12 A Short Introduction to Diodes

v1.4 – June 2021

Prerequisite knowledge required: Ohm and Kirchhoff's Laws, Resistor Networks

12.1 Introduction

Diodes are very useful electronic components. They act like valves: they let current flow in one direction, but not the other (or at least that's what an ideal diode does; real diodes are a little more complicated). Diodes allow circuits to be designed that can determine the amplitude of AC signals, produce DC voltage from AC voltage sources, produce outputs proportional to the logarithm of the input voltage, divide two voltages, and many more.

Designing circuits using diodes requires an understanding of what they do and how they work, and that's what this chapter is about: it introduces diodes and how they behave in circuits, including their non-linear behaviour, and the concepts of dynamic and static resistance. We'll also briefly look at the phenomenon of reverse breakdown in diodes and how this effect is usefully employed in Zener diodes.

(What this chapter doesn't do is look at how diodes work: for more information about this please see the chapter about "Electrons in Solids".)

The circuit symbol for a diode is shown below. Note that unlike the circuit symbol for a capacitor or a resistor it's not reversible; there is a very definite direction implied.

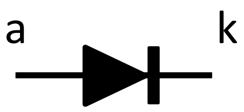

Figure 12.1 Circuit symbol for a normal diode

The two ends of the diode are known as the anode (denoted 'a') and the cathode (denoted by a 'k')[1]. (Note that you don't usually write the 'a' and the 'k' on a circuit diagram since it's obvious from the shape of the triangle which end is which.) The anode is at a higher voltage than the cathode in normal operation when current is flowing in the diode in the forward (intended) direction.

Actual diodes are usually labelled with a black band at the cathode (perhaps reflecting the thick black line at the cathode end in the circuit symbol).

12.2 Ideal diodes

The term *forward current* is used when the current is flowing in the direction of the arrow in the diode symbol (you can think of the symbol as an arrow with a bar across the point). When the current is flowing in this direction the diode is said to be *forward biased* and large currents can flow:

[1] Why 'k'? Most likely explanation is that it derives from the Greek "kathodos" meaning "the way down", as compared to the anode, from the Greek "anodos" meaning "the way up". Current travels from the "way up" to the "way down".

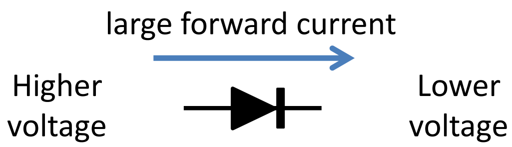

Figure 12.2 A forward-biased diode

The term *reverse biased* is used to refer to a diode in which the voltage at the cathode is greater than the voltage at the anode, for example:

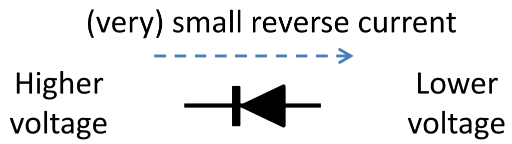

Figure 12.3 A reverse-biased diode

In one sense, an ideal diode would look like a piece of wire (a very low resistance) for current flowing in one direction, and an open circuit (an infinite resistor) for current trying to flow in the other direction. However no real diode behaves anything like this.

The term *ideal diode* is much more usefully applied any device which closely obeys a simple equation[2] that provides a good approximation to the performance of many diodes:

$$I = I_0\left(\exp\left(\frac{eV}{nkT}\right) - 1\right) \tag{12.1}$$

where e is the charge on an electron (1.6 x 10^{-19} C), k is Boltzmann's constant (1.38 x 10^{-23} J/K) and T is the absolute temperature in Kelvins. I_0 is known as the *saturation current* and n is the *ideality factor*; both are function of the type of diode being used. Since e and k are constants, and T is usually approximately equal to room temperature (around 290 K), this formula can be re-written in the slightly-easier-to-remember form of:

$$I = I_0\left(\exp\left(\frac{V}{n \times 25\text{ mV}}\right) - 1\right) \tag{12.2}$$

Sometimes the ideality factor n is assumed to be one as well[3], which leads to the even simpler formula:

$$I = I_0\left(\exp\left(\frac{V}{25\text{ mV}}\right) - 1\right) \tag{12.3}$$

Looking at the values of current for a few voltages illustrates the performance of real diodes.

