_{ADC}.

The output of each comparator is 1 if V_{ADC} is higher than its subdivided reference voltage and 0 if the input is lower. Look at the outputs of the set of comparators as V_{ADC} is raised from zero to V_{ref}.

- All comparators give 0 if $V_{ADC} < \frac{1}{2}$LSB.
- Comparator 1 gives 1 and all others remain at 0 if $\frac{1}{2}$LSB $< V_{ADC} < \frac{3}{2}$LSB.

- Comparators 1 and 2 give 1 and all others stay at 0 if $\frac{3}{2}\text{LSB} < V_{\text{ADC}} < \frac{5}{2}\text{LSB}$.

... and so on until

- All comparators give 1 if $\frac{13}{2}\text{LSB} < V_{\text{ADC}}$.

Thus the output from the comparators typically has a set of 1s from the lower values and 0s from the upper values, which is aptly called *thermometer code*. This is an inefficient way of representing the value because it requires one bit per nonzero value, so 7 bits here. Combinational logic is therefore used to convert the thermometer code into a more compact form, such as straightforward binary code. This needs to pick out the highest nonzero input, which is a standard function called a *priority encoder*. You should recall this from Digital Electronics 2. I have shown a latch on the output of the priority encoder, which is driven by a clock to synchronize the ADC with the digital system.

Flash ADCs are fast. The comparators give only one delay because they are in parallel, plus further contributions from the (simple) priority encoder and latch. The obvious disadvantage is complexity because so many comparators are needed. This typically limits flash ADCs to 8 bits (255 comparators). Another issue is the power consumption of the comparators because they need a high current for a fast response. The voltage input must drive the load presented by all the comparators in parallel, which can be a problem too. Flash ADCs are therefore restricted to applications that need high speed and low precision, and where the large power consumption can be tolerated.

A feature of the resistor chain in figure 3.4 on the facing page may have caught your eye. Most of the resistors have the same value R but the lowest is $\frac{1}{2}R$ and the highest is $\frac{3}{2}R$. This is necessary to give the usual transfer function, which was shown in figure 2.1. Remember that the bottom step is narrower than usual and the top step is wider.

Straightforward flash converters are rarely used now. The large number of comparators means that they are expensive to make and consume a lot of power. Newer devices reduce the number of comparators by using techniques called interpolation and folding. An example is the National Semiconductor ADC083000, an 8-bit converter that can produce 3×10^9 samples per second (3 Gsps). At that rate it could digitize directly the lower frequency used by GSM mobile phones, 900 MHz, and is almost fast enough for the higher frequency of 1.8 GHz. (Of course you have to get all that data out of the ADC and process it in some way...) Its data sheet is on the web if you want to find out more. It dissipates nearly 2 W from a 1.9 V supply.

Mobile phone systems illustrate the sort of application for which the ADC083000 is intended. The idea is to build radio systems where the signal from the antenna is digitized without any processing in its analogue form, except perhaps amplification. All the processing is done digitally in such a *software radio*. A great advantage is that they can 'trivially' be reprogrammed to switch between the two widely used systems for mobile phones, GSM and CDMA. Even better, a software radio can process the two simultaneously! This technology is already being tested in base stations for mobile phones. It may arrive in your handset before too long, given the general advance of digital processing, but the power consumption remains a serious problem.

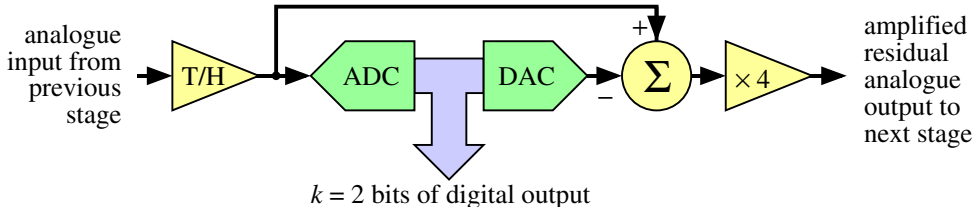


Figure 3.5 One stage of a pipeline ADC that produces $k = 2$ bits.

3.4 Pipeline converters

Pipeline converters are fast but much smaller than full flash converters. The basic idea is to make a coarse flash conversion of the input (only a few bits) and subtract the converted part of the voltage to leave a remainder. This is passed to the next stage of the pipeline, where the process is repeated to improve the resolution. There are as many stages as needed to get the desired number of bits. It is something like doing division by hand: You first operate on the most significant digit, then on the remainder (the next most significant digit) and so on, until the complete division has been performed down to the least significant digit.

A defect of pipeline converters is *latency*: The complete output is not available until the signal has passed through all stages of the pipeline. However, once the first value has been completed, subsequent outputs appear as fast as each stage can run. Pipelines are used where a continuous stream of conversions is needed and video signals are a good example.

Figure 3.5 shows the principle in detail. This is a stage in the middle of a pipeline; the final one would need only an ADC.

1. The track-and-hold circuit (T/H) holds the input to the ADC constant during each conversion. It opens to transfer the signal from the previous stage when the conversion in this stage has been completed and it is ready for the next sample.

2. The ADC makes a flash conversion of the input and produces a k-bit output. This may be only a single bit but is usually more (often '$1\frac{1}{2}$', which is tricky to explain). I have chosen $k = 2$ for the sketch.

3. The DAC converts the k-bit code back to an analogue voltage. This is the part of the analogue input voltage that is included in the k-bit digital output.

4. The subtractor gives the difference between the input and the output of the DAC. This is the residual part of the analogue input that is *not* included in the digital output and is passed to the next stage.

5. The amplifier boosts the residual signal back to the range of the input. For $k = 2$ the average magnitude of the residual is $\frac{1}{4}$ that of the input. An amplifier with a gain of 4 compensates for this and allow the circuits for each stage to be identical.

The operation is illustrated in figure 3.6 on the facing page for a two-stage pipeline and an input of $0.4V_{FS}$. Each flash converter uses voltage levels of $\frac{1}{4}V_{FS}$, $\frac{1}{2}V_{FS}$ and $\frac{3}{4}V_{FS}$.

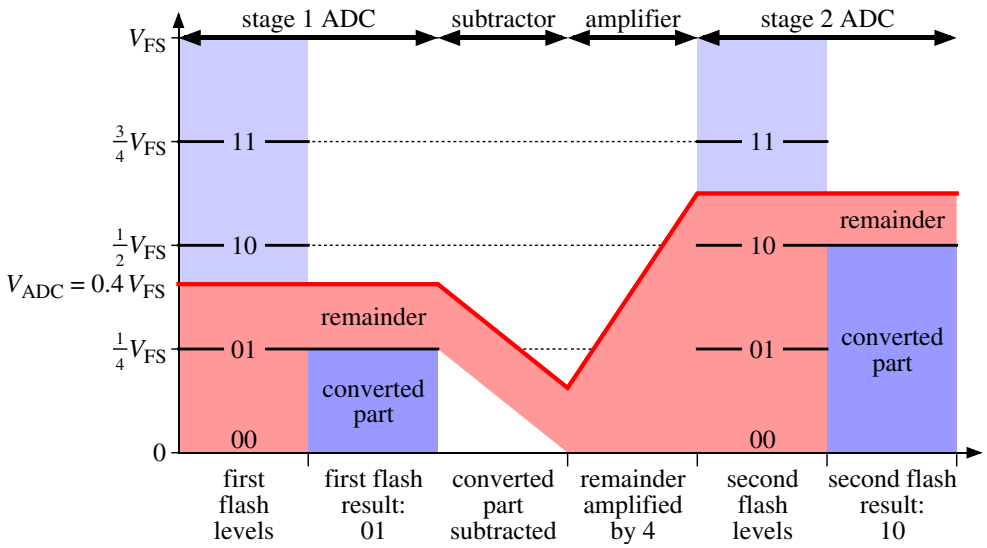


Figure 3.6 Operation of a 2-stage pipeline ADC with an input of $V_{ADC} = 0.4V_{FS}$ and a 2-bit flash ADC in each stage.

1. The input is compared with the voltage levels of the first flash converter.

2. The voltage lies between $\frac{1}{4}V_{FS}$ and $\frac{1}{2}V_{FS}$ so the output of this stage is 01, which gives the most significant pair of bits.

3. The difference betwen the input value and the nearest voltage level below is amplified by 4 for the next stage. Here the difference is $0.4V_{FS} - \frac{1}{4}V_{FS} = 0.15V_{FS}$, which is amplified to $0.6V_{FS}$.

4. The amplified remainder is compared with the usual set of voltage levels in the second flash converter.

5. The voltage lies between $\frac{1}{2}V_{FS}$ and $\frac{3}{4}V_{FS}$ so the output of this stage is 10, which gives the least significant pair of bits.

6. The remainder of $0.1V_{FS}$ would be amplified and passed to the next stage in a longer pipeline.

The overall converter includes logic to accumulate the bits produced by each stage and assemble the complete digital output. In practice this needs heavy error correction because of the DAC, subtractor and amplifier in each stage. A 'small' error in the first stage would almost certainly be larger than the value represented by the bits further downstream. Nasty errors such as lack of monotonicity and missing codes would arise if there were no correction.

How much simpler is a pipeline than a full flash converter? Suppose that we want 12 bits.

- A full flash converter would need $2^{12} - 1 = 4095$ comparators, which is a frighteningly large number.

- Assume that the pipeline uses 2-bit converters. It therefore needs 6 stages to produce 12 bits in total. Each 2-bit flash converter gives 4 possible outputs and therefore needs 3 comparators. This requires 18 comparators in total.

The pipeline requires a dramatically smaller number of comparators. Of course it is not this simple because the pipeline needs DACs, subtractors and amplifiers as well. A real pipeline converter also needs more stages to produce extra bits for the error-correcting logic.

Table 3.2 on page 15 shows typical specifications. Pipeline ADCs produce 8–14 bits at 50–500 Msps (megasamples per second). The inputs may be single-ended or differential. They are often used in image processing and high-frequency wireless communication systems. Specialized circuit design is needed to get good results at these high frequencies, which goes far beyond this course.

3.5 Successive-approximation (SAR) converters

Successive-approximation converters are currently the standard choice for a general-purpose ADC. Their resolution is typically 8–16 bits and the speed reaches megasamples per second (Msps), as shown in table 3.2 on page 15. For some reason the name is expanded into 'successive-approximation register' so that it can be contracted into SAR.

The operation is conceptually related to that of a pipeline converter, but a pipeline has numerous sets of hardware so that it can operate like a production line on numerous samples at the same time, while a SAR makes successive operations using a single set of hardware for economy at the expense of speed.

You are more likely to use a SAR ADC than any other type of ADC so it gets the most space in these notes. I'll go through their operation in this section and describe some practical aspects in the next. They are widely available as discrete components or built into microcontrollers and other systems-on-chip. The NXP LPC1768 microcontroller in the mbed module has an 8-channel, 12-bit SAR ADC.

SAR ADCs work by 'homing in' on the result using binary chopping, which is a standard way of finding solutions to equations of the form $f(x) = 0$. This is illustrated for a 4-bit SAR ADC in figure 3.7 on the next page. Here is the sequence of operations for an input voltage of $V_{\text{ADC}} = 0.4V_{\text{FS}}$.

1. The input voltage V_{ADC} is compared with the midpoint $\frac{1}{2}V_{\text{FS}}$ of the full range. In this case $V_{\text{ADC}} < \frac{1}{2}V_{\text{FS}}$ so the most significant bit (msb) = 0.

2. We now know that the input lies between 0 and $\frac{1}{2}V_{\text{FS}}$. The input is next compared with the midpoint of this range, $\frac{1}{4}V_{\text{FS}}$. We find $V_{\text{ADC}} > \frac{1}{4}V_{\text{FS}}$ so the next bit is 1.

3. Now we know that the input lies between $\frac{1}{4}V_{\text{FS}}$ and $\frac{1}{2}V_{\text{FS}}$. The input is next compared with the midpoint of this range, $\frac{3}{8}V_{\text{FS}}$. We find $V_{\text{ADC}} > \frac{3}{8}V_{\text{FS}}$ so this bit is 1 again.

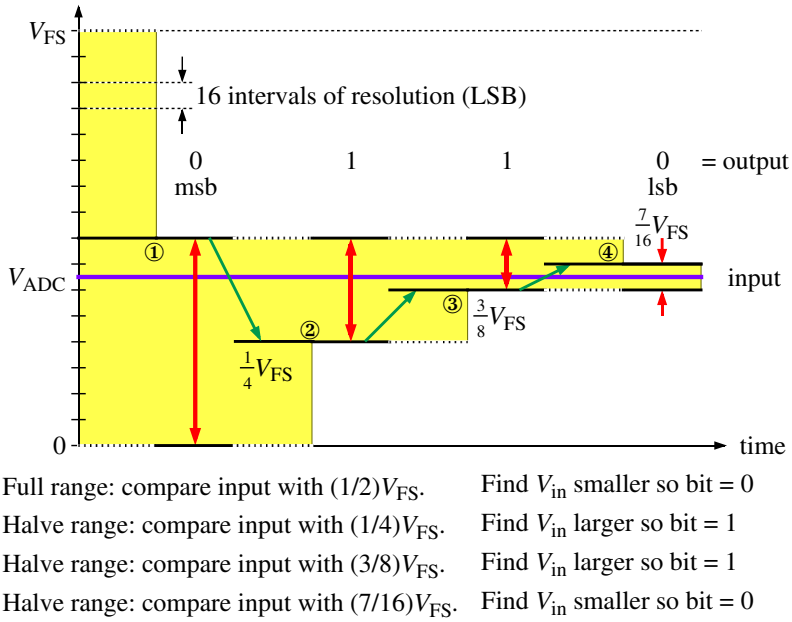


Figure 3.7 Operation of a 4-bit successive approximation ADC with an input of $V_{ADC} = 0.4V_{FS}$. The behaviour at the extreme values has been simplified.

4. Now we know that the input lies between $\frac{3}{8}V_{FS}$ and $\frac{1}{2}V_{FS}$. The input is therefore compared with the midpoint of this range, $\frac{7}{16}V_{FS}$. This time $V_{ADC} < \frac{7}{16}V_{FS}$ so this bit is 0. It is the least significant bit (lsb) for a 4-bit converter.

This process is repeated for each bit, halving the range each time. Each step typically requires one clock cycle (sometimes two) to make a comparison and set up the new voltage. An overhead is required to start the conversion, particularly to sample the input voltage. The bits of output are generated in sequence, starting with the most significant bit, msb. This fits naturally into an ADC with serial output.

Operation of a switched-capacitor SAR ADC

Three main functions are required inside a SAR ADC: logic to control the operation, some way of generating the voltages for comparison and a comparator. Back in the bipolar days the voltages were generated using a DAC, often with an R–$2R$ ladder (to be explained in section 8.4 on page 73). Ingenious arrangements of switched capacitors are now used instead to store the input, generate the voltages and make the comparisons. If you would like to know the detail, take a look at an application note from Texas Instruments [16] or section 12 of the Freescale M68HC11 Reference Manual [15].

A typical, modern SAR converter uses a set of capacitors and switches to redistribute their charge, as shown in figure 3.8 on the next page. This is a 4-bit converter to match the operations shown in figure 3.7. I'll assume the same input voltage as well, $V_{ADC} = 0.4\,V_{FS}$. The values of

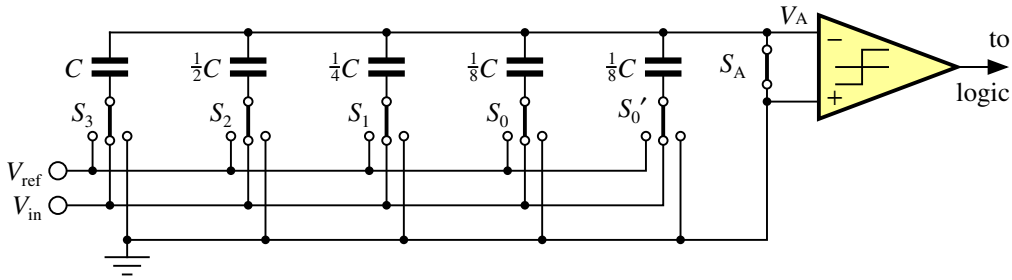


Figure 3.8 Circuit of a 4-bit, charge-redistribution SAR ADC. The switches are in the positions to sample the input.

the capacitors are C, $\frac{1}{2}C$, $\frac{1}{4}C$ and so on to give the binary chopping sequence. Their switches are labelled with the number of the corresponding bit in the output. An extra capacitor with the smallest value at switch S_0' brings the total capacitance to $2C$, which is important for the detailed operation. The reference voltage is equal to the full-scale voltage for this circuit, $V_{FS} = V_{ref}$. The switches are constructed from MOSFETs. The only other analogue component is a comparator, which is connected the wrong way around in figure 1 of the application note [16]. I have not shown the clock nor the logic needed to operate the switches and store the result. This includes the successive-approximation register that gives the converter its name.

The first step is to sample the input. All capacitors are connected to V_{ADC} and switch A is closed to connect the negative input of the comparator to ground as in figure 3.8. The 'top' plates of the capacitors are grounded and their 'bottom' plates are charged to V_{ADC}. The vital feature is that *the input of a SAR ADC is a capacitance*. I'll explain why this is important in section 3.6.

In the second step, switch A is opened to disconnect ground from the top plates of the capacitors. The individual switches are then moved so that the bottom plates of the capacitors are connected to ground instead. The resulting circuit is shown in figure 3.9. The charge on the capacitors remains fixed because no currents flow. This means that the voltage across the capacitors also remains the same. Thus the potential difference between the bottom and top plates is still V_{ADC} but now the bottom plates are grounded so the top plates are forced to

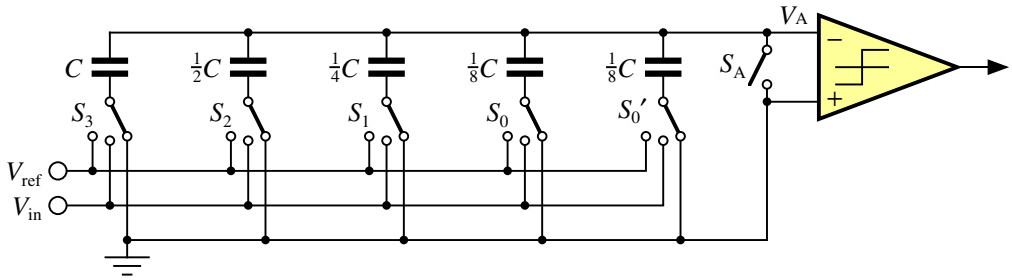


Figure 3.9 Holding phase of a 4-bit, charge-redistribution SAR ADC.

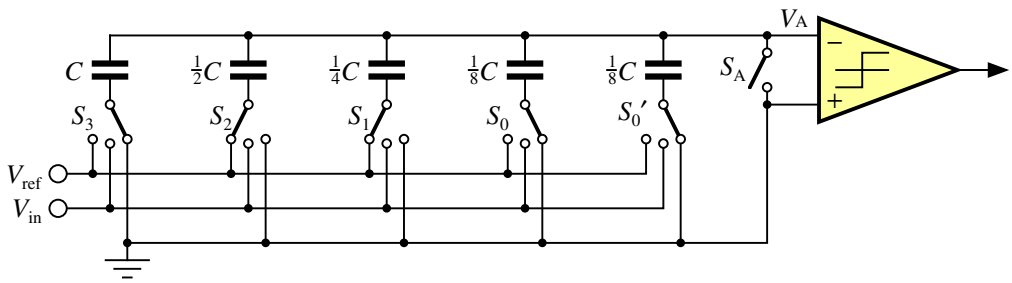


Figure 3.10 Final configuration of a 4-bit, charge-redistribution SAR ADC for an input of $V_{ADC} = 0.4\,V_{FS}$, which gives a binary output of 0110.

$V_A = -V_{ADC}$.

Next comes the conversion itself. Each of the successive approximations is made by moving one of the switches from ground to V_{ref}. The first step uses the largest capacitor. Moving S_3 from ground to V_{ref} adds a voltage of $\frac{1}{2}V_{ref}$ to V_A because the network of capacitors acts like a potential divider. (I'm not going to explain the details.) Thus the overall voltage on the top plates becomes $V_A = \frac{1}{2}V_{ref} - V_{ADC}$. The comparator compares this with ground, which is effectively the same as comparing V_{ADC} with $\frac{1}{2}V_{ref}$. This is just what we want for the first step of the binary chopping. Here we find that $V_{ADC} < \frac{1}{2}V_{ref}$ so S_3 is moved back to ground.

This is repeated for the remaining bits. Moving S_2 from ground to V_{ref} compares V_{ADC} with $\frac{1}{4}V_{ref}$ because the capacitance is only $\frac{1}{2}C$, and so on. Figure 3.10 shows the final configuration of the switches for $V_{ADC} = 0.4\,V_{FS}$.

3.6 Practical issues with SAR ADCs

SAR ADCs are probably the most straightforward type of ADC. Frequently you won't have to worry about some of the details because the converter is buried in a microcontroller, but even in these cases you must read the data sheet thoroughly! I haven't provided an example of a data sheet because you are more likely to use the SAR ADC in a microcontroller.

The inputs are usually single-ended but sometimes differential. The full-scale voltage is usually the same as the reference voltage, $V_{FS} = V_{ref}$, but occasionally $V_{FS} = 2V_{ref}$. Outputs may be parallel or serial, often using SPI or I²C. Serial output starting with the msb is most common and conveniently matches the operation of the ADC. Of course you just read a register if the ADC is in a microcontroller.

Choose the frequency of operation

As we have just seen, the operation of a SAR ADC relies on the storage of charge on switched capacitors. This leads to two limits on the frequency of the clock because none of the components are ideal.

- The conversion must be completed before significant charge has leaked away from the capacitors. The input of the comparator draws a small current, open switches do not have

infinite resistance and nor do the capacitors themselves. The clock must therefore not be too slow.

- On the other hand, closed switches do not have zero resistance. Time must therefore be allowed for charge to redistribute through them so that the voltages have stabilized and the output of the comparator is valid. Thus the clock cannot be too fast either.

The clock frequency therefore has upper and lower limits. Some discrete SAR ADCs have built-in clocks so this is not an issue. Others, particularly with serial output, need an external clock to synchronize the transmission of the result and this must lie in a specified range.

Put a capacitor across the analogue input

The data sheet may advise you to connect a capacitor between the analogue input and ground for two main reasons.

- The capacitor acts as a reservoir and makes it faster to charge the internal capacitance, especially if the source has a high internal resistance. I'll explain this shortly.

- It suppresses noise. This applies both to external noise, which would disturb the measurement. It also suppresses noise generated by the sampling process, which could affect the sensor and other sensitive electronics. This is not a trivial point: some early digital signal processing systems proved unusable because so much switching noise came out of their inputs.

The following advice is taken from the data sheet for the Freescale MC9S08QG and is typical.

> Empirical data shows that capacitors on the analog inputs improve performance in the presence of noise or when the source impedance is high. Use of 0.01 µF capacitors with good high-frequency characteristics is sufficient. These capacitors are not necessary in all cases, but when used they must be placed as near as possible to the package pins and be referenced to V_{SSA}.

There is always a disadvantage: a capacitor slows the response to changes in the input. Another potential problem is that the analogue input to the ADC may come from the output of an amplifier. Most op-amps do not like driving capacitative loads, for reasons that you will encounter in Control EE3 and Electronic System Design 3. A small resistance, typically 10–50 Ω, should therefore be installed between the output of the amplifier and the capacitor across the input. You may remember a resistor on the output of the op-amp in the microphone amplifier in Electronic Engineering 1Y. This was included for the same reason.

Allow sufficient time for the input to charge the capacitance

The input to a charge-redistribution SAR ADC is a capacitor. This must be charged 'fully' before the conversion starts. The time required commonly sets the maximum speed of conversions. The staff at the Texas Instruments European Product Information Centre told me that one of their most common problems is that engineers do not allow enough time for sampling the input by an ADC and I have heard the same from Applications Engineers at Freescale.

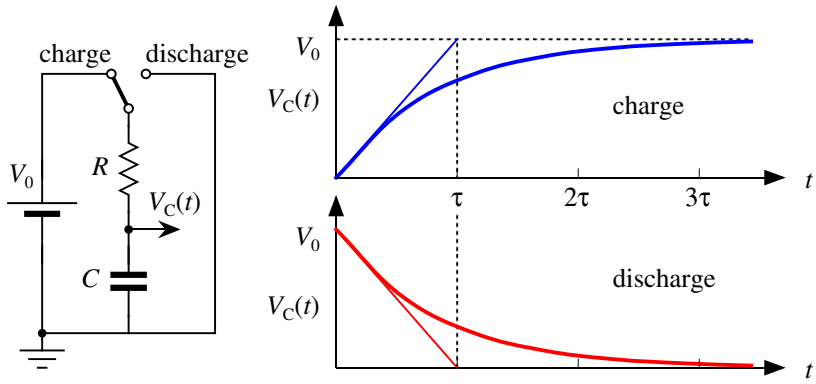


Figure 3.11 Charging and discharging an RC circuit.

How long should be allowed for this? You should of course know the equation for charging a capacitor by heart. Just in case you don't, here are the equations for charging an initially uncharged capacitor C through a resistor R from a supply at V_0 and discharging it from an initial voltage of V_0:

$$V_{\text{charge}}(t) = V_0\left[1-\exp\left(-\frac{t}{\tau}\right)\right] \tag{3.1}$$

$$V_{\text{discharge}}(t) = V_0\exp\left(-\frac{t}{\tau}\right). \tag{3.2}$$

The time-constant $\tau = RC$. Figure 3.11 should remind you of these curves.

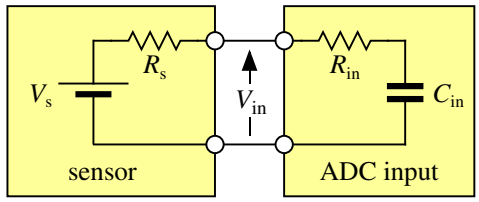
Now apply these equations to an ADC. Its input capacitance must be charged by the applied voltage. The worst case is when the capacitance is initially uncharged and the applied voltage is the maximum value. We can therefore use equation (3.1) with $V_0 = V_{\text{FS}}$. The exponential function inside the square brackets shows the difference between the present voltage and its final value, when the capacitor is fully charged, and is often called the *charging defect*. This is the error in voltage due to incomplete charging for the ADC and can never be eliminated completely. The usual criterion is that the error should be reduced below $\frac{1}{2}$ LSB so that it will not affect the digital output in most cases. Mathematically this requires that the time t_{charge} allowed for charging the input capacitance obeys

$$V_{\text{FS}}\exp\left(-\frac{t_{\text{charge}}}{\tau}\right) < \frac{\text{LSB}}{2} = \frac{1}{2}\frac{V_{\text{FS}}}{2^N} = 2^{-(N+1)}V_{\text{FS}}. \tag{3.3}$$

The full-scale voltage cancels from both sides to leave a requirement for the fractional error due to incomplete charging for an N-bit ADC,

$$\exp\left(-\frac{t_{\text{charge}}}{\tau}\right) < 2^{-(N+1)}. \tag{3.4}$$

(a) Thévenin equivalent circuit of sensor and input to ADC



(b) Potential divider to Thévenin equivalent

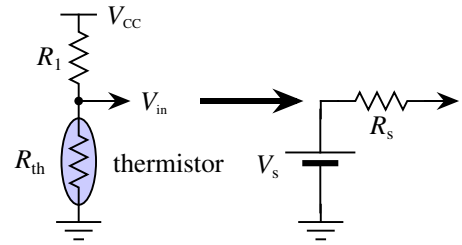


Figure 3.12 (a) Thévenin equivalent circuit of a sensor, feeding the equivalent circuit for the input of a SAR ADC. (b) Many simple sensors take the form of a potential divider, which can be converted to its Thévenin equivalent circuit.

Take natural logarithms of both sides to eliminate the exponential function:

$$-\frac{t_{\text{charge}}}{\tau} < -(N+1)\log_e 2. \tag{3.5}$$

Multiply throughout by τ and cancel the minus signs, remembering to reverse the inequality sign as well. This gives the final result

$$t_{\text{charge}} > \left[(N+1)\log_e 2\right]\tau. \tag{3.6}$$

Suppose that we are using a typical SAR ADC with a resolution of 10 bits. Then $(N+1)\log_e 2 = 7.6$ so $t_{\text{charge}} > 7.6\tau$. In words, at least 7.6 time-constants should be allowed for the capacitor to charge. This is a great deal longer than the usual 'rule of thumb' of 3τ because that would leave the capacitor only 95% charged, which isn't good enough for a 10-bit conversion: it must be at least 99.95% charged.

The next problem is to identify the values that enter the usual expression $\tau = RC$ for the time-constant. The capacitance is the input capacitance of the SAR network, C_{ADC}, which is given in the data sheet. The resistance comes from the switches inside the converter R_{ADC}, which is also in the data sheet, plus the output resistance R_s of the source that drives the ADC. This is illustrated in figure 3.12(a). You may need to do a little circuit analysis to convert the actual circuit of the sensor into its Thévenin equivalent (Electronic Engineering 1X). For example, temperature may be measured using a thermistor, which is typically connected in a potential divider as in figure 3.12(b). See chapter 7 on page 53.

The other question is how much time the system allows for the capacitor to charge. It is *not* the full conversion time but only the time for which the capacitor is connected to the input in the sample mode (figure 3.9). This may be only a few cycles of the ADC clock and is again specified in the data sheet. Sometimes the user can configure the sampling time of an ADC.

3.7 Integrating converters

These are the traditional high-resolution ADCs. They were widely found in multimeters, for instance, but have now been largely superseded by sigma–delta converters. Their operation

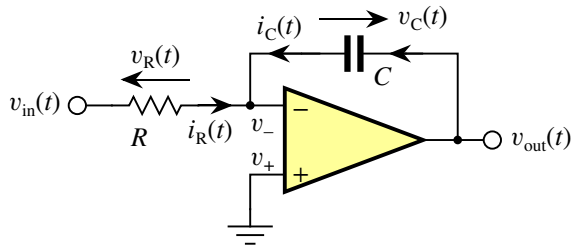


Figure 3.13 Op-amp connected as an integrator.

depends on analogue integration, which is based on the circuit shown in figure 3.13. You'll study this in Analogue Electronics 2 but here is a quick description of how it works. It all follows from the usual rules for analysing circuits with ideal op-amps and negative feedback.

0. Check that negative feedback is present: it is, although through a capacitor rather than the usual resistor.
1. The noninverting input terminal of the op-amp is connected to ground so $v_+ = 0$.
2. The inverting input terminal of the op-amp is therefore a virtual ground (virtual earth) because the op-amp tries to keep its inputs at the same voltage so $v_- = v_+ = 0$.
3. No current flows into the inverting input terminal so $i_R(t) + i_C(t) = 0$.

The voltages across the resistor and capacitor are given by

$$v_R = v_{in} - v_- = v_{in} \tag{3.7}$$
$$v_C = v_{out} - v_- = v_{out} \tag{3.8}$$

so the currents through them are

$$i_R = \frac{v_R}{R} = \frac{v_{in}}{R} \tag{3.9}$$
$$i_C = C\frac{dv_C}{dt} = \frac{dv_{out}}{dt} \tag{3.10}$$

The currents at the inverting input are therefore related by

$$0 = i_R(t) + i_C(t) = \frac{v_{in}}{R} + C\frac{dv_{out}}{dt} \tag{3.11}$$

so

$$\frac{dv_{out}}{dt} = -\frac{v_{in}}{RC}. \tag{3.12}$$

Integrating this with respect to time gives

$$v_{out}(t) = -\frac{1}{RC}\int^t v_{in}(t')dt'. \tag{3.13}$$

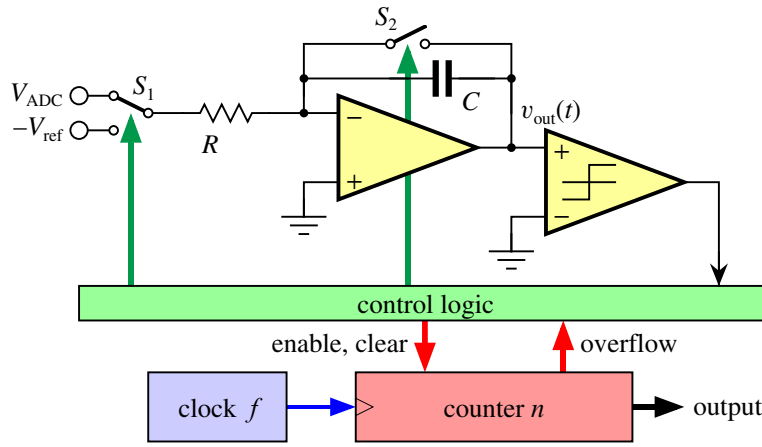


Figure 3.14 Dual-slope integrating ADC.

The factor of RC is the usual time constant.

This analysis shows that the output voltage is proportional to the integral of the input voltage. It needs a boundary condition or constant of integration, like all indefinite integrals. This is applied by short-circuiting the capacitor to discharge it, which sets $v_{out} = 0$ at a chosen time.

The integrator works for any variation of $v_{in}(t)$ on its input but I shall assume that it is constant to analyse the integrating ADC and will write $v_{in}(t) = V_{ADC}$ below.

Figure 3.14 shows how the integrator is used in an ADC. This is called a *dual-slope* converter for reasons that will become obvious. The clock has frequency f and period $T = 1/f$. Here is the sequence of operations, illustrated in figure 3.15 on the facing page for two values of the input voltage V_{ADC}.

1. Switch S_2 is closed to zero the integrator and the counter is cleared.

2. Switch S_1 is connected to the input voltage V_{ADC}, S_2 is opened and the counter is enabled at $t = 0$. The output voltage of the integrator falls linearly with time to give

$$v_{out}(t) = -\frac{V_{ADC}t}{RC}. \tag{3.14}$$

3. This continues until the counter has gone through its range N and overflowed after time NT. The output voltage of the integrator is now $v_{out}(NT) = -V_{ADC}NT/(RC)$.

4. Switch S_1 is than changed to take the input from the reference voltage $-V_{ref}$.

5. The counter restarts from zero and the output voltage of the integrator now rises with slope $+V_{ref}/RC$.

6. The comparator detects when the output of the integrator reaches zero and stops the counter at n so $\Delta t = nT$. The change in voltage at the output of the integrator is therefore $\Delta v_{out} = +V_{ref}nT/(RC)$.

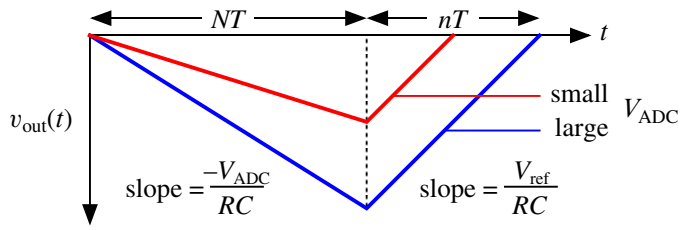


Figure 3.15 Operation of dual-slope integrating ADC for two values of the input voltage V_{ADC}.

Equating the change in voltage at the output of the integrator during the two phases shows that

$$\frac{V_{ADC} N T}{RC} = \frac{V_{ref} n T}{RC} \quad \text{so that} \quad V_{ADC} = \frac{n}{N} V_{ref} \,. \tag{3.15}$$

Thus the value n in the counter is the converted value within a scaling factor. This shows very clearly how the converter gives the ratio of the input to the reference voltage.

Some clever features of this design assist it to give precise results.

- The values of R, C and T cancel in the converted result. Thus none need be particularly accurate, which is why the dual-slope method is used. The values must all be *stable*, however, which means that they must remain constant during the measurement. In particular, the capacitor must not leak and its value must not change as a function of voltage. Imperfections of the capacitor often limit the performance of the converter.

- The input is sampled for the fixed time required for the counter to roll over. The frequency of the clock can be chosen so that this sampling time cancels interference of a particular frequency in the input. For example, pickup from the 50 Hz mains is a major problem in multimeters. We therefore arrange for the integration time to be 20 ms, the period of the mains. This will include one full cycle of the mains and multiple full cycles of its harmonics. The average of a sine wave over any number of complete cycles is zero so this technique cancels interference with a frequency of 50 Hz and harmonics. Unfortunately you will still have problems if you move to a country with 60 Hz mains!

Further tricks give extra accuracy but I shan't go into them because integrating converters are largely obsolete. Sigma–delta converters provide similar precision without such demands on the analogue components. However, plenty of integrating converters are still in use and they provide a good illustration of how op-amps can be used.

3.8 Summary of classical ADCs

This has been a lengthy chapter so I'll repeat the main points.

- Flash converters have the highest speed but are expensive and consume a high power.

- Pipeline converters are used for higher sampling rates than SAR ADCs; they have high throughput but suffer from latency.

- In practice you are most likely to use a successive-approximation (SAR) ADC, probably integrated into a microcontroller.
- Modern SAR ADCs work by redistributing charge around a network of switched capacitors.
- It is generally straightforward to use a SAR ADC but the capacitative nature of its input may cause problems: Ensure that you allow sufficient time for charging.
- Integrating converters may be useful for slow, high-precision measurements but have largely been displaced by sigma–delta converters.

Read the data sheet carefully before you use any ADC (or any component, come to that).

3.9 Examples

Example 3.1 A 4-bit flash ADC has a full-scale range of 0–5 V. How many comparators does it contain and what voltages are applied to them?

Example 3.2 Explain the operation of the above converter with an input of 3 V. Calculate the thermometer code and final binary output. [10 (decimal)]

Example 3.3 Repeat this for a 4-bit successive-approximation ADC with the same voltages. Work out the sequence of comparisons and the binary output. What is wrong with the result?

Example 3.4 Why do serial SAR ADCs send their most significant bit (msb) first?

Example 3.5 A 12-bit SAR ADC is specified as having an input capacitance of 40 pF and input resistance of 2 kΩ. What is the minimum sampling time that must be allowed to ensure that the error due to incomplete charging is less than $\frac{1}{2}$LSB, assuming that it is connected to an ideal voltage source? The ADC is then connected to a sensor with an output resistance of 10 kΩ. How does this affect the sampling time?

Example 3.6 An integrating dual-slope converter uses a 20-bit counter and the voltage reference has a magnitude of 10.0 V. At the end of a conversion the counter reads 838 859 (decimal). What was the analog input voltage? What is the resolution of this value? What temperature coefficient should the reference voltage have if the converter is to be accurate over the range 10–35°C? [7.999 98 V, 6 digits (1 in 10^6 or 10 µV), 0.04 ppm/°C]

4

Sampling, oversampling and sigma–delta converters

4.1 Sampling rate and the Nyquist frequency

We have already looked at precision and accuracy in the amplitude of sampling. (What's the difference?) Another important aspect is the relation between the frequency of the signal and the rate at which it is sampled. Remember that one aspect of using an ADC is that it converts a continuous function of time $v(t)$ to a discrete sequence of samples $v[n]$. This clearly may 'damage' the signal in some way and this section shows the problems that can arise. Let

- f be the frequency of the signal, assumed to be a simple sine wave
- f_s be the rate at which it is sampled, and $T_s = 1/f_s$ is the interval between samples

The sampling frequency f_s is often quoted with units of 'samples per second' (sps) rather than hertz (Hz) but the meaning is the same. Suppose that $f_s = 1$ ksps to keep the numbers simple and consider a signal with frequency $f = 310$ Hz. Figure 4.1(a) on the next page shows a plot of the continuous signal and the discrete samples every 1 ms. The samples look like a reasonable representation of the sine wave.

Now suppose that the frequency is raised to 690 Hz, chosen because $690 = 1000 - 310$. The outcome is shown in figure 4.1(b). The continuous input clearly has a higher frequency but the discrete values of the samples are *exactly the same* as those for the 310 Hz input! In other words, you cannot tell the difference between the signals after sampling. This called *aliasing* and is one of the fundamental problems of sampling data in time. The same happens for frequencies of $1000 + 310 = 1310$ Hz, 2000 ± 310 Hz and so on.

Aliasing is illustrated in another way by figure 4.2 on the following page, again for sampling at 1 kHz. Frequencies from zero to 500 Hz are sampled faithfully. Frequencies in the next zone, 500 to 1 kHz, are folded down below 500 Hz. This means that the samples from an input at 850 Hz cannot be distinguished from those due to an input of 150 Hz. In the next zone, a signal at 1150 Hz gives the same samples as those from 150 Hz, and so on. The folded ribbon shows how different input frequencies give identical samples.

Usually we would like the data to be sampled faithfully, meaning that there is no ambiguity in the sampled signal. We must therefore avoid aliasing. Suppose that the maximum frequency

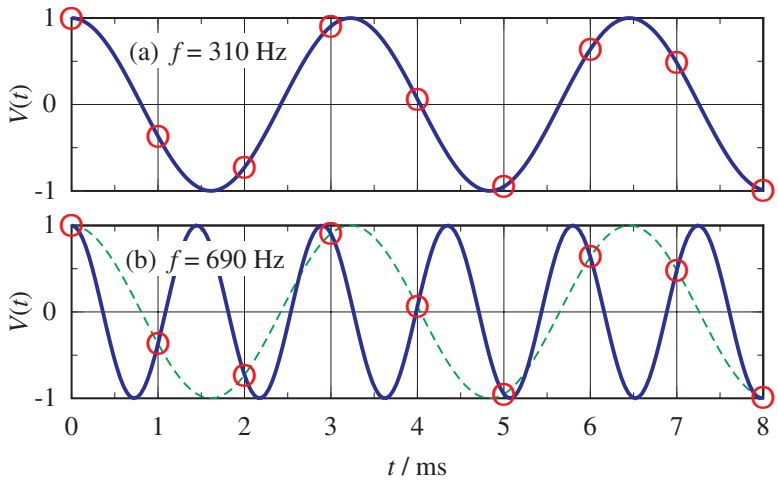


Figure 4.1 (a) A sine wave $v(t)$ at 310 Hz with samples $v[n]$ taken every 1 ms shown by the circles. (b) Sine waves at 310 Hz and 690 Hz sampled every 1 ms. The higher frequency is *aliased* – its samples are identical to those of the lower frequency.

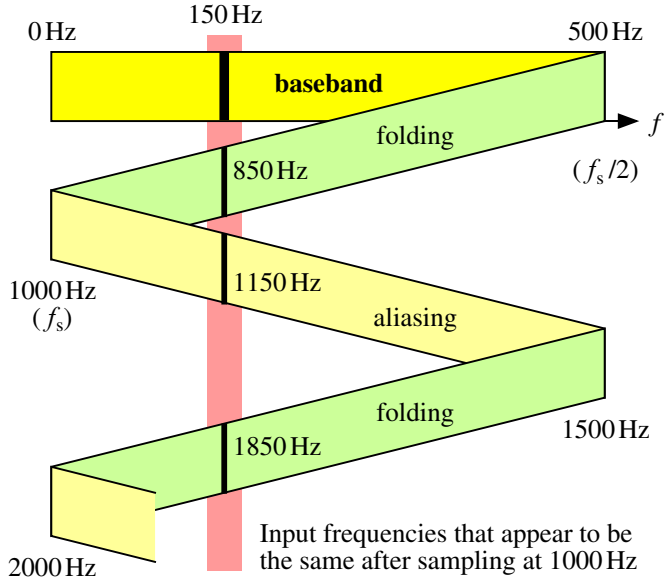


Figure 4.2 Aliasing and folding of frequencies after sampling at $f_s = 1$ ksps. The ribbon shows the frequency of the continuous input signal and the horizontal scale shows the apparent frequency of the sampled output. Input frequencies of 150, 850, 1150 Hz and so on cannot be distinguished after sampling. [Adapted from *The essential gude to data conversion* poster by Analog Devices.]

in the signal is f_{max}. Then the *Shannon sampling theorem* states that the signal can be reconstructed perfectly from discrete samples provided that the sampling rate f_s obeys

$$f_s \geq f_N = 2f_{max}. \tag{4.1}$$

The minimum acceptable sampling rate is called the *Nyquist rate* f_N and is twice the maximum frequency in the signal. Turning this around, the highest frequency that can be sampled without aliasing is half of the sampling frequency: $f \leq f_{max} = \frac{1}{2} f_s$. In the example above we sampled the data at 1 ksps so the frequency of the input must be kept below 500 Hz to avoid aliasing. (Strictly, the condition is that the sampling rate should be at least twice the *bandwidth* of the input but I'll assume that we are dealing with baseband signals, which go down to zero frequency.)

As another example, we might want to record audio frequencies in the range 0 –20 kHz. The sound must therefore be sampled at a frequency of at least 40 kHz. In practice, compact discs (CDs) use 44.1 kHz and digital audio tape (DAT) uses 48 kHz. Both obey this criterion. On the other hand, some systems such as NICAM (near instantaneous companded audio multiplex) and DAB (digital audio broadcasting) take samples at 32 kHz and therefore cannot reproduce the same range, but only up to 16 kHz. These high rates of sampling would create problems with the limited memory of small embedded systems so the bandwidth is usually reduced. A range from 300–3000 Hz is considered sufficient for 'telephone quality' sound and requires sampling at only 6 ksps.

An important part of most systems that sample data is therefore an *anti-aliasing filter* on the input to suppress all frequencies that suffer aliasing, $f > \frac{1}{2} f_s$. Analogue filters are hard to design, particularly if a sharp cutoff is needed. For example, the filter on a CD recorder should ideally pass frequencies up to 20 kHz or so, but strongly attenuate those above 22.05 kHz. It is practically impossible to design analogue filters with such a sharp cutoff. (Do you remember how slowly the output of a RC low-pass filter falls as a function of frequency?)

4.2 Sigma–delta converters

The basic idea behind sigma–delta ($\Sigma\Delta$ but often the other way round, delta–sigma) converters is to take a large number of samples at low resolution and to average them to obtain a final result with a much higher resolution. This means that the ADC in the core of the converter is very simple, often producing only a single bit as output, but it must run much faster than the final output. The ratio of the frequency at which the input is sampled to the frequency at which the outputs are produced is called the oversampling frequency, OSR. It is not unusual to have OSR = 1024 so an output frequency of 1 kHz requires the input to be sampled at about 1 MHz.

The analogue part of the system is made as simple as possible but the digital part, which carried out the averaging, is more complicated. It is called a digital filter. Sigma–delta ADCs were expensive in the past but the size of digital circuits keeps shrinking and sigma–delta converters are now more common. They are almost universally used for audio applications and general-purpose 16-bit $\Sigma\Delta$ ADCs are now available in £1 microcontrollers (table 3.1 on page 13).

Figure 4.3 on the following page shows the main blocks of a sigma–delta converter.

- The analogue input goes into a difference amplifier, which subtracts (hence 'delta') the current value of the output to leave the error.

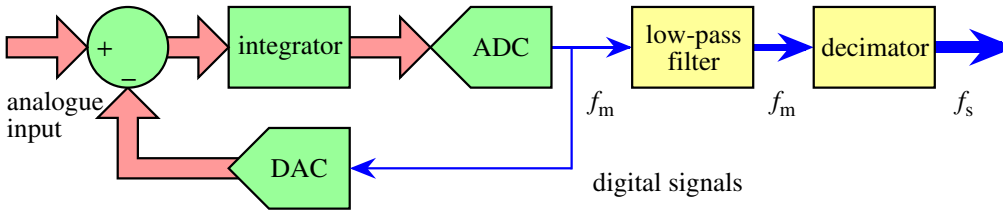


Figure 4.3 Block diagram of a sigma–delta analogue-to-digital converter.

- This error is integrated (much the same as summation, hence 'sigma').
- The output of the integrator is converted from analogue to digital in an ADC at a frequency f_m, the *modulator* or *oversampling* frequency. This is performed by a one-bit ADC, which is just a comparator.
- This digital signal is converted back to analogue in a DAC so that it can be subtracted from the input, forming a feedback loop. The one-bit DAC is no more than a switch.

These first four components in the loop form the sigma–delta modulator. The second part of the ADC handles purely digital signals. Its job is to take the fast stream of single bits from the modulator and convert them to a slower stream of multi-bit samples. In principle this is done in two stages.

- The digital signal is processed by a low-pass filter. This is needed because the stream of samples from the modulator can represent frequencies up to $\frac{1}{2} f_m$ but the slower, final output can represent frequencies only up to $\frac{1}{2} f_s$. Thus we must remove frequencies above $\frac{1}{2} f_s$ to avoid aliasing at the final sampling rate.
- The filtered digital signal is then decimated to reduce the rate of samples from f_m to f_s.

The effective number of bits rises in both these steps, which is why the lines get thicker in figure 4.3. In practice the filtering and decimation are combined in the digital filter.

4.3 Practical issues with sigma–delta converters

Sigma–delta converters have some special features that must be taken into account in a real system.

Input characteristics

The circuit of the input is very similar to that of a SAR ADC (section 3.6 on page 23). It is again a capacitance, which it must be possible to recharge within the period of the modulator, $1/f_m$. For example, the input of the sigma–delta ADC in the MSP430 is modelled with a capacitance of 10 pF in series with a resistance of 1 kΩ. This has a time-constant $\tau = RC = 10^{-8}$ s. The maximum frequency of the modulator is 1 MHz, which corresponds to 100 time-constants. This looks like plenty of time for the capacitance to charge but remember the external resistance and that you want a very accurate value.

Differential inputs

Most sigma–delta ADCs have *differential* inputs. This means that the ADC operates on the voltage between the inputs, $V_+ - V_-$, rather than the voltage between a single input and ground. You can always tie V_- to ground if this feature is not wanted but it is often helpful. For example, the weighing machine back in figure 1.2 on page 3 uses a bridge for its sensor. This gives differential outputs naturally, which could be connected directly to a sigma–delta ADC. We'll look at this in section 7.7 on page 62. Of course there must be sufficient gain, which brings us to the next point.

Programmable gain amplifier

Many sigma–delta ADCs have a *programmable gain amplifier* (PGA) on their inputs. This typically has a fairly modest gain, perhaps up to 32, but it may be sufficient to avoid the need for an external op-amp.

These amplifiers are nothing like a classic op-amp with feedback resistors. They amplify packets of charge rather than voltage and their input is like a switched-capacitor SAR ADC, described in section 3.6 on page 23. Do not expect a high input impedance as if there were a traditional instrumentation amplifier on the input (section 6.1 on page 39). A separate analogue buffer based on an op-amp may be provided to boost the input impedance.

Example

A wide range of sigma–delta converters is available with an broad spread of sampling frequencies, resolution and types of filter. A few examples are listed in table 3.2 on page 15. Rather than pick out a 'typical' device I've gone for an extreme. This is the AD7788 from Analog Devices, which has 16-bit resolution. A 24-bit version, the AD7789, is also available. I've attached a few pages from their data sheet [31]. The output data rate is 16.6 Hz to provide simultaneous rejection of 50 and 60 Hz. These may be the slowest sigma–delta converters available but you may need something like them to make high-precision measurements from sensors in future projects.

Audio converters are highly specialized and I don't have time to talk about them, unfortunately. They tend to use high-order (fourth or fifth) modulators and sophisticated digital filters to ensure a flat frequency response over the audible range.

4.4 Summary of sigma–delta converters

You aren't expected to learn the details of sigma–delta converters but they are so widely used that I have to include them. These are the main points that you should know.

- Sigma–delta converters oversample the input by a very large factor.
- This is followed by digital filtering and decimation to give good resolution but low speed.
- Their analogue parts are simple while the digital filter is complicated but easy to fabricate in VLSI.
- They have high resolution and differential inputs, which suit many types of sensor.

- The oversampling rate must be large enough for the desired number of bits.

4.5 Reflection on ADCs

It's interesting to note that the ADCs that I have chosen as examples have data rates from 16.6 sps to 3 Gsps, a factor of nearly 10^9. The power consumption reflects this: the 3 Gsps ADC083000 consumes nearly 2 W while the slow AD7788 needs only about 200 µW, a factor of 10^4. The resolution goes from 8 to 24 bits, which sounds like a factor of only 3 but is better viewed as $2^{24}/2^8 = 2^{16} \approx 10^5$. These figures highlight the *enormous* range of ADCs available. The range of prices is large too. I still find it hard to believe that a 24-bit ADC, whose resolution exceeds 1 part in 10^7, can be bought for £3!

4.6 Examples

Example 4.1 What is meant by *aliasing* and the *Nyquist frequency* f_N? How does f_N depend on the rate of sampling?

Example 4.2 An ADC is required to convert an input whose frequency may go from DC up to 50 kHz. What is the minimum rate of sampling that should be used? What would happen if a lower rate were used?

5

Summary: Selection of an ADC

Many aspects must be considered when choosing an analogue to digital converter. Here is a summary (more than long enough!) to help you when the need arises.

- **Precision** – The number of bits in the output. Alternatively, the resolution of the system can be defined as LSB, the change in the input that corresponds to one bit in the output.
- **Speed** – How many samples per second do you need? Some converters work over a wide range of sampling rates but some have only one.
- **Number of input channels** – How many do you need? Must they be sampled simultaneously (at exactly the same time) or can they be treated sequentially (one after the other)?
- **Characteristics of input** – Several issues arise here.
  - What is the range of input voltages? Can you connect the input directly or will an amplifier be needed?
  - Do you need a single-ended or differential input?
  - Is the input impedance high enough not to load your source or will a buffer be needed?
  - Is a sample-and-hold circuit necessary, either internal (usually) or external?
- **Type of converter** – In most cases this will fall out from the choices already made but there are some applications for which particular types of converter are especially suitable.
- **Voltage reference** – Is there an internal voltage reference or must you provide an external one?
- **Filtering** – Is filtering necessary? You almost always need a filter to remove noise and for anti-aliasing but there may be more specific requirements, to reject interference from the mains at 50 or 60 Hz for instance. Some types of converter do this intrinsically.

- **Accuracy** – Not the same as precision! Usually the total unadjusted error is the most appropriate number but the effective number of bits (ENOB) may be more relevant for sigma–delta ADCs.

  Accuracy depends on the overall system, not just the ADC itself. The voltage reference or gain of an amplifier may be less accurate than the ADC.

- **Power supply** – Particularly important in equipment powered by batteries.

  - What voltage does the converter need to power it? Some have two power supply pins, one for analogue and one for digital, with two grounds to correspond.
  - How much power does the converter use? Can be be shut down to save power?

- **Interface for digital output** – Will often be defined by the digital system to which the converter is connected. Common options are:

  - parallel output (usually as many bits as there are in the ADC's output)
  - serial peripheral interface (SPI). This is a simple and common way of connecting a single peripheral to a microcontroller; 3-wire and Microwire are similar.
  - inter-integrated circuit bus (I²C). This is another common way of connecting peripherals to a microcontroller but this is a bus and can be shared by several devices; 2-wire and SMBus are similar.

- **Package** – If the board is to be assembled by hand you should look for a plastic dual-in-line package (PDIP), which is easy to solder, but for production you would probably want a smaller, surface-mount package.

- **Price** – How much does it cost? Very important!

Clearly there are a lot of questions to ask, although some will be far more important than others in a given application. Look at the manufacturers' web sites. They have selectors where you type in the major specifications and receive a list of recommended devices. Follow this by checking the web page for likely components and downloading the data sheet.

Manufacturers also publish application notes to help you use (and buy!) their products. These can be extremely helpful and are often more up to date than textbooks. Older data sheets and application notes tend to be more informative, a sad reflection on cost-cutting.

6

Signal conditioning

It is rarely possible to connect a sensor or another source directly to an ADC. An amplifier will often be needed and usually a filter as well. This processing of the signal before it is converted by the ADC is called *signal conditioning*.

6.1 Amplification

An amplifier will be needed if the range of voltages from the sensor does not match the input of the ADC. Often the signal must be both amplified and shifted to match. You already know the basic circuits constructed using op-amps from Electronic Engineering 1Y and will study them further in Analogue Electronics 2.

Sometimes there is no need to amplify the signal but the source resistance is too high for the capacitance of the input to charge in in time. This was discussed in section 3.6 and a simple buffer or voltage follower on the input eliminates the problem. They are built into some ADCs and microcontrollers.

You are expected to be able to analyse and design standard circuits with op-amps. Professor Weaver's recognition chart from Electronic Engineering 1Y is helpful. Often you have to choose between inverting and noninverting configurations. These are the key features of the basic circuits.

- A noninverting amplifier has a high input resistance because the signal goes directly to a terminal of the op-amp.

- An inverting amplifier has a virtual ground, which enables the circuit to carry out simple sums and other mathematical operations. Its input resistance is determined by the resistors and is therefore lower than a noninverting amplifier.

Inverting amplifier circuits generally place less demand on the op-amp and should therefore be chosen if a high input resistance is not required. It usually doesn't matter if the signal to an ADC gets inverted because it is trivial to change the sign of the digital value.

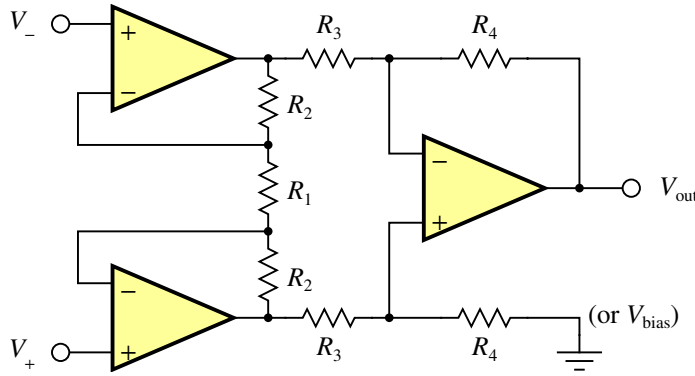


Figure 6.1 Standard circuit of an instrumentation amplifier.

Instrumentation amplifier

Often the desired input is the small difference between two large voltages – a *differential* signal. A circuit is needed to amplify the difference while rejecting the average voltage on the two wires, called the *common-mode* voltage. The output of a bridge is a typical example, as in the weighing machine shown in figure 1.2 on page 3. Sometimes you can use a single op-amp configured as a difference amplifier but the input resistance is often too low; a low input resistance acts like a potential divider and reduces the magnitude of the measured voltage. The desired characteristics are therefore

- amplification of differential part of input signal
- rejection of common-mode part of input signal
- high input resistance

The solution is a standard circuit with three op-amps called an *instrumentation amplifier*, shown in figure 6.1. Most of the resistors are in matched pairs. Both inputs are connected directly to non-inverting inputs of op-amps. Ideally these draw no current and in practice the input resistance is very high. The output voltage of this circuit is

$$V_{\text{out}} = \left(1 + 2\frac{R_2}{R_1}\right)\frac{R_4}{R_3}(V_+ - V_-). \tag{6.1}$$

Often $R_4 = R_3$, in which case the gain is determined just by R_2/R_1. The ground connection on R_4 can be replaced by a constant voltage V_{bias} to shift the output if necessary.

Complete instrumentation amplifiers can be bought in a single package. Sometimes the gain is fixed but often R_1 is external so that the gain can be changed. You will see more of this circuit in Analogue Electronics 2 and will learn how to get the best out of instrumentation amplifiers in Electronic System Design 3.

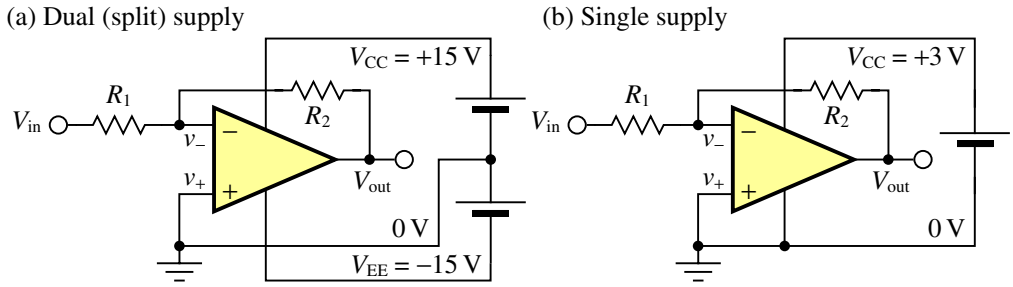


Figure 6.2 Standard inverting amplifiers using (a) split (dual) and (b) single power supplies.

6.2 Single-supply op-amps

All the circuits with op-amps that you have studied in the past have run from split (dual) supply voltages, typically ±15 V. These are often called the power *rails* and I'll use the notation V_{EE} (negative) and V_{CC} (positive), which is traditional for bipolar circuits. The power supply actually has three connections including the ground rail at 0 V. I've drawn the familiar circuit of an inverting amplifier in figure 6.2(a), including its power supplies. (You can see why we don't normally bother to show these!) The voltage gain is ideally $-R_2/R_1$.

Split ±15 V supplies are fine if you have a bench power supply available but are a nuisance in battery-powered equipment. The voltage is too large, for a start. Besides, digital systems manage with a single, positive supply so why should we have to provide a second, negative supply for the analogue components? It is much more convenient to use a *single-supply op-amp* with a lower voltage as in figure 6.2(b). The positive supply connection of the op-amp, V_{CC}, goes to the battery as usual but the negative supply goes to ground. Only two connections now go to the power supply rather than three.

In some ways it is better not to concentrate on the 'single supply' aspect but on the lack of a ground rail in the middle of the supply voltages. This is what causes most of the problems when designing circuits with single-supply op-amps. Before looking at these circuits, a few characteristics of single-supply op-amps themselves are important.

Supply voltage

Most traditional op-amps were designed to work from a ±15 V supply, a total range of 30 V. Modern devices are typically specifed at a total supply voltage of 3 V so that they work well from a single Li-ion cell. I've attached the data sheet of the ST Microelectronics TS951 [32] as an example. Its performance isn't particularly special but it comes in a PDIP, which makes it easy to solder (most modern devices are produced only in impossibly small packages). A '741, TL071 or similar op-amp will not work at all from a single 3 V supply.

Rail-to-rail input

The inputs of an old-fashioned op-amp must not be allowed too close to either V_{EE} or V_{CC} to ensure normal behaviour. More modern components may permit inputs to reach either one of

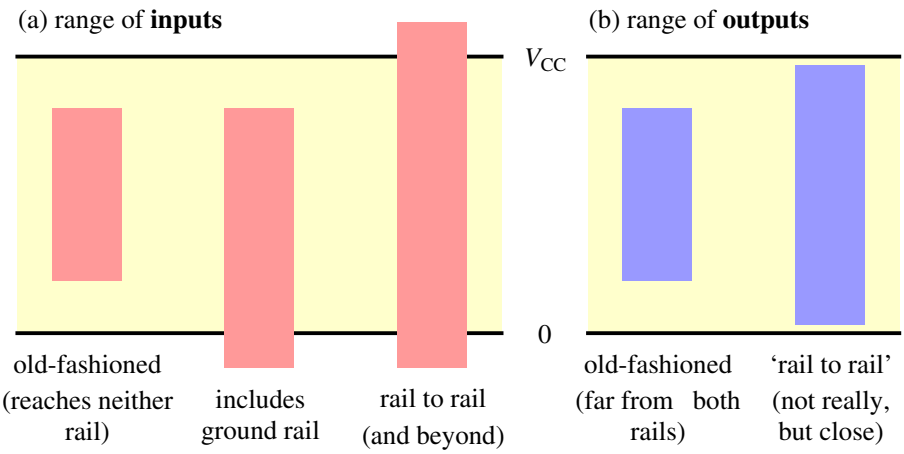


Figure 6.3 Illustration of different ranges of (a) inputs and (b) outputs available from op-amps, *not to scale*.

the supply rails or both. In fact they can often go about 0.2 V *outside* the supplies. The term *rail-to-rail input* therefore means what it says. The options are illustrated in figure 6.3(a).

You might think that it would always be a good idea to choose an amplifier with a rail-to-rail input. However, they are better avoided unless really necessary. The reason is that it is not possible to build a single input stage that works to both rails. An op-amp with rail-to-rail input therefore needs to include two input stages or to use other tricks, which may have undesirable side-effects.

Rail-to-rail input is not always needed during normal operation. The input to a circuit with gain must be smaller in magnitude than the output and may therefore stay clear of the supply voltages. Sometimes the input may go to zero voltage, which requires an input range that includes the ground rail but not the supply rail. An exception is a buffer with unity gain (voltage follower), which the range of both the input and output may span the supplies. Rail-to-rail inputs may also be needed if the voltage on an input goes to one of the supplies when the circuit is turned on. This was the case with the microphone preamplifier in Electronic Engineering 1Y.

Rail-to-rail output

If you remember your first experiments on op-amps, the output of an old-fashioned device like the OPA177 or '741 cannot get within 1 V or so of the supply rails. This would be a serious restriction with a 3 V supply! Most modern op-amps therefore feature a so-called *rail-to-rail output*. The reason for the 'so-called' is that the output can't actually reach ground or V_{CC}: *Rail-to-rail output is an advertising term only*. In practice the output gets within ± 0.1 V of the rails or closer. Again this is sketched in figure 6.3.

To make the behaviour of a rail-to-rail output clearer, figure 6.4 on the next page shows the input and output voltages of a single-supply voltage follower as the input voltage is reduced to zero. Ideally $V_{out} = V_{in}$ but this fails as the input falls below about 0.2 V. This particular

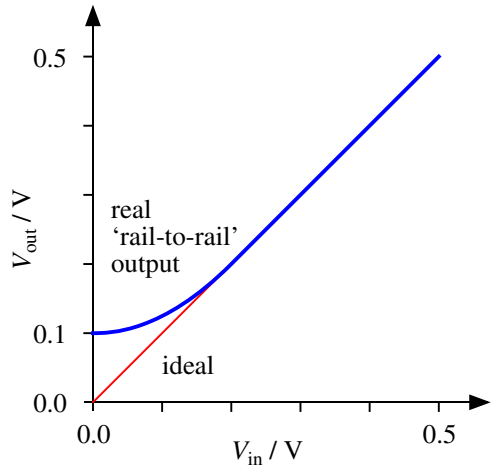


Figure 6.4 Input and output of a single-supply voltage follower for voltages near ground.

op-amp is unable to pull the output below 0.1 V so an input of zero cannot give an output of zero. Many modern op-amps can do much better than this – perhaps 0.01 V – but their outputs cannot actually reach the rails.

If the output of an op-amp *must* go all the way down to ground, a negative supply for V_{EE} is unavoidable. Don't worry, the manufacturers realise this and special ICs are available to produce a small, negative supply voltage for the op-amp from the single positive supply to the rest of the system. For example, the National Semiconductor LM7705 produces −0.23 V with low noise. It is based on a charge pump, described in section 14.1 on page 113.

6.3 Circuits with single-supply op-amps

Why should circuits with single-supply op-amps be any different from those with two supplies? Well, just look back at the standard inverting amplifier redrawn for a single supply in figure 6.2(b) on page 41. The input V_{in} must be positive because there are no negative voltages available. Unfortunately this means that the inverting amplifier wants to produce a negative output, which it cannot do for the same reason. Clearly this will not work as expected.

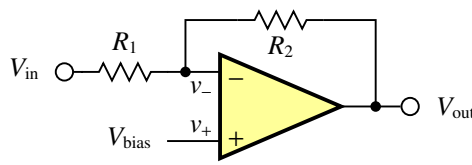


Figure 6.5 Inverting amplifier circuit with the noninverting input of the opamp connected to a bias voltage V_{bias} rather than ground.

The basic problem is that there is no reference voltage available between the two supply voltages, like the ground rail with dual supplies. There are two ways around this.

- Redesign the circuit so that it works correctly, with all voltages positive.
- Provide a reference voltage, midway between the power supplies, and use this like the ground rail in a circuit with split supplies. It is often called V_{mid}.

The standard inverting amplifier can be redesigned as in figure 6.5 on the preceding page so that its noninverting input is connected to a positive bias voltage V_{bias} rather than ground. It's easy to analyse this circuit using the usual three or four steps (good revision!), assuming an ideal opamp. *Always use the approach that we taught you in Electronic Engineering 1Y. Forget the rubbish from Higher Physics because it is* ***wrong****.*

0. Confirm that negative feedback is present.
1. The voltage at the noninverting input is $v_+ = V_{\text{bias}}$.
2. Negative feedback and the infinite gain of the opamp cause the two inputs of the opamp to come to the same potential, so $v_- = v_+ = V_{\text{bias}}$.
3. The final step, as always, is nodal analysis at v_-. The input to the opamp draws no current because it is ideal. Thus

$$\frac{V_{\text{in}} - V_{\text{bias}}}{R_1} + \frac{V_{\text{out}} - V_{\text{bias}}}{R_2} + 0 = 0. \tag{6.2}$$

This can be rearranged into different expressions:

$$V_{\text{out}} = \left(1 + \frac{R_2}{R_1}\right) V_{\text{bias}} - \frac{R_2}{R_1} V_{\text{in}} \tag{6.3}$$

$$= V_{\text{bias}} + \frac{R_2}{R_1}(V_{\text{bias}} - V_{\text{in}}). \tag{6.4}$$

Choose V_{bias} so that the output remains positive over the desired range of inputs. The gain still has its usual value. The inversion and shift are not a problem if the signal goes to an ADC because the original signal can be recovered by trivial arithmetic on the digital value. Noninverting amplifiers with an offset voltage can also be designed but it is a little more tricky [22]. Mathematically, the inclusion of V_{bias} allows the circuit to perform the operation $y = mx + c$ rather than just $y = mx$ for a simple amplifier.

The alternative approach is to generate a ground reference, V_{mid}. A potential divider between ground and V_{CC} is the simplest method. Sadly it is often not good enough because the voltage must remain constant, whatever we connect to it. This may need a 'stiff' divider with very small resistors, which wastes a lot of current. The simplest solution is to add an op-amp as a voltage follower, as shown in figure 6.6 on the next page. The capacitor on the potential divider is to suppress noise. This may seem like a waste of an op-amp but they usually come in multiple packages. Application note [23] suggests better circuits for ground references.

Note that the output voltage from the ground reference varies with V_{CC}. The same is true for the potential divider in figure 6.7. Often this is desirable, as we shall see in section 7.2 on page 55. If an absolute voltage is needed, which should not change if V_{CC} varies, a voltage reference should be used instead. These are described in section 7.1.

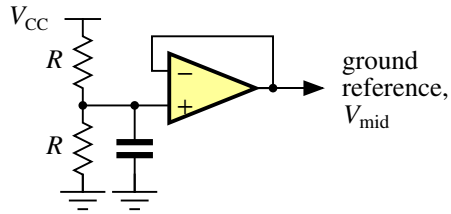


Figure 6.6 Ground reference generated by buffering a potential divider.

Worked example: Analysis of a single-supply opamp circuit

Analyse the single-supply amplifier in figure 6.7. Calculate the range of inputs over which this circuit works correctly. Does the op-amp need rail-to-rail inputs? What is the purpose of this circuit and why is a straightforward inverting amplifier not used?

The bias voltage V_{bias} is provided by a potential divider with a decoupling capacitor to remove noise. You could just put numbers into equation (6.3) to analyse the circuit but it's better to do the calculation from scratch because it's straightforward and avoids any need to memorise equations. Follow the usual steps.

0. Negative feedback is present.

1. Solving the potential divider shows that the voltage at the noninverting input is $v_+ = V_{\text{bias}} = 1\,\text{V}$.

2. Negative feedback and the infinite gain of the opamp cause the two inputs of the opamp to come to the same potential, so $v_- = v_+ = 1\,\text{V}$.

3. Use nodal analysis at v_-, which gives

$$\frac{V_{\text{in}} - 1\,\text{V}}{20\,\text{k}\Omega} + \frac{V_{\text{out}} - 1\,\text{V}}{100\,\text{k}\Omega} + 0 = 0,$$

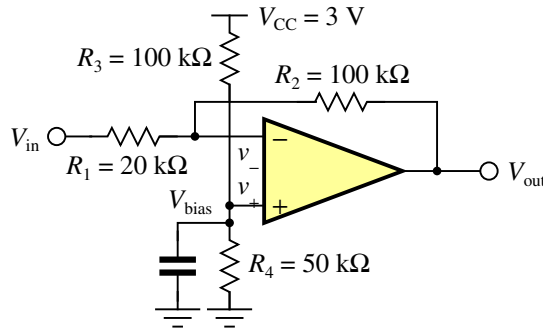


Figure 6.7 An amplifier based on an op-amp with a single supply.

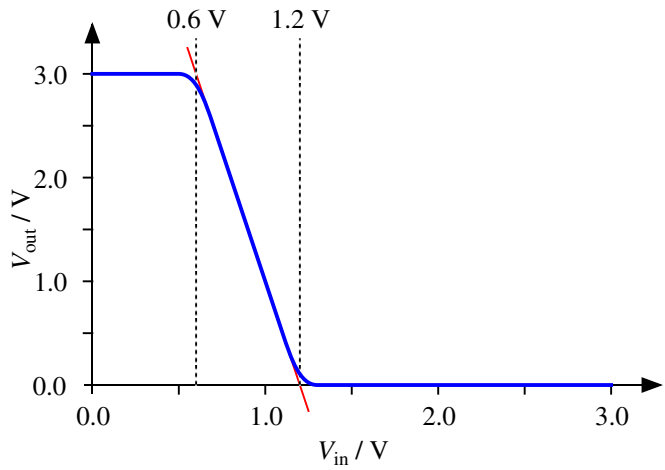


Figure 6.8 Transfer function for the single-supply amplifer in figure 6.7. The rounding of the output near the supply rails is exaggerated and the difference between the saturated output and the supply rails is too small to be seen.

$$5(V_{in} - 1\,V) + (V_{out} - 1\,V) = 0,$$
$$V_{out} = 6\,V - 5V_{in}\,. \tag{6.5}$$

This completes the analysis.

The next task is to find the range of inputs over which this amplifier works correctly. Both the input and output voltages must lie within the supply rails. Check the output voltage first.

- The output voltage cannot not go below ground so $V_{out} \geq 0$. This means that $6\,V - 5V_{in} \geq 0$ or $V_{in} \leq 1.2\,V$.
- Similarly, the output voltage cannot not exceed the supply so $V_{out} \leq V_{CC} = 3\,V$. This means that $6\,V - 5V_{in} \leq 3\,V$ or $V_{in} \geq 0.6\,V$.

The input voltage must therefore lie in the range $0.6\,V \leq V_{in} \leq 1.2\,V$ to avoid saturation of the output. This range does not go to either rail so rail-to-rail inputs would be pointless. The amplifier's behaviour is sketched in figure 6.8. This shows clearly the range of input voltages over which the amplifier functions as desired.

You were given the circuit in this problem and asked to find its behaviour. In practice you are more likely to do the opposite, design an amplifier to perform a given operation. For example, you might be told that the input voltage lies between 0.2 V and 0.3 V and asked to produce an output from 0 V to 3 V. In this case it is probably easier to start from equation (6.3) or (6.4).

6.4 Filters

I have mentioned already that (almost) all circuits for data acquisition need a low-pass filter to remove noise and for anti-aliasing. The trend is now to do as much filtering as possible in the

digital domain but it is always better to get rid of undesired signals as early as possible. Good layout of the circuit can go a long way to prevent noise being picked up in the first place.

The only relevant filters that you have studied are simple RC low-pass filters. These are often sufficient if the signal of interest is at such a low frequency that anti-aliasing is not an issue. The problem is that the amplitude falls off slowly with frequency, only as $1/f$. It is often said that the response has a *single pole*, which you will learn about in Communication Systems 3 and Control 3. A better filter, meaning one whose response falls more rapidly with frequency, is often needed for anti-aliasing. There are many classic filters with different characteristics that go by the names of Butterworth, Tchebychev (many spellings), Bessel, elliptic and so on. The problem is that no filter is ideal. For example, the Tchebychev filter gives the best response as a function of frequency, meaning that its amplitude falls off most rapidly with frequency, but it also distorts the shape of a square pulse in time most severely. These filters usually include opamps and are called *active filters*. You will study them in Electronic System Design 4.

6.5 Comparators and Schmitt triggers

I couldn't think of a logical place for this section but wanted to include a little more detail than in section 3.2 because comparators are often needed to clean up the signal for a standard digital input. The underlying problem is that real signals are always analogue so this connection acts as a sort of implicit analogue-to-digital conversion. Two issues are common.

- The voltage on a digital input should either be near ground for a logical zero or near V_{CC} for a logical 1; it should change rapidly from one of these definite values to another. Voltages around $\frac{1}{2}V_{CC}$ give unpredictable behaviour and the circuit may even oscillate. This situation arises if the input changes slowly – if there is a lot of capacitance, for example.
- Real signals always contain noise, which can cause multiple logical transitions when there should be only one.

The standard solution to these problems is to connect a comparator as a *Schmitt trigger*.

Schmitt trigger

Fgure 6.9(a) on the next page shows a straightforward comparator working from a single supply. The signal is connected to the inverting input of the comparator, which will make it easier to turn the circuit into a Schmitt trigger. The noninverting input is connected to a fixed voltage derived from a potential divider, which gives $\frac{1}{2}V_{CC}$ here. This the *threshold voltage* for the comparator, meaning that its output switches when the signal passes through this voltage.

The idea of a Schmitt trigger is to add feedback from the output to give two values for the threshold voltage, one for rising voltages and one for falling voltages. The circuit of an inverting Schmitt trigger is shown in figure 6.9(b). Note that the feedback is *positive*, so don't try to apply the rules for analysing circuits with opamps!

The operation is illustrated in figure 6.10 on page 49. I've assumed a 3 V supply and that the acceptable input voltages are the typical values for CMOS logic:

- Voltages below $\frac{1}{3}V_{CC}$ give logic 0

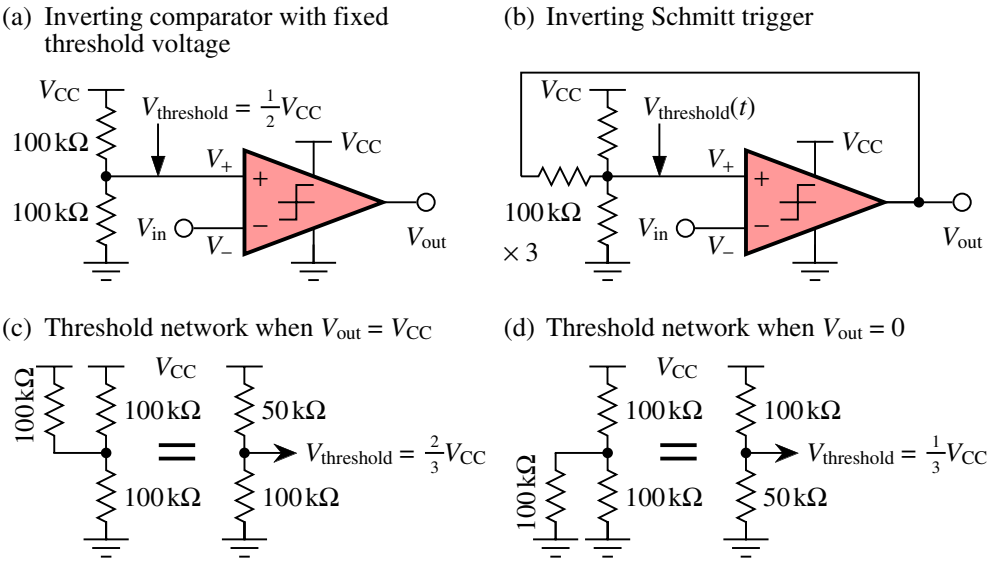


Figure 6.9 (a) Circuit of an inverting comparator with a fixed threshold voltage. (b) Inverting Schmitt trigger, with feedback from the output to the threshold voltage. The network for the threshold voltage gives (c) $V_{threshold} = \frac{2}{3}V_{CC}$ when $V_{out} = V_{CC}$ and (d) $V_{threshold} = \frac{1}{3}V_{CC}$ when $V_{out} = V_{CC}$.

- Voltages above $\frac{2}{3}V_{CC}$ give logic 1
- Voltages between $\frac{1}{3}V_{CC}$ and $\frac{2}{3}V_{CC}$ give undefined values.

This is discussed further in Digital Electronics 2. The conventional inverter in figure 6.10(a) gives output voltages in the undefined range for a range of input voltages. (You will analyse its behaviour in Electronic Circuit Design 3). Now compare the Schmitt trigger. It's easiest to consider an input voltage that rises steadily from zero as in figure 6.10(c).

(i) With $V_{in} = 0$ the output is definitely high, $V_{out} = V_{CC}$. The feedback network can be redrawn as in figure 6.9(c). Two of the resistors are connected to V_{CC} and are therefore effectively in parallel. The join is at $V_{T+} = \frac{2}{3}V_{CC}$, which sets the threshold voltage on V_+.

(ii) The output remains close to V_{CC} until the input rises through the threshold voltage, $\frac{2}{3}V_{CC}$. The output now falls to $V_{out} = 0$.

This changes the behaviour of the network for the threshold voltage, which now behaves as in figure 6.9(d). Two of the resistors are now connected to ground and the voltage at the join falls to $V_{T-} = \frac{1}{3}V_{CC}$. The input voltage is far above this new threshold voltage, which is why the output changes so abruptly and passes rapidly through the undefined range of voltages. This is the effect of positive feedback.

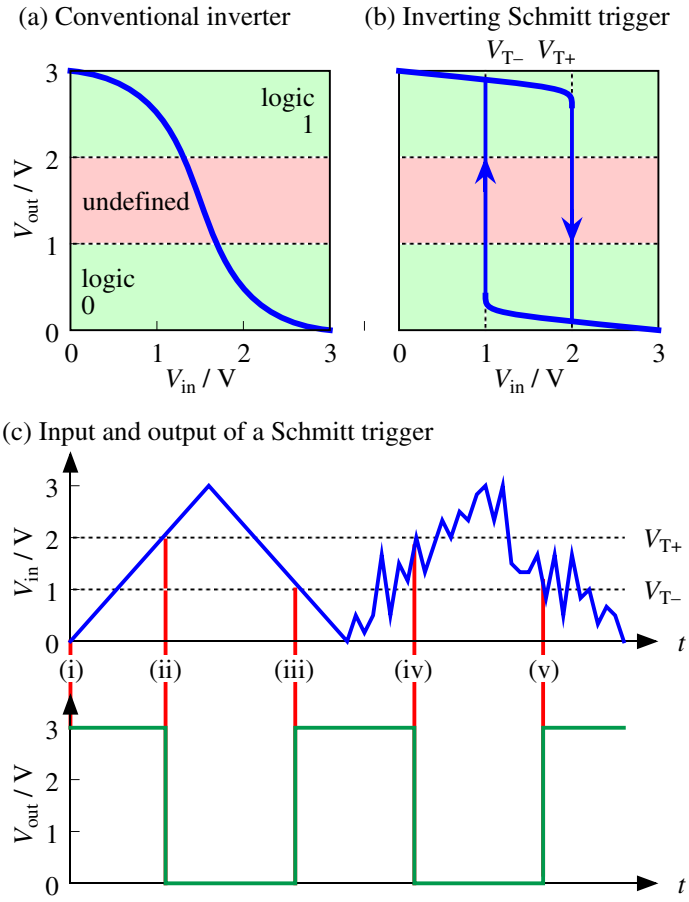


Figure 6.10 Transfer characteristic (output voltage as a function of input voltage) for (a) conventional inverter and (b) Schmitt trigger. The Schmitt trigger shows hysteresis and never gives an output with an undefined logic value. (c) Input to and output from a Schmitt trigger as a function of time. The trigger turns a slowly varying input into sharp transitions and eliminates noise.

(iii) The input voltage rises to V_{CC}, then falls again. Nothing happens as it passes through V_{T+}; the output does not switch until the input falls below the lower threshold voltage, V_{T-}.

The two threshold voltages, one for rising and another for falling input voltages, give the input–output characteristic sketched in figure 6.10(b). This type of behaviour is called *hysteresis*.

(iv) The second half of the trace shows the effect of a (rather fanciful) noisy signal. Going upward, the output switches when a spike of noise on the input first goes above V_{T+}.

(v) The output remains low until another spike on the falling signal goes below V_{T-}. The Schmitt trigger gives a clean output from this noisy input signal.

The resistors in the threshold network don't have to be equal, as I have assumed here – it was just to make the arithmetic simple.

Schmitt triggers have many other applications. A simple relaxation oscillator can be made by adding a resistor and capacitor, for instance. You will see this in Analogue Electronics 2.

An op-amp is not a comparator

The symbols for an op-amp and a comparator are similar (sometimes identical) so you might be tempted to use an op-amp as a comparator. *Do not.* Their internal circuits are significantly different because of their different functions.

- An op-amp is designed to work with negative feedback. As you know well, the feedback brings the inputs to almost the same potential, $V_+ \approx V_-$. The inputs of op-amps are designed with this in mind and many devices are damaged if $|V_+ - V_-| > 1\,\text{V}$.

  Comparators have no such restriction; the inputs are not 'tied together' in the same way and can be driven independently between V_{EE} and V_{CC}.

- An op-amp is used in linear circuits and its output can take any voltage between V_{EE} to V_{CC} (limited by saturation). Typically the output can only change rather slowly (the *slew rate*).

  Comparators are nonlinear devices and their output is designed to provide either V_{EE} or V_{CC}, not voltages in between, and to switch between these values as rapidly as possible.

- Some comparators have an *open-collector* output. This means that the output stage can pull the load down to V_{EE} but not drive it up to V_{CC}. A pullup resistor may be needed. Check this carefully when you use a comparator. Read the data sheet and see *The Art of Electronics* [4] for further details.

  Students are often caught out by open-collector outputs in projects! It isn't an issue for comparators built into larger systems such as microcontrollers.

Use the correct component for the job. Many microcontrollers contain analogue comparators or offer Schmitt triggers on their inputs to reduce the impact of noise or slowly-varying signals.

6.6 Sample-and-hold circuit

In theory a sample-and-hold circuit is a critical part of any analogue-to-digital converter. In practice they are incorporated into most types of ADC and you don't have to worry about them, which is why I've left them until the end.

A sample-and-hold (S/H) or track-and-hold (T/H) circuit is like the analogue equivalent of a transparent latch in digital electronics. Figure 6.11 on the next page shows an outline of the circuit and its operation. The two op-amps are connected as unity gain buffers (voltage followers). The first charges the capacitor to the input voltage as long as the switch is closed, in which case the output follows the input. When the switch is opened the capacitor is isolated

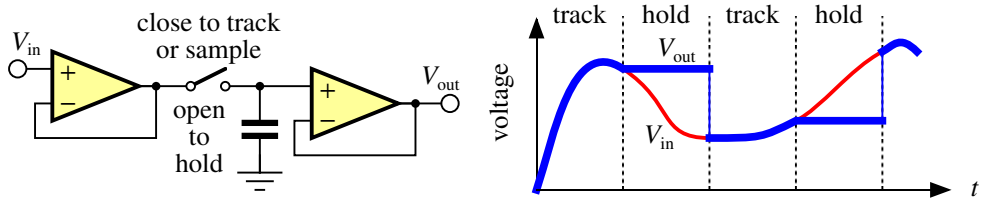


Figure 6.11 Basic track-and-hold (or sample-and-hold) circuit and its operation.

from the input and holds its voltage, provided that the input of the second op-amp does not draw too much current. (The input current would be zero for an ideal op-amp but nothing is ideal!)

Sometimes these are called track-and-hold, sometimes sample-and hold circuits. In principle a sample-and-hold circuit should be more like an edge-triggered flip-flop and sample its input only at a particular point in time (but in reality during a short interval called the aperture). My impression is that the names are used interchangeably in practice and that most sample-and-holds are really track-and-holds. They are tricky circuits to design but most ADCs do not need them nowadays. For example, the network of capacitors in a SAR ADC acts as an intrinsic sample-and-hold circuit.

6.7 Summary of signal conditioning

- Most data acquisition systems need to condition the signal in some way. The aim is to remove frequencies that would suffer aliasing, suppress noise and match the range of voltages to the converter.
- Some standard circuits need to be redesigned for single-supply op-amps.
- A comparator is not the same as an op-amp and should be used to clean noisy signals to avoid spurious transitions on digital inputs.

6.8 Examples

Example 6.1 You could write down equation (6.3) from the standard results for circuits with opamps using superposition – do you see how?

Example 6.2 What should be connected to the input of every ADC?

Example 6.3 An ADC is needed for an application on an aeroplane. The signal suffers from interference from the power supply, which runs at 400 Hz. What type of ADC and what sampling frequency would you recommend to minimize interference from the power supply?

Example 6.4 What are the characteristics of an *instrumentation amplifier*?

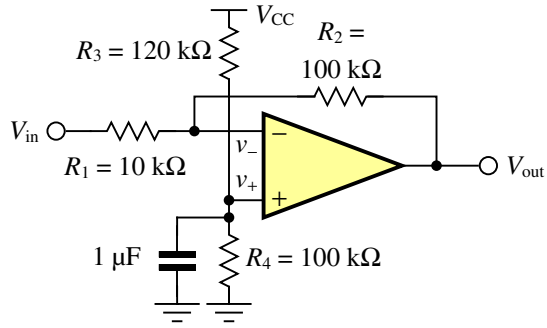


Figure 6.12 An amplifier based on an op-amp with a single supply.

Example 6.5 *Analyse* the amplifier in figure 6.12 based on an op-amp with a single-supply and show that

$$V_{out} = -10(V_{in} - \tfrac{1}{2}V_{CC}). \tag{6.6}$$

Hint: assume an ideal op-amp and use the usual three or four steps that you learnt in Electronic Engineering 1Y. Don't worry about the capacitor, which is included to suppress noise. Calculate the acceptable range of inputs to this amplifier assuming a supply of $V_{CC} = 3\,V$. Does the op-amp need rail-to-rail inputs? What is the purpose of this circuit – why is a straightforward inverting amplifier not used?

Example 6.6 Design an amplifier to work from a single 3 V supply. It should take an input voltage between 0.2 V and 0.3 V and amplify it to the full range of output voltage. It may be inverting or non-inverting at your option.

7

Complete systems with ADCs

Now it is time to put everything together to make a complete system with an analogue-to-digital converter. We shall first look at voltage references and how to avoid them before examining a few examples of sensor and how they can be connected to an ADC to meet a given specification.