---

[2] This equation is known as the Shockley equation, after William Shockley, one of the inventors of the transistor.

[3] It's worth noting that this is not always a good assumption.

12.2.1 A typical example: the 1N4148

The 1N4148 is a very common small-signal[4] silicon diode. A typical 1N4148 has an ideality factor[5] of around 2, and a saturation current of around 3 nA (3 x 10^{-9} amps). Putting some numbers into the equation for such a diode might help to give a feel for how real diodes perform:

Firstly, when the voltage across the diode $V = 0$,

$$I = I_0\left(\exp\left(\frac{0}{25\text{ mV}}\right) - 1\right) = I_0\left(\exp(0) - 1\right) = I_0\left(1-1\right) = 0 \qquad (12.4)$$

therefore there is no current flowing. This is expected: it would be very odd for a passive device like a diode to have any current flowing through it if there was no voltage placed across it.

Next, when the diode is heavily reverse-biased, so that the voltage across the diode V is large and negative, ($-V$ / 25mV) becomes a very large negative number, and since:

$$\exp(x) \rightarrow 0 \quad \textit{as} \quad x \rightarrow -\infty \qquad (12.5)$$

the current becomes approximately equal to $-I_0$, since:

$$I = I_0\left(\exp\left(\frac{V}{n\times 25\text{ mV}}\right) - 1\right) \rightarrow I_0\left(\exp(-\infty) - 1\right) = I_0\left(0-1\right) = -I_0 \quad \textit{as}\ V \rightarrow -\infty \qquad (12.6)$$

This is why I_0 is known as the *saturation current*: it's the maximum current that can flow through an ideal diode in the reverse direction[6]. The value specified for the 1N4148 of a few nanoamps is typical for many diodes.

Finally, when the voltage V is around one volt, exp(1 / 0.025) = 4.85 x 10^8 times the saturation current, which for a typical device with a saturation current of 3 nA would suggest a current of around[7] 1.5 amps. (This is a very large amount of current for a small-signal diode, and is likely to damage the diode. It's certainly likely to heat it up so you can no longer assume that T = 290K.)

$$I = I_0\left(\exp\left(\frac{1}{n\times 25\text{ mV}}\right) - 1\right) \approx I_0\left(\exp(20) - 1\right) = 4.85\times 10^8\, I_0 \qquad (12.7)$$

---

[4] "small signal" since they are designed to take comparatively small currents through them, currents associated with signals. The other major type of diode is the rectifier diodes used in power supplies, and designed to take the large currents associated with powering whole circuits.

[5] Although very common and popular, the 1N4148 is quite unusual in having a rather high ideality factor. The ideality factor for most other common diodes is closer to one.

[6] However please note that all real diodes have a maximum reverse voltage which if exceeded leads to an effect called *breakdown* and very large currents flowing.

[7] For real diodes it's likely to be less than this, since there is a small *parasitic resistance* associated with most real diodes, and the device behaves as if it has this parasitic resistance in series with an ideal diode. As the current increases, the voltage across this parasitic resistance increases, and this tends to limit the voltage drop across the diode.

This suggests that in normal operation, the voltage across this diode conducting forward current will be less than one volt, and this is certainly true for most real diodes.

For a more typical current of 1 mA (and still with n = 2 and I_0 = 3 nA), the voltage across this diode is given by:

$$\begin{aligned} 0.001 &= I_0\left(\exp\left(\frac{V}{n\times 25\text{ mV}}\right)-1\right) \\ V &= n\times 25\text{ mV}\times\ln\left(\frac{0.001}{I_0}+1\right)\approx 50m\times\ln(333334) \\ &= 0.64\text{ V} \end{aligned} \tag{12.8}$$

and this is a typical forward voltage drop cross a small-signal silicon diode when used in a typical circuit.