7.1 Voltage reference

Recall (again!) that the digital output of an ADC is the *ratio* of the analogue input voltage to a reference voltage, set out in equation (2.2):

$$N_{\mathrm{ADC}} = \mathrm{nint}\left(2^N \frac{V_{\mathrm{in}}}{V_{\mathrm{FS}}}\right). \tag{7.1}$$

The full-scale voltage V_{FS} is usually the same as the reference voltage V_{ref}. All ADCs therefore need a voltage reference, although this is sometimes hidden. The same is true of DACs.

In many cases the (analogue) supply voltage is used as the reference. This is almost always done in small microcontrollers, although larger ones offer an external connection or provide an internal reference. Small, discrete ADCs often use the supply as reference too. Care must be taken to keep the supply quiet in these cases.

Alternatively, a special voltage reference is used to give a precise, accurate value. The simplest circuit is a Zener diode with a resistor in series to give a suitable current. Unfortunately its performance is poor.

- The voltage alters if the current through the diode varies, because of changes in the load or the voltage that supplies the resistor.
- Even if the current is kept constant, the voltage changes as a function of temperature.

The dependence on temperature can be cancelled by placing a normal, forward-biassed diode in series with the (reverse-biassed) Zener. However, you don't have to worry about this because, as usual, manufacturers produce special voltage references. Cheaper ones are based on a different principle called a *bandgap reference* but special, buried Zener diodes are still in use. The data sheets are complex but two specifications are particularly important.

- **Initial accuracy** – how close is the voltage to the specified value.
- **Temperature coefficient** – by how much does the voltage change (drift) with temperature. It is often quoted in units of ppm/°C, where ppm is a standard abbreviation for 'parts per million'.

To give you an idea of what is available, here is a selection of bandgap references from Texas Instruments.[1]

- They come in voltages of 1.25, 2.048, 2.5, 3.0, 3.3 and 4.096 V.
- For about $0.50 you get the REF29xx. Its initial accuracy is 2% with a temperature coefficient of 100 ppm/°C.
- You must spend more money if this isn't good enough. Four grades of reference are offered, of which the best is the REF32xx with an initial accuracy of 0.2% and drift of 7 ppm/°C. The output is less noisy as well. Of course you pay more, nearly $2, which may be more than the ADC!

Consider the temperature coefficient a little further. The voltage from the REF29xx varies by 100 ppm, 10^{-4} or 0.01% if the temperature changes by 1°C. A change in temperature of 100°C causes the reference voltage to change by 10000 ppm or 1%. This may degrade the overall accuracy of the system seriously. However, it is always best to study the data sheet thoroughly rather than rely on a single number. A plot shows that the dependence of voltage on temperature is far from linear. In fact it's more like a parabola with a maximum at about 60°C.

These references are three-terminal or *series* devices. This means that they have input, output and ground connections are are used in the same way as a linear regulator. The difference is that you should not draw a large current from them. Two-terminal or *shunt* devices are used in the same way as a simple Zener diode.

To see the importance of the temperature coefficient, consider the reference required for a 10-bit ADC that must remain accurate between -50 and $+100$°C. Assume that 'accurate' means that any change in V_{ref} should produce less than 1 bit error in the converted value. A change of 1 bit in 2^{10} is about 10^{-3} and this is a range of 150°C so the temperature coefficient must be less than $10^{-3}/150 \approx 7 \times 10^{-6}$/°C or 7 ppm/°C. We would need to splash out on the most expensive reference to meet the specification. Actually we shouldn't do this without further thought: we should check the data sheet more carefully because of the nonlinear dependence on temperature. Even the best reference in this range would not meet the specification for a 12-bit converter.

The moral is that the voltage reference must be chosen carefully if you want a system with high, absolute accuracy. Please make certain that you really need such high *accuracy*: Would high *precision* be good enough? Suppose that you are designing a heating system, for instance. You might want to resolve changes of ± 0.1°C to give good control of the temperature. On the other hand, it might not matter if the selection of the absolute temperature is no better than ± 0.5°C. This example needs good resolution but is less demanding on accuracy.

The best sort of voltage reference is one that you don't need at all. We'll explore this next.

[1] A little history: These components were formerly sold by Burr–Brown, perhaps the premier brand in analogue electronics. The company was acquired by Texas Instruments and the original name has now been eliminated from the documentation. Sad. National Semiconductor has now gone the same way.

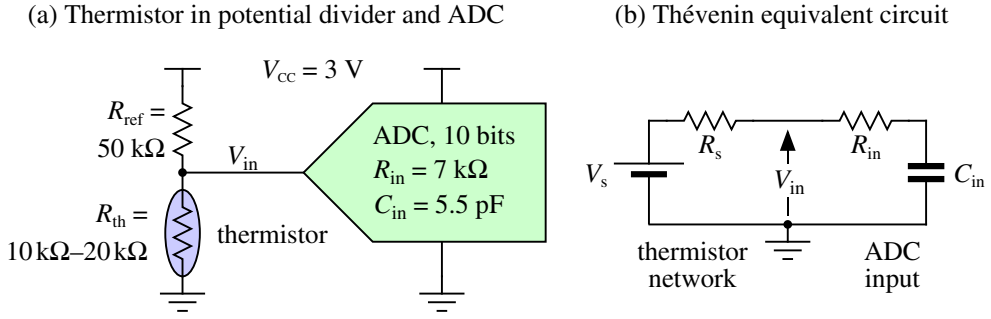


Figure 7.1 (a) Thermistor in a potential divider connected to a SAR ADC. (b) Thévenin equivalent circuit.

7.2 Ratiometric measurements

I have mentioned several times that systems can be designed to render a voltage reference unnecessary. As an example, suppose that a thermistor to measure temperature using the circuit shown in figure 7.1, whose operation will be described in section 7.5 on page 58. The voltage from the potential divider is

$$V_{in} = \frac{R_{th}}{R_{ref} + R_{th}} V_{CC}. \tag{7.2}$$

Here, yet again, is the relation between the input voltage and output of the ADC, *assuming that it uses the supply voltage V_{CC} as its reference*:

$$N_{ADC} = \text{nint}\left(2^N \frac{V_{in}}{V_{CC}}\right). \tag{7.3}$$

Putting these together gives

$$N_{ADC} = \text{nint}\left(\frac{2^N}{V_{CC}} \frac{R_{th} V_{CC}}{R_{ref} + R_{th}}\right) = \text{nint}\left(2^N \frac{R_{th}}{R_{ref} + R_{th}}\right). \tag{7.4}$$

The critical feature is that *the reference voltage V_{CC} has vanished* from the overall behaviour because it affects the signal and ADC equally. If V_{CC} falls, the output of the potential divider falls but so does the reference voltage of the ADC and the change cancels out in N_{ADC}. This circuit does not need a voltage reference.

If you look a little more closely, you will see that the output N_{ADC} depends on the ratio of R_{th} to R_{ref}. This is therefore called a *ratiometric* measurement. It is often possible to design systems in this way without a voltage reference. We'll see another example shortly in section 7.7 on page 62.

7.3 Measurement of absolute voltages with a simple ADC

The two preceding sections have shown how to design two simple systems with an ADC.

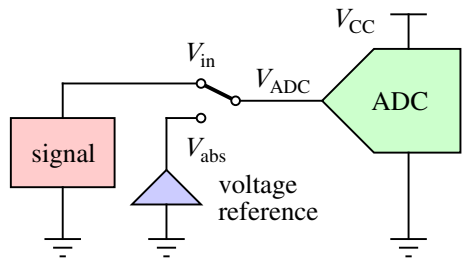


Figure 7.2 ADC with two inputs, the signal of interest V_{in} and an absolute voltage V_{abs}. The reference voltage for the ADC is taken from the power supply at V_{CC}.

- If the voltage from the sensor is proportional to V_{CC}, the reference voltage for the ADC should be taken from V_{CC}.
- If the sensor produces an absolute voltage, the reference voltage of the ADC should be taken from an absolute voltage reference.

Unfortunately you often need to measure an absolute voltage but the ADC takes its reference voltage from V_{CC} and this cannot be changed. An example of this is given in the following section. The output of the ADC is given by equation (7.3) and will change if V_{CC} changes, even if the voltage V_{in} from the sensor remains the same. How should the system be designed to eliminate this error?

The solution requires an absolute voltage reference, which cannot be avoided, but it must be connected to a signal input of the ADC rather than its reference input. The system is shown in figure 7.2. Use the ADC to measure the two input voltages:

$$N_{in} = \text{nint}\left(2^N \frac{V_{in}}{V_{CC}}\right) \quad \text{and} \quad N_{abs} = \text{nint}\left(2^N \frac{V_{abs}}{V_{CC}}\right). \tag{7.5}$$

Dividing the two and dropping the nint() function gives

$$\frac{N_{in}}{N_{abs}} = \frac{V_{in}}{V_{abs}} \quad \text{so} \quad V_{in} = \frac{N_{in}}{N_{abs}} V_{abs}\,. \tag{7.6}$$

Thus the absolute value of the input voltage can be found from the two measurements. Effectively we are using V_{abs} to calibrate V_{CC}.

The same method is often used to measure V_{CC} to check the health of the power supply. It still requires an absolute voltage reference although it might not need particularly high quality.

7.4 Worked example: Temperature sensor with LM35 and 8-bit ADC

We'll now look at a given sensor and ADC and investigate how the complete system should be designed to meet a specification. Suppose first that we want to measure the temperature of a room with the following components.

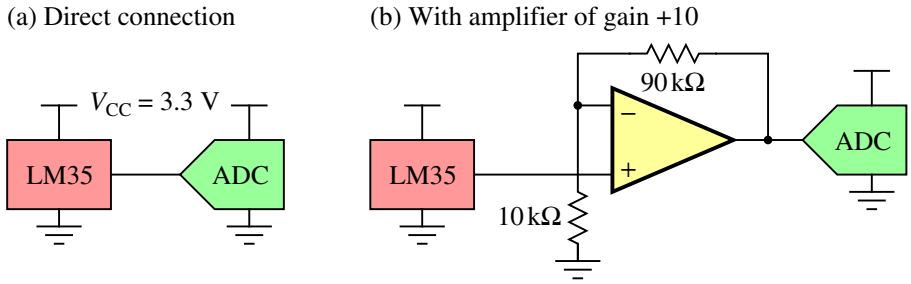


Figure 7.3 An LM35 temperature sensor connected to an ADC (a) directly and (b) through a noninverting amplifier of gain +10.

- The sensor is the LM35, which gives a voltage proportional to temperature in degrees celsius, so that 0°C gives zero voltage. The scaling factor is 10 mV/°C, so the LM35 gives 200 mV at 20°C and so on. The sensor is a silicon IC, not a thermistor. Its output is a linear function of temperature, which is convenient, but the voltage is small.
- The ADC is an 8-bit SAR, common in small microcontrollers, supplied at 3.3 V.

We wish to resolve changes of ±0.1°C over a range of 5°C to 30°C. Is it possible to meet this specification without further components? Only 250 values are needed so it sounds trivial – but is not.

Suppose first that we connect the output of the LM35 directly to the input of the ADC as in figure 7.3(a). What voltages must we measure?

- The range in temperature is 5–30°C and the scaling factor is 10 mV/°C so the output varies from 50–300 mV.
- A change in temperature of ±0.1°C scales to a change of ±1 mV.

Thus we need to measure 50–300 mV with a resolution of 1 mV. The full-scale range of the ADC is 3.3 V. Suppose that we use the maximum resolution of 10 bits. This gives $2^{10} = 1024$ possible outputs so the resolution on the input is LSB $= (3.3\,\text{V})/1024 \approx 3\,\text{mV}$. This is too large so we cannot meet the specification. The smallest change in temperature that could be detected is 0.3°C.

The basic problem is that we are using only a small part of the ADC's range, just 0.25 V out of 3.3 V. Over 90% of the range is wasted so this sensor is a poor match to the ADC. This mismatch is illustrated in figure 7.4(a) on the following page.

The conclusion is that a 10-bit ADC cannot meet the specification, although only 250 values are needed and you might have hoped that an 8-bit ADC would be sufficient. Here are two solutions to this problem.

- Use an external ADC with better resolution. Assume that it works over the same range of voltage, 3.3 V. We need to resolve 1 mV, which needs at least 3300 output values. The next power of two above this is $4096 = 2^{12}$. A 12-bit ADC would therefore meet the specification.

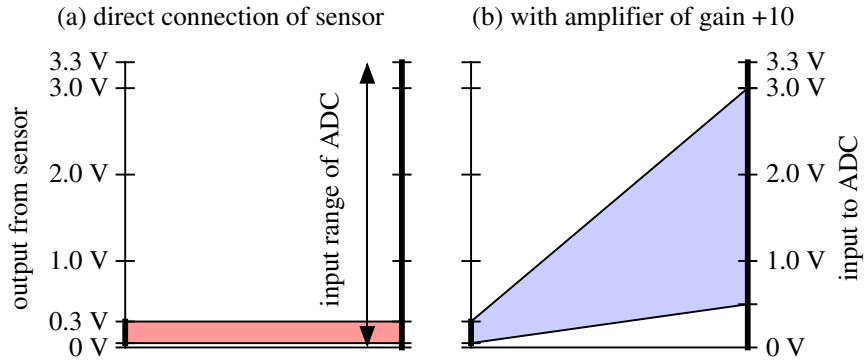


Figure 7.4 Matching of output of LM35 to input of ADC (a) directly and (b) through a noninverting amplifier of gain +10.

- Insert an amplifier with a gain of +10 between the sensor and the ADC as shown in figure 7.3(b). The sensor now presents an amplified input of 500 mV–3000 mV to the ADC for the range 5°C–30°C as shown in figure 7.4(b). The ADC need only resolve a change of 10 mV, which is trivial in 10-bit mode and almost possible in 8-bit mode.

We can therefore meet the specification for resolution with an extra component but not without. This is a simple example of how it is usually necessary to condition a signal before converting it. We should also include a capacitor between the output of the LM35 and ground to reduce noise picked up from the environment. The data sheet suggests a 1 µF capacitor in series with a 75 Ω resistor for the LM35, but other components have different requirements.

A serious defect of this system is that the output of the LM35 is an absolute voltage but the ADC uses V_{CC} as its reference. The output of the ADC will therefore change if V_{CC} changes, even if the temperature stays the same. The previous section explains how to eliminate this error.

7.5 Worked example: Measurement of temperature using a thermistor

Return now to the circuit in figure 7.1 on page 55, which is used to measure temperature using a *thermistor*. This is a resistor made of a special material whose resistance changes strongly with temperature (the opposite of the quality usually desired). Thermistors are widely used because they are simple, cheap and can be encapsulated in a rugged package. They give a large change in resistance, so that they can often be connected directly to an ADC; most other sensors require amplifiers or more complicated signal conditioning. Unfortunately the resistance is a strongly nonlinear function of temperature so a lookup table may be needed. A common application is to measure the temperature of the coolant in a car engine.

A thermistor is usually connected in a simple potential divider as in figure 7.1. (There doesn't seem to be a standard symbol for thermistors.) Suppose that the resistance of the thermistor varies between 10 kΩ–20 kΩ over the operating range. We shall analyse two aspects of this system:

- the time that should be allowed for the ADC to sample the input
- the range of voltages presented to the ADC and whether this circuit can provide a given resolution.

Sampling time

First, turn the potential divider into its Thévenin equivalent circuit, as in figure 7.1(b). I won't explain this because you should know – look back at Electronic Engineering 1X if not.

$$V_s = \frac{R_{th}}{R_{th} + R_{ref}} V_{CC}, \tag{7.7}$$

$$R_s = R_{th} \parallel R_{ref} = \frac{R_{th} R_{ref}}{R_{th} + R_{ref}}. \tag{7.8}$$

Both of these depend on the resistance of the thermistor. Start with the highest value, $R_{th} = 20\,\text{k}\Omega$.

1. The potential divider gives $V_s = 0.86\,\text{V}$ and $R_s = 14\,\text{k}\Omega$.
2. The total resistance in the equivalent circuit is $R = R_s + R_{ADC} = 14 + 7 = 21\,\text{k}\Omega$.
3. The capacitance $C = C_{ADC} = 5.5\,\text{pF}$.
4. Therefore the time-constant is $\tau = RC = 21\,\text{k}\Omega \times 5.5\,\text{pF} = 0.12\,\mu\text{s}$.
5. We found earlier that the input of a 10-bit ADC should be allowed to charge for at least $7.6\,\tau = 0.9\,\mu\text{s}$; say $1\,\mu\text{s}$ for safety.

This must be repeated for the lowest value of resistance, $10\,\text{k}\Omega$, which gives $V_s = 0.50\,\text{V}$ and $R_s = 8.3\,\text{k}\Omega$. The lower resistance allows the input of the ADC to charge more quickly, so in general we should use the longer sampling time calculated for $20\,\text{k}\Omega$.

Resolution

The previous calculation shows that the output voltage ranges from 0.50 V–0.86 V, a span of 0.36 V. This voltage enters an 10-bit ADC, whose range of inputs is 0.0–3.0 V, so the thermistor uses only 12% of the possible values. This is roughly $\frac{1}{8} = 2^{-3}$ of the range so it is like throwing away 3 of the ADC's bits and reducing it from a 10-bit to a 7-bit device.

Another way of looking at this is to calculate $\text{LSB} = (3\,\text{V})/2^{10} \approx 3\,\text{mV}$. The number of outputs over the operating range is given by $0.36\,\text{V}/\text{LSB} = 120$. An amplifier is required to make best use of the ADC.

Consider again a straightforward noninverting amplifier. The maximum input voltage is 0.86 V. The op-amp is supplied from 3.0 V but it is best to keep about 0.1 V away from the supply rails even if the op-amp has a so-called rail-to-rail output. The maximum gain permitted is therefore $2.9/0.86 \approx 3.4$. Use a gain of 3 for simplicity. The range of input voltages is now from 1.50 V–2.58 V, varying by 1.08 V. this is a big improvement but we still waste more than half of the ADC's range.

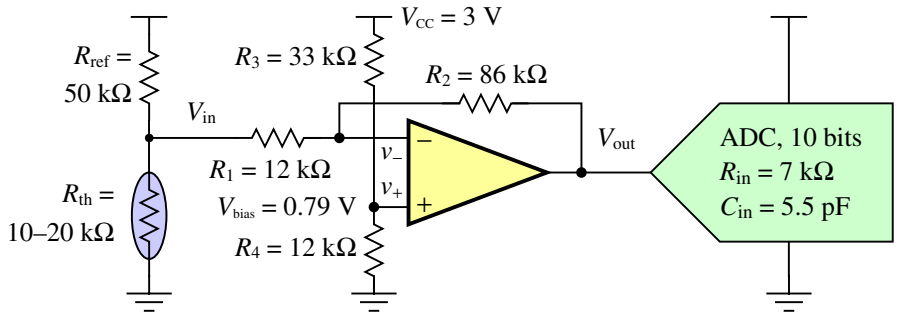


Figure 7.5 Inverting amplifier with bias to match a thermistor to an ADC.

An inverting amplifier with a bias voltage should be used if it is important to fill the range of the ADC. The circuit is shown in figure 7.5 and was analysed in section 6.3 on page 43. It should take an input from 0.50–0.86 V and convert it to an output from 2.9–0.1 V, which is safely clear of the supply rails. Its gain should be

$$\frac{2.9-0.1}{0.50-0.86} \approx -7.8. \tag{7.9}$$

The gain doesn't have to be an integer. It is set by the usual pair of resistors, $-R_2/R_1$ in figure 6.5 on page 43. You will have to experiment if you wish to get near this value of gain with standard values. For example, $R_1 = 8.6\,\text{k}\Omega$ and $R_2 = 68\,\text{k}\Omega$ give a gain of -7.9 or $R_1 = 12\,\text{k}\Omega$ and $R_2 = 86\,\text{k}\Omega$ give -7.2. The second choice is safer.

Having found the gain, the next step is to calculate the bias voltage. Here is a reminder of equation (6.3):

$$V_{out} = \left(1 + \frac{R_2}{R_1}\right) V_{bias} - \frac{R_2}{R_1} V_{in}. \tag{7.10}$$

Put $V_{in} = 0.5\,\text{V}$, $V_{out} = 2.9\,\text{V}$, $R_1 = 12\,\text{k}\Omega$ and $R_2 = 86\,\text{k}\Omega$ into this:

$$2.9\,\text{V} = \left(1 + \frac{86}{12}\right) V_{bias} - \frac{86}{12} \times 0.5\,\text{V}, \tag{7.11}$$

which gives $V_{bias} = 0.79\,\text{V}$.

Finally, we need a potential divider to give 0.79 V from 3.0 V. Again you must experiment and I found that 12 kΩ and 33 kΩ gave 0.80 V, which is pretty good. These values are shown in figure 7.5.

We seem to have a satisfactory system but sadly it has a major problem. We found earlier that the potential divider has an output resistance of about 14 kΩ. The input resistance of the inverting amplifier is given by $R_1 = 12\,\text{k}\Omega$. The amplifier will therefore load the potential divider severely and have a large effect on the voltage measured, up to a factor of 2. This is an unacceptable error but it is possible to calculate the error and apply a correction. Alternatively a voltage follower (unity-gain buffer), meaning an opamp connected to give a gain of $+1$, can be connected between the potential divider and the inverting amplifier.

7.6 Worked example: sensor with given range of voltages.

This is a typical example from a past test paper. The output of a temperature sensor is a voltage proportional to absolute temperature with a scale of $10\,\text{mV}\,\text{K}^{-1}$. It is required to work over the range $-40°\text{C}$ to $+75°\text{C}$. (Take $0°\text{C} = 273\,\text{K}$.) The output is connected directly to an ADC, whose full-scale voltage is set by the single power supply at 5.0 V.

(a) What range of voltages must be converted by the ADC?

(b) The system is required to resolve 0.5°C or better. How many bits of output must the ADC produce?

(c) The digital system chosen for this application has only an 8-bit ADC. Is it possible to meet the specification of the system by including some type of amplifying circuit? If so, explain what type of circuit should be used and give its specification but do not design the circuit in detail.

(d) A reviewer suggests that it would be better to use an ADC with a separate reference voltage, rather than a reference derived from the power supply. Explain whether this is good advice or not.

(e) The designer accepts the advice and chooses a reference with a voltage drift of 100 ppm/°C. Explain whether this is a good choice or not.

The lowest temperature is $-40°\text{C} = 233\,\text{K}$ giving 2.330 V; the maximum temperature is $+75°\text{C} = 348\,\text{K}$, giving 3.480 V. The range is therefore 2.330 V to 3.480 V.

A resolution of 0.5°C means 5 mV so the ADC must resolve at least $5.0\,\text{V}/5\,\text{mV} = 1000$ values. The next power of 2 up is 1024, giving a 10-bit ADC. Note that you must use the full range of the ADC, not just the range of inputs from the sensor.

The number of values to be resolved in the range of interest is $(75°\text{C}--40°\text{C})/0.5°\text{C} = 231$ including both endpoints. This is possible with an 8-bit ADC, which has 256 output values. The 8-bit ADC has $\text{LSB} = 5.0\,V/256 \approx 19.53\,\text{mV}$ so we need an amplifier with a gain of $19.53/5 = 3.906$ to match the change in output from the sensor for a change of 0.5°C to LSB of the ADC. A simple noninverting amplifier will not work because this gain would increase the voltage at +75°C to 13.6 V, which is far above the supply voltage. We must therefore use an inverting amplifier with an offset voltage to keep the output voltage within the supply rails.

It is good advice to use a separate reference voltage. The current system has an absolute input voltage to the ADC but the reference voltage is taken from the power supply, which may be poorly stabilised. It is not a ratiometric measurement. The reference voltage should also be an absolute value, which requires a 'real' reference.

The range of temperatures is 115°C, giving a fractional change in reference voltage of $115 \times 100 = 11500\,\text{ppm} = 0.0115 = 1.15\%$. The fractional resolution of the ADC is $1/256 \approx 0.0039 = 0.39\%$ so the change in reference voltage is equivalent to a change in output of 3 for the maximum input accepted by the ADC. Maybe not a good choice.

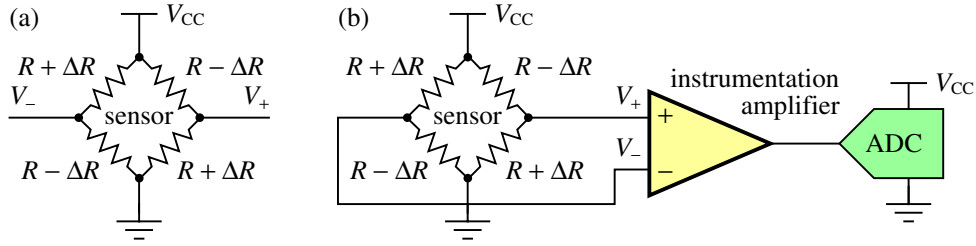


Figure 7.6 (a) Weight sensor with four elements connected as a Wheatstone bridge. (b) Sensor connected to an instrumentation amplifier and ADC.

7.7 Worked example: Sensor for a weighing machine

The final example is the sensor and ADC for an electronic weighing machine. The sensors are usually based on the *piezoresistive effect*, which means that the resistance of a material changes when it is strained (distorted mechanically). Typically the sensor is is a thin diaphragm of silicon, four regions of which act as piezoresistive sensors. They are connected as a Wheatstone bridge, which is arranged so that the resistance of two elements goes up when a weight is applied and the other two goes down. The circuit is shown in figure 7.6(a).

Look at the voltage V_+ first. Ideally $V_+ = \frac{1}{2}V_{CC}$ and $\Delta R = 0$ when no weight is present. When a weight is applied,

$$\frac{V_+}{V_{CC}} = \frac{R + \Delta R}{(R + \Delta R) + (R - \Delta R)} = \frac{1}{2} + \frac{\Delta R}{2R}. \tag{7.12}$$

The largest part of this is the $\frac{1}{2}$ because $\Delta R/R$ is small. Similarly,

$$\frac{V_-}{V_{CC}} = \frac{R - \Delta R}{(R - \Delta R) + (R + \Delta R)} = \frac{1}{2} - \frac{\Delta R}{2R}. \tag{7.13}$$

Again, the largest part is the $\frac{1}{2}$. Instead of looking at V_+ and V_- themselves, it is more illuminating to work with their average value and difference. The average is called the *common-mode* voltage and is given by

$$V_{CM} \equiv \frac{V_+ + V_-}{2} = \frac{1}{2}V_{CC}, \tag{7.14}$$

while the difference is

$$\Delta V \equiv V_+ - V_- = \frac{\Delta R}{R}V_{CC}. \tag{7.15}$$

The common-mode voltage is large but boring and we are interested only in the small difference ΔV, which is a common situation.

A typical value for the sensitivity is $\Delta R/R = 1\%$ for a full load of 1000 g. Suppose that we wish to resolve differences of 10 g, which might be good enough for domestic kitchen scales (actually mine work to 5 g). This requires only 100 intervals, which sounds trivial. Unfortunately it is trivial only if the range of outputs from the sensor matches the range of inputs to the ADC perfectly. If possible we'll use a simple 8-bit ADC.

Try first to convert V_+ and V_- separately and use subtraction to find the difference of the digital values. What is the range of V_+ and by how much does it change as a function of weight?

- $V_+ = 0.500\,V_{CC}$ with no weight
- $V_+ = 0.505\,V_{CC}$ with the maximum weight of 1 kg, a change of $0.005\,V_{CC}$
- a change in weight of 10 g therefore causes a change of $\Delta V_+ = 0.000\,05\,V_{CC}$

If the ADC works between 0 and V_{CC}, it needs to resolve $1/0.000\,05 = 20\,000$ values. This would need a 15-bit ADC, which seems crazy when we need only 100 final values! The problem is that we are using only the range from 0.500 to 0.505 of V_{CC}, which is $1/200$ of the range of inputs to the ADC. Over 99% of its range is wasted.

Clearly we are approaching this system the wrong way. This is a job for an instrumentation amplifier (section 6.1 on page 40) as shown in figure 7.6(b). It amplifies the desired difference $\Delta V = V_+ - V_-$ but suppresses the common-mode voltage $V_{CM} = \frac{1}{2}V_{CC}$. The maximum difference voltage is $\Delta V = 0.01\,V_{CC}$ so a gain of 50 (say) would bring this to $0.5\,V_{CC}$. We need to resolve 100 values within the range from 0 to $0.5\,V_{CC}$, which means 200 values within the full range from 0 to V_{CC}. This requires only an 8-bit ADC as desired.

Why not amplify the output of the bridge all the way to V_{CC}? The reason is that practical strain sensors have a large *offset*. This means that the four resistances in the bridge are not all equal when no weight is applied. In turn, this implies that $\Delta V \neq 0$ and it may be positive or negative. We must compensate for the offset by adding an offset voltage to the circuit of the instrumentation amplifier (V_{bias} in figure 6.1 on page 40). This keeps the input to the ADC positive, which it needs for valid conversions. Thus we need an amplifier with a potentiometer for the offset voltage. The potentiometer is called a *tare* control and is used to adjust the scales so that they read zero when the pan is empty.

A good feature of this design is that both the input voltages and the reference voltage for the ADC are taken from V_{CC}. This makes it a *ratiometric* measurement as described in section 7.2 on page 55. No voltage reference is therefore required.

It would be even better if we could connect the outputs of the bridge directly to an ADC. We need two vital features:

- differential inputs for V_+ and V_- to reject the common-mode voltage
- high resolution because of the small changes in voltage

This suits a sigma–delta ADC perfectly. For example, the ADC in the MSP430F2003 (table 3.1 on page 13) has differential inputs with a range of ± 0.6 V and 16-bit resolution. Its LSB is therefore $(1.2\,\text{V})/2^{16} = 18\,\mu\text{V}$. For comparison, a change in weight of 10 g changes $\Delta R/R$ by 0.0001. This is equal to $\Delta V/V_{CC}$ so ΔV changes by 0.0003 V or 300 µV if $V_{CC} = 3$ V. In fact we could measure to a resolution of 1 g. This ADC includes a programmable gain amplifier, which could improve resolution further. The only catch is that the measurement is no longer ratiometric because the ADC in the MSP430F2003 does not use V_{CC} for its reference.

This shows the advantage of using the correct type of ADC for the job. The sigma–delta ADC can also handle negative differences in voltage from the offset and we could subtract the offset voltage digitally. Thus we don't need a potentiometer – just a Tare button.

This might sound a bit specialised but bridge sensors are widely used, not just in weighing machines. Pressure sensors and strain gauges are very similar. Another example is the mass air flow (MAF) sensor in a car. This measures the rate at which air enters the engine, which is one of the key inputs to the engine management system.

7.8 Examples

Example 7.1 A reference diode has a temperature coefficient of 0.005%/°C. Over what temperature range could it be used in conjunction with (i) an 8-bit converter (ii) a 12-bit converter? [78°C, 5°C]

Example 7.2 What temperature coefficient is required for a reference voltage source which is to be used with a converter operating over a temperature range of 0°C to 70°C if the converter has (i) 8 bits, (ii) 12 bits, (iii) 16 bits? [5×10^{-3}%/°C, 3×10^{-4}%/°C, 2×10^{-5}%/°C]

Example 7.3 What value should be chosen for the reference resistor R_{ref} to get the maximum change in voltage for the system in section 7.2 on page 55, where R_{th} varies between 10 kΩ and 20 kΩ? What is the new change in voltage? Hint: find an expression for the change in voltage and find its maximum as a function of R_{ref}. It's straightforward but messy. [About 14 kΩ]

Example 7.4 What change in temperature could be resolved with the system in section 7.4 on page 56 using the ADC in 8-bit mode?

Example 7.5 Is a gain of 10 the best choice to resolve 0.1°C for the system in section 7.4?

Example 7.6 Suppose that the range of temperatures were extended to 0–30°C for the system in section 7.4. Is the design with the amplifier still satisfactory? (The answer is not trivial.)

Example 7.7 A sensor produces an output between 0.5 V and 1.0 V, which is to be must be digitized with a precision of 1 mV or better. The ADC has a full-scale range from 0.0 V–2.5 V. How many bits are needed if the sensor is connected directly to the ADC and what would the actual resolution be? How would this change if an amplifier were used to match the output of the sensor to the full range of inputs of the ADC? Give a specification for the amplifier.

Example 7.8 Figure 7.7 shows a light-dependent resistor (LDR) connected in a potential divider. The LDR has resistance 1.6 kΩ in the light and 100 kΩ in the dark. Find the voltage on the potential divider and the output resistance at these two extremes. (In other words, find its Thévenin equivalent circuit.)

The ADC is in the LPC1768 and has a resolution of 12 bits, an input resistance of 7.5 kΩ and capacitance of 15 pF. For how long should it sample its input to ensure that errors due to incomplete charging are negligible? How many clock cycles is this for an ADC clock at 13 MHz?

If the system had only to distinguish between light and dark, what simpler component could be used?

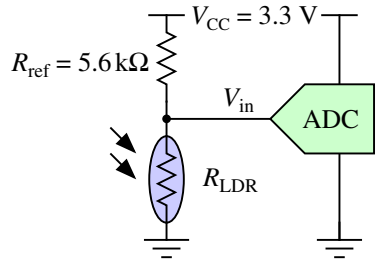


Figure 7.7 A light-dependent resistor (LDR) connected in a potential divider and monitored by an ADC.

Example 7.9 [Hard] Calculate the error due to the input resistance of the inverting amplifier in figure 7.5. Hint: the voltage at the inverting input of the op-amp is held fixed by the negative feedback. This is similar to a virtual ground but the voltage is V_{bias} rather than ground. Use nodal analysis to find V_{in}.

Suggest in principle how the error could be corrected (details are not required).

8

Digital-to-analogue converters

8.1 Introduction

Digital-to-analogue converters or DACs are used much less than ADCs, which is why I've put them at the end. Few microcontrollers provide DACs, for instance, although the LPC1768 has one. The main reason is that pulse-width modulation (PWM) is good enough for many applications and needs only a timer, which is purely digital and therefore much simpler than a true DAC.

Of course it is not always acceptable to use digital output to simulate an analogue signal. The most common exception in consumer products is audio, where power amplifiers have almost always been linear circuits. Afficianados of the 'valve sound', who like to see the anodes of their output amplifiers glowing red-hot, are unlikely to switch allegiance but many audio power systems now use PWM or something similar. The main drive is the efficiency required if an audio power amplifier is to be squeezed into a flat-screen television. It is simply not possible to cool a linear amplifier effectively in a thin casing without a fan, whose noise would be intrusive.

8.2 General features of digital to analogue converters

A DAC takes in a digital value and produces an analogue output. The output of an ideal 3-bit DAC is shown in figure 8.1 on the facing page, compared with an ideal 3-bit ADC. I've assumed that the analogue signals are all voltages although this is often not true for DACs. An important difference between the two plots is that the DAC has only a discrete number of input values while the input to the ADC is a continuous function. Thus the plot for the DAC has just 8 points, although I have joined them with a line for clarity. The output voltage V_{out} of an ideal N-bit DAC is given by

$$V_{\text{out}} = \frac{N_{\text{DAC}}}{2^N} V_{\text{FS}} = N_{\text{DAC}}\text{LSB}, \tag{8.1}$$

where the digital input is N_{DAC} and LSB is defined in exactly the same way as for ADCs (equation 2.3). This is the analogue of equation (2.2) on page 7 but doesn't need the nint() function because it is a one-to-one relation between input and output.

A digital input of zero gives an analogue output of zero. The only surprise is at the the top end of the range. The maximum input to a 3-bit DAC is 0b111 = 7 and there are $2^3 = 8$

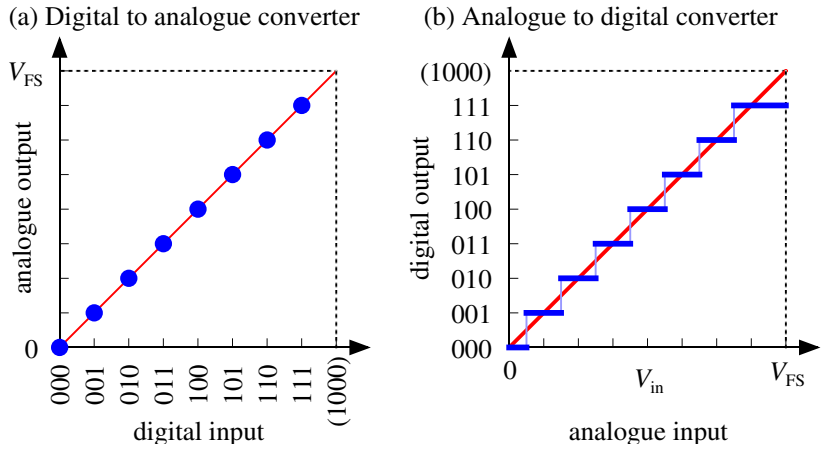


Figure 8.1 Output as a function of input (transfer function) for an ideal 3-bit DAC and an ADC.

possible values so the maximum output is $\frac{7}{8}V_{FS}$. It is not possible to get the full-scale output V_{FS} from a DAC. The same problem causes the extra-long step at the top of the range of an ADC (figure 8.1(b)).

No DAC is ideal and the errors are specified in the same way as for ADCs, so I won't say much more. A straightforward number to check is the *integral nonlinearity*, which gives the maximum deviation in LSBs between the real output points and the ideal straight line.

Like ADCs, DACs do not produce 'absolute' outputs but need a reference, which may be internal, external or taken from V_{CC}. The same advice applies as for ADCs.

The maximum sampling frequency was one of the important specifications of an ADC. This is slightly more complicated in DACs because two numbers may be quoted. The update rate specifies how often the digital input may be loaded but you are probably more interested in how rapidly the output changes. This is given by the *output voltage settling time*, which may be a lot longer. It may also depend on the size of the change in the input.

Another issue, because the output is analogue, is its *compliance*. This means how much current you can draw without affecting the voltage 'significantly'. The output of some DACs is a current, in which case the maximum voltage is specified instead. The digital inputs are usually serial, typically SPI or I²C, but may be parallel.

I'll now run through some of the more common types of DAC and the methods that are used as substitutes for them. To be logical I ought to start with a 1-bit DAC. This has an output that can be on or off – in other words a switch. A straightforward digital output does this, switching its voltage between V_{SS} and V_{DD}. We have already encountered a 1-bit DAC in the loop of a sigma–delta modulator. It is the opposite of a comparator, which is effectively a 1-bit ADC.

8.3 Pulse width modulation

Pulse width modulation is a common substitute for real digital-to-analogue conversion. The load is switched on and off at a fixed frequency, which can be performed with a purely digital

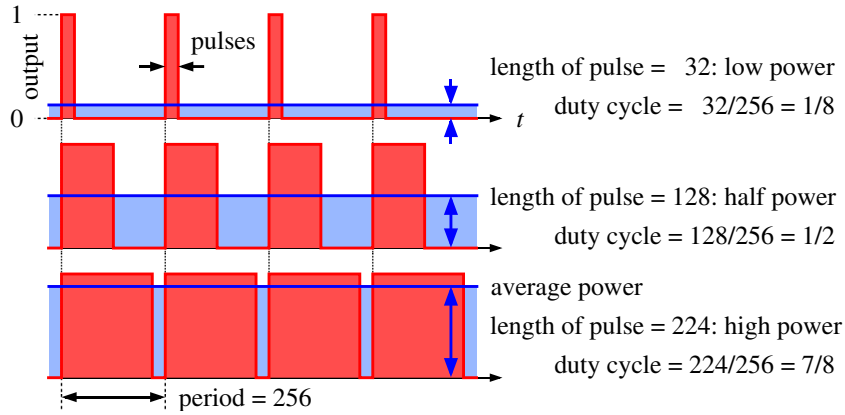


Figure 8.2 Output as a function of time for pulse width modulation, showing three power levels.

system. The fraction of time for which the load is switched on is called the duty cycle ('duty fraction' might be more accurate) and is adjusted to give the desired average value. The output is sketched for three duty cycles in figure 8.2.

If no smoothing is applied, the frequency of the square wave must be high enough not to be noticeable. LEDs should be modulated at 100 Hz or more, for instance, so that the eye does not readily discern the flashing. Many loads provide their own smoothing. For example, heaters often have a large thermal mass and their temperature only responds slowly, smoothing out the changes in heat input. Many loads such as motors are inductive and act as low-pass filters themselves. Remember that an inductor obeys

$$V_\mathrm{L} = L\frac{\mathrm{d}I_\mathrm{L}}{\mathrm{d}t} \qquad \text{so} \qquad I_\mathrm{L} = \frac{1}{L}\int V_\mathrm{L}(t)\,\mathrm{d}t. \tag{8.2}$$

The torque produced by a simple motor is proportional to current and the inductance converts the square wave in voltage into a triangular wave in current. This is less disruptive but you can often hear the switching frequency for the traction motors on trains. The PWM frequency must be fast enough not to produce mechanical resonances or anything nasty. It is possible to filter the PWM output if a steady (so-called DC) voltage is really needed. However, a real DAC may be a better solution in such cases.

Almost all microcontrollers, including the LPC1768, contain one or more timers – often many – to generate waveforms for PWM in hardware, independently of the main processor. They are controlled by special function registers in the usual way and run automatically after the period and length of pulse have been loaded.

There are variations on this conventional form of PWM. Sometimes the length of the pulse is kept constant and the repetition frequency is varied to control the average power. Amplifiers that use PWM or variations are known as *Class D* in audio applications.

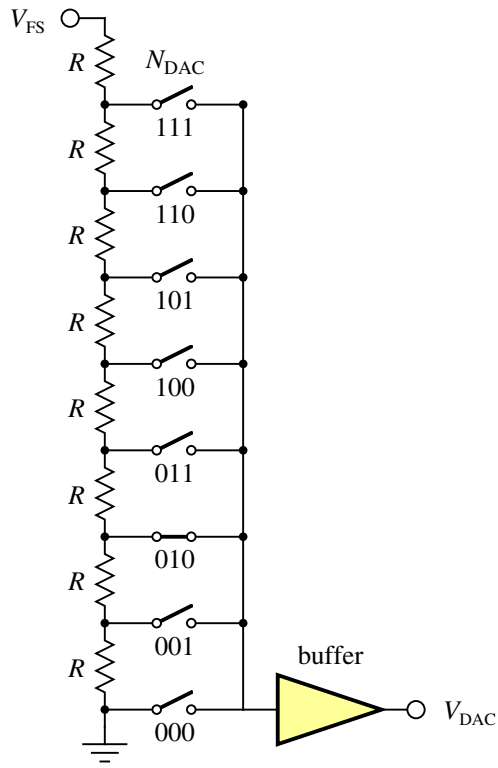


Figure 8.3 Simple string DAC. A 3 to 8 decoder is also needed to control the switches. The input is $N_{DAC} = 0b010$.

8.4 Types of digital to analogue converter

I'll now run through some of the more common types of DAC. Far fewer DACs than ADCs are listed in the catalogues and many are intended for specific applications. Although some produce voltages, the core of many DACs produce currents instead. The output from the IC may be a current or an internal amplifier may be used to turn the current into a voltage.

String DACs

These are perhaps the simplest DACs and the circuit is roughly equivalent to a flash ADC (section 3.3). They are also known as *voltage segment* DACs and the circuit is shown in figure 8.3. Voltages are tapped off a string of equal resistors between V_{ref} and ground, called a *Kelvin divider*. This circuit has a couple of important features, simple as it is.

- The resistors are all equal, even the two at the ends, in contrast to the chain in the flash ADC.
- The bottom of the ladder has a tap, which can be selected to give an output of zero.

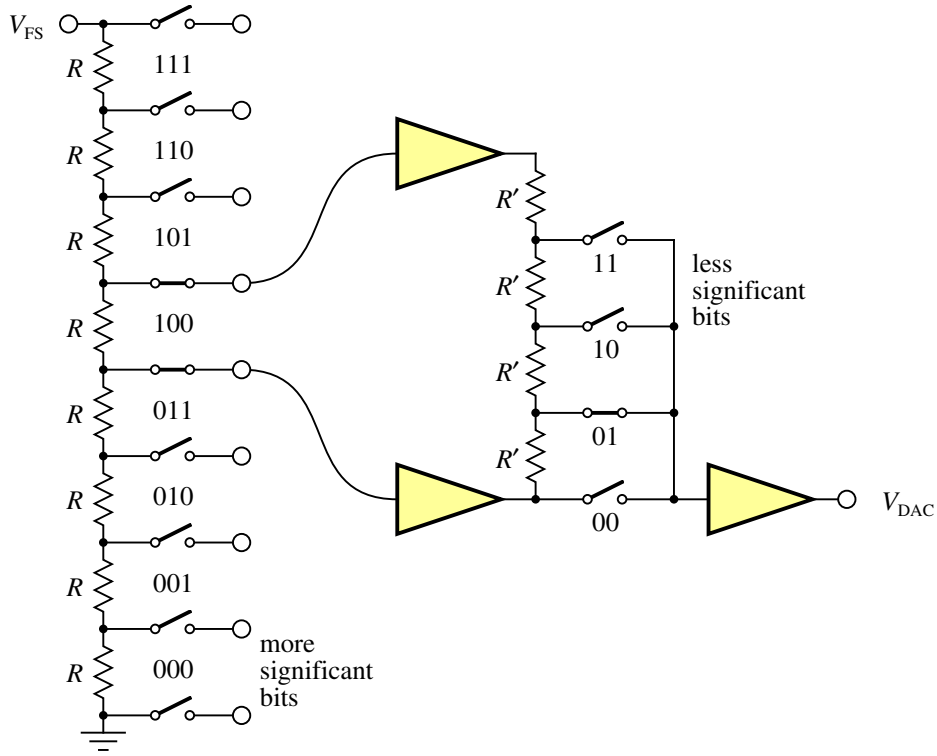


Figure 8.4 Interpolating string DAC. A 3 to 8 decoder is needed to control the switches on the main string and a 2 to 4 decoder for the interpolator. The input is $N_{DAC} = 0b10001$.

> However, the top has *no* tap. This is because of the transfer function of a DAC, which can never reach its full-scale value of V_{ref} (figure 8.1 on page 67).

A simple string DAC has a single chain of resistors as shown in figure 8.3(a). I have included a buffer (voltage follower), which ensures that the output does not load the resistors and affect the voltage. This architecture gives a monotonic output because a higher tap can never give a lower voltage than a lower tap. The resistors must be identical to give the ideal output. The DAC needs a decoder to take a 3-bit digital input and select one of the 8 switches, which are MOSFETs as usual. This is the opposite of the priority encoder used in the flash ADC.

An obvious problem with this circuit is the number of resistors needed, 2^N for an N-bit DAC. This can be reduced by using two chains as shown in figure 8.4. The second chain is connected across two adjacent taps on the first chain and interpolates between the two voltages. I have shown 4 resistors in the second chain, which raises the total number of bits from 3 to 5. This DAC is slower because the voltages must pass through two sets of switches and buffers.

You might be surprised that so primitive-looking an architecture is widely used. Resolution varies from 8 to 16 bits and the LPC1768 offers a 10-bit string DAC. Most include a buffer amplifier on the output as shown in figure 8.3. Decoding the input and changing the internal

switches is fast so the speed is generally limited by the *slew rate* of the buffer. You will learn about this in Analogue Electronics 2. It means that the output changes rapidly for small steps in the digital input but more slowly for large steps. Typical settling times are a few µs.

Digital potentiometers

Analogue potentiometers or *trimmers* are often used to make small adjustments to circuits to bring them into specification. This requires a human to perform the calibration, which is expensive, and the contacts on potentiometers degrade over time. Trimmers are therefore best avoided.

A *digital potentiometer* is a possible substitute. Its circuit is virtually the same as a simple string DAC without a buffer. It has one fewer resistor and the top tap is connected directly to the V_{ref} input (although it is no longer called that). Thus the output can be connected to either one of the inputs or to any of the taps between. Another difference is that the setting may be stored in non-volatile memory so that it is retained permanently. Otherwise the calibration would have to be repeated every time that power was applied.

Digital potentiometers are sometimes called digipots or RDACs. They typically offer 32–1024 settings with volatile, one-time-programmable or flash memory. Some can be connected to mechanical pushbuttons to replace a potentiometer on a control panel but most have serial interfaces for in-circuit control. They consume current, unlike mechanical potentiometers, but this may be below 1 µA. The price isn't particularly low so this may be one example where the analogue approach is cheaper! A basic, volatile, 32-position device costs about $0.40 but fancier, non-volatile ones may cost over $5. My impression is that they have not caught on to the extent that the manufacturers hoped.

Thermometer DACs

These have nothing to do with temperature! The name refers to the code that is used to control the switches inside the DAC, which is the same as in a flash ADC. They are also called fully-decoded DACs. This is the first example of a DAC that produces current rather than voltage. The idea of a thermometer DAC is simple: it contains a set of identical current sources and the required number of these is selected in parallel. The digital input is converted into thermometer code to drive the switches.

Figure 8.5 on the next page shows some of the circuits that can be used to make a thermometer DAC. The simplest is a set of equal current sources in parallel. If the digital input is 2, then 2 sources are connected to the output. Note that only 3 sources are needed for a 2-bit DAC because the highest input is 3! It is much better to switch the output of the sources between two buses as in figure 8.5(b) on the following page because a current source does not like to be disconnected from a load – it is the equivalent of short-circuiting a voltage source. This provides a differential output: current is switched from one to the other according to the digital input. The 'negative' output I_- can be grounded if it is not wanted.

How do we get all these current sources? A straightforward way is to use a classic circuit called a *current mirror*, which you will analyse in Electronic Circuit Design 3. The simplest,

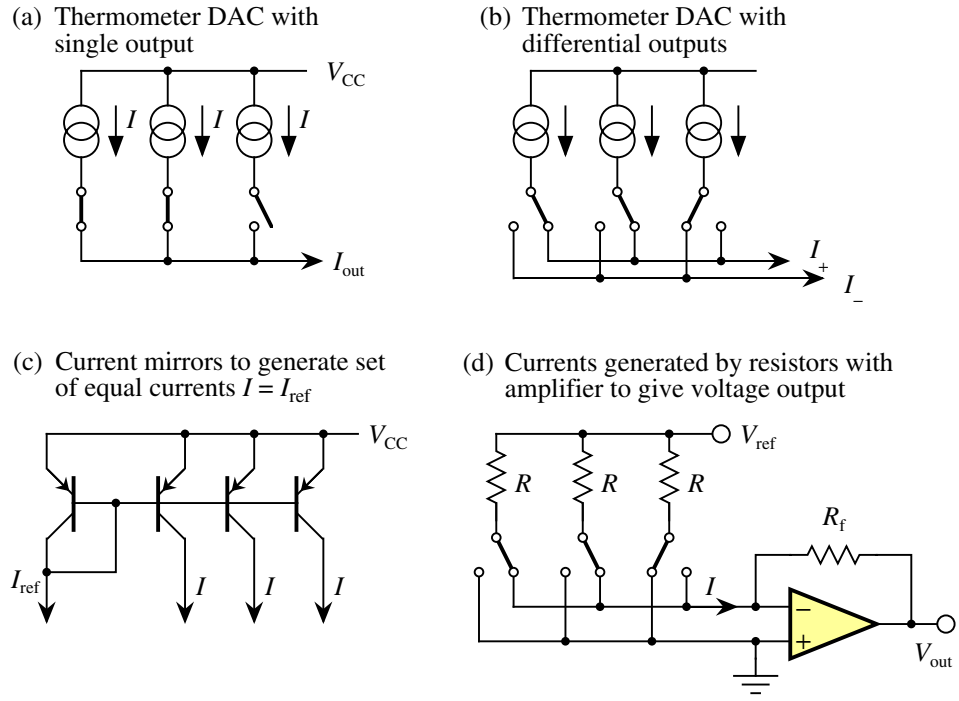


Figure 8.5 A selection of circuits that function as 2-bit thermometer DACs: (a) current sources and single (current) output, (b) current sources with differential outputs, (c) bipolar transistor mirrors as current sources and (d) resistors with an op-amp to give voltage output.

reliable equation for the collector current through a bipolar transistor in active mode is

$$I_c = I_s \exp\left(\frac{V_{be}}{V_T}\right). \tag{8.3}$$

Active mode requires a sufficiently large collector–emitter voltage, roughly $V_{ce} > 0.3\,\text{V}$. This is a bare-bones version of the *Ebers–Moll* equation. The prefactor I_s is called the scale or saturation current; it depends on the area of the transistor and the way in which it is made. The other constant is the thermal voltage, $V_T = k_B T/e \approx 26\,\text{mV}$ at room temperature. The collector current does not depend on the collector voltage in this simple model – the transistor acts like a current source, which is just what we want. (Of course this picture is full of approximations.)

The main point is that the collector current is controlled by the base–emitter voltage V_{be}. This means that if we collect a set of identical transistors with their bases and emitters in parallel, they feel the same V_{be} and their collector currents are the same. If we apply a known current I_{ref} to one transistor, its V_{be} adjusts to the appropriate value for this current, the other transistors experience the same value of V_{be} and therefore pass the same current, $I = I_{ref}$. This is shown in figure 8.5(c). I have used pnp transistors and drawn them upside down so the emitters (with the arrows) are at the top. This is so that the positive supply is at the top of the drawing, which

is conventional.

The same method can be used with MOSFETs by connecting their sources and gates in parallel. The corresponding equation for the drain current in terms of the gate–source voltage is

$$I_{\mathrm{d}} = K(V_{\mathrm{gs}} - V_{\mathrm{t}})^2, \tag{8.4}$$

where V_{t} is the threshold voltage at which the transistor switches on (not to be confused with the thermal voltage). The MOSFET must be in saturation mode, which is equivalent to active mode for a bipolar transistor – a confusing nomenclature. This needs $V_{\mathrm{ds}} > V_{\mathrm{gs}} - V_{\mathrm{t}}$.

Larger currents can be obtained by connecting transistors in parallel. This is sometimes used to produce sources with binary-weighted currents of I, $2I$, $4I$ and so on. These can be combined to give the desired current by switching each source according to the value of the corresponding bit in the digital input. No decoding is required.

The simplest way of generating a known current is to connect a known voltage across a resistor and this can be used in a thermometer DAC as well. The circuit is shown in figure 8.5(d). You may think that it is simpler to make resistors than current sources, but this is not obvious for an integrated circuit! Each current is switched to the output if required or to ground if it is not needed. The voltage across the resistors must be kept constant for accurate currents. This is achieved by feeding the output into the inverting input of an op-amp. The non-inverting terminal is grounded and negative feedback keeps the inverting terminal at the same potential: a virtual ground (remember that?). Thus the voltage across the resistors is held at V_{ref} whether they are connected to the output or ground.

Another advantage of switching unwanted resistors to ground rather than disconnecting them is that the total current drawn remains constant. This reduces errors due to the dependence of V_{ref} on current.

The amplifier has a second function, which is to turn the current I into a voltage. This current cannot flow into the op-amp's input so it must flow through the feedback resistor, giving $V_{\mathrm{out}} = -R_{\mathrm{f}}I$. Now, I contains a contribution V_{ref}/R from each of the n resistors connected to the output, so

$$V_{\mathrm{out}} = -\frac{R_{\mathrm{f}}}{R} V_{\mathrm{ref}}\, n. \tag{8.5}$$

The full-scale output is therefore $-2^N (R_{\mathrm{f}}/R) V_{\mathrm{ref}}$. It is negative for the circuit as drawn, which is a problem in a system with only a single supply.

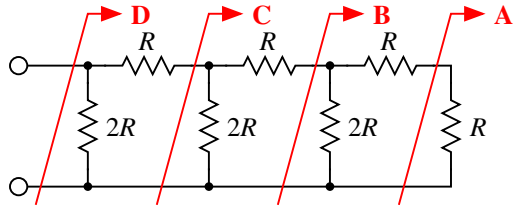
Another way of analysing the circuit is to treat V_{ref} as a voltage input. Then the circuit looks like a straightforward inverting amplifier and its gain is given by the usual ratio of resistances, $-R_{\mathrm{f}}/(R/n) = -(R_{\mathrm{f}}/R)n$.

DACs based on the R–2R chain

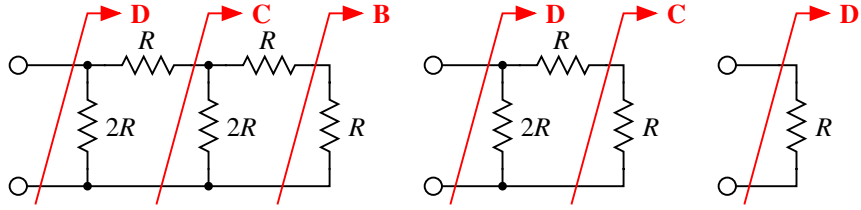
Figure 8.6 on the next page shows a classic circuit, the R–$2R$ chain, which is used in many DACs and formerly in successive-approximation ADCs. It is elegant to analyse. Suppose that the circuit is cut along the lines labelled A–D and we connect an ohmmeter to the part that remains on the right: what would it measure?

A. This is trivial: the resistance is just R.

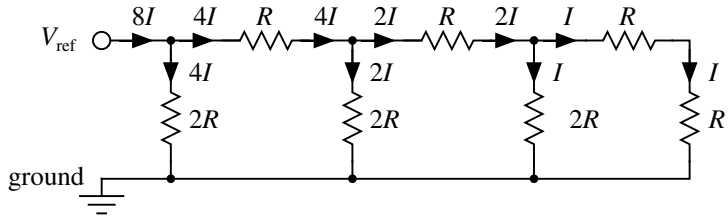
(a) Classic *R*–2*R* chain



(b)–(d) Steps in the analysis of a *R*–2*R* chain



(e) *R*–2*R* chain in 'current mode'



(f) Complete 3-bit DAC based on *R*–2*R* chain with input of 110

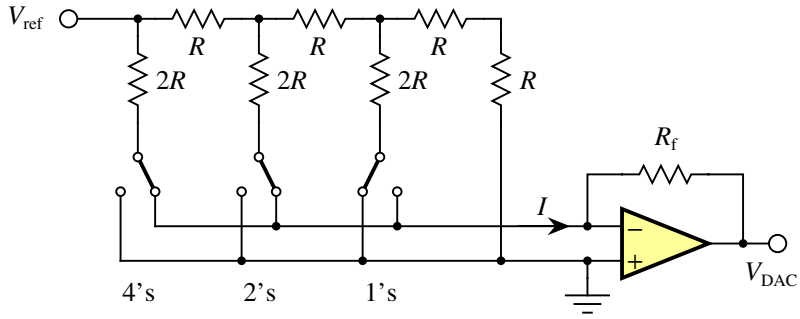


Figure 8.6 The classic *R*–2*R* chain: (a) basic circuit, (b)–(d) steps in the reduction of the circuit, showing that the sections to the right can be replaced by a resistance of *R* at each stage, (e) currents through the chain, which divide by a factor of 2 in each stage and (f) complete 3-bit DAC with transresistance amplifier to give voltage output.

B. Now we have all three resistors to the right of line B, supposing that the wires cut for A have been soldered back together. Two resistors of R in series give $2R$, which is in parallel with another resistor of $2R$, and $2R \parallel 2R = R$ so we would measure R again.

C. Now we have a more complicated circuit. However, we have just shown that we can replace everything to the right of B with a single resistor R, as shown in figure 8.6(b). This makes it trivial to work out the resistance seen looking into C because it is the same as the last calculation and the result is again R.

D. Again, we can replace everything to the right of C by a single resistor R as in (c). The calculation is the same again and the apparent resistance is R.

Thus the resistance is simply R between the external terminals, as shown in figure 8.6(d). The $2R$ resistors are usually made from two resistors of R in series so that all resistors have the same value, which makes it easier to fabricate them accurately.

This is very pretty but what is the use? Suppose that we connect the upper terminal to a voltage reference, V_{ref} as in figure 8.6(e). This feeds a current into the network, which I'll call $8I$ for convenience. Its value is $8I = V_{\text{ref}}/R$ because we have just shown that the overall resistance of the network is R. How does the current flow inside the network?

- At the first node the current can go either through the $2R$ resistor to ground or through the rest of the network to the right. Figure 8.6(c) shows that the rest of the network behaves like a resistance of $2R$ so the current splits equally with $4I$ down each branch.
- A current of $4I$ reaches the next node in the network and again splits equally for the same reason, giving $2I$ in each branch.
- The same happens at the third and final node.

Thus the currents that flow through the $2R$ resistors to ground are divided by two at each stage. There is an extra current with the smallest value, just as there was an extra capacitor with the smallest value in the charge-redistribution network of a SAR ADC (figure 3.8 on page 22).

These binary-weighted currents can be used to make a DAC by switching and combining them. This is similar to a thermometer DAC but the binary weighting means that each bit in the digital input drives a switch directly; the thermometer code is not needed. A transresistance amplifier is again used to collect the currents at a constant voltage (ground) and transform the currents into a voltage.

I have described the 'current mode' of operation of an R–$2R$ chain. It can also be operated in 'voltage mode', where the output is a voltage rather than a current. This is a bit trickier to explain so I won't bother. This circuit provided the voltages used for comparisons in SAR ADCs in the days before MOSFETs and capacitors displaced bipolar transistors and resistors.

Segmented DACs

Many DACs are made using a combination of circuits rather than a single architecture for the whole converter. It is simple to combine the resistive thermometer DAC in figure 8.5(d) with the R–$2R$ chain in figure 8.6(e), for instance. The thermometer DAC is used for the more significant bits and the R–$2R$ chain for the less significant bits. This is called a *segmented*

DAC. A commercial example is the 18-bit Texas Instruments DAC9881, whose architecture is described as 'an R–$2R$ ladder configuration with the four MSBs segmented'. It is linear to within ± 2 LSB, which is seriously impressive.

Another type of segmented DAC uses several sets of current sources, each as in figure 8.5(b). A larger current is used for the most significant bits and a smaller current for less significant bits. There are plenty of variations on these themes!

Sigma–delta DACs

These are very similar to sigma–delta ADCs and are based on the same sort of modulator. The converter takes in digital data with N bits at the final sampling frequency f_s. It produces a stream of single bits at the faster, oversampling (modulator) frequency f_m, whose average value matches that of the input. The sigma–delta modulator pushes the fluctuations to high frequency so that the analogue output can easily be filtered to keep only the components below the Nyquist frequency. The jargon is that the modulator 'shapes the noise' to enable easy filtering.

Sigma–delta DACs dominate particular applications, notable audio, and are spreading into other fields that require high resolution but relatively slow analogue outputs.

Multiplying DACs

The output of all DACs depends on a reference input but usually this must lie within a fairly narrow range. *Multiplying DACs* accept a wide range of reference inputs and the output is the product of the analogue reference and the digital input. The circuits based on resistors can be used as multiplying DACs provided that the switches and amplifiers can handle the range of voltages. This sounds a daft requirement for the switches but remember that they are made from MOSFETs, not metal contacts!

Amplifiers in DACs

Several of the circuits described above include an amplifier. This is a problem if the output should go from rail to rail because no real amplifier can do this if it shares the same power supply as the DAC itself (section 6.2). The system may therefore be linear over most of its range but with much poorer performance at the two extremes.

8.5 Summary of DACs

- Pulse-width modulation (PWM) is a widely used substitute for a 'real' DAC. Most microcontrollers contain hardware to drive PWM automatically.
- A wide range of DACs is available, many of which produce a current rather than a voltage.
- Multiplying DACs accept a wide range of reference inputs and effectively multiply the analogue reference by the digital input.
- Digital potentiometers work in the same way as string DACs and may be useful for trimming circuits.

I have not mentioned one important point, which is smoothing the output. The output of an DAC changes in multiples of LSB but a smooth signal is generally desirable. A particular type of lowpass filter is required, called a *reconstruction filter*, and needs more sophisticated design than I can describe here.

8.6 Examples

Example 8.1 Calculate the output voltage from an 8-bit DAC converter with a full-scale output of 5.00 V when the digital input in decimal is (i) 10 (ii) 150 (iii) 200. What is the maximum output voltage? [0.20 V, 2.93 V, 3.91 V]

Example 8.2 Calculate the output voltage from a 12-bit DAC with a full scale output of 10.00 V when the digital input in hexadecimal is (i) 0xBAD (ii) 0xACE (iii) 0x0FF. [7.30 V]

Example 8.3 An output needs to be driven by a microcontroller whose PWM modules are based on a 16-bit timer. The output should provide an average voltage from 0 to 5.00 V that can be adjusted in steps of 10 mV. It does not need a particularly steady voltage and PWM is therefore acceptable. Suggest a suitable format for the PWM waveform.

Example 8.4 How many resistors would be needed for an 12-bit simple string DAC? Would they all have the same value? How would this change if the DAC were segmented into a 6-bit string and a 6-bit interpolator?

Example 8.5 A 8-bit thermometer DAC has resistors of 100 kΩ and a voltage reference of 2.5 V. The transresistance amplifier on the output has a feedback resistor of 1 kΩ. What compliance is needed for the reference voltage (in other words, how much current need it provide)? What output voltages does the DAC produce?

Example 8.6 A segmented 8-bit DAC uses individual current sources based on resistors for the most significant 4 bits and a R–$2R$ ladder for the least significant 4 bits. How many resistors are needed if all are of value R?

Part II

Power supplies and passive components

9

Power supplies

9.1 Introduction

All electronic and electrical equipment needs a source of energy: a power supply. Examples include:

- **battery**: chemical → electrical energy, reversible if the battery can be recharged
- **converter**: electrical → electrical, with different possible sources
  - AC input → DC output, e.g. 230 V, 50 Hz → 5 V DC, as used for almost all electronic systems and chargers
  - DC input → DC output, e.g. 12 V DC → 5 V DC, often used for individual ICs within a system, where they are called *point of load* or POL supplies
  - DC input → AC output (inverter), e.g. 12 V DC → 230 V AC, 50 Hz, as might be used to power mains appliances in a caravan.
  - AC input → AC output, used to change the frequency or voltage
- **energy harvesting** from the environment
  - solar cell: light → electrical energy
  - mechanical energy → electrical energy, such as . . . ?

When analysing a circuit we assume a perfect (ideal) supply that has a constant voltage independent of the load current (or input voltage). This implies that the supply has *zero* source impedance, and can deliver any magnitude of current – up to infinity! In practice all sources have some internal resistance which limits the current. This internal or source resistance (in general an impedance) may affect the operation of the circuit unless it is included in the theoretical analysis. See figure 9.1 on the next page. Remember Thévenin's theorem? The resistance is almost never a real component but represents the resistance of the windings of a transformer, plates of a capacitor and so on. Most power supplies have electronic circuits to regulate their output, in which case R_s is a characteristic of the whole circuit rather than a particular component.

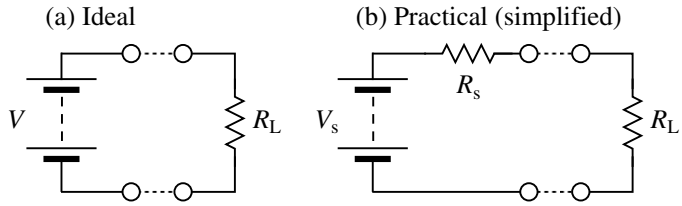


Figure 9.1 An ideal power supply and simplified Thévenin equivalent circuit of a practical power supply with internal resistance R_s, connected to a load R_L.

The difference between the EMF of a battery and the potential at its terminals is called the *lost volts* in Higher Physics. It is the voltage dropped across R_s. In many applications the value of R_s must be 'small' at all frequencies of interest or changes in load current will affect the voltage. This variation of terminal voltage with current is called *regulation*.

Regulation can be defined in different ways, shown in figure 9.2. One is in terms of the voltage at no load and full load. Reminder:

- **no load** means zero current, nothing connected, open circuit, $R_L = \infty$
- **full load** means the maximum rated current, lowest permitted value of load resistance

Then

$$\text{regulation} = \frac{V_{\text{no load}} - V_{\text{full load}}}{V_{\text{no load}}} \times 100\%. \tag{9.1}$$

(Sometimes the denominator is $V_{\text{full load}}$ instead; it makes little difference if regulation is good.) We would like as small a value as possible. The regulation is a pure number by this definition, usually expressed as a percentage.

Alternatively, regulation can be defined in terms of small changes about the rated output,

$$\text{regulation} = -\left.\frac{\delta V}{\delta I}\right|_{\text{rated output}}. \tag{9.2}$$

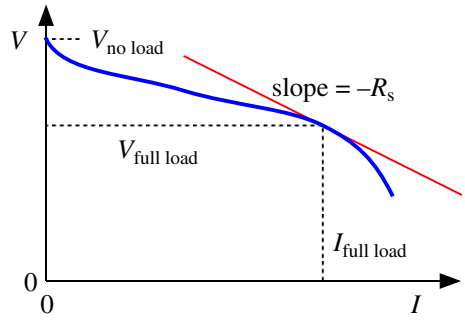


Figure 9.2 Output voltage of a real power supply as a function of current, showing the two ways in which regulation is often defined.

This is just the Thévenin resistance R_s evaluated at the rated output. As the plot shows, the apparent value of R_s varies with current. It is therefore not a constant in a real circuit, although we often assume this to make calculations simpler. Again we would like the variation to be as small as possible, which requires R_s to be small. It is measured in ohms because it is a resistance.

The regulation of cheap plug-in dc supplies ('wall warts') can be spectacularly bad. I measured a 9 V supply once and found that its no-load voltage was 13 V. Regulation can be significant even in mains supplies if the consumer is at the end of a long supply line: lights dim when a heavy load such as a cooker is switched on or when the motor in a vacuum cleaner starts.

Power supplies are big business. Here is a quotation from reference [25]. 'Since power supplies are so widely used in electronic equipment, these devices now comprise a worldwide segment of the electronics market in excess of $5 billion annually.' That was in 2002 and the figure has grown since then. Texas Instruments (formerly National Semiconductor) has a fabrication plant in Greenock, which has recently been extended at great expense, and their main business is in power supplies.

Example 9.1 A dc power supply produces 9.0 V with no load. This drops to 8.1 V when a 50 Ω load is attached. Calculate the current that flows, the regulation and the internal resistance R_s (assumed constant). What would be the output voltage with a load of 10 Ω?

Example 9.2 A battery has an open circuit voltage of 9 V and an internal resistance of 2 Ω. Calculate the output voltage and percentage regulation when the load is (i) 100 Ω, (ii) 50 Ω, (iii) 10 Ω. [8.82 V, 2.0%]

Example 9.3 Prove that the regulation due to series resistance R_S is $R_S/(R_L + R_S)$ for a load of resistance R_L.

Example 9.4 A 12 V car battery has an internal resistance of 0.01 Ω. What is the terminal voltage when starting the car if the starter motor takes 300 A? What would happen if the internal resistance rose to 0.1 Ω on a really cold morning? [9 V]

Example 9.5 A transformer has an open circuit voltage of 6.00 V, which falls to 5.62 V when a load taking 50 mA is connected. What is its effective source resistance? [7.6 Ω]

10
Batteries

10.1 Introduction

These are the most common source of power in portable equipment and come in two main categories. The performance of all types of cell has advanced substantially during my lifetime but sadly the rate of development is much slower than advances in semiconductors.

Primary cells

These are used once only, then discarded. Here are the more common types but many exotic varieties are also available. Extracts from data sheets are attached.

- **Alkaline** (AAA, AA, C and so on), the most widely used. They produce about 1.5 V when fresh but this declines during their lifetime [34].
- **Lithium coin cells** (CR2032 for example), common for backup applications and in small portable products. They contain lithium and manganese dioxide, which produce about 3.0 V [35].
- **Lithium cylindrical cells** such as AA use different materials to produce 1.5 V so that they can be interchanged with alkaline cells. Their anode is lithium metal and the cathode is iron disulphide. They have about double the energy density per unit mass of alkaline cells and work much better at low temperature.
- **Button cells** are used in calculators, watches and the like. They are often based on silver oxide but cheaper versions are alkaline. Usually 1.5 V.

Secondary cells

These are rechargable and therefore more economical and environmentally friendly but of course a charger is required. They used to suffer from much lower capacity than primary cells but that less true nowadays [36].

- **Nickel metal hydride** (NiMH) cells are the most common type for general use, producing 1.2 V [37]. Nickel–cadmium (NiCd) cells remain widely available but should be avoided

because of the environmental impact of cadmium and because NiMH generally performs better.

- **Lithium-ion** cells (*not* just 'lithium' because they do not contain lithium metal) now dominate integrated systems such as computers and mobile phones. Their energy density is much higher and they produce around 3.6 V when fresh, depending on the chemistry [38]. This declines to about 3.0 V when discharged. The positive electrode is based on a transition metal, typically Co, Mn or Fe as an oxide or phosphate while the negative electrode is a form of carbon – graphite or even nanotubes. Most cells include a polymer to separate the anode and cathode and may be called Li-polymer or LiPo cells.

- **Lead–acid** batteries are an old technology but still used where resistance to abuse and a harsh environment is important, such as in cars. They have a low internal resistance, which means that they can provide a high current for starter motors. Each cell gives about 2 V but of course they are usually packaged into batteries for 6 or 12 V.

The performance of Li-ion cells is seductive but unfortunately they also have famous incendiary tendencies and therefore need expert care and attention. The Panasonic web site says:

> In order to ensure the use of properly designed safety circuits with lithium ion battery packs, Panasonic lithium ion cells are not sold as 'off the shelf' products and are not available as a standard product from distributors. Lithium ion cells, however can be assembled into packs by authorized pack assembly centers that have been approved for safety circuit assembly and lithium ion pack design.

The safety aspect was made abundantly clear in 2013 by the fires in Boeing 787 aeroplanes caused by faults in their Li-ion batteries.

Li-ion cells must never be connected to a system directly but always through a supervisory circuit. This includes a 'fuel gauge' to ensure that the cells are never charged or discharged excessively. The current must be limited and the cell should be shut down if its temperature rises outside specification. The supervisor must be carefully matched to the precise chemistry of the cell because a difference of 0.1 V between the ratings of the cell and charger can lead to an explosion.

10.2 Capacity of batteries

Capacity is usually quoted as the amount of *charge* that can be supplied by a battery before it is exhausted or 'flat' – the voltage is too low to be useful. This is just the product of current and time if the current is constant or an integral in general. The SI unit of charge is coulombs but is never used for the capacity of batteries. Instead it is measured in ampère–hours (A h) or mA h. For example, the capacity of a car battery might be quoted as 50 A h. This means that it can supply

- 50 A for 1 hour
- 10 A for 5 hours
- 500 A for 0.1 hour (in principle!)

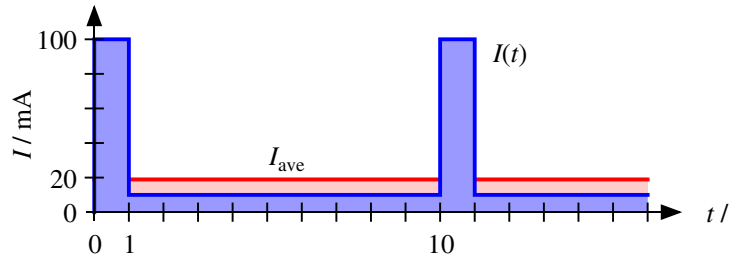


Figure 10.1 Variable current as a function of time, $I(t)$, and its average value I_{ave}.

In practice the capacity depends on the current drain, so that manufacturers specify the capacity at a particular current or with a particular load as in the data sheet [34]. The symbol C is often used for capacity, so that a current of $C/10$ means the capacity divided by 10 hours, which would be 5 A for this car battery. Capacity also depends on temperature, being lower at low temperatures. The open-circuit voltage drops with temperature too. (Put lithium-based batteries in your camera if you go somewhere cold – their performance is much better than alkalines at low temperature.)

The *energy* capacity of a battery is also important. This is the product of the power dissipated in the load and time, assuming that the power is constant. It is quoted in watt–hours (W h) and depends on the voltage produced by the battery as well as the charge. Li-ion cells give about 3.6 V compared with 1.2 V for NiMH, which boosts their energy capacity substantially.

The charge capacity of a battery is increased by putting cells in parallel: Two cells in parallel have twice the capacity of one, and so on. However, putting cells in series has *no* effect on the charge capacity: It increases the voltage instead. It doubles the energy capacity through doubling the voltage rather than the charge. It is not a good idea to put some types of cell in series or parallel because the load may not be shared equally among them, leading to nasty side-effects.

Typical (charge) capacities for modern AA cells are 2500 mA h (alkaline primary), 2000 mA h (NiMH secondary) but with a wide range. The energy capacity is about 2.5 W h. For comparison, the battery of my antique laptop has a capacity of 4400 mA h at 14.8 V, giving 65 W h. An interesting example is the battery in the Toyota Prius (data from wikipedia). It produces about 280 V with a capacity of 6.5 A h or 1.8 kW h.

Divide the capacity by the average current to find the lifetime of a battery. This is trivial if the current is constant but it usually varies according to the state of the system. Also, many systems switch on and off so their current is far from constant. Often there is a low current while the system sleeps with brief pulses of high current when activity is required. Recall that the average of a periodic function can be found from the integral

$$I_{\text{ave}} = \frac{1}{T}\int I(t)\,\mathrm{d}t\,, \tag{10.1}$$

where the integral is taken over any period T. For example, suppose that a system draws 100 mA for 1 s, then 10 mA for 9 s, and repeats this pattern, which is illustrated in figure 10.1. This sort of profile is common with high, short peaks of current separated by longer intervals with a much

lower current.

The period is 10 s and the integral can be done by calculating the area under the curve. This shows that the average current is

$$I_{\text{ave}} = \frac{100\,\text{mA} \times 1\,\text{s} + 10\,\text{mA} \times 9\,\text{s}}{1\,\text{s} + 9\,\text{s}} = \frac{(100 + 90)\,\text{mA s}}{10\,\text{s}} = \frac{190\,\text{mA s}}{10\,\text{s}} = 19\,\text{mA}. \quad (10.2)$$

Note that the average current is a *current* and is therefore measured in A (or mA etc); it doesn't involve time.

This system is powered by two AA cells in series, each of 2500 mA h capacity. The capacity of the battery is *not* increased by putting cells in series so the lifetime is

$$\text{lifetime} = \frac{\text{capacity}}{\text{average current}} = \frac{2500\,\text{mA h}}{19\,\text{mA}} = 132\,\text{h} \approx 5.5\ \text{days}. \quad (10.3)$$

This isn't very long: Suppose that the system had to run for at least a week without a change of battery. Is it more important to reduce the current at the peaks or between them? In this case the two regions of current make almost equal contributions to the average so we probably need to work on both. Don't assume that the larger current makes the bigger contribution to the average! This depends on its duration as well. Often one part of the cycle gives a much larger contribution to the average and should receive most attention, even if it is the smaller current.

10.3 How should you choose a battery?

These are some of the main points to consider.

- What voltage is needed?
- What capacity is needed?
- How portable is the product?
- Is it possible to use a rechargable system rather than primary cells?
- It is a good idea to choose a battery that is easy to replace:
  - alkaline batteries can be bought at any corner shop
  - rechargeable NiMH are also easy to buy, as are chargers
  - some lithium coin cells and button cells are fairly easy to find but exotic batteries are a nuisance, even if they have wonderful electrical properties!

Here are some useful web sites for data on batteries.

- www.duracell.com/oem/default.asp – mainly primary cells plus NiMH rechargables.
- data.energizer.com – wide range of primary cells and NiMH rechargables.
- www.panasonic.com/industrial/battery/oem – wide range of primary and secondary batteries.
- www.batteryuniversity.com – mainly rechargable batteries

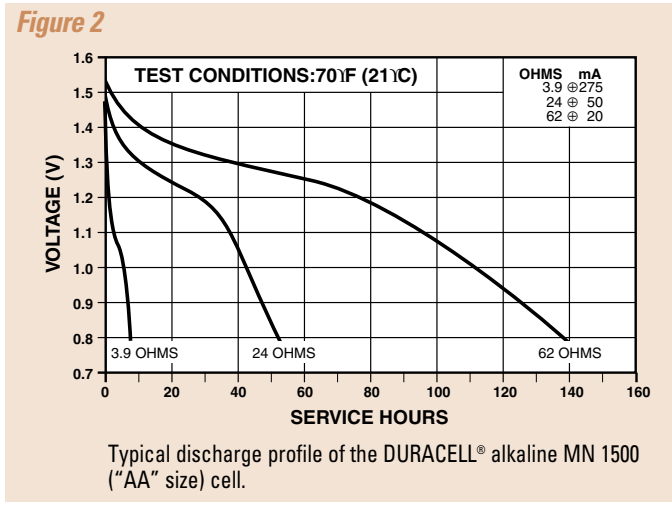


Figure 10.2 Discharge profiles of an alkaline AA cell, taken from the Duracell web site.

10.4 For what voltage should you design a battery-powered circuit?

This sounds like a no-brainer but the discharge curves in figure 10.2 show that it is not easy to choose the voltage at which a circuit should operate! A common choice is 2 × AA cells. This means that the supply is 3 V, isn't it? Well, no – not most of the time.

Alkaline cells produce over 1.5 V when they are new but this drops steadily through their life (which makes it easy to determine how much capacity is left). The end-point is often taken as 0.9 V, so 2 × AA cells produce only 1.8 V at the end of their life! (Battery manufacturers quote the capacity to 0.8 V to make the capacity seem a bit higher, which is even worse.)

Thus a well designed product that uses 2 × AA cells should be capable of operating from 1.8–3.1 V if primary cells are permitted – maybe even down to 1.6 V. Some products now come with instructions forbidding the use of alkaline cells, perhaps so that they can be designed for 2.4 V rather than 1.8 V.

Other cells have different characteristics. See the pages from data sheets for details.

- NiMH produce close to 1.2 V fairly constantly until the voltage plummets when they have been discharged. In this case 2 × AA cells produce 2.4 V throughout their useful life.

- Lithium coin cells, as used in the novelty lights in Electronic Engineering 1X, give a fairly constant 3.0 V for most of their life but have a high internal resistance.

- The behaviour of Li-ion secondary cells is between NiMH and alkaline cells. They can start as high as 4.0 V and the voltage declines steadily to about 3.5 V, after which it falls rapidly. They are typically rated down to 3.0 V.

(a) Direct connection of load to cell (b) Reservoir capacitor added to supply

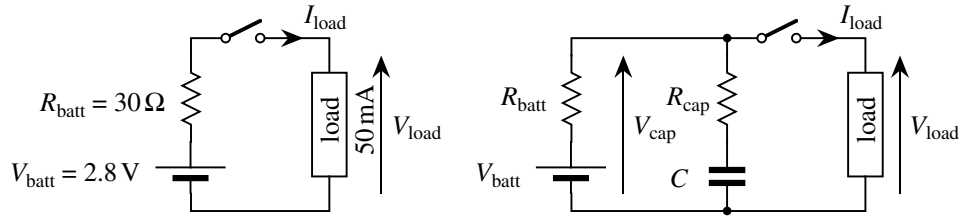


Figure 10.3 Equivalent circuits of (a) coin cell connected directly to pulsed load and (b) with reservoir capacitor.

10.5 Supercapacitors

These are a variety of capacitor that uses a special, double-layer construction to provide a high capacitance in a small volume. For example, the PC10 from Maxwell Technologies measures $30 \times 30 \times 5$ mm and has a capacitance of 10 F (that's farads, not microfarads). A snag is that the dielectric is very thin, which limits the working voltage to only 2.5 V in this case. Supercapacitors are widely used as 'buffers' between the main power source and a load to provide:

- backup power while the main source is unavailable – while changing a battery, for instance
- extra current when a heavy load is applied – the main supply need only provide the average current. We'll study an example shortly. Many electric vehicles use this approach to provide extra current for acceleration and store charge from active braking.

A problem is that supercapacitors are capacitors, so their voltage drops linearly as a function of charge ($V = Q/C$). Power electronics is usually required to give a constant output voltage.

10.6 Worked example: pulsed current drawn from a coin cell

A domestic electronic product draws a steady current of 3 µA at all times. It also draws pulses of 50 mA for 1 ms every 1 s. The voltage must remain above 2.4 V for correct operation. The designer wishes to use a CR2032 cell. Is this possible without extra components? If not, what else is needed? What is the expected lifetime?

A plot of current against time resembles that in figure 10.1 on page 84 but with different numerical values. The data sheet for the CR2032 [35] shows that it produces above 2.8 V with an internal resistance of below 30 Ω for most of its rated lifetime, giving a charge capacity of 200 mA h. These are somewhat arbitrary choices from the gradual curves. The analysis uses these values, rather than the starting values, because it should be based on the worst case and therefore the most restrictive values.

Ignore the steady current for now; it may affect the lifetime but won't otherwise be a problem. Coin cells have a high internal resistance and the current of 50 mA may therefore be a problem. Figure 10.3(a) shows the equivalent circuit of the cell connected directly to the load. Be-

cause of the internal resistance the voltage across the load is given by $V_{\text{load}} = V_{\text{batt}} - R_{\text{batt}} I_{\text{load}} = 2.8\,\text{V} - (30\,\Omega) \times (50\,\text{mA}) = 2.8 - 1.5 = 1.3\,\text{V}$. This is far below the limit of 2.4 V for correct operation so it is not possible to connect the system directly to the coin cell.

We need another component to store charge and release it to provide the 50 mA pulse: a capacitor connected in parallel with the coin cell as shown in figure 10.3(b). The resistor R_{cap} in series with the capacitor represents its internal resistance, which will be explained in section 15.2 on page 128; ignore it for now. Make the following assumptions to estimate the value C of capacitance required.

- The capacitor is fully charged to 2.8 V at the start of the pulse. We must check later that there is time for it to recharge from the cell between pulses.
- It provides the full 50 mA during the 1 ms pulse. This is a worst-case assumption because the cell provides part of the current directly.
- It is discharged to the minimum acceptable voltage of 2.4 V at the end of the pulse.

The capacitor must therefore provide a charge during the pulse of $\Delta Q = I\,\Delta T = (50\,\text{mA}) \times (1\,\text{ms}) = 50\,\mu\text{C}$. Its voltage must not fall by more than $2.8 - 2.4 = 0.4\,\text{V}$. The defining equation for a capacitor, $Q = CV$, also applies to changes, $\Delta Q = C\,\Delta V$, so $\Delta V = \Delta Q / C$. This means that we need

$$\Delta V = \frac{\Delta Q}{C} = \frac{50\,\mu\text{C}}{C} < 0.4\,\text{V} \tag{10.4}$$

so

$$C = \frac{\Delta Q}{\Delta V} > \frac{50\,\mu\text{C}}{0.4\,\text{V}} = 125\,\mu\text{F}. \tag{10.5}$$

This is a bit larger than the standard value of 100 µF so we would have to choose the next size up, 220 µF.

The next question is whether the capacitor can recharge between pulses. The load is disconnected, which leaves the classic RC charging circuit as in figure 3.11 on page 25. The time constant $\tau = (R_{\text{batt}} + R_{\text{cap}})C = (30\,\Omega) \times (220\,\mu\text{F}) = 7\,\text{ms}$. I have neglected R_{cap} in the calculation again. This is far shorter than the 1 s available between pulses so we conclude that there is plenty of time for recharging and that the circuit should operate correctly.

Finally, the lifetime. By current conservation the average current drawn from the cell is equal to that drawn by the load; current flows in and out of the capacitor but the average must be zero. It is much easier to calculate the average current drawn by the load. The pulsed part of the current has a period of 1 s = 1000 ms so its average value is

$$\langle I_{\text{pulse}} \rangle = \frac{(50\,\text{mA}) \times (1\,\text{ms}) + 0 \times 999\,\text{ms}}{1000\,\text{ms}} = 50\,\mu\text{A}. \tag{10.6}$$

Angle brackets such as $\langle x \rangle$ are often used to indicate an average or expected value. The steady current is 3 µA so the total is 53 µA and the lifetime is

$$T = \frac{Q}{\langle I \rangle} = \frac{200\,\text{mA h}}{53\,\mu\text{A}} = 3800\,\text{h} = 160\,\text{days}. \tag{10.7}$$

Clearly the pulsed part of the current is much more significant than the steady current.

This example shows how a capacitor can be used to work around the problem of the high internal resistance of a battery. This principle is widely used in other types of power supply as well.

10.7 Examples

Example 10.1 Roughly how much petrol (gasoline) is needed to drive the car the same distance as the energy in the battery of a Prius?

Example 10.2 A 12 V car battery has a capacity of 50 A h. For how long will it supply the following loads?

(a) 4 side lights (5 W each) plus a 4 W lamp for the number plate? [25 h]

(b) These lights plus two 60 W headlights? [4.1 h]

(c) All the lights above plus a 100 W heated rear window? [2.5 h]

Example 10.3 Many products are designed to run for at least a year (sometimes 10) on a single CR2032 lithium coin cell. What average current is permitted? What resistance of load could be used if the current was steady?

Example 10.4 A temperature sensor transmits data once per minute. It draws 40 mA while it runs at full power during each transmission, which lasts 10 ms. Between transmissions it goes into a low-power mode and draws only 0.1 mA. How long will it last using a 500 mA h battery? Which has the more significant effect on the lifetime, the low-power or the active mode?

Example 10.5 Computer memory takes a standby current of 10 µA and will retain the data if the voltage is held above 3.3 V. The standby voltage is provided by a 1 F capacitor, which is initially charged to 4.5 V. For how long will the data be held in the memory? [30 h]

Example 10.6 The resistance R_{cap} of the capacitor was neglected in section 10.6 on page 87. A small electrolytic capacitor may have a series resistance of a few ohms, say 3 Ω. Will this have a significant impact? [Potentially yes]

Example 10.7 Sketch the current through the load and that drawn from the coin cell as a function of time for the system in section 10.6 on page 87. Sketch also the voltage across the capacitor. Accurate values are not required, just the shapes of the curves and rough scales.

11

Rectifiers

The most common type of power supply unit (PSU) for electronic equipment produces a low DC voltage (typically 3–15 V) from the AC mains at 230 V, 50 Hz. (This applies in Europe. What is the mains supply in North America, for instance?) We would like to have a supply with zero source impedance, so that the output voltage does not depend on the load current; and to give a constant voltage when the input voltage (mains) varies.