12.2.2 The I-V characteristic

The current-voltage curve (or I-V characteristic) is the most common way of graphically showing the current through a diode for different levels of voltage dropped across the diode. The I-V characteristic of a typical diode is shown below in Figure 12.4:

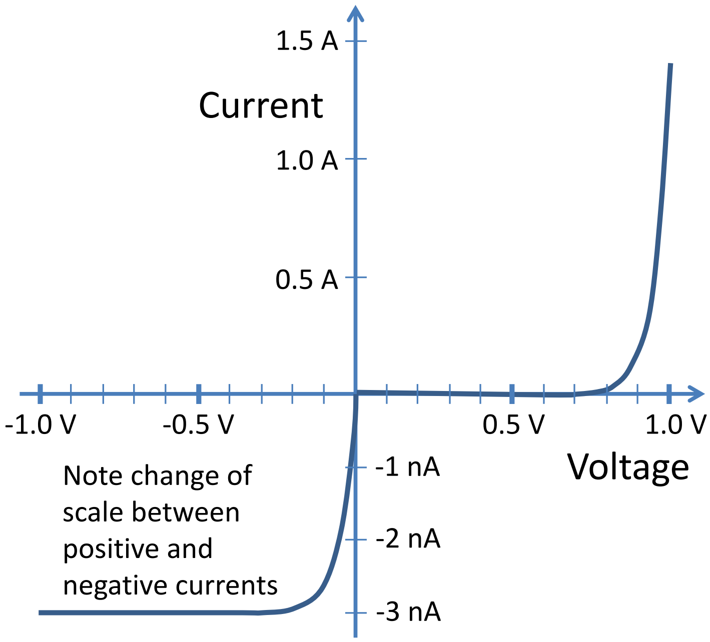

Figure 12.4 Current versus voltage characteristic of "ideal" diode

Looking at the shape of the whole curve, it can be seen that the current changes rapidly between very small values (a few micro-amps) and very large values (up to one amp) over the range of forward voltages from 350 mV to 920 mV. For a current of 1 milliamp (a more typical current for a small signal diode), the forward voltage drop across this diode would be expected to be around 650 mV. (In fact, to a first approximation when designing circuits, the voltage across a silicon diode is often assumed to be around 650 mV.)

Note the shape of the curve for reverse voltages: the current almost reaches the saturation current for a reverse voltage of less than one volt and doesn't change much if the reverse voltage is increased further from that point (until breakdown is reached).

12.3 Non-linearity, dynamic and static resistance

Diodes do not obey Ohm's law: the current through the diode is not proportional to the voltage across the diode. This means that diodes are not linear devices (in other words if you double the voltage across the diode, the current through the diode does not double).

For an ideal resistor, the resistance (defined as the ratio of the voltage across the component to the current through the component) is a constant. For a diode, this is not true: the ratio of the voltage across the diode to the current through the diode is a function of the voltage across the diode.

On a graph, the effective resistance (defined as the ratio of the voltage across the diode to the current flowing through it) is the inverse of the gradient of the straight line from the origin to the point on the diode characteristic curve. This is more usually known as the *static resistance* of the diode at this operating point (and it's important to say what the operating point is, since the static resistance can change from several megaohms (at low voltage levels) to a few ohms (at higher voltages)).