The change from mains to DC requires several functions. Traditionally they were performed in the following sequence:

- **isolation** between the equipment and the mains for safety
- **reduction in voltage** with a transformer
- **rectification** to convert AC to DC
- **smoothing** to make the raw DC waveform from the rectifier into something closer to smooth DC
- **regulation** to refine the output by removing ripple and making the voltage less sensitive to the load

A problem is that the transformer must operate at 50 Hz, which makes it large, heavy and expensive. Nowadays the mains is usually rectified directly, smoothed and supplied to an electronic circuit to reduce its voltage, isolate and regulate it. A transformer is still needed but can be small because it operates at high frequency. Much of the PSU works at mains voltage so *keep your fingers out*. We'll look at the traditional approach first. It is now almost obsolete for commercial products but remains attractive for individual designs.

11.1 Transformers

A simple transformer (figure 11.1 on the next page) consists of two inductors wound on the same core so that the windings experience the same magnetic flux. The input is applied to the

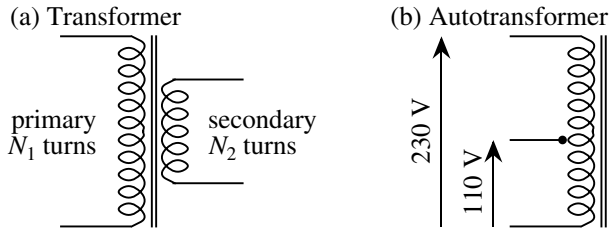


Figure 11.1 (a) Simple transformer with primary (input) and secondary (output) windings, and (b) autotransformer, where the windings are shared.

primary winding and the output is taken from the secondary winding. You will study transformers in Engineering Electromagnetics 2. The voltages across the two windings are (ideally) proportional to the number of turns:

$$\frac{V_{\text{secondary}}}{V_{\text{primary}}} = \frac{N_2}{N_1}. \tag{11.1}$$

The core is made of soft iron for a mains transformer so they are weighty items. A second important function of a transformer is to isolate the two circuits, which is essential for safety in equipment connected to the mains.

An *autotransformer* shares the windings as in figure 11.1(b). These are widely used to convert between 230 V and 110 V so that North American equipment can be used in Europe and vice versa. The shared windings reduce the weight, size and cost but *they do not provide isolation*. Thus the important safety feature of a full transformer is lost.

11.2 Rectifiers

Traditional rectifiers are based on diodes. Ideally these allow current to pass only in one direction, shown by the arrow in its symbol. A practical silicon diode drops about 0.7–0.8 V when conducting (forward biassed), as sketched in figure 11.2. Very little current passes in the opposite direction (reverse bias) unless the voltage exceeds the breakdown voltage, which must obviously be avoided in normal operation. *Schottky* diodes are made in a different way and have a lower forward voltage drop of about 0.3 V.

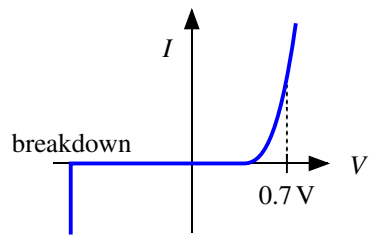


Figure 11.2 Sketch of $I(V)$ for a conventional silicon diode.

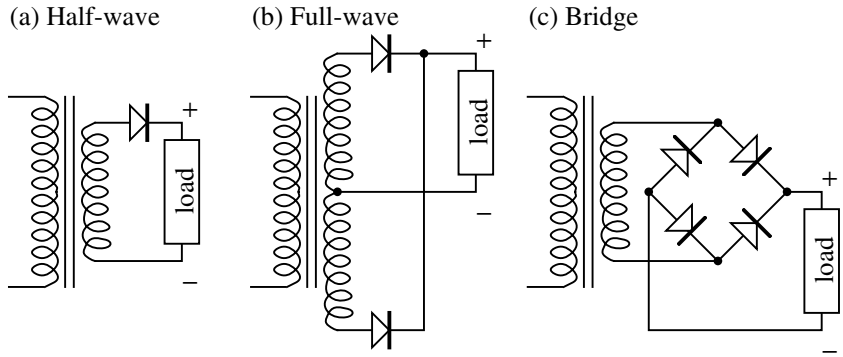


Figure 11.3 Standard rectifier circuits: (a) half-wave, (b) full-wave and (c) bridge.

Three standard circuits are used for rectifiers, shown in figure 11.3.

(a) The simplest, **half-wave** circuit uses a single diode, which conducts only for the positive half-cycles. This gives an output that is 'DC' in the sense that current flows only in one direction, but the voltage is nowhere near constant. It uses only half the waveform, which gives a poor quality supply that requires heavy smoothing.

(b) The **full-wave** rectifier uses both halves of the cycle but requires double the winding on the transformer.

(c) The 4-diode **bridge** rectifier is common, with the diodes often supplied as a single package. It is also a full-wave rectifier, meaning that it uses both halves of the wavefunction, and therefore makes best use of the transformer. It has the disadvantage that the voltage suffers from two diode drops. This is an increasing nuisance because of the low voltage used by modern electronic circuits.

The sketches in figure 11.4 on the facing page show the output voltage for half-wave and full-wave or bridge rectification. I've included a small voltage drop across the diodes because it makes the curves clearer. All components are supposed to be ideal apart from that (no voltage drop in transformer for instance). These unsmoothed waveforms are useless for almost all applications. In particular, the half-wave rectifier produces no voltage at all for half of each period T.

The voltage drop across a diode, or two diodes for a bridge, is unacceptable in many low-voltage systems. Instead they use *synchronous rectification*, where the diodes are replaced by MOSFETs that act as switches. These switches are closed for the appropriate parts of each cycle to achieve rectification. An IC is needed to control the system, which increases complexity, but the advantage is higher efficiency.

11.3 Smoothing

The rectified waveforms must be smoothed to bring them closer to ideal DC, whose voltage is strictly constant in time. The principle of smoothing is to store energy when the voltage

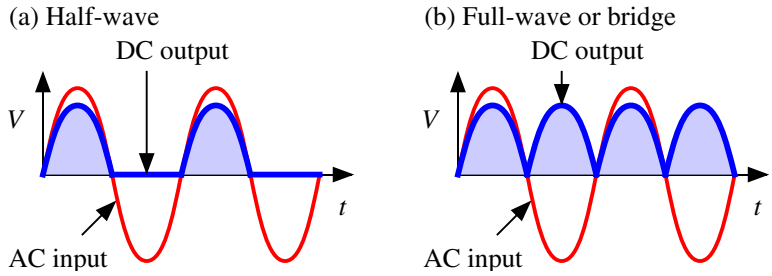


Figure 11.4 Unsmoothed output of (a) half-wave and (b) full-wave or bridge rectifier.

from the rectifier is high and to release it again when the voltage from the rectifier falls. Two basic components store energy: capacitors and inductors. Both are used, often together, but the simplest power supplies have only a *reservoir* or *smoothing* capacitor connected across their output. Figure 11.5 shows a half-wave rectifier with a smoothing capacitor.

The capacitor needs to have a large value, as we shall see shortly, and is usually an electrolytic type. These are *polarized*, which means that they must be connected the right way round. They explode if not! We'll look at types of capacitor in section 15.2 on page 128.

Figure 11.6 on the next page shows the effect of smoothing on the voltages and currents in the circuit. The bottom curve shows the current that flows into the positive plate of the capacitor as a function of time. It has two distinct regions.

- For much of the time, while the voltage from the rectifier is low, the capacitor supplies all of the current to the load. Thus the current $I_{\text{cap}} = -I_{\text{load}}$. It is negative because charge flows out of the capacitor into the load. The capacitor supplies the load for over twice as long with only a half-wave rectifier.

- Between these intervals are short, positive peaks of current when the capacitor recharges up to the peak voltage from the rectifier. These peaks are narrower and higher than the flat regions, so the magnitude of the current is much larger than I_{load}.

The charge that flows into the capacitor during the peaks must balance the charge that flows out to the load between the peaks so that the average charge on the capacitor remains constant over

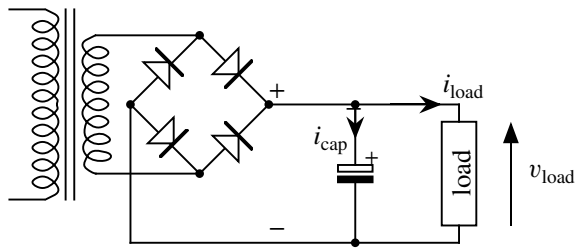


Figure 11.5 Bridge rectifier with a smoothing capactor and load.

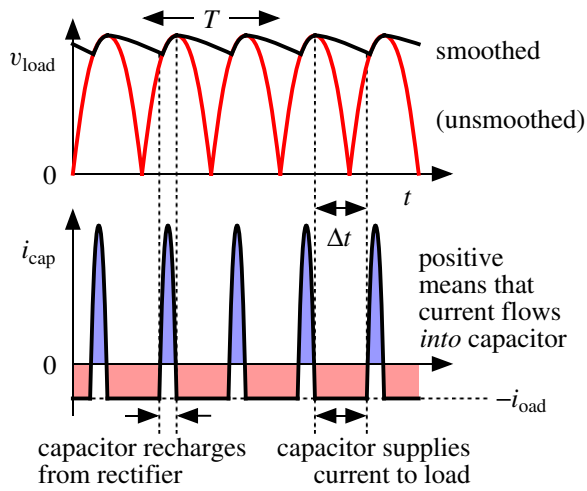


Figure 11.6 Output voltage from a full-wave or bridge rectifier with smoothing, showing current flow in and out of capacitor.

a complete cycle. Remember that charge and current are related in general by

$$Q = \int I(t)\,\mathrm{d}t. \tag{11.2}$$

This means that the area between the peak and the time axis, while the capacitor recharges, should be the same as the area between the axis and $-I_{\text{load}}$, while the capacitor supplied the load.

The output voltage falls while the capacitor is discharging and supplying current to the load. This leads to a 'sawtooth' variation in output voltage called *ripple*. It depends on the load current, the size of the capacitor and the time for which the current flows. The variation is called the ripple voltage V_{ripple} and is always quoted as a *peak-to-peak* value.

Remember the basic equation for a capacitor, $V = Q/C$. This also works for changes, $\Delta V = \Delta Q/C$. To calculate the ripple, ΔQ is the charge that the capacitor supplies to the load while the input voltage is low. If the voltage change ΔV is small then we can assume a constant current into the load. This gives

$$\Delta Q = \int I(t)\mathrm{d}t \approx I_{\text{load}}\Delta t, \tag{11.3}$$

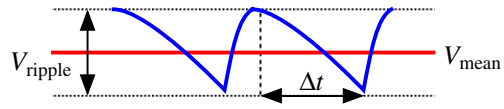


Figure 11.7 Ripple voltage with mean value V_{mean} and peak-to-peak value V_{ripple}.

where Δt is the time for which the capacitor supplies the load with current I_{load}. The ripple voltage is then given by

$$V_{\text{ripple}} = \Delta V = \frac{I_{\text{load}} \Delta t}{C}. \tag{11.4}$$

Unfortunately it is hard to estimate Δt by hand. Each period of the voltage from a full-wave rectifier is $\frac{1}{2}T = 1/(2f)$, half the period T of the input AC. Clearly Δt is less than this but it is difficult to determine the exact value. Simulation shows that $0.6/(2f)$ is about right. We can use this to estimate the size of capacitor needed.

For example, suppose that we want a 5 V supply to deliver 100 mA with a peak-to-peak ripple of 0.5 V. Then

$$C = \frac{I \Delta t}{\Delta V} \approx \frac{0.6I}{2f\Delta V} = \frac{0.6 \times 0.1}{2 \times 50 \times 0.5} = 1200\,\mu\text{F}. \tag{11.5}$$

This is a big capacitor!

11.4 Calculation of input voltage

The output voltage will obviously depend on the voltage at the input. To determine this, work back from the output, adding up all the voltage drops to get the *peak* value at the input.

1. The peak value at the load in the above example is $V_{\text{mean}} = 5$ V plus half of $V_{\text{ripple}} = 0.5$ V, giving a peak output voltage of 5.25 V.

2. To this must be added two diode drops, each about 0.8 V, giving a total of 6.8 V (to a precision of 0.1 V).

3. The drop across the source resistance of the transformer must finally be added. This is tricky because of the waveform of the current, shown in figure 11.6(c) on the facing page. Typically its peak value is roughly three times the load current, or about 300 mA. The internal resistance of a small transformer might be around 7 Ω, so the voltage dropped is about 2.1 V (again *very* roughly). This raises the total voltage to 8.9 V.

4. The output voltage of a transformer is always quoted as an rms value. The mains is ideally a sine wave so we can divide by the usual factor of $\sqrt{2}$ to get 6.3 V. You should buy a transformer for this voltage.

You can see that the source resistance of the transformer has a large effect on the performance. It is hard to predict the voltages dropped by the diodes and transformer because of the peaked nature of the current, so the result is unlikely to be accurate. Simulation may help.

Figure 11.8 on the next page shows oscilloscope traces from measurements on a power supply with a transformer, bridge rectifier and smoothing capacitor (formerly an experiment for this course). The peaked nature of the current is particularly clear in the figure 11.8(b), where the current and voltage can be compared on the input. The distorted shape of the mains voltage in the Rankine Building is also clear.

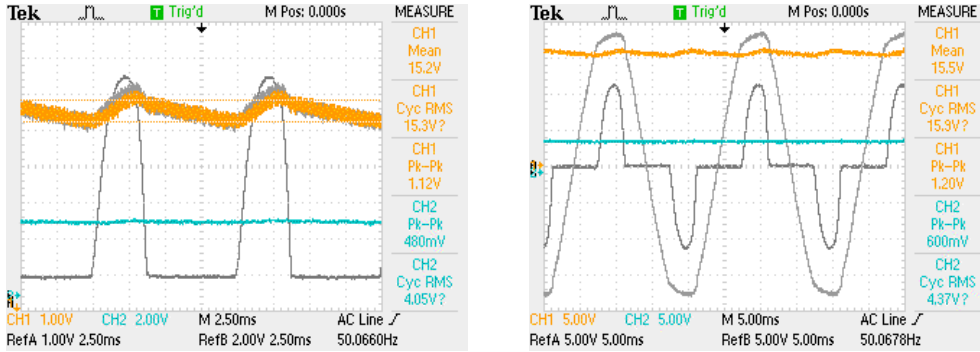


Figure 11.8 Oscilloscope traces from a power supply with transformer, bridge rectifier and smoothing capacitor. (a) Top curves (orange and light grey): output voltage showing ripple of about 0.6 V; bottom curves: current into load (blue, nearly constant) and into capacitor (dark grey, strongly peaked). (b) Top curves: AC input voltage (light grey, roughly a sine wave) and smoothed output voltage (orange, nearly constant but ripple visible); bottom curves: AC input current (dark grey, strongly peaked) and current into load (blue, nearly constant).

11.5 Conclusion

This simple supply described in this chapter, with a rectifier and smoothing capacitor, has poor performance. Its output has a lot of ripple, despite a large capacitor, and the output voltage depends on the load and also on the input voltage. It is adequate for undemanding applications, usually electrical rather than electronic, but is not suitable for powering integrated circuits or transistors. A regulator is needed as well.

Another problem is the shape of the current waveform on the input. This is far from sinusoidal: No current flows for much of the cycle, with strong peaks as the capacitor recharges around the peaks of the voltage. The mains supply in the Rankine Building suffers severely from this because of all the computers, as you can see from figure 11.8. Legislation has made this illegal for larger power supplies and sophisticated *power factor correction* is needed to control the shape of the current waveform.

11.6 Examples

Example 11.1 A transformer takes its input from the 230 V mains and is required to produce 5.5 V peak output as the input to a 5.0 V DC supply. The primary has 600 turns. How many turns should there be on the secondary winding? (Remember to convert both voltages to rms before taking the ratio.)

Example 11.2 Draw up a table to compare half-wave, full-wave and bridge rectifiers with the headings: windings needed on transformer; utilization of transformer; number of diodes needed; reverse voltage rating of diodes; voltage drop across diodes (assuming typical silicon components); frequency of ripple; ease of smoothing.

Example 11.3 How large a capacitor would be needed to reduce the ripple to 0.1 V for the example in section 11.3?

Example 11.4 A bridge rectifier is required to supply an average current of 200 mA with ripple not to exceed 0.5 V. What size of smoothing capacitor is needed? The AC supply is at 50 Hz. [around 2400 µF]

How would your answer change if the supply were designed for North America rather than Europe?

Example 11.5 Design a power supply with a transformer, bridge rectifier and smoothing capacitor to work from the AC mains in Europe and supply 9 V at 200 mA. Ripple on the output should not exceed 0.5 V. Assume that the secondary winding of the transformer has a resistance of 10 Ω. Calculate the value of smoothing capacitor and the (rms) output voltage of the transformer needed. What voltage rating should be specified for the capacitor, considering both no load and full load? Estimate the no-load voltage and regulation.

[Roughly 2400 µF, 12 V; 17 V; 40% or 30 Ω]

12

Linear regulators

Linear regulators take in current at a higher, variable voltage and supply it at a lower, regulated voltage. The difference in voltage is lost in a semiconductor device and the difference in power is dissipated as heat. Thus they are intrinsically inefficient. The advantages are that they are simple, cheap and electrically quiet. Another disadvantage is that they require a DC input at a relatively low voltage, which typically comes from a transformer, rectifier and smoothing capacitor. The transformer makes the overall supply large and heavy. You can see this in 'wall warts' or mobile phone chargers: older ones that use linear regulators are larger and heavy, while newer ones that use switching supplies instead are often smaller than British mains plugs.

Linear regulators are also widely used within equipment to deliver the voltages required by different components. Mobile phones might have around ten, for instance. A system might have its main power supply at 5.0 V DC, which is reduced to 3.3 V, 2.5 V, 1.8 V or less for individual ICs.

12.1 Zener diode regulator

A Zener diode is a special type of diode, which is designed to have a sharp reverse breakdown characteristic at a specified voltage V_Z, as shown in figure 12.1(a) on the facing page. This means that the voltage across the diode remains almost constant if it is reverse biassed. In reality it is not quite constant and the data sheet gives a recommended value of current for the best performance. Remember that Zener diodes are operated in *reverse* bias!

A simple circuit that uses this feature is shown in figure 12.1(b). This circuit is called a *shunt* regulator because the diode is connected *across* the load. It would be a good idea to connect a capacitor across the load as well. (We'll next look at regulators where the active component is in series with the load.) A resistor R_s must always be connected between the diode and supply.

- When there is no load connected ($I_L = 0$), $I_s = I_Z = (V_{in} - V_Z)/R_s$. This current flows through the resistor and Zener diode and all its energy is wasted as heat.
- Now connect a load with a high resistance. The voltage $V_{out} = V_Z$ so the current I_s drawn from the supply remains the same. However, some of this current now flows through the load as I_L and the current I_Z through the Zener diode is reduced. Thus current is diverted away from the Zener diode and into the load but the total current remains constant.

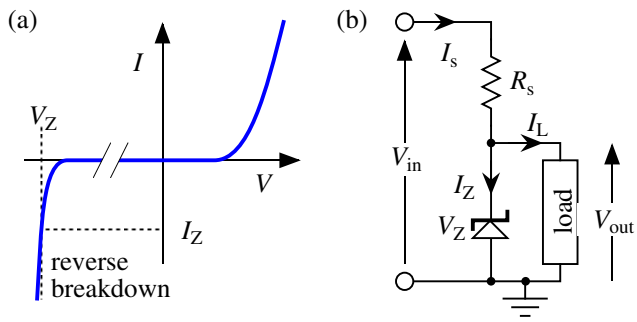


Figure 12.1 (a) The current–voltage characteristic of a Zener diode and (b) a simple shunt regulator.

- This continues as the resistance of the load is reduced until *all* the current Is flows through the load and there is none left for the diode, so $I_Z = 0$.
- The diode ceases to function once there is no current flowing through it, so the voltage V_{out} falls below V_Z if the resistance of the load is reduced any further.

This circuit therefore acts as a regulator for currents through the load up to I_s. There is another limitation as well. The power dissipated in the Zener diode, given by $I_Z V_Z$, must not exceed its rating. This sets a maximum value for I_Z and therefore a minimum value for I_L.

This simple regulator is adequte when the changes in I_L are small and V_{in} does not change too much. Real Zener diodes show a small change in V_Z with current I_Z, which can be reduced by using one regulator as the source voltage for a second regulator.

Figure 12.2 shows a rough analogy with a water tank that may make the operation clearer. The aim is to supply water with a constant pressure (head), equivalent to voltage. Water flows into a tank from an inlet. The tank has an overflow, which stops the level rising above a set value. The outlet is at the bottom of the tank. The system will work provided that the flow from the outlet is less than that from the inlet. If this is not true, the level of water will fall, the overflow will stop running and the pressure goes down.

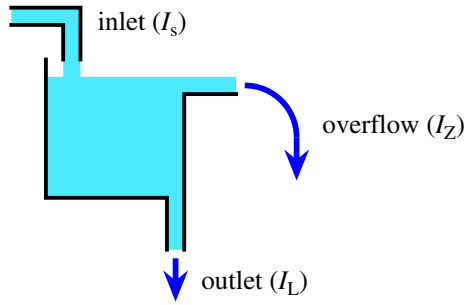


Figure 12.2 A water tank as an analogy to a shunt regulator.

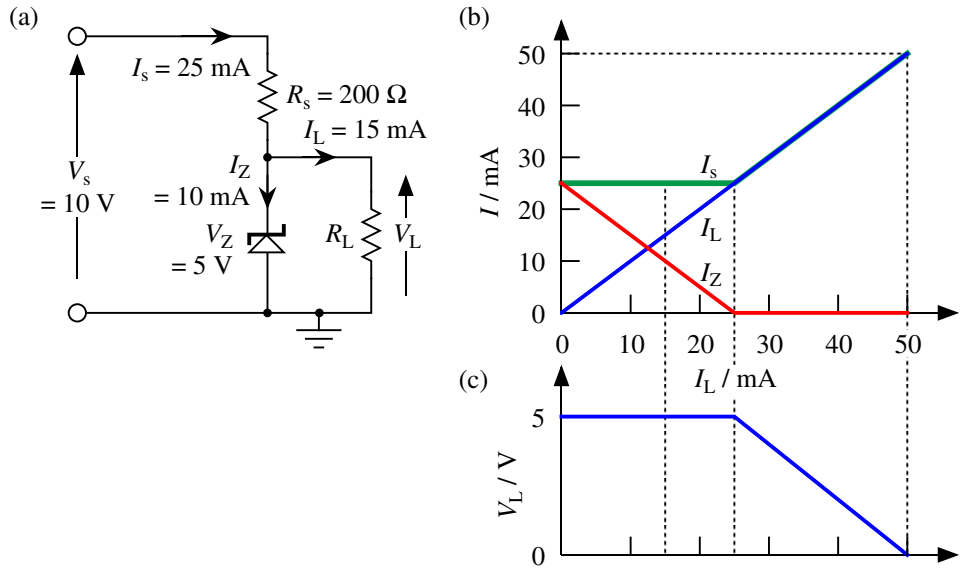


Figure 12.3 Analysis of a regulator with a Zener diode. (a) Circuit and conditions at specified operating point. (b) Currents as a function of current through the load. (c) Voltage across load as a function of current through it.

Zener diodes are now rarely used for most applications. They have been replaced by *bandgap references*, whose performance is better in almost all respects. A straightforward bandgap reference gives a voltage of about 1.2 V. This is related to an electronic property of silicon called its bandgap, which you will study in Electronic Devices 2. Integrated references are available for a range of common voltages. The accuracy depends on how much you wish to pay, as does the stability against temperature, output resistance and so on. This is discussed in section 7.1 on page 53 in connection with data conversion.

Worked example

A 5 V Zener diode is specified for a current of 10 mA and is required to regulate a load of 15 mA. It is fed from a 10 V supply. Complete the design of the circuit and analyse its behaviour as a function of the current through the load.

Consider the operating point first. The total current through the load and diode is 25 mA and the voltage across the diode is 5 V. The supply is at 10 V, leaving 5 V to be dropped across the series resistor R_s. Thus $R_s = (5\,\text{V})/(25\,\text{mA}) = 200\,\Omega$. The circuit is shown in figure 12.3.

Now vary the current through the load (by changing its resistance). First, suppose that the load becomes an open circuit, $R_L = \infty$, so that $I_L = 0$. Now all the 25 mA flows through the Zener diode. The power dissipated in the diode is $5\,\text{V} \times 25\,\text{mA} = 125\,\text{mW}$. It should be rated to dissipate at least this power, which is the maximum possible in this circuit.

Increase the current I_L through the load. The current I_s from the supply remains the same because it is given by $(V_s - V_Z)/R_s = 25\,\text{mA}$. KCL gives $I_s = I_Z + I_L$ so the current through

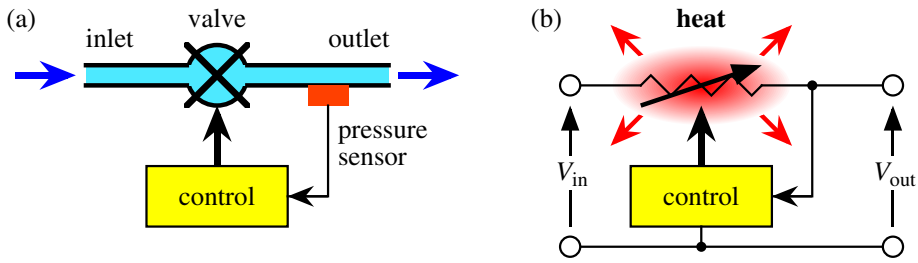


Figure 12.4 A series regulator in a (a) plumbing and (b) electrical system.

the diode falls as that through the load rises. The Zener diode still acts as a regulator provided that some current flows through it so $V_L = V_Z = 5\,V$.

This continues until all the current flows through the load, so $I_L = I_s = 25\,mA$ and $I_Z = 0$. At this special point the Zener diode has just stopped conducting so it no longer regulates but the voltage across the load remains 5 V.

Increase the current through the load further. The current drawn from the supply now exceeds the design value of 25 mA so the voltage dropped across R_s increases, which means that the voltage across the load falls. In fact R_s and R_L form a simple potential divider. No current flows through the Zener diode; it all flows through the load, so $I_L = I_s$. The voltage across the load is given by

$$V_L = V_s - R_s I_s = V_s - R_s I_L \,. \tag{12.1}$$

Thus V_L falls as I_L increases. This continues until the load draws its maximum possible current, which means that the load has become a short circuit. In this limit, $V_L = 0$ so all of V_s appears across R_s and $I_L = I_s = V_s/R_s = (10\,V)/(200\,\Omega) = 50\,mA$.

All these results are plotted in figure 12.3. The maximum current that can be drawn by the load before regulation fails is equal to the total current from the supply while the load is regulated.

12.2 Series transistor regulator

This is the most commonly used type of linear regulator. I'll start with a plumbing analogy again, shown in figure 12.4(a). This time the flow is throttled by a valve, which is controlled so that the outlet remains at a constant pressure as measured by a sensor. This is still wasteful because the energy is dissipated in the throttle valve but it is a great deal better than the tank with the overflow because the flows in the inlet and outlet are equal (unless some water is needed to power a hydraulic control system).

The electrical analogy is to use a controllable resistor instead of the valve as I have shown in figure 12.4(b). The 'controllable resistor' is really a transistor, either a bipolar junction transistor (BJT) or a metal–oxide–silicon field-effect transistor (MOSFET). It gets hot and often needs a heatsink, which we'll investigate in section 15.5 on page 133.

A more complete circuit for a traditional (high-dropout) regulator is shown in figure 12.5 on the next page. Power is supplied at V_{in} and the regulated output is delivered at V_{out}. The

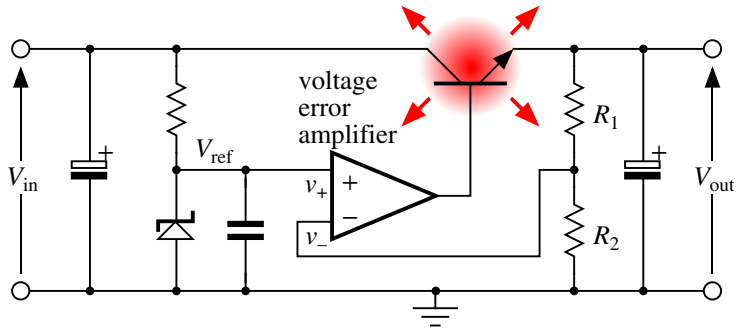


Figure 12.5 Simplified circuit of a series regulator with a bipolar junction pass transistor and error amplifier. The power supply to the amplifier is omitted.

usage '3 V regulator' means that the *output* is at 3 V. Linear regulators need $V_{\text{in}} > V_{\text{out}}$ and the difference must exceed a parameter called the dropout voltage, which I'll define later. The control is performed by a *voltage error amplifier*, which is really just a type of operational amplifier.

- Its non-inverting input is connected to a reference voltage. I've shown the previous circuit with a Zener diode although a superior reference would be used nowadays.
- The inverting input is connected to the output voltage, reduced by a potential divider.
- The output drives the base of a *pass transistor*, which controls the flow of current from the input to the output. I've shown an npn bipolar transistor but it would be an n-channel MOSFET nowadays.

It looks as though it is difficult to analyse this circuit but negative feedback makes it easy. Remember the basic rule of operation for an operational amplifier with negative feedback:

- The amplifier does whatever it can to bring its inverting and noninverting inputs to the same potential.

Here this means that it tries to get

$$V_{\text{ref}} = \frac{R_2}{R_1 + R_2} V_{\text{out}} \qquad \text{or} \qquad V_{\text{out}} = \frac{R_1 + R_2}{R_2} V_{\text{ref}}. \tag{12.2}$$

The voltage divider is required because otherwise we would need $V_{\text{ref}} = V_{\text{out}}$, which could be awkward. It also gives a way to vary the output voltage, should this be needed.

In words, the circuit operates as follows. Suppose that the output voltage drops. This causes v_- to drop at the opamp, so $v_+ > v_-$ and the output of the opamp becomes more positive. This drives more current into the base of the pass transistor, which increases the output voltage. The error is therefore corrected by negative feedback.

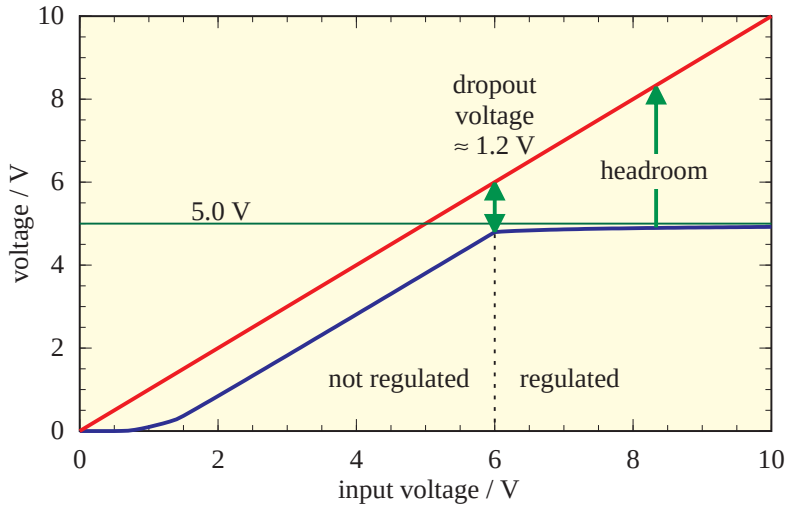


Figure 12.6 Simulated output voltage as a function of input voltage for the 5 V regulator in figure 12.5 on the preceding page.

Headroom and dropout voltage

The loss in voltage in the regulator, $V_{\text{in}} - V_{\text{out}}$, is called the input–output voltage differential or *headroom* voltage. The regulator will operate correctly only if the headroom is greater than a minimum value called the *dropout voltage*. This is one of the most important parameters on the data sheet.

Figure 12.6 shows a simulation of the output and input voltages for the simple regulator in figure 12.5 on the preceding page configured for a 5 V output. It requires $V_{\text{in}} > 6.0$ V to regulate and has a dropout voltage of 1.2 V. You can also see that it is not a perfect regulator because the output voltage rises noticeably as the input is increased above 6 V. It gives an output of only 4.8 V when it starts to regulate. A commercial IC has far better performance than this.

The dropout voltage of a real regulator is rather higher than in this example because it has a more elaborate circuit to drive the pass transistor. A traditional regulator may therefore have a dropout voltage of 2–3 V.

Power dissipation

The voltage reference, op-amp and output divider use a small amount of power but most is dissipated in the pass transistor. It is given by the product of the current and the headroom voltage,

$$P_{\text{diss}} = I_{\text{out}}(V_{\text{in}} - V_{\text{out}}). \tag{12.3}$$

Input smoothing capacitor, input voltage and efficiency

The input to the regulator in a mains supply comes from a transformer, rectifier and smoothing capacitor. We saw that a large capacitor was needed to reduce ripple in a simple supply, without

a regulator. How does this change when a regulator is used?

The critical requirement is that the input voltage to the regulator is always kept high enough for correct operation. In other words, it must stay above the output voltage plus the dropout voltage. Let's look at an example.

A 5.0 V, 100 mA power supply uses a regulator with a dropout voltage of 2.0 V. The designer wishes to get away with a 100 µF smoothing capacitor on the output of a bridge rectifier, which is the input to the regulator. What voltage should the transformer produce, assuming again that it has an internal resistance of 7 Ω?

1. The input to the regulator must remain above 5.0 + 2.0 = 7.0 V for correct operation. This determines the bottom of the ripple waveform.

2. The ripple voltage is given by equation (11.4) on page 95. Here the current is 100 mA (ignoring the current used by the regulator itself), $\Delta t \approx 0.6/(2f)$ as usual and we are given $C = 100\,\mu\text{F}$. These give $V_{\text{ripple}} = 6.0\,\text{V}$. The peak of the ripple is therefore at 7.0 + 6.0 = 13.0 V.

3. Add two diode drops from the bridge rectifier as usual to get 14.6 V.

4. The voltage drop in the transformer is the same as in section 11.4 on page 95 because the current is the same. That was about 2.1 V so the final result is 16.7 V peak.

5. This is equivalent to 11.8 V rms, which would be rounded up to 12 V or more in practice. (It is rounded up because this gives more headroom; rounding down could give too low a voltage for the regulator to operate.)

This PSU therefore needs 12 V rms input to produce a 5 V DC output, which shows that most of the input power is wasted in the regulator. The input and output currents are roughly the same on average and $P = VI$ so the efficiency is very roughly $V_{\text{rms,out}}/V_{\text{rms,in}} = 5/12 \approx 40\%$, which is poor. You can see why cheap power supplies get hot.

12.3 Low dropout regulators (LDOs)

A typical dropout voltage is around 2 V for the conventional regulator that I described above. This loss is not a problem for a high output voltage, but what about a 3 V output for instance? The input voltage would have to be above 5 V and around half the energy would be dissipated in the regulator rather than supplied to the load. This is not acceptable, particularly in portable applications. *Low dropout regulators* (LDOs) are therefore available with dropout voltages of 0.3 V or less. The circuit differs from figure 12.5 on page 102 in one critical respect: The transistor is pnp rather than npn and the output is taken from the collector rather than the emitter.

You might reasonably ask: why would anybody ever use a high dropout regulator? The reason is that the circuit in figure 12.5 has two desirable characteristics.

- It gives a **low output resistance**, although this depends on the circuit as a whole, not just the transistor. Remember that an ideal supply has zero output resistance.

- It has good **stability**, meaning that it is unlikely to oscillate.

Why should oscillation occur? This will be explained in Control 3 but here is a rough picture. The controller uses an op-amp with negative feedback, which is equivalent to a phase change of 180° for a sine wave. Other components in the circuit also give phase changes, which vary with frequency. Disaster strikes if these other components give a further phase change of 180° at some frequency. The total phase change becomes 360°, the negative feedback is now positive feedback and the whole circuit oscillates violently.

Most low-dropout regulators need to be protected against oscillation by conecting a capacitor across their output. The value and even the type of capacitor are specified in the data sheet and this advice *must* be followed. I'll say more about this when we look at an example of a data sheet for the LM2931 in chapter 13.

Modern regulators, like most analogue circuits, are built from metal–oxide–silicon field-effect transistors (MOSFETs) rather than bipolar transistors. The pass transistor of an LDO becomes a p-MOSFET rather than a pnp bipolar transistor. I described MOSFETs briefly in Electronic Engineering 1Y because they have dominated digital electronics for much longer. They have two particular advantages for LDOs.

- The equivalent of the base of a bipolar transistor is the gate of a MOSFET. This is the electrode that controls the flow of current between the other two terminals, the source and drain. The critical feature is that the gate looks like a capacitor and draws no current in a steady state, which reduces the power consumption.

- The voltage between the collector and emitter of a bipolar transistor cannot fall below a value called the saturation voltage in normal operation. This will be explained in Analogue Electronics 2. Its magnitude is determined by the physics of silicon and it is hard to make it much lower than 0.2 V. On the other hand, the channel of a MOSFET is more like a resistor, whose value can be reduced by design. Thus it is possible to reduce the dropout voltage further. For example, the FDS9926A MOSFET that we may use in the project has an 'on' resistance of 40 mΩ and more modern devices have even lower resistance.

You will study MOSFETs in Electronic Devices 2, Power Electronics 2 and Electronic Circuit Design 3.

12.4 Packaged regulators

It is very unlikely that you will need to build a regulator from separate components. A vast range of packaged regulators is available with the following features:

- fixed or variable output voltage (remember that a '5 V regulator' means one with an *output* of 5 V, not input)

- positive or negative supply

- conventional or low dropout (LDO)

- wide choice of power ratings (and packages to suit)

- current limiting and thermal shutdown for safety

- shutdown inputs, 'power good' outputs and other optional features

The most straightforward devices give a fixed output voltage and come in packages with 3 pins: input, output and common (ground). In fact they look just like simple transistors and you may remember one like this on the novelty lights that you built in Electronic Engineering 1X. I'll go through an example in the next chapter.

12.5 Examples

Example 12.1 A zener diode has a breakdown voltage of 4.7 V at a current of 10 mA. Determine a suitable series resistor if it is to regulate a load current of 50 mA from a supply at 10 V. How much power is dissipated in the diode? What happens if the load is removed? [88 Ω, 47 mW]

Example 12.2 Calculate the diode current, output voltage and whether the regulation is still effective if the load current in the previous question changes to (i) 10 mA and (ii) 70 mA. [50 mA, 4.7 V, regulation working; 0 A, 3.84 V, regulation fails]

Example 12.3 How could the output voltage be made adjustable?

Example 12.4 Suppose that you want an adjustable power supply for an undergraduate electronics laboratory whose output is variable from 3 V for logic circuits to 20 V for analogue circuits, with a maximum current of 1 A. The input voltage is chosen to be 25 V to give plenty of headroom at all output voltages. Under what conditions is the maximum power dissipated in the pass transistor and what is its value? [22 W]

Example 12.5 The bench PSUs in Rankine 709 have adjustable, bipolar, regulated outputs from 0 to ± 15 V at 200 mA. Estimate the maximum power dissipated. [7.2 W]

Why are these linear rather than switched regulators?

Example 12.6 The designer is appalled by the figure of 40% for the example in section 12.2 on page 101 and tries to redesign the supply for 75% efficiency. Can this be done and, if so, how? If not, how about 50%?

13

How to read a data sheet: the LM2931 low-dropout regulator

A data 'sheet' is available for every electronic component. I have put sheet in quotation marks because some of them run to hundreds of pages. A good data sheet doesn't provide only the specifications but gives plenty of ideas for how to use the component. Often you find the solution to your problem in the applications information. Manufacturers want you to buy their components and it is in their interest to make them easy to use.

We are lucky to have good data sheets available from most manufacturers in electronics. In my experience it is much harder to get such helpful information for mechanical components. Always get the data sheet from the manufacturer's web site. Many other pages come up if you use Google to find a component but they are best avoided. In the era before the web, data sheets were published in books and we keep a selection of these in a small library near the electronics stores. Some books contain useful introductions but frankly most of the material is obsolete.

I'll use the data sheet for the Texas Instruments (formerly National Semiconductor) LM2931 low-dropout regulator as an example. I've chosen this because it is kept in our electronic component stores and has a good data sheet. It is also 'automotive qualified', which makes it useful for projects such as the Formula Student racing car. Against this, it is now an old-fashioned component. More modern components have far superior performance but generally come in tiny packages that we cannot assemble.

I'll go through the main headings on the data sheet, which vary a little by manufacturer.

13.1 General description

This is on the front page and gives a general description of the device, as you might expect. It will be a summary of its most relevant properties, inevitably presented in the best light.

- Output voltage (two fixed values or adjustable with on/off).
- Range of input voltage (up to 26 V, but probably limited by power dissipation).
- Dropout voltage (0.2–0.6 V, depending on conditions).

- Current rating (rather vague, probably due to the different versions and packages).
- Quiescent current I_Q. This means the current consumed by the regulator itself, which flows out of the ground pin rather than the output. It is described as 'low' but is high by modern standards at around 10% of I_{out}. This is a side-effect of the low dropout voltage. Effectively it is wasted current. (A more modern device with a MOSFET rather than a bipolar pass transistor would waste *much* less current.)
- Packages, which include a large power package, a conventional small transistor outline (TO92) and tiny solder-bump surface mount device (SMD).
- Protection against a wide range of faults. This device was designed for automotive applications, which present a notoriously hostile electrical environment. For example, the device can withstand transient voltages of $\pm 50\,\mathrm{V}$ on its input. You could describe it as student-proof but that might be going a bit too far.

This section is useful to give you an immediate idea of whether the component will meet your needs. However, you always need to go further to check the details.

13.2 Connection diagrams

You'll need these when you come to lay out the circuit if you do it by hand (although this information should be in the OrCAD libraries). Note that the TO-220 and TO-263 packages have metal tabs, which are connected internally to ground. This is important if you bolt the regulators to the chassis of your equipment. There will be no problem if the chassis is also at ground, otherwise – bang!

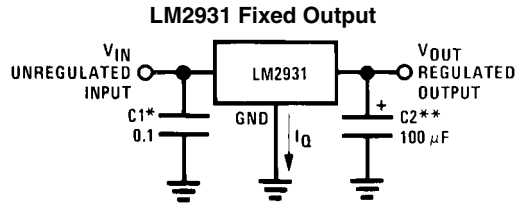
This device can be bought in a surface mount package. Many components can *only* be bought in such packages nowadays. Be very careful that you do not order surface mount packages by mistake! Some catalogues show 'SMD' symbols to highlight these, but not always. Students order a lot of surface mount devices in error, where larger packages are available and would have been much easier to use.

13.3 Ordering information

You don't normally need this because we can order only the varieties listed in the major catalogues (preferably RS or Farnell, but also Rapid, CPC, Maplin, DigiKey, ...).

13.4 Typical applications

Now we have reached the single most important section, where the manufacturer tells you how to use the product. I've extracted the most relevant part in figure 13.1 on the next page. It shows the circuit, which is simple in this case. *Follow these instructions.* It's fairly straightforward for a simple, 3-pin device like the fixed-voltage versions of the LM2931. However, you *must* do what it says. Here you are told that 'C2 must be at least 100 µF to maintain stability'. It means it! I can assure you that the regulator will not work if you omit this capacitor – plenty of students have tested this over the years and found that the data sheet is correct. You are further



*Required if regulator is located far from power supply filter.
**C2 must be at least 100 µF to maintain stability. May be increased without bound to maintain regulation during transients. Locate as close as possible to the regulator. This capacitor must be rated over the same operating temperature range as the regulator. The equivalent series resistance (ESR) of this capacitor is critical; see curve.

Figure 13.1 Extract from 'Typical Applications' on data sheet for National Semiconductor LM29312.

advised that the ESR of this capacitor (equivalent series resistance, explained in section 15.2 on page 128) is critical and are referred to a curve later in the data sheet.

13.5 Absolute maximum ratings

This means what it says: exceed these ratings and you will destroy or damage the device. An interesting feature is the limit on internal power dissipation, because there isn't one. It is internally limited, which means that the device shuts itself down if it detects that it is overheating. The maximum junction temperature is needed to calculate the size of a heat sink.

13.6 Electrical characteristics

Three sets of characteristics are tabulated in this data sheet because the LM2931 comes in versions for fixed 3.3 V and 5.0 V outputs and an adjustable version. I'll focus on the 3.3 V device.

You will almost certainly need to study this section carefully to find out whether a device is suitable for your particular application. This data sheet is a little unusual because there are usually three columns of numbers: minimum, typical and maximum (although one entry is often missing for each row). Here there is only a single 'Limit' column instead of minimum and maximum, but it sometimes has two numbers! The Limit heading refers us to Note 3, which explains the conditions under which the limits hold. Let's look at a few parameters in detail. The thermal resistance is curiously buried in note 4, possibly because it depends on the package rather than the output voltage.

Output voltage

Look at the first row, for output voltage. The typical value is 3.3 V, which you would probably have guessed. The limits are 3.135 and 3.465 V and are in normal type, so they hold at 25°C

junction temperature.

Given this range of parameters, which should you use when you design a circuit?

- If you are designing a single piece of equipment and can afford to buy a few extra components, use the 'typical' value and test that the component performs close enough to this.
- This won't work if you are designing a product for mass production – thousands or more. Sooner or later you will inevitably get a component whose performance is close to one of the limits. In this case you must design for the whole range of possible parameters.

This means that you should generally design for the worst case. Will your circuit work correctly over the full range of 3.135–3.465 V? It's no good if you are using a microcontroller whose specification for V_{DD} is 3.3 ± 0.1 V: you will have to find another regulator.

In fact it is even worse than this because the range 3.135–3.465 V applies only at 25°C. If you include the full spread of operating conditions the specification widens to 2.970–3.630 V.

Quiescent current

This time we are given a typical value of 0.4 mA and a single limit of 1.0 mA_{max} for a small load, $I_o \leq 10$ mA. Obviously the worst case corresponds to the maximum quiescent current because it is wasted. There would be no point in giving the opposite limit of the minimum current; ideally it would be zero, which no real device can ever deliver. You should design for the worst case, which is 1.0 mA.

There is a second row for $I_o = 100$ mA, which gives a typical value of 15 mA. There really ought to be a maximum as well (and there is for the 5.0 V version).

Dropout voltage

Data for two output currents are provided in this case. The range for 10 mA current seems large. As usual, you should design for the worst case of 0.2 V. Further insight over the range of dropout voltage is given in the plots so we'll look at them next.

13.7 Typical performance characteristics

This section contains a large number of plots that show the performance of a *typical* device. (In most cases it would not be practicable to show the limiting cases as well.) You might be surprised that so much data is provided for a simple power regulator!

The first two plots are for the dropout voltage and illuminate the values in the table. The very first plot shows that the dropout voltage rises with temperature, from 0.05 V to 0.10 V in the case of 10 mA load. This explains a large part of the range shown in the table. The second plot shows the dependence on current at a fixed temperature, presumably 25°C. The numbers don't seem quite consistent with the first plot. . . .

The plot for 'Output at voltage extremes' show how the device shuts down if the input voltage becomes too large, which is a good safety feature. The plots for the quiescent current help to explain the ranges listed the the earlier table. The plots of power dissipation assist you to design a suitable heat sink.

Finally comes the promised plot of the 'Output capacitor ESR', which shows the region within which the device is stable. Interestingly there is a lower limit as well as an upper, so you shouldn't spend too much money on a capacitor with a very low ESR! I looked up the ESR of electrolytic capacitors in the data sheet that I provided for the microphone amplifier in Electronic Engineering 1Y. It quoted 2.7 Ω for a 100 µF, 16 VW general-purpose capacitor. This does not meet the specification for the LM2931. You should look for a capacitor that is specially designed for power supplies. (More modern LDOs have less demanding requirements and some will work without a capacitor.)

13.8 Schematic diagram

You don't often see this on modern data sheets, partly because of trade secrets but also because they are usually too complicated to be of much practical use to anybody but a professional circuit designer. Still, it's interesting to pick out the major components. Can you spot the pass transistor? Several transistors have two collectors, many with numbers next to the separate connections. These give the relative areas and the currents are normally in the same ratio.

13.9 Application hints

Here again the manufacturer helps you to use the device and avoid common problems. It starts again with the capacitor across the output, which is a critical requirement. There is some interesting material on use at low temperatures, which can cause aluminium electrolytic capacitors to freeze and lose their capacitance! Probably you won't be designing circuits for temperatures below −30°C in the near future but, who knows, you might take a job with the British Antarctic Survey.

13.10 Definition of terms

This is a fairly uncommon feature but useful for unfamiliar components. Some analogue-to-digital converters have excellent sections that define the technical terms used. Note the precise definition of the dropout voltage.

13.11 Physical dimensions

These isn't of much use in most cases because the footprints should already be in OrCAD. On the other hand, the drawings show the precise size of the package and the spacing of the pins, which isn't always obvious. It is particularly helpful for surface-mount packages, which come in a huge variety of sizes whose names are not always standard.

Example 13.1 This question refers to the data sheet for the LM2931. Assume that we are using a device with a 5 V fixed output and a TO-92 package.

(a) What capacitor is required on the output?

(b) What range of input voltages should be used?

(c) What happens if the component is connected back to front?

(d) What is the dropout voltage at 100 mA?

(e) What happens if the input voltage drops too low?

(f) What is the maximum output current?

(g) Suppose that the device supplies 100 mA at the maximum recommended input voltage. How much power is dissipated? Is this acceptable? What would happen?

(h) What is meant by the term 'ripple rejection'?

Example 13.2 An LM2931 is used to supply a system with 100 mA at 5 V. Use worst-case design to specify the input that must be available to the regulator to ensure that it can supply the load under all conditions permitted. By how much does the output voltage change if the load is removed?

14

Switching power supplies

Linear regulators have two serious disadvantages, their poor efficiency and the requirement for the input voltage to be higher than the output voltage. Switching regulators work in an entirely different way: they store the energy at the input voltage and release it at the output voltage. The output voltage can be smaller than the input voltage as in linear regulators but it can instead be larger or even of the opposite sign for a switching regulator. They are *very* versatile and can give much higher efficiency than linear regulators. Energy can be stored in either capacitors or inductors and both are used. Inductors are much more widely employed but capacitors have advantages for some applications.

14.1 Switched-capacitor, charge-pump or flying-capacitor converters

These produce outputs whose voltage is given by a multiple or simple fraction of the input voltage, including negative values. The simplest is an inverter: $V_{\text{out}} = -V_{\text{in}}$. Here is how the circuit in figure 14.1 works.

- The two left-hand switches are closed on one phase of the clock (figure 14.1(a)) and the capacitor C_1 charges to the input voltage. It's 'top' plate is positive.
- During the other phase (figure 14.1(b)) the top plate of the capacitor is connected to ground. The voltage across the capacitor does not change because it is determined by the

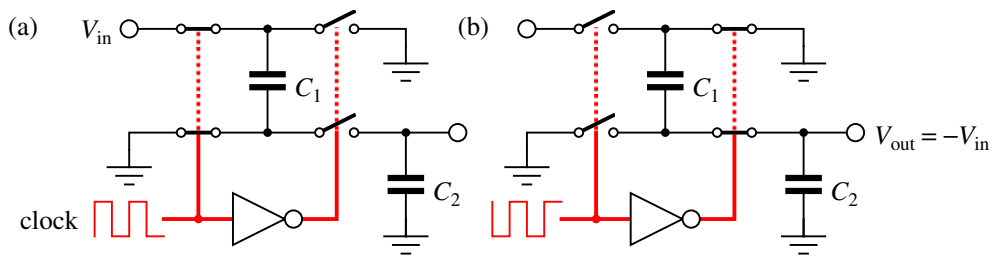


Figure 14.1 A switched-capacitor voltage inverter.

charge, which has not moved, so the bottom plate is forced to $-V_{\text{in}}$. This is connected to the output so $V_{\text{out}} = -V_{\text{in}}$.

The second capacitor C_2 is a reservoir or smoothing capacitor as in a simple rectifier. It supplies the load while C_1 is recharging from the input. The active capacitor C_1 is whimsically called a *flying capacitor* and the switches are really MOSFETs.

The circuit can easily be modified to produce $V_{\text{out}} = 2V_{\text{in}}$. By charging two capacitors in parallel but discharging them in series it is possible to produce $V_{\text{out}} = -2V_{\text{in}}$. More complicated circuits with networks of capacitors are needed for $V_{\text{out}} = \frac{3}{2}V_{\text{in}}$ and fractional ratios.

Switched-capacitor converters have the advantage that they are simple, needing only the regulator chip and external capacitors. They produce a little electromagnetic interference because of the switching but it is usually not serious. The disadvantages are ripple on the output and poor regulation in simple converters. The switching frequency is often around 1 MHz nowadays, which reduces both the ripple and size of capacitors needed (why?). They are restricted to simple ratios of input and output voltage or the efficiency suffers. The disadvantages are serious and switched-capacitor converters are used only in particular niches. One is to generate a negative bias so that a single-supply opamp can drive its output all the way down to zero (section 6.2 on page 41). Here are two others that you might encounter.

RS-232 interface drivers

Digital systems often need to communication with computers and a serial (COM) port is simple to use. COM ports have now vanished from standard desktop computers but remain widely used in commerce and industry. A problem is that the COM port uses an ancient protocol called RS-232, which employs a voltage between -15 and -3 V to represent a 1 and $+3$ to $+15$ V to represent a 0. The official voltage levels used to be ± 12 V but many systems now use lower voltages, often ± 5 V. This is closer to the voltages used for digital electronics but the negative voltage is a particular nuisance.

The solution is to use an interface driver that includes a switched-capacitor converter. The most well-known is the Maxim MAX232, now rather old, which needs three external capacitors to produce levels of $\pm 2V_{\text{supply}}$ from 5 V. The newer MAX3232 is designed for 3 V supplies and every major company offers equivalent devices.

Drivers for white and blue LEDs

Most modern digital components work with supplies of 3 V (or less). Red, yellow or green LEDs are easy to drive because they need about 1.8 V. White and blue LEDs are based on different semiconductors and need a higher voltage, around 3 to 4 V, so they cannot be driven directly. Special drivers are therefore needed to supply small blue and white LEDs from 3 V in products such as mobile phones. Dimmable lights, 'fun' lights (toys, children's shoes…) and high-power white LEDs for camera flashes in mobile phones are other applications.

Wherever there is a need, the manufacturers will provide a solution. A wide range of LED drivers is available, many of which use switched capacitors to step up the voltage. They may also regulate the current rather than the voltage, which is done because brightness $\propto$ current. (Many drivers use inductors rather than capacitors for their greater efficiency, particularly for higher current.)

14.2 Switched inductor converters

Most switching converters use inductors and the word 'inductor' is usually dropped – in fact they are often just called 'switchers'. Portable computers and even desktop computers of reasonable size would not be possible without switching converters. I'll describe the basic types, which convert DC from one voltage to another, and say a little about mains supplies at the end. Switchers come in three basic varieties:

- **buck** – steps down voltage, same sign
- **boost** – steps up voltage, same sign
- **buck/boost** or **inverting** – changes sign of voltage, can step up or down

Thus you can get any magnitude and either sign of output voltage (in principle)! The **flyback** converter combines an inverting converter with isolation between input and output. More complicated topologies are often used in practice.

An inductor stores energy in its magnetic field, which is proportional to current. The current must therefore rise and fall as the energy is stored and released. The details follow from the basic equation for an inductor,

$$v_L(t) = L\frac{\mathrm{d}i_L}{\mathrm{d}t}. \tag{14.1}$$

A voltage is generated by a *change* in current, not a steady value. You can imagine that the inductor tries to resist changes in current. Here are two important cases.

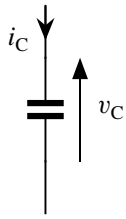
- If the current through an inductor is constant, there is zero voltage across it. (Zero voltage does *not* imply zero current!)
- If the voltage across an inductor is constant, the current rises or falls steadily depending on the sign of the voltage.

The traditional way of analysing AC circuits with inductance, using phasors and impedance, is useless for these applications because the waveforms are nothing like sine waves. The circuits must be analysed in time rather than frequency.

The general idea is that an inductor is switched so that energy is stored in its magnetic field from the input during one phase and released to the load during the other phase. It is rather like the smoothing capacitor in a linear supply, which stores and releases charge. That caused its voltage to go up and down during each cycle. In the same way, the current through the inductor ramps up and down as energy is stored and released. The confusing feature is that the voltage across the inductor changes sign between the two phases because of equation (14.1). The equivalent feature of the storage capacitor is that its current changes sign between storage and release, but that seems far more natural. Figure 14.2 on the next page shows a comparison of these two components.

14.3 Basic buck converter

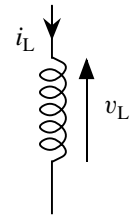
The *buck* converter is a step-down supply so $V_{\text{out}} < V_{\text{in}}$, like the linear regulator. The circuit in figure 14.3 on page 117 has two switches driven in antiphase and an inductor in series with the



$$v_C = \frac{q}{C} = \frac{1}{C}\int i_C(t)\,\mathrm{d}t$$

$$i_C = C\frac{\mathrm{d}v_C}{\mathrm{d}t}$$

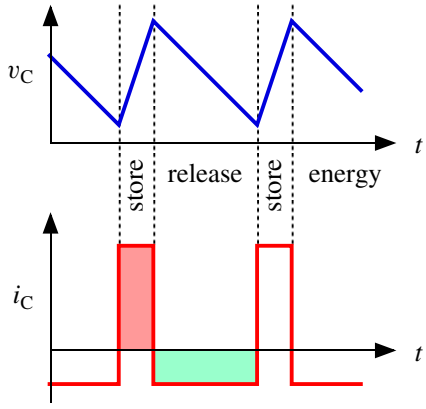
Energy stored in *electric* field,
generated by *voltage*
Cannot change *voltage* instantly



Periodic storage and release of energy in a power supply

Assume constant currents for capacitor and constant voltages for inductor

Reservoir capacitor with ripple voltage



Equal areas: charges balance so that
average voltage remains constant.

$$\langle i_C \rangle = 0$$

Switched inductor with ripple current

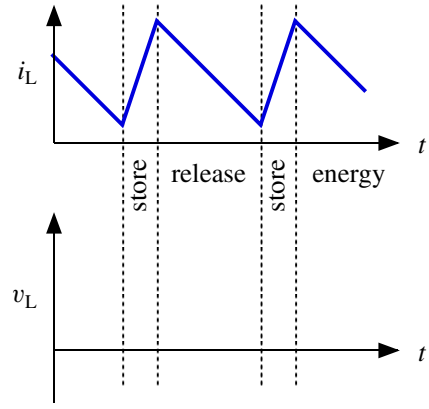


Figure 14.2 Comparison of stored energy in capacitors and inductors. Note carefully the directions of the currents and voltages.

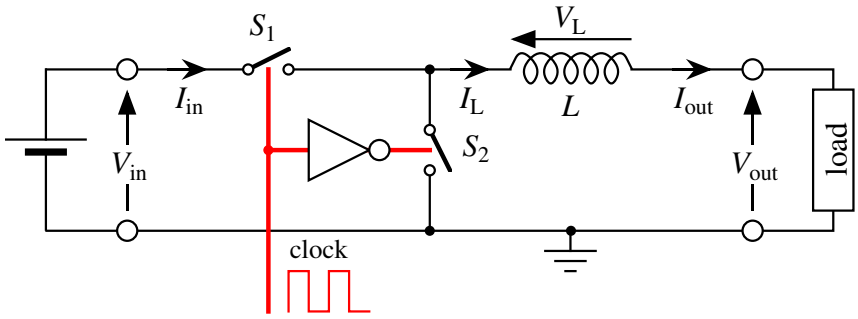


Figure 14.3 Circuit of a basic buck converter.

load. (One of the switches can be replaced by a diode but it's easier to analyse it like this.) Let's look at the two phases separately, where the inductor stores and releases energy, or charges and discharges. To make life easy I'll assume that the input and output voltages remain constant. This needs reservoir capacitors across the load and input, which are not shown. Let the duty cycle be D, which means the fraction of each cycle that is spent in the charging phase. It lies in the range $0 \leq D \leq 1$. If the overall period is T, the converter spends DT in each charging phase and $(1 - D)T$ in the discharging phase.

Charging or storage phase

In the storage phase, S_1 is closed and S_2 is open. The circuit can be simplified to figure 14.4 with the input supply, inductor and load in series. The currents are all the same, $i_{in} = i_L = i_{out}$, and the voltages obey $v_{in} = v_L + v_{out}$. Check the signs carefully! The rate of change of current is given by the equation for an inductor,

$$\frac{\mathrm{d}i_L}{\mathrm{d}t} = \frac{v_L}{L} = \frac{v_{in} - v_{out}}{L} > 0. \tag{14.2}$$

Thus the current *increases* steadily provided that $v_{in} > v_{out}$, which is what we want during the storage phase: Energy is drawn from the input and stored in the magnetic field of the inductor.

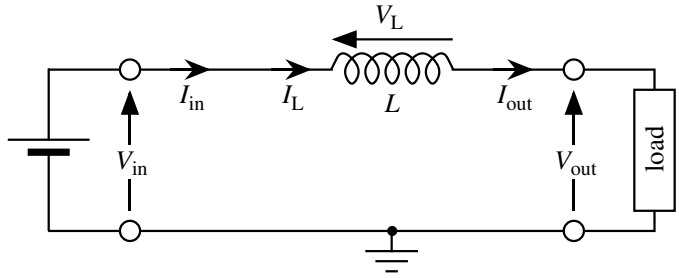


Figure 14.4 Effective circuit of a basic buck converter in the charging or storage phase.

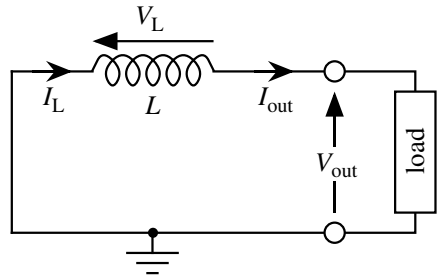


Figure 14.5 Effective circuit of a basic buck converter in the discharging or release phase. Note that $v_L < 0$.

The inductor is charged with energy.

Release phase

Figure 14.5 shows the circuit in the 'release' phase, with S_1 open and S_2 closed. Only the inductor and load are left in the circuit. The input is disconnected and $i_{in} = 0$. The sum of the EMFs around the circuit must be zero by Kirchoff's law so now $v_L = -v_{out}$. Thus v_L has changed sign and become negative. The rate of change of current is given by

$$\frac{\mathrm{d}i_L}{\mathrm{d}t} = \frac{v_L}{L} = \frac{-v_{out}}{L} < 0. \tag{14.3}$$

The current *decreases* steadily as energy is released from the inductor to the load. The inductor is being discharged.

Current and voltage waveforms

The waveforms are plotted in figure 14.6 on the next page. A central relation between them follows from the average value over each cycle of the voltage across the inductor:

$$\langle v_L \rangle = L \left\langle \frac{\mathrm{d}i_L}{\mathrm{d}t} \right\rangle. \tag{14.4}$$

This is just the average of the basic equation (14.1) for the voltage across an inductor. The average value of the current remains constant over a long time: it ramps up and down within each cycle but stays the same on average. Thus the right-hand side is also zero and it follows that $\langle v_L \rangle = 0$ as well. This means that the average voltage across the inductor is zero over each cycle of operation. (The equivalent relation for a smoothing capacitor is that $\langle i_C \rangle = 0$ so that $\langle v_C \rangle$ remains constant.)

We know the voltages across the inductor in the two phases so we can work out the average:

$$0 = \langle V_L \rangle = T_{store}(V_{in} - V_{out}) + T_{release}(-V_{out}) = [DT](V_{in} - V_{out}) + [(1-D)T](-V_{out}) = (DV_{in} - V_{out})T. \tag{14.5}$$

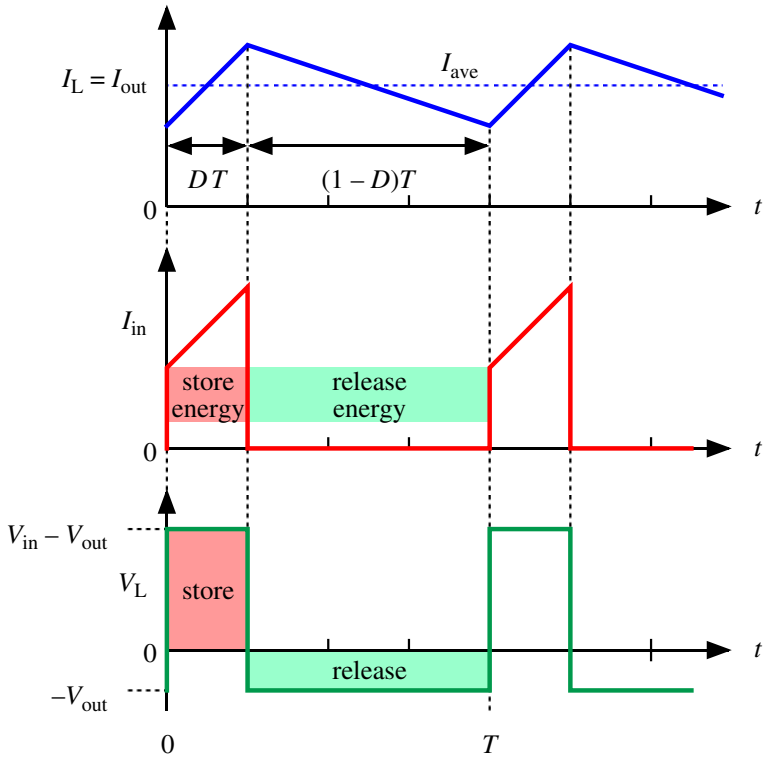


Figure 14.6 Currents and voltage across the inductor in a buck regulator with duty cycle $D = \frac{1}{4}$.

This gives the relation between the input and output voltages,

$$V_{out} = DV_{in}\,. \tag{14.6}$$

The output voltage is always less than the input voltage and can be controlled by varying D. This is an example of pulse width modulation (section 8.3 on page 67). Figure 14.6 shows $D = \frac{1}{4}$.

Switch S_2 is often replaced by a diode, as shown in figure 14.7. Confirm for yourself that this conducts at the correct times. A diode is not always used, although it is simpler than a switch; why might a switch be preferred?

Summary of buck converters

The main properties of buck converters are:

- output voltage < input voltage
- output current flows through inductor; it flows all the time so it is relatively easy to smooth
- input current is pulsed

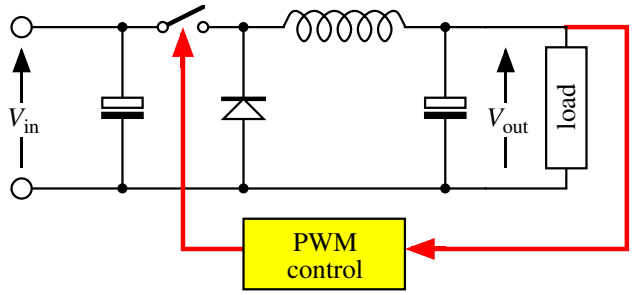


Figure 14.7 Block diagram of a complete buck regulator.

The inductors and capacitors can be made smaller if the frequency is increased because less energy needs to be stored and released in each cycle. Frequencies used to be in the kHz range but now may be in MHz.

A complete converter needs a feedback loop to control D and maintain a constant output voltage. A block diagram is shown in figure 14.7. You buy an integrated circuit, of course, and I've provided the data sheet for the LM3100 as an example [40]. Buck converters are widely used to supply low-voltage digital systems (5.0, 3.3, 2.5 and 1.8 V) from higher DC voltages, such as 12 V in a car or from a battery of several cells in series. They do not provide isolation between the input and output so they cannot be used to power equipment from the mains (which also needs a rectifier).

14.4 Boost converter

The two switches and the inductor form a 'Y' in the buck converter and this Y can be arranged in three ways. The other two ways give the other two types of converter. We'll next look at the *boost* converter, shown in figure 14.8. I have shown a diode rather than a second switch. Its output voltage is given by

$$V_{out} = \frac{V_{in}}{1 - D}. \tag{14.7}$$

The main characteristics are:

- Output voltage > input voltage, same sign, hence the name.

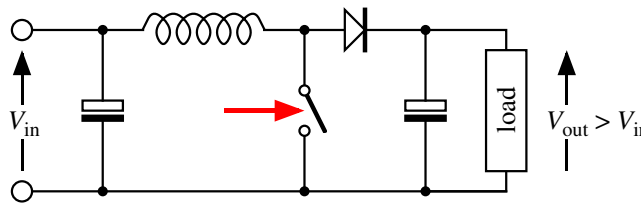


Figure 14.8 Circuit of a basic boost converter.

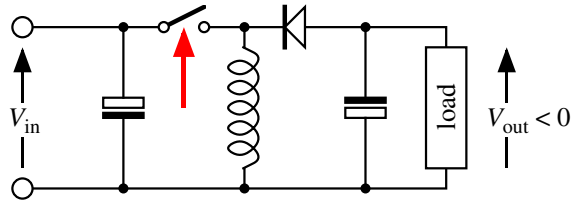


Figure 14.9 Circuit of a basic inverting or buck/boost converter. The output has the opposite polarity to the input; note the orientation of the capacitors and diode.

- Output current is pulsed and needs good smoothing.

Boost converters are useful in equipment that works off a single AA cell or the like. This includes a lot of portable electronics, such as MP3 players. They are also used to generate voltages for LEDs, often several LEDs in series. For example, the National Semiconductor LP5526 [41] provides all of these functions for backlights and a camera flash in a mobile phone.

Another example of a modern boost converter is the Texas Instruments TPS61200. This can work from a 0.3 V input, admittedly not with wonderful efficiency. This low input voltage allows it to operate from a single solar cell, which typically produces about 0.3–0.4 V; several cells in series are needed to get sufficient voltage to be useful without such a regulator.

14.5 Inverting (buck/boost) converter

The circuit is shown in figure 14.9. Note the orientation of the diode and electrolytic capacitor on the output. Its output voltage is given by

$$V_{out} = -\frac{D}{1-D} V_{in} \,. \tag{14.8}$$

The main characteristics are:

- Output voltage has *opposite sign* to input voltage and may be larger or smaller in magnitude.
- Both input and output currents are pulsed, so heavy smoothing is needed.

Effectively the inductor is ‘charged’ from the input then ‘discharged’ into the output. This action is analogous to the reservoir capacitor in a linear power supply. The current rises during charging so the voltage across the inductor is positive; the current falls during discharging so the voltage changes sign.

These converters are useful where a wide range of input voltages must be tolerated, either lower or higher than the output. They are often used in modern products powered by Li-ion cells, whose voltage declines roughly from 4 V to 3 V or below as they discharge. Many circuits are designed to work from 3.3 V and therefore need a supply that can step the voltage up or down. Buck/boost converters are ideal.

Buck/boost converters can be made to give the same sign of output voltage as the input by using two switches and two diodes, but that’s getting rather complicated (look at the data sheet for the Linear Technology LTC3454 if you are interested).

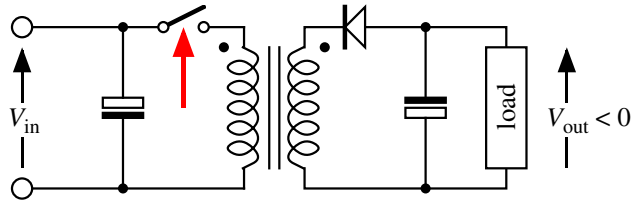


Figure 14.10 Circuit of a basic flyback converter. Many more components are needed in practice.

14.6 Flyback converter

The inverting converter can be modified so that the inductor has two windings, one for the input and a separate one for the output. This is called a *flyback converter* for historical reasons. The circuit is shown in figure 14.10.

The inductor now looks like a sort of transformer but does not work in anything like the same way as a traditional transformer with steady sine waves on input and output. In fact it is better to think of it as a coupled inductor. The dots show the 'sense' of the windings, the starts if they are both wound in a chosen direction. The circuit has the same two phases of operation as the other switching converters.

- In the storage phase, current flows from the input through the primary winding (on the left) and builds up the magnetic flux and energy in the core of the transformer; no current flows in the secondary winding because of the diode.

- In the release phase, the magnetic flux decays and drives a current through the secondary winding (on the right) and the forward-biassed diode into the load; no current flows in the primary winding because the switch is open.

Thus energy is stored in the core of the inductor in one phase and released in the other as before. The new feature is that one coil is used to feed in energy and the other to extract it. This isolates the output from the input, which is essential in a mains power supply. Further secondary windings can be added to get multiple outputs, although only one can be regulated by the PWM action.

14.7 Complete mains switching power supply unit

The flyback converter can be used as the core of a complete power supply unit to give a low DC voltage from the AC mains. (More complicated circuits are used in practice for higher power and efficiency.) It is often called an'off-line' supply because the electronics works directly from the 'line' (American usage), not via a transformer. These are the functions of the blocks in figure 14.11 on the facing page.

- The electromagnetic interference (EMI) filter keeps switching noise generated by the PSU out of the mains. It should also protect the system from incoming noise and spikes.

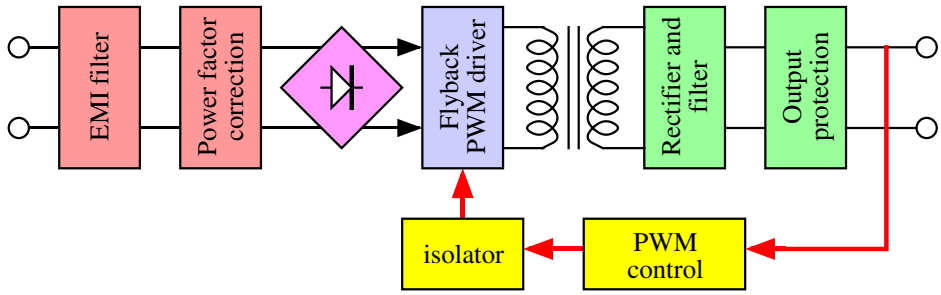


Figure 14.11 A complete mains switching power supply.

- Power factor correction is needed to avoid the problem of drawing current only at the peaks of the voltage and is required by the EU for powers above a certain rating (60 W?). Correction is getting more demanding as legislation becomes more stringent.
- The mains is then rectified directly, smoothed (not shown) and fed to a flyback converter.
- The flyback driver switches the current through the transformer winding on and off
- The output of the transformer is rectified and smoothed.
- Any practical product requires protection against overload, short circuit and the like.
- The supply must be regulated safely so the feedback control must include an isolator between the input and output. This might be optical (diode and detector) or a transformer.

You are not advised to poke inside one of these! Their mains-powered parts reach the peak voltage of the AC input and it is not pleasant to touch one of these components. (Yes, I've tried.)

14.8 General features of switching converters

Let's start with the good features.

- Efficient, over 90% possible.
- Versatile, wide range of currents, input voltages and output voltages.
- Lightweight and compact – make laptops with compact chargers practicable.
- Boost converters enable operation from very low input voltages, such as an MP3 player from a single AA cell.
- Buck/boost converters can provide 3.3 V output while compensating for the decline in voltage of a Li-ion cell as it discharges.

And now for the less attractive issues.

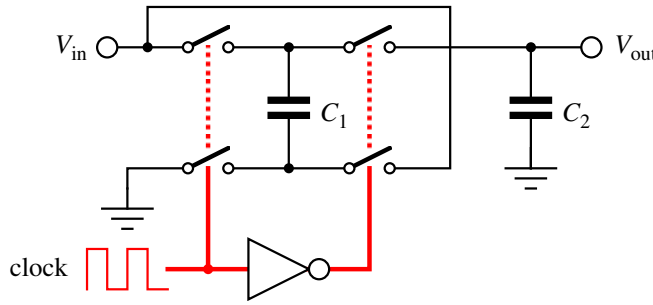


Figure 14.12 What is the function of this switched-capacitor converter?

- Create strong electromagnetic interference (EMI), which must be painstakingly suppressed. Analogue circuits are particularly sensitive and need a quiet supply.
- Need careful design, particularly the selection of the inductors and the layout of the printed circuit board. Manufacturers provide detailed advice in data sheets.
- Unreliable – less so than in the past, but often seem to be the part that breaks in domestic products. The off-line components are highly stressed and particularly vulnerable.

14.9 Will I need to design one of these?

You will almost certainly have to design a small, linear supply as part of Team Design Project 3. The hardware used in the current project runs from a 12 V supply, which is needed to drive two DC motors. However, the microcontroller uses 3.3 V and other parts of the electronics may need 5 V or even ±15 V! You cannot escape from power supplies in any practical design.

It is less likely that you will have to design a switching power supply. However, they have often been required in Team Project 4 and several students have needed them for Individual Project 4. Read the data sheet *extremely* carefully and follow its recommendations to the letter. The details of the inductor, capacitor and layout of the PCB are critical.

14.10 Examples

Example 14.1 What is the function of the switched-capacitor converter in figure 14.12?

Example 14.2 A buck (step-down) switched-mode converter operates at 100 kHz and provides a 5 V output from a 15 V input. The average output current is 1 A and must not fluctuate by more than ±10% during each cycle.

(a) At what duty cycle D does the converter operate?

(b) Sketch the waveforms for the input current, output current and voltage across the inductor, making the same simplifications as in the lecture. Your plots should show the scales.

(c) What is the average input current?

(d) What value of inductor is needed? This is not covered in the notes but follows simply from $v_L = L\,\mathrm{d}i_L/\mathrm{d}t$.

(e) Suppose that the switching supply is 95% efficient. How does this compare with a linear regulator for the same job?

Example 14.3 What is the output from a boost (step-up) converter with an input of 5 V and $D = \frac{1}{2}$? What would happen if D were raised to 0.95? [10 V]

Example 14.4 The input to an inverting (buck/boost) converter is 10 V. What values of D are needed to get output voltages of -5, -10 and -20 V? [$\frac{1}{3}$]

Example 14.5 Suggest suitable types of power supply for the following applications and explain your choice. Detailed designs are not required.

(a) A mains-powered hi-fi amplifier, which runs from supplies of ± 30 V.

(b) A mains-powered embedded digital system with supplies of 5.0 V and 3.3 V, both at several amperes.

(c) The display of a basic mobile phone, which is powered by a 3.6 V battery. The backlight for its display uses six green LEDs and can be switched on–off, not dimmed. You may connect the LEDs in any way that you wish.

(d) The processor in a digital camera, which is designed to work at 2.4 ± 0.1 V. The power comes from two AA batteries, which may be of any common type that you may specify. Is a regulator required at all and, if so, what type should be used?

15

Passive components, heatsinks and printed circuit boards

I shall first review some properties of common passive components that you will encounter – resistors, capacitors and inductors. They are distinguished from active components, which can amplify a signal and are usually made from semiconductors. The images are taken from the RS Components web site, rswww.com. We'll then look at heat sinks and finally printed circuit boards (PCBs).

15.1 Resistors

What do you need to specify for a resistor? Well resistance is obvious! However, it not the only parameter – the list is surprisingly long. Here is an abbreviated catalogue.

- **Resistance** – Values commonly range between an ohm and a few megohms. You have to worry about the resistance of the leads and joints for low values, and leakage around the resistor at high values. Avoid extreme values where possible.
- **Tolerance** – No 10 kΩ resistor has a resistance of exactly 10 kΩ because of small fluctuations in manufacturing. Most of our resistors have a tolerance of ś2% but 5% components are also common and 1% too. You can get selected values with better accuracy but it costs money. Look for a better solution, such as using an integrated circuit that includes matched resistors to set the gain of an opamp, for instance.
- **Power rating** – What happens if you put 10 V across a standard 10 Ω resistor from stores, which is probably rated for a power dissipation of $\frac{1}{10}$ W? Not what you might want...
- **Construction** – Resistors are fabricated in many ways. Two common methods are to use a solid carbon composite material or to cut a helical track in a metal film on the surface of a cylindrical body. Wirewound resistors are used for high powers and are constructed like an old-fashioned electrical bar fire (but encapsulated).
- **Package** – Traditional axial leads, surface mount packages, special casings to dissipate the heat produced by high-power resistors...

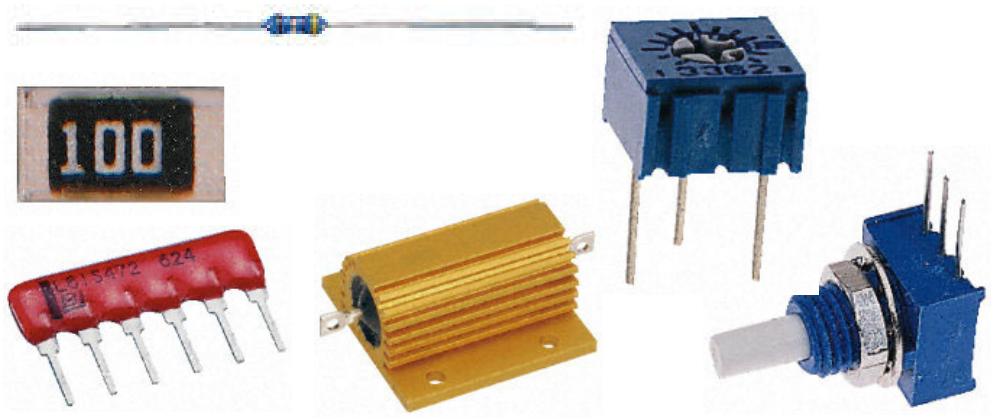


Figure 15.1 A selection of resistors: axial, surface mount (*much* smaller than the others), single-in-line (SIL) pack of 5 resistors, power resistor, trimmer and potentiometer (not to the same scale).

- **More** – Temperature coefficient of resistance, long term stability, maximum voltage, noise, ...

Figure 15.1 shows images of several types of resistor. Often you need several resistors in a circuit and manufacturers produce several types of packs for different applications. A few symbols from Capture are shown in figure 15.2 on the next page.

- **R1** – The gain of a simple amplifier based on an opamp is set by the ratio of two resistors. You can buy packs of two connected resistors (three pins) whose ratio is trimmed to meet a specificed accuracy.

- **R2** – Often you need several resistors with the same value, such as when you drive several LEDs from a digital circuit. If the resistors can all be connected between the LEDs and ground (or V_{DD}) you can choose a pack with one end of all the resistors connected to a common pin. These typically come in single-in-line packages (SIL or SIP).

- **RN1** and **RN2** – These are different symbols for packs of identical resistors with independent connections, used where it is not possible to connect them all to a common pin. This is also common when driving displays. They often come in dual-in-line packages (DIL or DIP) as the symbol for RN2 makes clear, but SIL packages are also available. (RN stands for Resistor Network.)

You must be *very* careful to check that the numbering of the pins in Capture match that of the package when using any of these multiple resistors.

Potentiometers are used where a variable resistor is needed. They have a third connection called the *slider* that can be moved from one end of the resistor to the other, forming a potential divider. Some are designed to be installed on front panels and turned by knobs but this is now old-fashioned. Others, often called *trimmers*, are mounted on a PCB and adjusted with

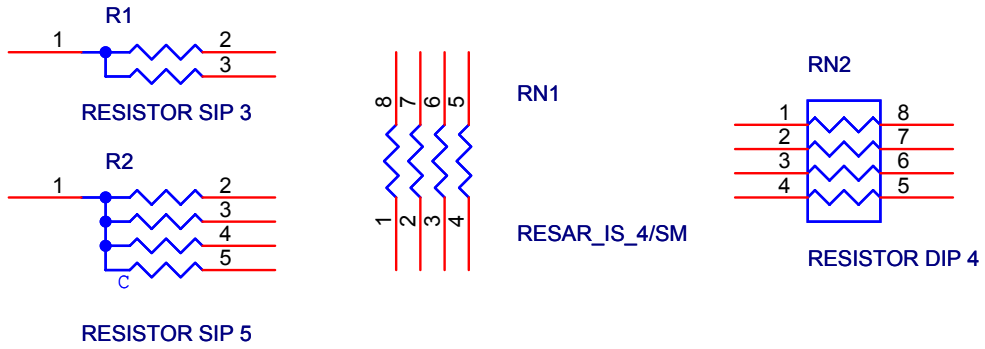


Figure 15.2 Selection of resistor packs from the DISCRETE library in OrCAD Capture.

a screwdriver to achieve the desired performance from a circuit – the gain of an amplifier, for instance. Any component with a moving part is less reliable than a fixed one so you should aim to avoid trimmers. A PCB with a large number of trimmers is almost always a poor design.

Figure 15.3(a) shows a simplified diagram of an idealised resistor. It is formed from a piece of resistive material of length d and constant cross-section A with a contact at each end. Its resistance is

$$R = \frac{\rho d}{A}, \tag{15.1}$$

where ρ is called the *resistivity* of the material. The conductivity σ is often used as well: $\rho = 1/\sigma$. The electrical behaviour is not as straightforward as this when you look in detail. In particular, a resistor has inductance as well as resistance. This becomes significant at high frequencies because the impedance of an inductor rises with frequency. You may need to select a type of resistor that has low inductance (carbon composition rather than metal film) for high-frequency circuits.

15.2 Capacitors

These come in a huge range of values from pF to F with numerous different types. The principle is straightforward: a capacitor has two metal plates separated by a dielectric, as shown in

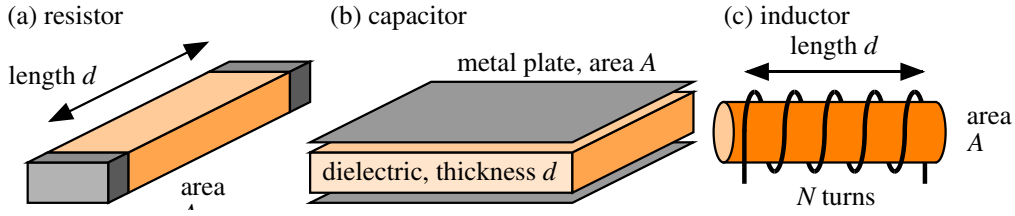


Figure 15.3 Diagram of a theoretical (a) resistor, (b) capacitor and (c) inductor (as you might remember from school!).

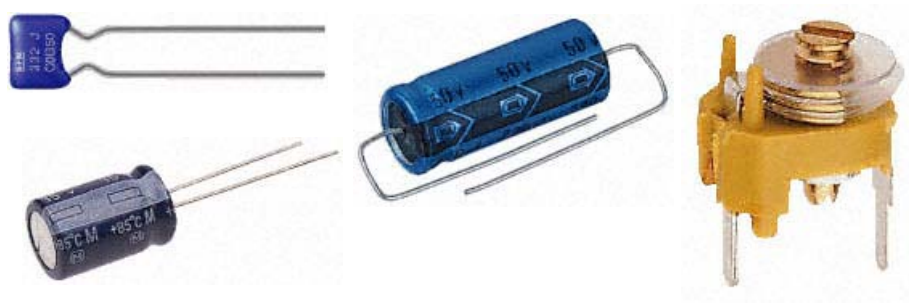


Figure 15.4 A selection of capacitors: ceramic multilayer, two aluminium electrolytics and a padder.

figure 15.3(b) on the preceding page. The capacitance is given by

$$C = \frac{\epsilon_0 \epsilon_r A}{d}, \tag{15.2}$$

where A is the area of each plate, d is the thickness of the dielectric that separates the plates, $\epsilon_0 = 8.854 \times 10^{-12}\,\mathrm{F\,m^{-1}}$ is the permittivity of free space and ϵ_r is the *relative permittivity* of the dielectric or the dielectric constant. This is unity for free space by definition, around 10 for many plastics and over 100 for some ceramics. A large value gives a compact capacitor.

The main distinction between capacitors is the type of dielectric. Most are plastics: polyester, polypropylene, polycarbonate, polystyrene, teflon (rare). Some work well at high frequency (polystyrene, for instance) but give large packages, while others give compact capacitors but are restricted to low frequency. Ceramic capactors are particularly small because of the high dielectric constant but this comes at a price: their capacitance varies strongly with temperature and applied voltage (changes of 50% are not unusual; see below). I've shown a few types in figure 15.4. Many modern devices come in surface mount packages and are virtually indistinguishable from other components with two leads, such as resistors. They are just anonymous, rectangular, black patches!

In practice many capacitors do not have flat plates as in the sketch. Often the 'plates' and dielectric are wound into a cylinder like a Swiss roll but this increases the series resistance and inductance (see below). You may encounter a small adjustable capacitor called a *padder*, analogous to a trimmer resistor. Old-fashioned radios were tuned with large, air-spaced variable capacitors but these are long obsolete.

No electronic components are ideal. Capacitors have two resistances associated with them: one in series and one in parallel as shown in figure 15.5(a) on the next page. (They exhibit inductance as well, which becomes important at high frequency.) The parallel resistance represents leakage between the plates and is particularly significant with electrolytic capacitors (below). The series resistance limits the speed at which the capacitor can charge – remember the RC time constant? Capacitors that must respond quickly, such as the decoupling capacitors across digital ICs or in switching power supplies, must be chosen to have a low value of R_{series}. It is also known as the *equivalent series resistance* or ESR.

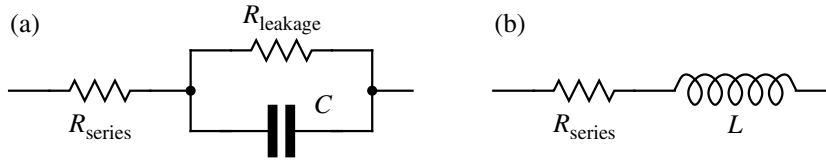


Figure 15.5 Simplified equivalent circuit of a realistic (a) capacitor and (b) inductor.

Electrolytic capacitors

Electrolytic capacitors are made in a rather different way from the simple sketch. The plates are usually aluminium and are separated by a conducting paste, the electrolyte. A current is passed between the plates, which causes an oxide layer to grow on the anode. This is the same process of anodization that is used to give a shiny, protective finish to aluminium products. The oxide is a good insulator and acts as the dielectric layer of the capacitor. It is also thin, which gives a large capacitance. Thus electrolytic capacitors are used where a high value is needed.

Electrolytic capacitors need a permanent DC bias across them to maintain the oxide film. This means that they must be connected the correct way round. If this is not done the oxide thins, breaks down, allows a current to pass and the capacitor explodes with an unpleasant smell! The polarity is always shown on the package. Tantalum electrolytic capactors offer superior performance at a higher cost.

The tolerance of electrolytic capacitors is poor, typically $\pm 20\%$ but sometimes $-50/+100\%$, which is a fancy way of saying a factor of 2! They leak badly and their range of temperature is restricted. Their lifetime is limited and they are often responsible for the death of equipment from old age.

A second parameter that must always be specified for an electrolytic capacitor is its working voltage. You will have seen this as CMAX in OrCAD. A typical value is 16 V for a small component but obviously much higher working voltages might be needed for a power supply.

Power supplies usually contain large, aluminium, electrolytic capacitors for smoothing. The ripple current flows in and out of the capacitor on each cycle and cause it to heat up because of the power dissipated in the series resistance R_{series}. Such capacitors therefore have a ripple current rating that must be respected.

Decoupling capacitors

All digital ICs should have a decoupling capacitor connected across them to reduce the spread of noise from them. This is because CMOS circuits draw a pulse of current at every clock transition, which may be very strong for a large IC (100 A for a fancy microprocessor). Manufacturers give detailed recommendations, which should be followed. This advice is taken from the data sheet for the Freescale MC9S08QG8, a small 8-bit microcontroller, and you will find something similar in every data sheet.

> Typically, application systems have two separate capacitors across the power pins: a bulk electrolytic capacitor, such as a 10 µF tantalum capacitor, to provide bulk charge storage for the overall system, and a bypass capacitor, such as a 0.1 µF

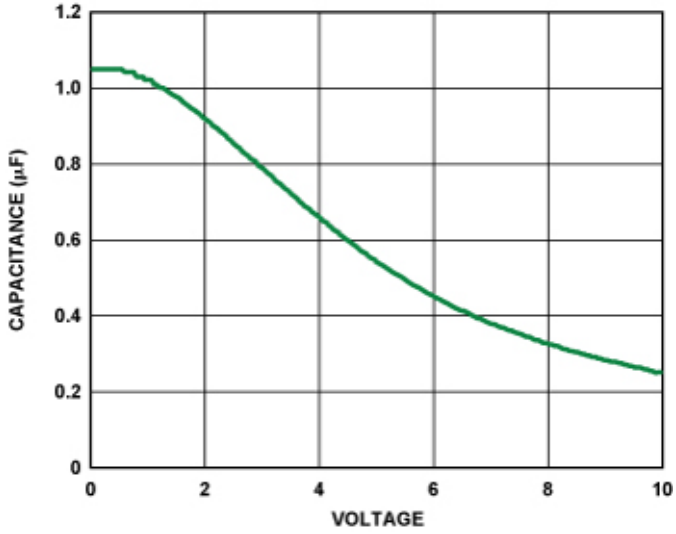


Figure 15.6 Capacitance as a function of voltage for a 1 µF, 10 V, X5R multilayer ceramic capacitor [29].

ceramic capacitor, located as near to the MCU power pins as practical to suppress high-frequency noise.

It seems strange to put a 0.1 µF capacitor in parallel with a 10 µF capacitor. Why bother, when the smaller one makes a tiny contribution to the overall capacitance? The difference is the ESR. Electrolytic capacitors have relatively high resistances, which gives them a long time-constant $\tau = R_{\text{series}}C$. This means that they cannot charge and discharge quickly. The smaller capacitor is usually specified as a multilayer ceramic component, which has a very low ESR. It can therefore respond quickly to the spikes of current drawn by the microcontroller. The larger capacitor acts as a reservoir for slower changes.

Decoupling capacitors should also be used for analogue components, such as op-amps, in a mixed signal system. In this case the purpose is to keep noise out of the IC.

Low-dropout power supplies need a capacitor on their output for stability and these are usually specified as multilayer ceramic types for modern ICs. Ceramic capacitors of 1 µF are now readily obtainable although they were unfeasibly large in the past. The compact size comes at a penalty, mentioned above. To make this clear, figure 15.6 shows the capacitance of a nominally 1 µF, 10 V multilayer ceramic capacitor as a function of voltage [29]. The capacitance falls rapidly as a function of voltage and is down to about 25% of its nominal value at the rated voltage. These components do not obey $Q = CV$! The properties depend strongly on the specific ceramic used as the dielectric. These are denoted with codes such as X5R, which was used for the figure. The capacitance may also vary strongly with temperature.

These capacitors have other interesting properties. The dielectrics are *piezoelectric*, which means that a mechanical stress produces an electric field. In other words, the capacitor generates a voltage if you drop it. This can be put to good use for energy harvesting if the capacitor is

Figure 15.7 A selection of inductors: axial, toroidal, ferrite beads and transformer kit. These are not to scale and the kit on the right is *much* larger than the others.

subject to regular vibration. On the other hand, it would be hopeless to use such a capacitor on a wire carrying a weak signal.

15.3 Inductors

These are used less frequently but are unavoidable in power supplies and for keeping high-frequency noise out of circuits. They are usually made by winding turns of wire on a core as shown rather badly in figure 15.3(c) on page 128. The value of inductance L is given *roughly* by

$$L = \frac{\mu_0 \mu_r N^2 A}{d} \tag{15.3}$$

where $\mu_0 = 4\pi \times 10^{-7}\,\mathrm{H\,m^{-1}}$ is the permeability of free space, μ_r is the relative permeability of the material of the core (unity for air), N is the number of turns, A is the cross-sectional area of the core and d is the length of the coil. The values range from μH for high-frequency chokes to H for large inductors in low-frequency power supplies. Figure 15.7 shows some examples.

The core may be air for small values, ferrite dust for high frequencies or laminations (thin, insulated layers) of soft iron for low frequencies. The core must be an insulator to avoid losses by eddy current, which is why laminations or particles are used rather than a solid block of material (Engineering Electromagnetics 2).

The shape varies: inductors are often axial (like resistors). Toroidal (doughnut-shaped) inductors have the advantage that magnetic flux does not leak. Low-frequency transformers have laminated cores made of E and I-shaped stampings. The wire of the winding must be thick enough to carry the specified current. This wire has resistance, which makes real inductors far from ideal. This is one reason why they are avoided where possible. I have drawn the simplest equivalent circuit in figure 15.5(b) on page 130. Really there is capacitance as well, which causes resonance – inductors can be nasty. However, they are essential for switch-mode power supplies and detailed recommendations are made in the application sheet for ICs. Follow them carefully!

Small inductors called *chokes* are used to suppress high-frequency noise in many systems. Often the inductance is small enough that a ferrite core or bead can be placed around the wire rather than vice versa. You will find these all over computer systems – on the lead to the monitor and USB cables, for instance, often incorporated into the connector.

Table 15.1 Standard values of components. The row for E24 shows the *extra* values beyond E12.

E3	10				22				47			
E6	10		15		22		33		47		68	
E12	10	12	15	18	22	27	33	39	47	56	68	82
E24	11	13	16	20	24	30	36	43	51	62	75	91

Transformers

These are inductors with two or more windings as I mentioned in section 11.1 on page 90. They are used to change the voltage for AC and for isolation. Obviously both features are used in power supplies. Critical specifications are the voltages on primary and secondary and the power that can be transferred, measured in V A. (Why not watts? The difference is partly to emphasize that the tranformer transmits this power rather than dissipating it, and partly because the current and voltage may not be in phase. You'll encounter the issues in Power Engineering 3.) Isolation transformers are widely used in networks to protect the system in case of faults. This includes ethernet and the public telephone system; specialized devices are available for each application.

You may occasionally encounter adjustable transformers, often called Variacs. Usually the secondary 'winding' is just a tap (an intermediate connection) on the primary winding, which means *that they do not provide isolation*. This arrangement is called an autotransformer (figure 11.1 on page 91) and saves wire. Fixed autotransformers are widely used for converting 230 V to 110 V or vice versa.

15.4 Standard values of components

Components come in a restricted range of values. This is particularly true for capacitors because they cover such a wide range and come in so many varieties. The standard values are given in table 15.1 and may be multiplied by a power of 10. The reason for the apparently strange choice of numbers is that they give roughly equal ratios between values. (In other words, their logarithms are equally spaced.) Resistors are readily obtainable in E12 and E24 values, sometimes more. Many capacitors are available only in E3 values, often E6 but only a few types offer a wider choice of values.

15.5 Heat sinks

The heat generated in any component must be dissipated to prevent the component getting too hot. Heat is dissipated by

- radiation
- convection to the air surrounding the device
- conduction to another body, either a heatsink or the PCB to which the component is mounted

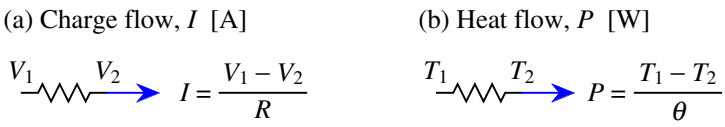


Figure 15.8 Flow of charge through an electrical resistance R compared with flow of heat through a thermal resistance θ.

The small size of most components means that special measures have to be taken to remove the heat if the power dissipation is above about $\frac{1}{4}$ W. The most common problem arises in power transistors, as in the series regulator. The temperature limit for the semiconductor junction is typically 125–150°C. The heat generated flows from the semiconductor to the case and then to the surrounding air. This flow is limited by the thermal resistance θ (ghastly notation but standard), measured in units of °C/W or °C W^{-1}. The flow of heat in watts (P) depends on the temperature difference and the thermal resistance, so that $P = (T_1 - T_2)/\theta$ or $(T_1 - T_2) = P\theta$. This is just like Ohm's law, $I = (V_1 - V_2)/R$ or $(V_1 - V_2) = IR$, and the same rules apply for combining resistances. Figure 15.8 illustrates the analogy.

If the ambient temperature (that of the surroundings) is T_a, the junction temperature is T_j and the case temperature is T_c, then $T_j = T_a + P(\theta_{jc} + \theta_{ca})$ where θ_{jc} is the resistance between junction and case and θ_{ca} is the resistance between case and the surrounding air. The two resistances are in series and therefore add, as shown in figure 15.9.

Let $\theta_{jc} = 1.5$°C/W and $\theta_{ca} = 100$°C/W, which are typical values for a small device. To dissipate 20 W with an ambient temperature of 40°C would give a junction temperature of $40 + 20(1.5 + 100) = 2070$°C. The device would not last long! (Nor might your eyes if you were not wearing safety goggles.)

The value of θ_{ca} can be reduced by attaching a *heat sink*. This is a piece of metal that improves the transfer of heat from the device to the air. Figure 15.10 on the next page shows a couple of examples. Sometimes the case of the equipment can be used; power transistors are often mounted on a metal rear panel.

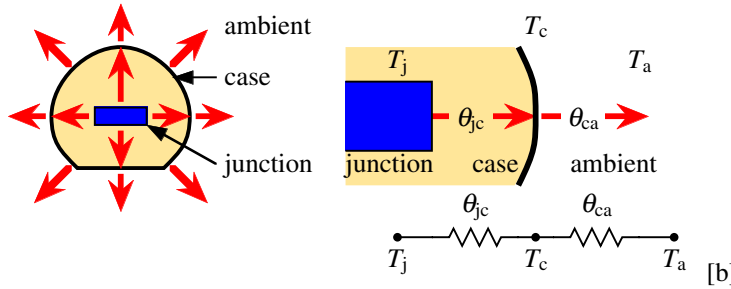


Figure 15.9 Flow of heat out of a transistor from the junction to the case, with resistance θ_{jc}, and from the case to the ambient, with resistance θ_{ca}. This is like current through resistors in series.

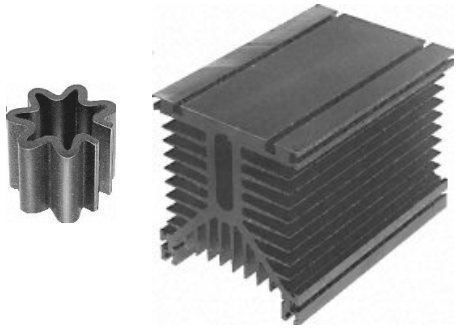


Figure 15.10 Heatsinks with thermal resistances of 50°C/W and 0.4°C/W.

When a heat sink is attached, θ_{ca} for the naked device is *replaced* by the value for the heat sink, θ_{hs}. This is illustrated in figure 15.11 on the following page. Do not add the resistance of the heat sink to θ_{ca} or the resistance will go up, not down! A third resistance is often added to model the flow from the case to the heatsink, θ_{ch}, but I'll assume that this has been included in θ_{hs}.

Common heat sinks have values of θ_{hs} ranging from 50°C/W for a small heat sink to below 1°C/W for a large one. If a simple heat sink is not sufficient then the component may have to force cooled with a fan to blow air across it, as in most PCs. Liquid cooling is required in extreme cases, often because of insufficient space for air to flow.

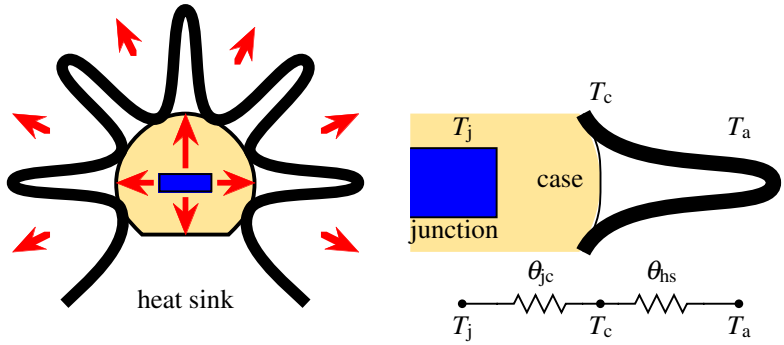
Suppose that the device in the example above were mounted on a heat sink with $\theta_{ca} = 4$°C/W. Then $T_j = 40 + 20(1.5 + 4) = 150$°C, which is just permissible. It might not be good for reliability, though, so a larger heat sink (smaller θ_{ca}) would be preferable.

Often the metal case or tab on the package of a transistor, which is used to bolt it to the heat sink, is also connected to one of the terminals of the transistor itself – usually the drain or collector, because this is where most energy is dissipated. An insulating washer must be used between the transistor and heatsink if this connection would cause a short circuit.

Surface-mount devices (section 15.7) are often designed to use an area of copper on the printed circuit board as their heat-sink. The data sheet shows the shape required to carry away the heat. For example, in the project we may use the Fairchild FDS9926A dual n-MOSFET, which somes in a SOIC-8 package. Its data sheet shows three examples of PCB layout with different thermal resistance, reproduced in figure 15.12 on the next page. A catch is that these require '2 oz copper', which is thicker than usual (typically 1 oz). Many surface-mount devices have large pads under the middle of the package to provide good thermal contact to the heatsink on the PCB. This works well but such packages are almost impossible to assemble by hand.

Power dissipation derating curve

Some manufacturers specify the value of θ_{jc} but it is also common to give a *derating curve* instead. This specifies the power dissipation as a function of case temperature (not ambient temperature). Figure 15.13 on page 137 shows an example. Alternatively the data is given in the form of an equivalent statement:



[b!]

Figure 15.11 Flow of heat out of a transistor from the junction to the case, with resistance θ_{jc}, and from the case to a heat sink with resistance θ_{hs}.

- Total dissipation at 25°C case temperature = 3 W.
- Derate at 20 mW/°C for higher temperatures.

To convert this information to a thermal resistance, go back to the general equation for heat flow between the junction and ambient for a bare device,

$$P = \frac{T_j - T_c}{\theta_{jc}} \tag{15.4}$$

The maximum power dissipation occurs when the junction reaches its maximum temperature, so

$$P_{max} = \frac{T_{j,max} - T_c}{\theta_{jc}} = \frac{T_{j,max}}{\theta_{jc}} - \frac{T_c}{\theta_{jc}} \tag{15.5}$$

If this is considered as a function $P_{max}(T_c)$, it is a straight line with slope $-1/\theta_{jc}$, which is negative. This is the same as the derating figure apart from the sign. Turning the expression for the slope around, the thermal resistance is the inverse of the negative slope of the graph.

In this example the permitted dissipation is 3 W at a case temperature of 25°C, falling to zero at 175°C. The derating figure is therefore $(3-0)/(175-25) = 3/150 = 0.02\,\text{W/°C} = 20\,\text{mW/°C}$. Taking the reciprocal gives a thermal resistance of $\theta_{jc} = (175-25)/3 = 50\,\text{°C/W}$.

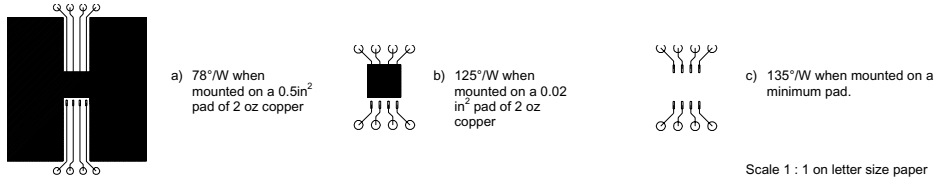


Figure 15.12 Shape of copper on PCB for Fairchild FDS9926A dual n-MOSFET to obtain different values of thermal resistance, taken from the data sheet.

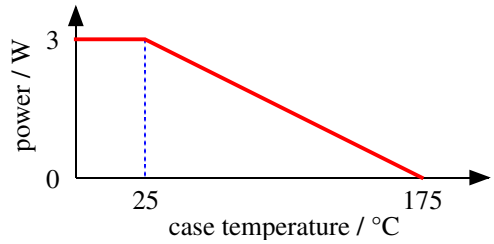


Figure 15.13 A typical derating curve for a small power transistor, showing the maximum permitted dissipation as a function of the temperature of the case.

These calculations are all for a steady state. The heatsink can be treated as a capacitor for brief pulses of power and the heat flow should be analysed in the same way as an RC circuit. See Power Electronics 2.

Here are some approximate expressions for the power dissipated in common components.

- Field-effect transistor: $V_{ds} \times I_d$.
- Bipolar transistor: $V_{ce} \times I_c$.
- Zener diode: $V_Z \times I_Z$.
- Linear regulator: $(V_{in} - V_{out}) \times I_{out}$.

Of course you have to analyse the circuit to find the voltages and currents needed for these expressions.

15.6 Printed circuit boards

Most circuits are built on printed circuit boards or PCBs. Other systems, such as stripboard (veroboard) are sometimes used for constructing prototypes, but these can be more trouble than they are worth. (But breadboards are even worse!) Electrical connections between components are provided by copper tracks on the PCB. This is called *etch* in PCB Designer because of a common manufacturing process.

1. The board starts with a complete layer of copper coated with a light-sensitive layer called photoresist.
2. The board is exposed to light through a mask that covers the area where the copper should remain.
3. A developing solution removes the exposed photoresist, revealing the copper underneath.
4. The board is placed in an etching solution, traditionally ferric chloride ($FeCl_3$) with a little hydrochloric acid, which strips the exposed copper to leave only the desired regions.

Double-sided boards are made in much the same way with two exposures but more complicated processes are needed to make boards with internal layers. A further step is also needed to add plating through the holes in most commercial boards. Read Maxfield's book [9] to learn more. Boards produced in the department do not have plated-through holes, so vias must be installed by soldering wires through the hole.

In the distant past the layout was done by hand and photoreduced onto the board. Now the layout is done by CAD tools, as you know. Several layers must be specified in addition to the tracks themselves.

- **Silkscreen** – mainly text to help the user of the board, particularly by the connectors; often only on the top
- **Assembly outlines** – identifies each component for assembly
- **Solder mask** – restricts solder to joints and prevents it spreading over all coppered areas
- **Solder paste** – needed to attach surface-mount components

A further **drill file** specifies the coordinates and diameters of holes for mounting components and the board itself. At the end you normally send a set of files that specify all the layers and drill holes to a manufacturer. These are commonly called *Gerbers* after a common format. Prototype boards can be made for less than £25 and the cost falls rapidly with quantity. The manufacturing process in the department's electronic workshop is more basic. You must print the masks for the top and bottom of the board, there is no silkscreen. Holes are drilled by hand for one-off boards but an automatic drill is used for production runs.

The conducting tracks are made of copper, whose thickness is specified in ounces per square foot (sorry – the USA dominates). The thickness of '1 oz' copper is about 35 μm, which is useful to calculate the resistance of a track.

The material of the board itself must have good electrical properties (insulator and dielectric), be mechanically strong and heat-resistant. Most professional boards are made of a green fibreglass–epoxy laminate called FR-4. The department uses a cheaper, light brown material for less demanding applications, designated FR-16. The 'FR' stands for 'fire resistance', which seems a curious way to classify PCBs.

Boards are made with different numbers of layers, as you saw in the laboratory.

- **Single layer** – adequate for simple designs but rapidly become hard to route.
- **Double layer** – widely used for less demanding applications; much easier to route than single layer. The department can produce single and double-sided boards.
- **Four layer** – typically the signals run on the outer two layers while the inner ones are used for power and ground planes. This gives much better electrical performance. The planes give low impedance and help to screen signals in the tracks from one another (you'll learn about this in Electromagnetic Compatibility 3). It is easy to probe the board if the signals are on the outside.
- **Six layer** – usually have signals on the outsides, then power and ground planes, with two further layers of signals in the middle. The middle layers are particularly well screened but hard to probe, so test points must be added.

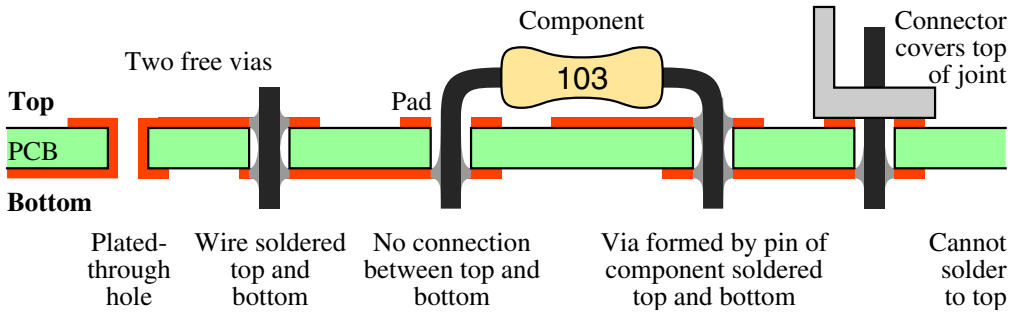


Figure 15.14 Cross-section of a double-sided printed circuit board (PCB) showing free vias formed by a plated-through hole and a wire through a non-plated hole soldered top and bottom. A via can also be formed using a pin-through-hole component but not at a connector because it covers the top pad.

- **Eight layers** or more – complicated! (And expensive.)

In a commercially produced, multi-layer board the copper plating extends through the holes, joining the pads on the two sides of the board. A plated-through hole that is used purely to move a track from one side of the board to the other is called a *via*. See the sketch in figure 15.14. Unfortunately we cannot produce plated-through holes in the department, which is why you had to insert wires for vias and solder some components top and bottom on the novelty lights in Electronic Engineering 1X. This is a nuisance, so try to avoid vias when you lay out your own PCBs. If vias are unavoidable, put them somewhere convenient – not under components, for instance. Follow the tips in the instructions on PCB Designer.

When a PCB is laid out, the CAD software needs to know the widths of tracks, spacing of tracks, size of vias and similar information. These constitute the *design rules* and the information must be entered by hand or read from a *technology* (or *tech*) file in PCB Designer.

Solder used to consist of a lead–tin alloy but most large-scale production must now use lead-free components and assembly. The European Union's Restriction of Hazardous Substances (RoHS) legislation has outlawed many other chemicals that were formerly used and similar restrictions are being imposed in other parts of the world. We still use solder that contains lead in the laboratory because lead-free solder is much harder to use in manual assembly, but we expect to be forced to change in a few years.

15.7 Component packages

The packages in which components are encapsulated have changed dramatically in the last two decades. The two general styles are shown in figure 15.15 on the following page for an 8-pin integrated circuit.

- Older components are designed to be mounted on the top of the board (usually). Their pins poke through holes to the opposite side of the board, where they are soldered to pads on the tracks. These are *pin-through-hole* or PTH components. The pins are usually laid out on a $0.1''$ grid, which makes them easy to solder by hand.

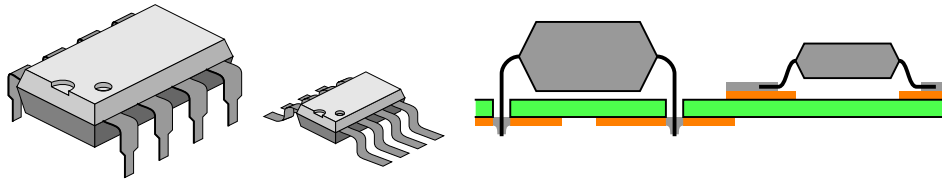


Figure 15.15 Pin-through-hole and surface-mount (gull wing) integrated circuits. Some surface-mount packages have solder pads underneath them, usually for heatsinks. There may also be an adhesive pad to mount the component until it has been soldered.

- Most modern components are designed to be mounted on the same side of the board as their tracks. These are *surface-mount devices* or SMDs. They do not need holes drilled through the board for pins, which allows the leads to be much closer together.

An increasing number of components have no pins or leads at all. They just have solder bumps or metal pads on their boundaries to make the connections. These need specialized assembly.

The standard packages for almost all components are now surface-mount, with the exception of large items such as sockets that need the security of pins through the board. Unfortunately the range of surface-mount packages is vast so I have provided a guide to the more common types of package in figure 15.16 on the next page.

- **Dual inline package (DIP)** – sometimes DIL, or PDIP for plastic DIP. This was the standard pin-through-hole package for integrated circuits for many years. Pins are $0.1''$ apart with rows are separated by $0.3''$. Wider packages were used for larger ICs with many pins but are now rare. However, modules such as the mbed have the same pinout so that they can be used easily with veroboard and breadboards (prototyping boards).

- **Small outline integrated circuit (SOIC)** – sometimes just SO. These were the earliest surface-mount packages. Their pads are $0.05''$ apart with rows about $0.2''$ apart. You may use an opamp or a dual MOSFET in a SOIC-8 package in the project. They are one of the few types of SMD that is relatively easy to solder by hand.

- **Thin shrink small outline package (TSSOP)** – 'thin' refers to the vertical dimension and 'shrink' to the separation of the pins. They are roughly twice as close as in a SOIC but dimensions were changing from imperial to metric at the time so the separation is officially 0.65 mm. (It just happens that this is close to $0.025''$.)

- **Quad flat pack (QFP)** – have pins on all four edges, unlike the packages described previously. The pitch (separation between centres of pins) is often 0.65 or 0.50 mm, which makes them 'challenging' to solder. Microcontrollers typically come in these packages. Plenty of variations are offered, such as LQFP for low profile QFP.

- **Quad flat-pack no-lead (QFN)** – one of the most popular packages at present. It has no leads at all, just metal patches on the underside of the package. Most devices also have a large central pad, typically used for ground or as a heatsink. They are extremely difficult to solder by hand and the central pad is close to impossible. Please avoid them.

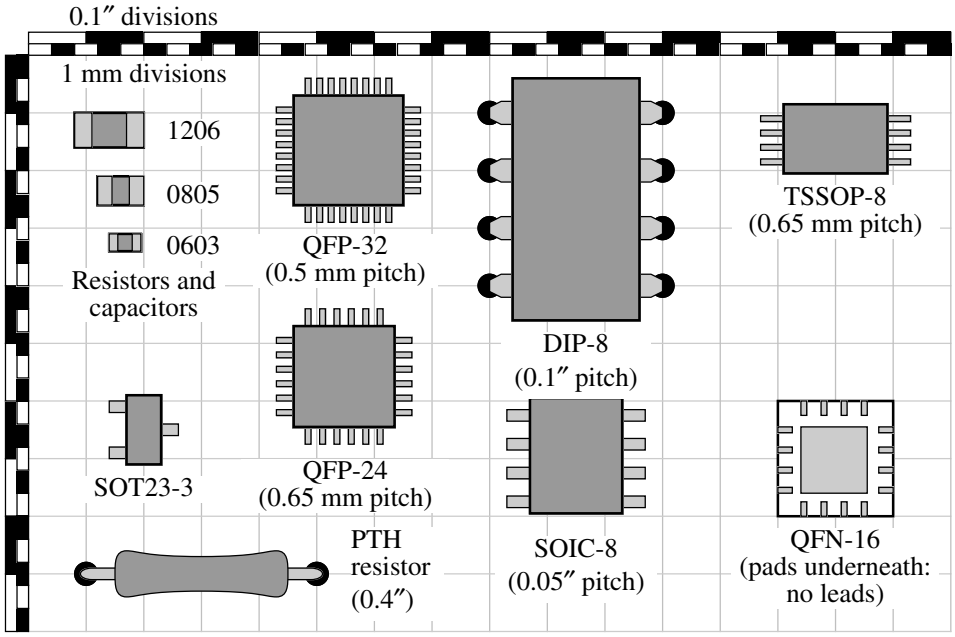


Figure 15.16 Outlines of a selection of surface-mount packages with a conventional resistor and DIP-8 IC for comparison. The suffix on the semiconductor packages shows the number of pins and the background grid has 0.1″ spacing.

- **Ball grid array (BGA)** – no pins, just a two-dimensional array of solder balls on the underside. Vital for fancy digital processors with hundreds of pins but definitely for automatic assembly. **Pin grid array** is similar but with through-hole pins on a 0.1″ grid and is now rare because the package is so much larger.
- **Plastic leaded chip carrier (PLCC)** – an early SMD, often used for microcontrollers. Has J-leads that curl under the body rather than gull-wing that stick out. The pitch is typically 0.05″ and special sockets are used.
- **Small outline transistor (SOT)** – many varieties, of which SOT23 is now common for discrete transistors. It is about 2 mm × 3 mm. The package is not restricted to transistors despite its name: Op-amps often come in SOT23-5 packages, which are the same size as SOT23-3 for transistors but with 5 leads instead of 3.

Passive components are available in surface-mount packages to match semiconductors but the geometry is simple because they have only two leads. The size is quoted as two double-digit figures such as 1206, which means 0.12″ × 0.06″. This particular size is not too awkward to handle but you must not breathe too heavily on an 0603 package or it will fly away! Even smaller 0402 packages are now in use for portable electronic products where a compact PCB is essential.

15.8 Examples

Example 15.1 What is the maximum continuous voltage that can safely be applied to a 120 Ω, quarter-watt resistor? Would this be a problem in a circuit with a 5 V supply? What is the maximum safe current? [5.5 V, 46 mA]

Example 15.2 Old-fashioned radios were tuned with air-spaced variable capacitors with a maximum value of around 300 pF. What total area is required, assuming that the plates are 0.5 mm apart?

Example 15.3 A typical value for a decoupling capacitor for a digital integrated circuit is 100 nF. Suppose that the IC runs at 10 MHz. What value of equivalent series resistance (ESR) is needed to ensure that the capacitor decouples digital switching noise effectively? Is this likely to be a problem in practice? Only a rough estimate is needed. What would happen if the IC ran at 1 GHz instead?

Example 15.4 A coil is 25 mm long with 25 turns wound on a core of 5 mm diameter and relative permeability 1000. Estimate its inductance. Estimate also its resistance assuming that the wire is copper with diameter 0.5 mm. [0.6 mH, 0.04 Ω]

Example 15.5 A perverse bureaucrat decrees that there should be only 10 standard values of resistor per decade instead of the traditional 12. What should they be? (In other words, what replaces the current values of 10, 12, 15, 18…?) Remember that they are separated by equal *ratios*.

Example 15.6 A transistor dissipates 2 W and has a thermal resistance of 5°C/W between junction and case. Calculate the junction temperature when the ambient temperature is 50°C with a heat sink of thermal resistance (i) 50°C/W and (ii) 10°C/W. [80°C, 160°C]

Example 15.7 A TIP120 transistor is required to dissipate 40 W in an ambient temperature of 30°C. Determine the thermal resistance of the heatsink required to keep the junction temperature below 150°C. There is an insulating washer of thermal resistance 0.5°C/W between the case and the heatsink. The transistor is specified for a dissipation of 65 W below 25°C, derated at 0.5 W/°C at higher temperatures. [0.5°C/W]

What size of heatsink would be needed if a TIP3055 were used instead? This is a larger transistor with a maximum dissipation of 90 W derated at 0.7 W/°C.

Example 15.8 What is the resistance of a 'typical' track on a PCB? Take it to be 20 mm long, 0.5 mm wide and made of '1 oz' copper. Take the resistivity of copper to be 2×10^{-8} Ω m. What would happen if the board were made with half-ounce copper instead? [Very roughly 0.02 Ω]

Further reading

This is a rather unsystematic collection of books, articles and application notes that may be useful for this course and particularly in subsequent projects. All manufacturers provide Application Notes to help you use (and therefore encourage you to buy) their components. Both these and the data sheets contain a wealth of information. The circuit that you need is probably in one of these documents unless you are tackling a genuinely new problem.

Books

[1] Bonnie Baker. *A Baker's Dozen: Real Analog Solutions for Digital Designers.* Newnes, 2005. (ISBN 0750678194)

This is close to being the 'book of the course' for the first half. It is aimed at digital engineers who need to handle the interface to analogue electronics. Much of the book is on ADCs and DACs and treats issues such as noise. The author is now at Texas Instruments, having been formerly at Microchip Technology and Burr–Brown, so she knows what she is writing about! The mathematics is occasionally a little flaky and she uses 'differentiate' where I would write 'subtract' (the usage comes from the term 'differential amplifier', which subtracts rather than differentiates).

[2] John H. Davies. *MSP430 Microcontroller Basics.* Newnes, 2008. (ISBN 9780750682763) See userweb.elec.gla.ac.uk/j/jdavies/mspbook for errata and downloads.

Naturally I recommend my own book! It covers a different microcontroller but the general aspects of embedded systems are equally applicable to the mbed. You might find some of the material familiar.

[3] D Fitzpatrick. *Analog Design and Simulation using OrCAD Capture and PSpice.* Newnes, 2011. (ISBN 9780080970950)

[4] Paul Horowitz and Winfield Hill. *The Art of Electronics.* Second edition, Cambridge University Press, 1989. (ISBN 0521370957)

No electronic engineer should be without this book, with its lucid coverage of all aspects of electronics. Many details are now out of date but the principles are as true as ever. The only problem is that it is *too* easy to read, so that the reader tends to fly through the text too fast to absorb it!

[5] Walt Kester (editor), Analog Devices. *Data Conversion Handbook.* Newnes, 2004. (ISBN 0750678410)

Nearly 1000 pages on all aspects of data conversion from their history to the design of PCBs. It goes into the theory more deeply than Baker's book but is less coherent because of its edited nature. The authors are from Analog Devices and the examples are taken from their range of products.

[6] Walt Kester (editor), Analog Devices. *Mixed-signal and DSP Design Techniques.* Newnes, 2002. (ISBN 0750676116)

[7] Walt Kester (editor), Analog Devices. *Op Amp Applications Handbook.* Newnes, 2004. (ISBN 0750678445)

[8] Ron Mancini (editor), Texas Instruments. *Op Amps for Everyone.* Third edition, Newnes, 2009. (ISBN 9780750677011)

Despite the different title, this covers much the same material as Baker's book although there is rather more about op-amps, as you might expect. This is a good place to look for information on single-supply op-amps. It suffers a little from having an editor rather than single author but is reasonably coherent. An earlier version can be downloaded from the TI web site as application note SLOD006b but it is nearly 500 pages long so the printed version might be worth the cost.

[9] Clive Maxfield. *Bebop to the Boolean Boogie: An unconventional guide to electronics.* Third edition, Newnes, 2009. (ISBN 9781856175074) See also www.maxmon.com/booginfo.htm.

This book fully lives up to its subtitle: It is nothing like a conventional textbook and covers a broader spectrum of electronics than you would guess from the title. It starts with the relation between analog and digital signals and the main body of the book concludes with useful material on components and construction. The appendices cover some fascinating topics followed by a comprehensive glossary of terms commonly used in electronics (and many more that are not). Study carefully the section on *How to become famous*.

[10] Kraig Mitzner. *Complete PCB Design using OrCAD Capture and PCB Editor.* Newnes, 2009. (ISBN 9780750689717)

An excellent book on OrCAD PCB Designer. It goes far beyond your introduction at the beginning of the year and explains numerous techniques. Highly recommended. The only drawback is that it concentrates on PCBs for modern, commercial production rather than our old-fashioned, in-house process.

[11] John Watkinson. *The Art of Digital Audio.* Third edition, Focal Press, 2000. (ISBN 9780240515878)

An excellent book on many aspects of analogue-to-digital and digital-to-analogue conversion with emphasis on audio applications (as you would expect from the title). The same author has written an *Introduction to Digital Audio*, second edition, Focal Press, 2002 (ISBN 9780240516431).

[12] R Toulson and Tim Wilmshurst. *Fast and Effective Embedded Systems Design: Applying the ARM mbed.* Newnes, 2012 (ISBN 9780080977683)

[13] Joseph Yiu. *The Definitive Guide to ARM Cortex-M3 and Cortex-M4 Processors.* 3rd Edition, Newnes, 2013. (ISBN 9780124080829)

Application notes on ADCs and DACs

[14] Nicholas Gray. *ABCs of ADCs.* National Semiconductor, 2004. Application note.

[15] Freescale (formerly Motorola) *M68HC11 Reference Manual* Motorola, 2002. Available on the web at www.freescale.com/files/microcontrollers/doc/ref_manual/M68HC11RM.pdf

Section 12 provides a good description of the operation of a successive-approximation ADC.

[16] Thomas Kugelstadt. *The operation of the SAR–ADC based on charge redistribution.* Texas Instruments, 2005. Application note SLYT176.

[17] *A Simple ADC Comparison Matrix.* This leads into a set of more detailed application notes on individual types of ADC. Maxim Integrated Products, 2003. Application note AN2094.

[18] Bonnie Baker. *Glossary of analog-to-digital specifications and performance characteristics.* Texas Instruments, 2006. Application report SBAA147.

[19] Ron Mancini, Texas Instruments. *Sensor to ADC—analog interface design.* Application note slyt173 (2000).

Explains how to match the span of a sensor's output voltage to the input of an analog-to-digital converter (ADC).

Application Notes on Amplifiers (mostly single-supply)

[20] Ron Mancini, Texas Instruments. *Single-supply op amp design.* Application note slyt189 (1999).

[21] Bruce Carter, Texas Instruments. *A Single-Supply Op-Amp Circuit Collection.* Application note sloa058 (2000).

[22] Bruce Carter, Texas Instruments. *Designing Gain and Offset in Thirty Seconds.* Application note sloa097 (2002).

[23] Kitchin, Charles. *Avoid common problems when designing amplifier circuits.* Analog Dialogue, volume 41, issue 08, pages 1–4, 2007 August. Available on the web at www.analog.com/library/analogDialogue/archives/41-08/amplifier_circuits.pdf

This is an excellent and concise summary of common problems, much of which applies to single-supply circuits. There is an earlier article on *Demystifying single-supply op-amp design* in Electronic Design News, 2002 March 21, pages 83–90.

[24] Ron Mancini, Texas Instruments. *How to read a semiconductor data sheet.* Electronic Design News, 2005 April 14. www.edn.com/article/CA514964.html.

This focusses on op-amps but the principles are more general – designing for the worst case of each parameter, for instance.

Application Notes on Power Supplies

[25] National Semiconductor. *Introduction to Power Supplies.* Application Note AN-556 (2002).

[26] Chester Simpson, National Semiconductor. *Linear Regulators: Theory of Operation and Compensation.* Application Note AN-1148 (2000).

Components

[27] Ron Mancini, Texas Instruments. *Understanding basic analog – passive devices.* Application report SLOA027

[28] Ron Mancini, Texas Instruments. *Understanding basic analog – active devices.* Application report SLOA026A

[29] Glenn Morita, Analog Devices. *Low dropout regulators—Why the choice of bypass capacitor matters.* Analog Dialogue 45–01 Back Burner, January 2011.

Magazines and articles

[30] *Circuit Cellar* is a monthly electronics magazine. Its subtitle is *The magazine for computer applications* but the computers are almost always embedded rather than on a desktop. It is published in the USA and the postage makes the printed version expensive in Britain but the online edition at www.circuitcellar.com is affordable. A good read.

Data sheets provided

I have attached a selection of data sheets as examples of the range of components available. Most are only extracts to save paper. Download the full sheet from the manufacturer's web site if you need it; the online version of these notes contains hyperlinks to the components or document itself.

[31] Analog Devices AD7788, a 16-bit sigma–delta ADC.

[32] ST Microelectronics TS951 low-power op-amp.

[33] Texas Instruments REF29xx voltage reference.

[34] Energizer alkaline–manganese dioxide LR03 cell (2012).

[35] Energizer CR2032 lithium coin cell (2009).

[36] Comparison of battery chemistries from Buchmann battery university (mainly on secondary cells), www.batteryuniversity.com (2005).

[37] Duracell nickel–metal hydride rechargable cells (1997).

[38] Panasonic Li-ion CGR18650E cell (2007).

[39] National Semiconductor LM2931 low-dropout regulator (2006).

[40] National Semiconductor LM3100 buck switching regulator (2006).

[41] National Semiconductor LP5526 lighting management unit with high voltage boost converter (2006).

Solutions to examples on data conversion

2.1 LSB $= (3.3\,\text{V})/2^{12} = 0.8\,\text{mV}$.

2.2 The output of an n-bit converter is given by $(V_{\text{ADC}}/V_{\text{FS}}) \times 2^n$ rounded to the nearest integer, except that the maximum value of 2^n is impossible. Here $n = 12$ and $2^n = 4096$ so we get (i) $0.1 \times 4096/5.0 = 81.92$, which rounds to 82; (ii) $1.0 \times 4096/5.0 = 819.2$, which rounds to 819; (iii) $4.0 \times 4096/5.0 = 3276.8$, which rounds to 3277; (iii) 5.0 is the full-scale input so the output is also the maximum, 4095 – it cannot be 4096! These are decimal; in hexadecimal they are 0x052, 0x333, 0xCCD, 0xFFF.

2.3 This is of course the inverse of the previous question but with the difference that the input voltage can lie in a *range*; it is not a single number because of the quantization. In most cases the input can lie in a range of $LSB = (2048\,\text{mV})/2^8 = 8\,\text{mV}$. The exceptions are the ranges at the ends. The analogue input voltage in the middle of the normal intervals is related to the digital output d by $V_{\text{ADC}} = V_{\text{FS}} \times (d/2^n) = 8d$ mV.

(i) Zero output corresponds to inputs of zero to $\frac{1}{2}$LSB, which is 0–4 mV. This has only half the width of the normal intervals.

(ii) The midpoint is 72 mV and this has the normal spread of $\pm\frac{1}{2}$LSB so the range of possible inputs is 68–76 mV.

(iii) 0xAB is 171 in decimal so the midpoint is 1368 mV and the range is 1364–1372 mV.

(iv) This is the top of the scale, which is the other abnormal interval. 0xFF is 255 in decimal so the usual midpoint would be 2040 mV. In this case the range goes down $\frac{1}{2}$LSB but up by a whole LSB, giving 2036–2048 mV.

2.4 The number of possible output values is $1.8\,\text{V}/1\,\text{mV} = 1800$. This must be rounded up to the next power of two, $2048 = 2^{11}$. Thus an 11-bit converter is needed. (These are unusual and you would probably have to go for 12 bits.)

2.5 It arises from the unavoidable difference between the staircase and straight line. Of course it reduces as the number of bits increases because LSB gets smaller.

3.1 A flash converter compares the input with voltages of $\frac{1}{2}$LSB, $\frac{3}{2}$LSB and so on. For a 4-bit converter LSB $= V_{\mathrm{FS}}/16$ so the 15 voltages for comparison are $\frac{1}{32}V_{\mathrm{FS}}$, $\frac{3}{32}V_{\mathrm{FS}}$ and so on up to $\frac{29}{32}V_{\mathrm{FS}}$. The wider top interval means that $\frac{31}{32}V_{\mathrm{FS}}$ is not required.

For a 5 V range the voltages are $\frac{5}{32}$ V, $\frac{15}{32}$ V and so on up to $\frac{145}{32}$ V or 0.15625 V, 0.46875 V ..., 4.53125 V.

3.2 The input voltage in terms of LSB is $3/(5/16) = 9.6$. The comparators have voltages of $\frac{1}{2}$LSB, $\frac{3}{2}$LSB and so on. The output will be 1 from the comparators up to $9\frac{1}{2}$LSB $= \frac{19}{2}$LSB, which is 10 comparators. The remaining 5 will give zero. Thus the thermometer code has 10 low bits of 1 and 5 high bits of 0: 0b0000011111111111. Decoding this gives a decimal value of 10 or normal binary of 0b1010. This is the nearest integer to 9.6, which is what we expect.

3.3 It's easier to write this out for $V_{\mathrm{ADC}}/V_{\mathrm{FS}} = 3/5 = 0.6$. Here is the sequence of four comparisons for a 4-bit ADC.

1. Compare with half full range: $0.6 > \frac{1}{2}$ so bit = 1.
2. Split upper range from $\frac{1}{2}$ to 1 in two to get $\frac{3}{4}$ and compare: $0.6 < \frac{3}{4}$ so bit = 0.
3. Split lower range from $\frac{1}{2}$ to $\frac{3}{4}$ in two to get $\frac{5}{8}$ and compare: $0.6 < \frac{5}{8}$ so bit = 0.
4. Split lower range from $\frac{1}{2}$ to $\frac{5}{8}$ in two to get $\frac{9}{16}$ and compare: $0.6 > \frac{9}{16}$ so bit = 1.

Thus the output is 0b1001 = 9 but we expect it to be 10! Oh dear....

The problem is that this method always gives the integer *below* the exact value, not the nearest integer. None of the application notes or books makes this clear! One way of solving this problem might be to add an offset to the capacitor network. Another is to perform one extra comparison:

5. Split upper range from $\frac{9}{16}$ to $\frac{5}{8}$ in two to get $\frac{19}{32}$ and compare: $0.6 > \frac{19}{32}$ so extra bit = 1.

This shows that the result should be rounded up from 9 to 10.

3.4 The first comparison is with $\frac{1}{2}V_{\mathrm{FS}}$, which gives the msb.

3.5 The fractional deviation from perfect charging must be less than $\frac{1}{2} \times 2^{-12} \approx \mathrm{e}^{-9}$. Thus 9 time-constants τ must be allowed. For the given values, $\tau = RC = 0.8\,\mu\mathrm{s}$ and $9\tau = 7\,\mu\mathrm{s}$.

The output resistance is in series with the internal resistance so the total rises from 2 kΩto 12 kΩ, a factor of 6. This affects τ with the same factor so the sampling time should be increased to about 43μs.

3.6 The ratio of the input voltage to the reference is given by the ratio of the count to the range of the counter, so $V_{\mathrm{ADC}} = 10 \times 838859/2^{20} = 10 \times 838859/1048576 = 7.99998$ V. The result must be quoted to this precision.

The resolution is 1 in 2^{20} = 1 in 1048576 ≈ 1 in 10^6. The corresponding voltage is LSB = 10 μV.

The reference voltage must not change by more than 1 part in 2^{20} over a range of 25°C, so the coefficient must not exceed $1/(2^{20} \times 25) = 4 \times 10^{-8}/°\mathrm{C} = 0.04\,\mathrm{ppm}/°\mathrm{C}$. It will not be easy to meet this specification!

4.1 The frequency of sampling must be at least the Nyquist frequency f_N for the sampled sequence to be a faithful representation of the continuous input. In other words, it is possible to reconstruct the input from the sequence. Higher frequencies are *aliased*, which means that they give the same sequence of samples as a lower frequency. The formula is $f_N = 2f_{max}$, where f_{max} is the maximum frequency in the input.

4.2 The signal must be sampled at double its maximum frequency or more, so the minimum rate of sampling is 100 kilosamples per second. Aliasing would occur if a lower rate were used: high-frequency signals would appear to have a lower frequency after sampling.

6.1 First imagine connecting the noninverting terminal of the opamp to ground instead of V_{bias}. This leaves a standard inverting amplifier, which gives the second term on the right-hand side of equation (6.3).

Next, restore V_{bias} but connect the input of the circuit to ground instead of V_{in}. This leaves a standard noninverting amplifier with V_{bias} as input, which gives the first term on the right-hand side of equation (6.3).

6.2 Well, something to provide a signal, of course. . . . More seriously, the intended answer is a low-pass filter to remove noise and prevent aliasing. An amplifier may also be needed.

6.3 An integrating or sigma–delta ADC should average over a whole number of cycles to reduce interference. This means sampling at 400 Hz or submultiples, 200, 133, 100 Hz and so on. None of this would help with a SAR ADC, which samples its input over a short interval. On the other hand, it would be possible to average the outputs from a SAR over a period of the power supply in software. This achieves the same effect.

6.4 Differential inputs, high input impedance on both inputs, 'high' gain for differential inputs, programmed by a single resistor, and low gain for common-mode inputs. It may also shift the output voltage by an offset.

6.5 I hope that you remember how to do analyse this sort of circuit! Here are the usual three steps.

1. Work out the voltage on the non-inverting input, v_+. This is set by the potential divider so $v_+ = V_{CC} \times R_3/(R_3 + R_4) = (5/11)V_{CC}$. We are assuming that no current flows into the non-inverting input because an ideal op-amp has infinite input resistance.

2. An ideal op-amp has infinite gain and does whatever is necessary with negative feedback to bring its inputs to the same potential, so $v_- = v_+$.

3. We now know v_- and can apply nodal analysis at the inverting input, again assuming that no current flows into the op-amp. We also assume that the op-amp has zero output resistance, which means that the load has no effect on the output voltage V_{out}. Here the load is just the feedback circuit.

The last step is, as usual, the only one to require any work. Nodal analysis gives

$$\frac{V_{in} - v_-}{R_1} + \frac{V_{out} - v_-}{R_2} + 0 = 0. \tag{15.6}$$

I've added a 0 as a reminder that no current flows into the op-amp. Rearranging this gives

$$\begin{aligned} V_{\text{out}} &= -\frac{R_2}{R_1}\left(V_{\text{in}} - \frac{R_1+R_2}{R_2}v_-\right) \\ &= -10(V_{\text{in}} - \tfrac{11}{10}v_-) \\ &= -10(V_{\text{in}} - \tfrac{1}{2}V_{\text{CC}}). \end{aligned} \tag{15.7}$$

The output must line between ground and V_{CC}, assuming true rail-to-rail output. This determines the range of inputs. Clearly $V_{\text{in}} = \frac{1}{2}V_{\text{CC}} = 1.5\,\text{V}$ gives $V_{\text{out}} = 0$. The other limit is $V_{\text{out}} = V_{\text{CC}}$, which needs $V_{\text{in}} = 0.4\,V_{\text{CC}} = 1.2\,\text{V}$. Thus the input must lie between 1.2 and 1.5 V for correct operation. There would be no point in specifying rail-to-rail inputs.

A simple inverting amplifier could not be used because a positive input would lead to a negative output, which cannot be provided with a single supply. An inverting amplifier must have a shift to work in a single-supply system.

6.6 I'll choose an inverting amplifier to match the text. The circuit will be the same as figure 6.7 on page 45 except for the values of the resistors.

1. The range of input voltage is 0.1 V and the range of output voltage is 3.0 V so the gain is −30. This is given by $-R_2/R_1$ in the usual way so we need $R_2 = 30\,R_1$. A suitable choice using standard values is $R_1 = 3.3\,\text{k}\Omega$ and $R_2 = 100\,\text{k}\Omega$.

2. Substitute a pair of values for V_{in} and V_{out} into equation (6.3) or (6.4) to find V_{bias}. Choose $V_{\text{in}} = 0.3\,\text{V}$ and $V_{\text{out}} = 0\,\text{V}$ in equation (6.3), which gives

$$0 = 31\,V_{\text{bias}} - 30 \times 0.3\,\text{V}. \tag{15.8}$$

   Thus $V_{\text{bias}} = (9/31)\,\text{V}$.

3. The potential divider formula for R_3 and R_4 gives

$$V_{\text{bias}} = \frac{R_4}{R_3+R_4}V_{\text{CC}} = \frac{9}{31}\,\text{V}. \tag{15.9}$$

   This can be rearranged to

$$\frac{R_3}{R_4} = \frac{28}{3}. \tag{15.10}$$

   It is not possible to find suitable numbers within the most common set of values (E12, described in section 15.4). The problem is that the ratio of successive values in E12 is about 1.2 but $3.0/2.8 \approx 1.07$, which inevitably lies between values. You could make a suitable value for R_3 by putting two values in series. Alternatively, try the E24 series, which offers twice as many values. I found that $R_4 = 16\,\text{k}\Omega$ and $R_3 = 150\,\text{k}\Omega$ gave the ratio very closely.

7.1 The fractional change in voltage must be smaller than $1/2^n$ for an n-bit converter to avoid errors of a single bit or more. For an 8-bit converter this needs a change of less than $1/256$ so the range of temperature ΔT must obey $5 \times 10^{-5}\Delta T < 1/256$ so $\Delta T < 78°\text{C}$. The range is reduced by $2^4 = 16$ to 2°C for a 12-bit converter.

7.2 This is the reverse of the previous question. The temperature coefficient C must obey $C\Delta T < 1/2^n$ or $C < 1/(2^n \Delta T)$. For 8 bits this needs $C < 1/(256 \times 70) = 5 \times 10^{-5}/°\mathrm{C} = 5 \times 10^{-3}\%/°\mathrm{C} = 50\,\mathrm{ppm}/°\mathrm{C}$. I have rounded the coefficient *down* to one significant figure. The other numbers follow in the same way.

7.3 Call the minimum and maximum resistance of the thermistor R_{min} and R_{max}. Then the range of voltage from the potential divider is

$$\Delta V = V_{\mathrm{CC}} \left(\frac{R_{\mathrm{max}}}{R_{\mathrm{ref}} + R_{\mathrm{max}}} - \frac{R_{\mathrm{min}}}{R_{\mathrm{ref}} + R_{\mathrm{min}}} \right) \tag{15.11}$$

$$= V_{\mathrm{CC}} \frac{R_{\mathrm{ref}}(R_{\mathrm{max}} - R_{\mathrm{min}})}{(R_{\mathrm{ref}} + R_{\mathrm{max}})(R_{\mathrm{ref}} + R_{\mathrm{min}})} \tag{15.12}$$

(a couple of terms cancelled from the numerator). Now we have to find the maximum as a function of R_{ref}, which requires differentiation and setting the result to zero. Clearly the result is going to be a mess! Fortunately we need only to know when the derivative is zero so we can forget about its denominator. In other words, we want the derivative of f/g and the usual rule gives $(f'g - fg')/g^2$, but we just need to find the value of R_{ref} that gives $f'g - fg' = 0$. Here goes.

$$(R_{\mathrm{max}} - R_{\mathrm{min}})(R_{\mathrm{ref}} + R_{\mathrm{max}})(R_{\mathrm{ref}} + R_{\mathrm{min}}) - R_{\mathrm{ref}}(R_{\mathrm{max}} - R_{\mathrm{min}})(2R_{\mathrm{ref}} + R_{\mathrm{max}} + R_{\mathrm{min}}) = 0. \tag{15.13}$$

I multiplied out the denominator to find its derivative. We can cancel the factor of $(R_{\mathrm{max}} - R_{\mathrm{min}})$ and multiply out the remaining brackets. Most terms cancel to leave

$$R_{\mathrm{ref}}^2 - R_{\mathrm{max}} R_{\mathrm{min}} = 0 \tag{15.14}$$

so

$$R_{\mathrm{ref}} = \sqrt{R_{\mathrm{max}} R_{\mathrm{min}}}. \tag{15.15}$$

The numbers here give $R_{\mathrm{ref}} = 14\,k\Omega$ to the nearest kilohm; 15 kΩ is the closest standard value.

With 15 kΩ and a 3 V supply the voltages from the potential divider go from 1.20 V to 1.71 V. The range is now 0.51 V compared with 0.36 V before. It's a significant improvement but an amplifer would do far better.

7.4 In 8-bit mode LSB $= (3.3\,\mathrm{V})/256 \approx 13\,\mathrm{mV}$ at the input to the ADC. This is 1.3 mV at the input to the amplifier, which is the output of the sensor, and therefore 0.13°C.

7.5 The answer to a question like this is inevitably No but it needs justification! Suppose that we designed the amplifier so that LSB corresponds to 0.1°C. The output from the LM35 is 1 mV and LSB $= (3.3\,\mathrm{V})/1024 = (3300\,\mathrm{mV})/1024$ so the gain should be $3300/1024 = 3.22$.

7.6 The trivial answer would be 'No problem, the amplifier now gives 0.0–3.0 V for this range of temperatures'. The difficulty is that even an amplifier with so-called rail-to-rail output cannot give an output of 0 V, so the system would fail at the lowest temperatures.

This is an ideal application for the LM7705 or a similar negative bias generator, which would enable the output of the noninverting amplifier to go down to zero.

Alternatively, add an offset voltage to the amplifier, which would have to be changed to an inverting configuration as in figure 6.5 on page 43. Choose a gain of -10 instead of $+10$ for convenience. The inverting amplifier could then map $0\,\text{V} \rightarrow 3.2\,\text{V}$ and $0.3\,\text{V} \rightarrow 0.2\,\text{V}$. Equation (6.3) shows that this needs $V_{\text{bias}} = (3.2/11) \approx 0.29\,\text{V}$.

7.7 With a direct connection we need 1 mV resolution in 2.5 V, which needs 2500 values. This is a bit too high for $2^{11} = 2048$ so 12 bits are needed. The resolution is $\text{LSB} = (2.5\,\text{V})/2^{12} = 0.6\,\text{mV}$.

Only 500 values are needed if the amplifier boosts the signal to the full range of the ADC. This number is just the range of inputs (0.5–1.0 V) divided by the resolution. This is conveniently just below $512 = 2^9$ so a 9-bit converter would be sufficient (if such a device exists). The actual resolution would be $(500\,\text{mV})/2^9 = 0.98\,\text{mV}$.

The amplifier needs to provide both gain and offset. To provide perfect matching it would need a gain of $2.5/0.5 = 5$ and an offset to bring the output to 0.0–2.5 V. If an inverting amplifier is used, equation (6.3) on page 44 shows that $V_{\text{bias}} = 2.5/6 \approx 0.42\,\text{V}$ is needed. However, problems may arise if the system works from a single supply because the output of the op-amp would be unable to go all the way down to 0 V. See section 7.5 on page 58 for how to solve this problem.

7.8 The potential divider gives

- Light: 0.73 V and 1.2 kΩ
- Dark: 3.12 V and 5.3 kΩ (not far from V_{CC} and 5.6 kΩ)

The worst case is in the dark, where the total resistance is $5.3\,\text{k}\Omega + 7.5\,\text{k}\Omega \approx 13\,\text{k}\Omega$ and $\tau = 13\,\text{k}\Omega \times 15\,\text{pF} \approx 0.2\,\mu\text{s}$. The usual calculation shows that the number of time constants should be greater than $13 \ln 2 \approx 9$, giving 1.8 µs. This is 234 cycles of a 13 MHz clock.

A comparator would be sufficient if the system had only to distinguish between light and dark.

7.9 Taking the hint in the question allows the circuit to be simplified as in figure 15.17. Nodal analysis at the central junction gives

$$\frac{V_{\text{in}} - V_{\text{CC}}}{R_{\text{ref}}} + \frac{V_{\text{in}} - 0}{R_{\text{th}}} + \frac{V_{\text{in}} - V_{\text{bias}}}{R_1} = 0. \tag{15.16}$$

Multiply out the fractions and collect terms. It's a bit of a mess but straightforward. The result is

$$V_{\text{in}} = \frac{R_1 R_{\text{th}} V_{\text{CC}} + R_{\text{ref}} R_{\text{th}} V_{\text{bias}}}{R_1 R_{\text{th}} + R_{\text{th}} R_{\text{ref}} + R_{\text{ref}} R_1}. \tag{15.17}$$

In fact this is not the best route. It is clearer to replace the potential divider formed by R_{th} and R_{ref} with its Thévenin equivalent. This gives the alternative expression

$$V_{\text{in}} = \frac{R_{\text{th}} V_{\text{CC}}}{R_{\text{ref}} + R_{\text{th}}} \frac{R_1}{R_1 + R_{\text{s}}} + \frac{R_{\text{s}} V_{\text{bias}}}{R_1 + R_{\text{s}}} \tag{15.18}$$

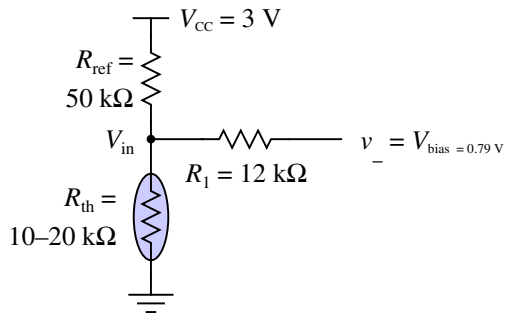


Figure 15.17 Analysis of thermistor and input resistance of inverting amplifier.

where $R_s = R_{ref} \parallel R_{th}$ is the Thévenin resistance of the divider. The first fraction in the first term is the potential from the unloaded divider and the second fraction shows how this is reduced by the input resistance of the amplifier, R_1. The second term arises from the bias.

It's probably best to substitute numbers now. For $R_{th} = 20\,\text{k}\Omega$ we found $V_{in} = 0.86\,\text{V}$ before, which falls to 0.82 V including R_1. That's about a 5% error. At the other end of the range, where $R_{th} = 10\,\text{k}\Omega$, V_{in} rises from 0.50 V to 0.62 V. This is an error of nearly 25%. Several solutions are possible. We could use a non-inverting amplifier with offset to raise the input resistance. Alternatively, we could stick with the circuit in figure 7.5, calculate the error and compensate for it in software.

8.1 The general formula is $V_{out} = V_{ref} \times d/256$ where d is the digital input. The maximum output is given by $d = 255$, not 256! It is 4.98 V.

8.2 Same as before.

8.3 The main point is that 5 V is 500 intervals of 10 mV so the period of the PWM signal should be 500. It can then be varied to give the desired resolution. This would be no problem with a a 16-bit timer because $500 < 2^9$.

8.4 A 12-bit simple string needs $2^12 = 4096$ resistors. They all have the same value, unlike the resistors in a flash ADC. If it were segmented, each string would need $2^6 = 64$ resistors giving a total of 128. Of course this would also need buffers, switches and logic to drive the interpolator.

8.5 The current through each resistor is $2.5\,\text{V}/100\,\text{k}\Omega = 25\,\mu\text{A}$ and there are $2^8 - 1 = 255$ of them so the reference must provide about 6.4 mA, which is rather high by modern standards – this is not a low-power device!

The current is 'multiplied' by the 1 kΩ feedback resistor to give 25 mV for each 25 µA. Thus the outputs are *negative* 0, 25 mV, 50 mV, ... up to a maximum of 6.375 V from all 255 sources. Remember that, like all DACs, it cannot produce 256 currents and 6.4 V!

8.6 The most significant 4 bits need 15 current sources and therefore 15 resistors. The R–$2R$ ladder has an extra section compared with that shown in the lecture notes and therefore requires

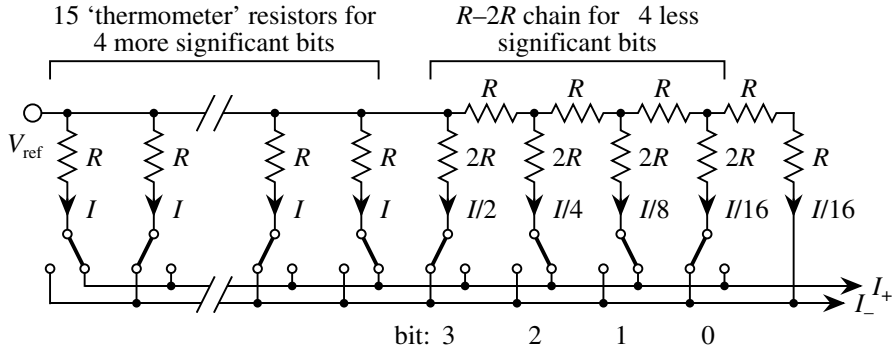


Figure 15.18 An 8-bit segmented DAC, where the 4 more significant bits use thermometer currents and the 4 less significant currents are generated by an R–$2R$ chain.

four resistors of $2R$ and five of R, giving 13 of R in total (each $2R = R + R$). Thus the overall total is 28 resistors. The part of the DAC that generates the current is shown in figure 15.18.

Solutions to examples on power supplies and passive components

9.1 There is 8.1 V across the 50 Ω load so the current is $I = V/R = 8.1\,\text{V}/50\,\Omega = 162\,\text{mA}$. The regulation is $(9.0 - 8.1)/9.0 = 0.9/9.0 = 10\%$. The lost voltage across the internal resistance is 0.9 V at 162 mA so the resistance is 5.6 Ω.

With a 10 Ω load the total resistance including R_s is 15.6 Ω so the current is $9.0/15.6 = 0.58$ A and the terminal voltage is 5.8 V. Alternatively, treat the load and internal resistances as a potential divider.

9.2 (i) $V_o = 9 \times 100/102 = 8.82\,V$; $Reg = 100 \times (9 - 8.82)/9 = 2.0\%$. (ii) $V_o = 9 \times 50/52 = 8.65\,V$; $Reg = 100 \times (9 - 8.65)/9 = 3.9\%$. (iii) $V_o = 9 \times 10/12 = 7.50\,V$; $Reg = 100 \times (9 - 7.50)/9 = 16.7\%$. These answers may have an excessive number of 'significant' figures.

9.3 $\text{Reg} = (V_{oc} - V_L)/V_{oc} = 1 - V_L/V_{oc}$. But $V_L = V_{oc} \times R_L/(R_L + R_S)$. Thus $\text{Reg} = 1 - R_L/(R_L + R_S) = (R_L + R_S - R_L)/(R_L + R_S) = R_S/(R_L + R_S)$.

9.4 Current of 300 A through 0.01 Ω gives a drop of 3 V, so terminal voltage is 9 V.

If the internal resistance rose to 0.1 Ω, the short-circuit current would be $12\,\text{V}/0.1\,\Omega = 12$ A so it would be impossible to supply 300 A. You would have to walk to university instead (well wrapped up).

9.5 The voltage drop across the source resistance is $6.00 - 5.62 = 0.38$ V. This is caused by a current of 50 mA, so the resistance is $0.38/0.050 = 7.6\,\Omega$.

10.1 It is best to convert everything to SI units, so $1.8\,\text{kW h} = 1.3 \times 10^3 \times 60 \times 60 = 6.5\,\text{MJ}$. Petrol (gasoline) has an energy density of about $45\,\text{MJ kg}^{-1}$ or $35\,\text{MJ L}^{-1}$ according to wikipedia and The Physics Factbook. It seems that we need only about 0.2 L petrol to provide the same amount of energy as a fully-charged Prius battery!

This isn't fair because the electrical system driving a motor is probably over 90% efficient while the internal combustion engine has a much lower efficiency. It's thermal efficiency is somewhere around 35% but a lot of the power is wasted in friction and driving accessories, so my guess is 20% overall (I must ask a mechanical engineer). A fairer figure for the fuel is therefore around 1.0 L. It's still a *much* smaller volume and mass than the battery.

10.2 (a) Total load is 24 W so the load current is 2 A. Life is $50/2 = 25$ h.

(b) Load is now $24 + 120 = 144$ W, current is 12 A, life is $50/12 = 4.2$ h.

(c) Load is 244 W, current is 20.3 A, life is 2.5 h.

10.3 The capacity of a CR2032 is 240 mA h according to the Energizer data sheet. The average current must not exceed $0.24/(365 \times 24)$ A $= 27\,\mu$A. The cell produces 3.0 V when it is new so the resistance must be at least 110 kΩ. It is interesting that the data sheet uses a load of about 15 kΩ.

10.4 The average current is

$$I_{\text{ave}} = \frac{40\,\text{mA} \times 10\,\text{ms} + 0.1\,\text{mA} \times (60 - 0.01)\,\text{s}}{60\,\text{s}} \tag{15.19}$$

$$= \frac{400 + 6000}{60}\mu\text{A} \tag{15.20}$$

$$= 7 + 100\,\mu\text{A} = 107\,\mu\text{A} \tag{15.21}$$

This shows that the current during the low-power mode dominates. The full current is much higher but is drawn for so short a time that it is less significant.

The lifetime is roughly $500\,\text{mA h}/0.1\,\text{mA} = 5000\,\text{h} \approx 200$ days.

10.5 Capacitor charge $Q = CV$, and $Q = IT$ where constant current I flows for time T. Initial $Q = 1\,\text{F} \times 4.5\,\text{V} = 4.5\,\text{C}$. The final charge at 3.3 V is 3.3 C. Therefore the current can take 1.2 C, which means a current of 10 µA for $1.2/10^{-5} = 1.2 \times 10^5$ s, 2000 minutes or 30 hours.

10.6 During the current pulses the resistance causes 'lost volts' in the same way as the internal resistance of a battery. The voltage dropped is $\Delta V = R_{\text{cap}} I_{\text{load}} = (3\,\Omega) \times (50\,\text{mA}) = 150\,\text{mV}$. This could be significant because the extra drop could cause the voltage on the load to fall below the limit of 2.4 V. The acceptable voltage drop as the capacitor discharges is reduced from 0.40 V to 0.25 V. This in turn raises the capacitance required from 125 µF to 200 µF. We are lucky because we had to choose a 220 µF capacitor so R_{cap} will have no impact but it could have been more serious.

While recharging, $R_{\text{cap}} = 3\,\Omega$ is in series with $R_{\text{batt}} = 30\,\Omega$. This is a change of only 10% and we found a huge margin for the recharging time so there is no impact.

10.7 See figure 15.19. The voltage falls linearly while the capacitor provides the current to the load, then recovers exponentially according to the usual RC charging curve.

11.1 The ratio of turns is the same as the ratio of voltages. The output voltage is given as a *peak* value but the input is *rms* so convert the input to peak as well: $230\sqrt{2} = 325$ V. The ratio is $325/5.5 = 59$ so the number of turns on the secondary is $600/59 = 10$ to the nearest turn.

11.2 Here's a suitable table.

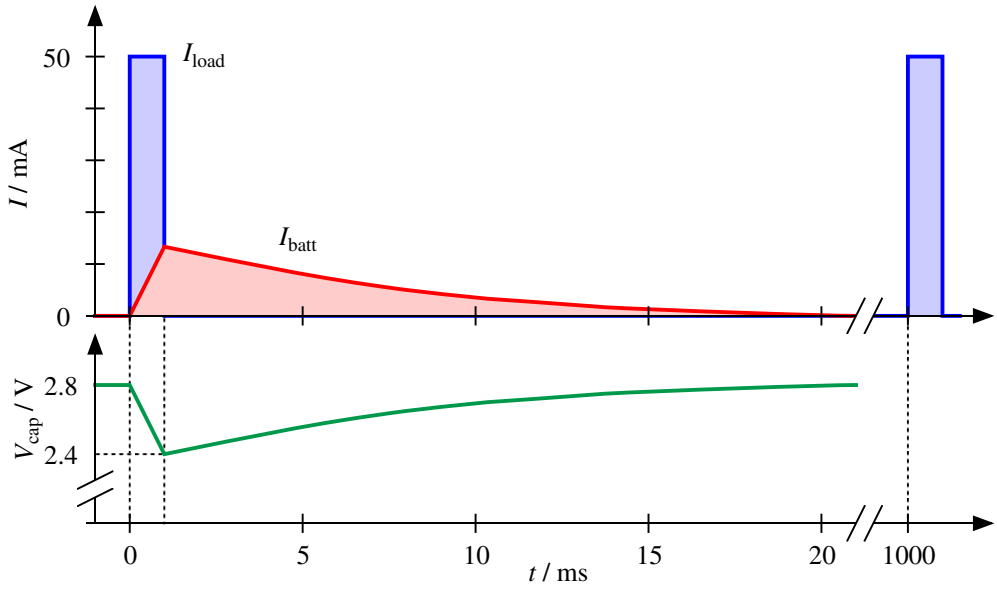


Figure 15.19 Current through the load, current drawn from coin cell and voltage across capacitor as a function of time.

	half-wave	full-wave	bridge
windings needed	one	two (centre-tapped)	one
utlilization	half	half (one winding at a time)	full
diodes needed	1	2	4
reverse rating	V_{peak}	$2V_{peak}$	V_{peak}
voltage drop	0.7 V	0.7 V	1.4 V
frequency of ripple	f_{input}	$2f_{input}$	$2f_{input}$
ease of smoothing	difficult (empty half cycles)	easier	easier

11.3 This is a factor of 5 reduction, so the capacitor needs to be 5 times as large or 6 mF.

11.4 A pessimistic approach is to assume that the capacitor must supply 200 mA for the full period of the ripple. This is half the period of the input or 10 ms in Europe. The charge required is $0.2 \times 0.01 = 0.002$ C. The change in voltage must not exceed 0.5 V and $Q = CV$ so $C = Q/V = 0.002/0.5 = 0.004$ F or 4000 µF.

It would be less pessimistic to assume that the capacitor supplies the current for only 60% of the time, which reduces C by the same factor to 2400 µF; the nearest conventional values are 2200 µF and 3300 µF.

The frequency of the mains (line) supply is 60 Hz rather than 50 Hz in North America, which reduces the period of each ripple and therefore the capacitance needed. The 'optimistic' estimate falls from 2400 µF to 2000 µF. Still big!

11.5 The 'AC mains in Europe' implies 230 VAC at 50 Hz. Calculate the smoothing capacitor needed first. This needs no effort at all because the current and ripple are the same as in the previous question. The average voltage has no effect on the value of capacitance; this depends only on the average current and tolerable ripple voltage. Thus $C = 2400\,\mu\text{F}$. For completeness we should also specify the voltage and ripple current ratings. The obvious peak voltage is the average output of 9 V plus half the ripple of 0.5 V (remember that this is a peak-to-peak value), giving 9.25 V. This is much too low, as we shall see shortly. The peak current into the capacitor every time it recharges is about 3 times the average current to the load or 600 mA and the ripple current rating must not be lower than this.

Now for the voltage required from the transformer. The peak value of output voltage is 9.25 V. Adding two diode drops from the bridge gives 10.8 V. The peak current from the transformer is around 600 mA, about 3 times the average output current, and the resistance of the transformer was given as 10 Ω, which yields a voltage drop of 6 V (*very* roughly). Adding this gives a peak voltage of 17 V to the nearest volt. Divide by $\sqrt{2}$ to find an rms voltage of 12 V. It is depressing that a 12 V transformer is needed for 9 V output.

Work forward from the input to find the no-load output voltage. The transformer provides 17 V peak without any drop in its winding because no current flows. The diodes reduce this to about 15.5 V. The smoothing capacitor will charge fully to the peak voltage and not discharge between cycles because no current is drawn from it. The no-load output voltage is therefore also 15.5 V. In fact it might well be higher because the diodes will drop a smaller voltage when so little current flows; the output could rise much closer to the peak voltage from the transformer of 17 V. The smoothing capacitor must therefore work safely at this voltage, much higher than the nominal 9 V output.

The regulation is clearly terrible, although not unrealistic for a cheap 'wall wart'. Substituting into equation (9.1), which uses the no-load and full-load voltages of 15.5 V and 9.0 V, gives about 40%. Alternatively, the Thévenin resistance from equation (9.2) is $R_s = 32.5\,\Omega$. This seems a large resistance: where does it come from?

First, the imperfect smoothing by the capacitor causes the average voltage to drop when more current is drawn. This is just half of the (peak-to-peak) ripple voltage. The ripple is zero when no current flows and 0.5 V at 200 mA so the average voltage drops by 0.25 V at 200 mA, which is equivalent to a resistance of 1.25 Ω. The other obvious resistance is 10 Ω from the winding of the transformer. Unfortunately this is multiplied because of the peaked nature of the current, which rises to 600 mA rather than the average of 200 mA. It therefore gives a much higher voltage drop then might be expected, and the factor of 3 in current multiplies its effective resistance to 30 Ω. The sum of these is 31.25 Ω, which differs from 32.5 Ωbecause of rounding – the final estimate of the voltage required from the transformer can be trusted only to the nearest volt (if that). The important finding is that the transformer dominates the regulation and should be replaced by a superior component, with a lower resistance, if the performance of the power supply needs to be improved.

12.1 The total current through the resistor is 60 mA. The voltage drop across the resistor is $10 - 4.7 = 5.3$ V. Thus the resistance must be $5.3/0.06 = 88\,\Omega$. (The closest standard value is 86 Ω.) A current of 10 mA passes through the diode, which drops 4.7 V, so the power is $4.7 \times 0.01 = 0.047$ W.

If the load is removed, all the 60 mA from the supply flows through the Zener diode. Its

power dissipation rises to about 300 mW and the diode should be rated to withstand this.

12.2 If the load current changes to 10 mA, then the other 40 mA must flow through the diode. Thus the diode current is 50 mA and the voltage drop stays at 4.7 V. The power dissipation in the diode increases.

The total current flowing in the original design was 60 mA so if the load current increases to 70 mA, the extra 20 mA *cannot* be taken from the current through the diode. The voltage drop across the resistor increases to $0.07 \times 88 = 6.2$ V, so the voltage across the diode is 3.8 V and no current will flow through the diode. The voltage is no longer controlled by the diode and regulation fails.

12.3 Replace R_1 and R_2 with a potentiometer.

12.4 Maximum power is dissipated when the current and headroom voltage take their maximum values. The maximum headroom voltage implies the minimum output voltage of 3 V, when the headroom is $25 - 3 = 22$ V. The maximum current is 1 A, giving a power dissipation of 22 W.

12.5 The input voltage to the regulators must be greater than 15 V to provide some headroom for regulation. Assume a value of 18 V. The maximum dissipation is when the output voltage is very low, say zero. Then the dissipation is $18\,\text{V} \times 0.2\,\text{A} = 3.6\,\text{W}$ for each channel. This must be doubled for a bipolar supply, giving 7.2 W total.

The regulators are linear to avoid noise in their output, which would confuse your experiments!

12.6 An efficiency of 75% means that $V_{\text{rms,in}} = V_{\text{rms,out}}/0.75 = 6.7$ V and the peak value is $6.7\sqrt{2} = 9.4$ V. Unfortunately the regulator needs 7.0 V, the diodes drop 1.6 V and the transformer drops 2.1 V. These add up to 10.7 V so it can't be done.

For an efficiency of 50% the input rms voltage is twice that of the output or 10 V. The peak value is therefore 14.1 V. This is safely above 10.7 V and the design is therefore possible. The ripple must be reduced to $14.1 - 10.7 = 3.4$ V. Equation 11.5 on page 95 shows that this needs $C > 180\,\mu\text{F}$, which would be rounded up to the nearest standard value of 220 µF.

The calculation should really take better account of the peaked nature of the current drawn from the AC supply but this is difficult other than by simulation.

13.1 (a) It must have a capacitance of at least 100 µF but there is no upper limit (page 4). Its equivalent series resistance (ESR) must lie between 0.03–1 Ω (page 11).

(b) 6–26 V (page 6).

(c) Nothing – it has reverse polarity protection (page 1 and 6).

(d) The range is specified as 0.3 V typical – 0.6 V limit (page 6).

(e) There is a plot on page 8. Regulation fails when the input is below about 5.3 V and the output voltage falls as the input voltage is reduced further.

(f) Trickier than you might expect! The front page suggests at least 100 mA. The plot on page 9 shows over 400 mA at room temperature but see the next part of this question.

(g) Taking the input to be 26 V and the output to be 5 V, the regulator drops 21 V at a current of 0.1 A so the power is 2.1 W. The maximum power dissipation for a TO-92 at room temperature is about 0.6 W so this would not be acceptable. The device would overheat and shut down (safely)!

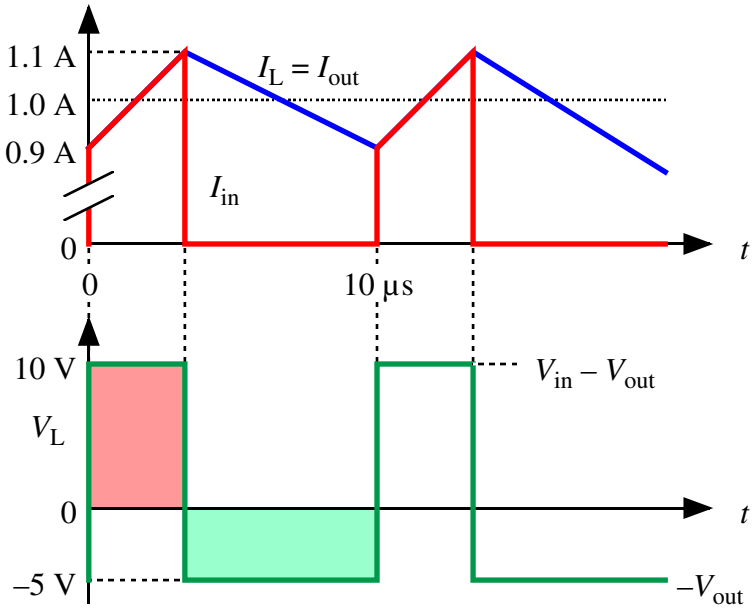
(h) See the Definition of Terms on page 13.

13.2 The worst-case dropout voltage is 0.6 V at 100 mA so the input must exceed 5.6 V. The worst-case quiescent current is shown as 30 mA, which seems astonishingly high, and implies that at least 130 mA must be available. The load regulation is given as 50 mV for currents up to 100 mA so the output voltage may increase by 50 mV if the 100 mA load is removed.

14.1 It's a voltage doubler: $V_{out} = 2V_{in}$. The charged capacitor C_1 is connected in series with the input when it supplies the output.

14.2 (a) The input and output voltages are related by $V_{out} = DV_{in}$ so $D = \frac{1}{3}$.

(b) The average output current is 1 A and fluctuations of ±10% mean that the peak is 1.1 A and the trough is 0.9 A. The input current is equal to the output current during the 'store' phase and zero during the 'release' phase. The voltage across the inductor is $V_{in} - V_{out} = 10$ V during the store and $-V_{out} = -5$ V during the release. The period of each cycle is 10 µs for a 100 kHz frequency. This gives all the numbers needed to plot the waveforms.



(c) The input current flows for only $\frac{1}{3}$ of the time and is therefore $\frac{1}{3}$ A.

(d) In the store phase, i_L increases from 0.9 A to 1.1 A in $\frac{1}{3}\times 1-$ µs, so $\mathrm{d}I/\mathrm{d}t = 6\times 10^4$ A s^{-1}. The voltage $v_L = 10$ V so $L = v_L/(\mathrm{d}i/\mathrm{d}t) = 1.7 \times 10^{-4}$ H = 170 µH.

(e) If the switching supply is 95% efficient, it dissipates about 5% of the power that it supplies. This is $5\,\text{V} \times 1\,\text{A} = 5\,\text{W}$ so the dissipated power is about 0.25 W.

In contrast, the linear regulator would dissipate power given by the full current and the dropped voltage or $10\,\text{V} \times 1\,\text{A} = 10\,\text{W}$. This is twice the useful power supplied and 40 times the dissipation of the switching regulator! The efficiency is only 33%.

14.3 The relation is $V_{\text{out}} = V_{\text{in}}/(1 - D)$ so $V_{\text{out}} = 2V_{\text{in}} = 10\,\text{V}$.
With $D = 0.95$, $V_{\text{out}} = 20V_{\text{in}} = 100\,\text{V}$ if the simplifications hold.

14.4 This time the relation is $V_{\text{out}} = -V_{\text{in}}D/(1 - D)$ so $D = -V_{\text{out}}/(V_{\text{in}} - V_{\text{out}})$. The values of D are therefore $\frac{1}{3}$, $\frac{1}{2}$ and $\frac{2}{3}$.

14.5 (a) The hi-fi amplifier would probably use a transformer, rectifier and linear regulator – all big, heavy and expensive. However, many modern amplifiers use 'Class D', which is pulse width modulation, in which case a switching supply is acceptable. The quality is probably not hi-fi though.

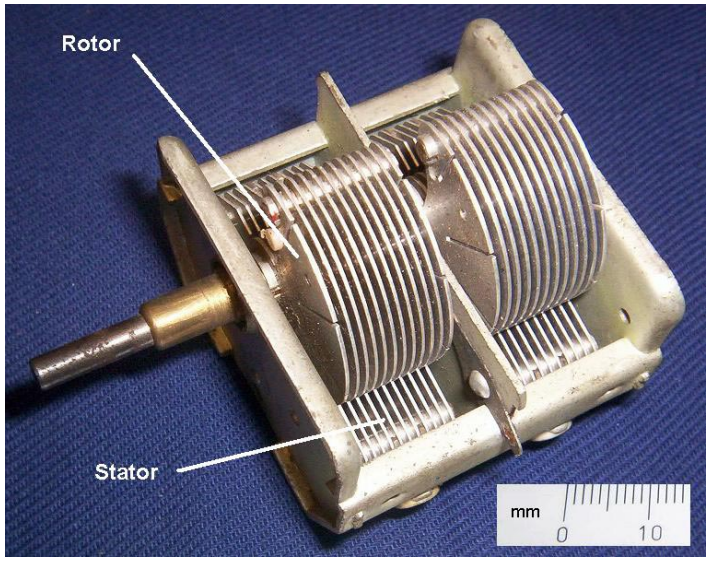
(b) The power supply for a mains-powered system must provide isolation so a flyback switcher is an obvious choice. However, the two voltages are a problem. It is possible to put two windings on the transformer but only one output can be regulated by the switcher itself. Another possibility would be to reduce 5.0 V to 3.3 V with a step-down (buck) switcher.

(c) Each LED needs about 1.8 V, which is a nuisance because there is not quite enough voltage for two in series but it is wasteful to put them in parallel and drop 1.5 V. One possibility is to put all in series, which requires 10.8 V, and use a step-up (boost) switcher. Another possibility is to make two chains of three LEDs, each requiring 5.4 V, and use a switched capacitor voltage doubler.

(d) Two rechargable NiMH cells in series produce 2.4 V over the whole of their useful life so no regulator would be needed if the user were restricted to these. Unfortunately the specification admits primary cells as well and the voltage from two alkaline cells in series drops from about 3.1 V when new to 1.8 or even 1.6 V at the end of their useful life, depending on how this is defined. We therefore need a regulator that can step up or down and an inverting (buck/boost) switcher could be used.

15.1 The power is given by $P = V^2/R$ so $V < \sqrt{PR} = \sqrt{\frac{1}{4} \times 120} = \sqrt{30} = 5.5\,\text{V}$. You are safe in a 5 V circuit but only just, which might come as a surprise!
Similarly, the current is limited by $P = RI^2$ so $I < \sqrt{P/R} = 46\,\text{mA}$, which is not large.

15.2 Capacitance is given by $C = \epsilon_0 A/d$ so $A = Cd/\epsilon_0 = (300 \times 10^{-12}) \times (0.5 \times 10^{-3})/(8.85 \times 10^{12}) = 0.017\,\text{m}^2$. Wow – pretty big!
Practical variable capacitors had about 10 pairs of interleaved plates, which effectively gave 20 capacitors in parallel and reduced the area of each to about $0.001\,\text{m}^2 = (30\,\text{mm})^2$. See the image below from wikipedia. Clearly it would not be possible to build radios that fit inside mobile phones using this technology!



15.3 We need to capacitor to charge between pulses of current drawn by the IC, which occur every 0.1 µs for a 10 MHz frequency. The time taken for a capacitor to charge is given roughly by R_sC, where R_s is the resistance in series with the capacitor. Here we need $R_sC < 0.1\,\mu s$ so $R_s < 0.1\,\mu s/0.1\,\mu F = 1\,\Omega$. This shouldn't be a problem.

At 1 GHz the time has gone down by a factor of 100 and so does the limit on the resistance: we need $R_s < 0.01\,\Omega$. This might be difficult because it includes the equivalent series resistance of the capacitor itself, the track on the PCB, lead of the IC and joints. High-speed design is difficult, even for decoupling capacitors!

15.4 The inductance is given approximately by $L = \mu_0\mu_r N^2 A/l$ from the lecture notes. Here this gives $(4\pi \times 10^{-7}) \times 1000 \times 25^2 \times (\pi \times 0.0025^2)/(0.025) = 0.6\,\text{mH}$.

The resistance is given by $R = \rho l/A$ as in the previous question where l is now the length of the wire rather than the inductor. Thus $R \approx (2\times10^{-8})\times25\times(\pi\times0.005)/(\pi\times0.00025^2) = 0.04\,\Omega$.

15.5 For n steps in a decade (factor of 10) the ratio r must obey $r^n = 10$, so $r = 10^{1/n} = 1.259$ for $n = 10$. The successive values starting with ten to two significant figures are 10, 13, 16, 20, 25, 32, 40, 50, 63 and 79. The next step gives 100 to start the next decade.

15.6 $T_j = T_a + P(\theta_{jc} + \theta_{ca})$ where $T_a = 50°C$ and P = 2 W. Then (i) $T_j = 50 + 2(5 + 50) = 160°C$ and (ii) $T_j = 50 + 2(5 + 10) = 80°C$.

15.7 The total thermal resistance must be less than $(T_j - T_a)/P = (150 - 30)/40 = 3°C/W$. The transistor is derated at 0.5 W/°C and its equivalent thermal resistance between junction and case is the reciprocal of this, 2.0°C/W. The overall thermal resistance $\theta_{ja} = \theta_{jc} + \theta_{ch} + \theta_{ha}$ where 'h' stands for heatsink and θ_{ch} is the thermal resistance of the washer between case and heatsink. Thus $2.0 + 0.5 + \theta_{ha} < 3.0$ and we need $\theta_{ha} < 0.5°C/W$.

The TIP3055 is derated at 0.7 W/°C, equivalent to a thermal resistance of 1.4°C/W. Now we need $1.4 + 0.5 + \theta_{ha} < 3.0$ so $\theta_{ha} < 1.1°C/W$. The transistor is more expensive but the heatsink can be smaller.

15.8 The resistance is given by $R = \rho l / A = [(2 \times 10^{-8}) \times 0.02]/[(5 \times 10^{-4}) \times (35 \times 10^{-6})] = 0.02\,\Omega$. This is low enough for most nets in a circuit but could be a problem for power, which is why wider tracks are used. The resistance would double with half-ounce copper.