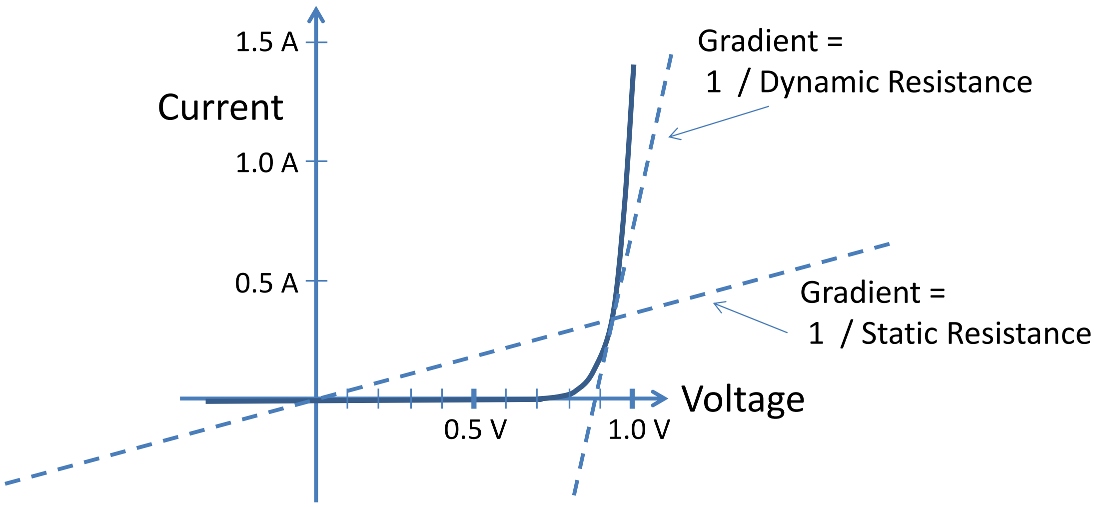

Figure 12.5 Determining the static and dynamic resistances

However there is another important resistance to consider when dealing with non-linear components such as diodes. The *dynamic resistance* is the ratio of a small change in voltage across the diode to the small change in current that results. On the graph of current against voltage, this is the inverse of the gradient of the curve at the operating point.

Dynamic resistance is often the more important of the two, since in many applications the diodes will be used with small signal voltages superimposed on larger, constant voltages (known as *bias voltages*).

Consider, for example, a small change in the voltage across a diode, and the small change in current that this produces:

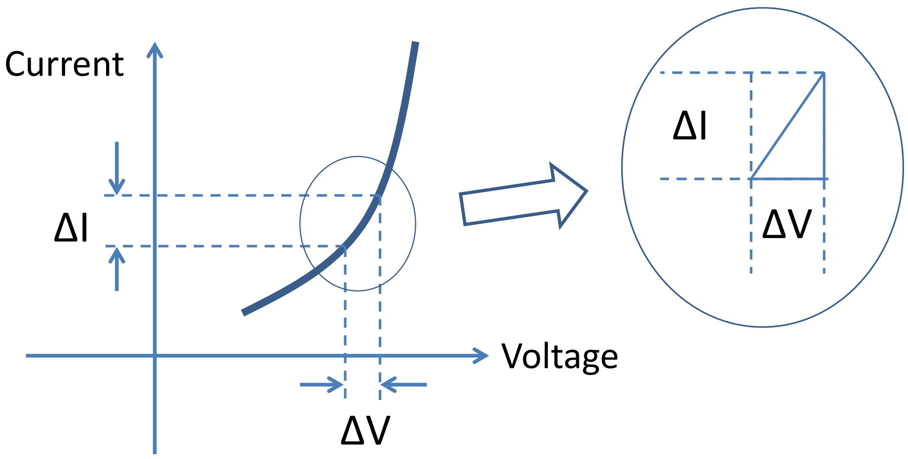

Figure 12.6 Differential gradient and the dynamic resistance

The dynamic resistance R_d is the gradient of the I-V curve, and approximately equal to the ratio of a small change in the voltage across the diode to the small change in the current through the diode that results. (In contrast the static resistance R_s is just the ratio of the total voltage across the diode to the total current flowing through it):

$$R_d = \frac{dV}{dI} \approx \frac{\Delta V}{\Delta I} \qquad R_s = \frac{V}{I} \tag{12.9}$$

Both are functions of the voltage across the diode and the current through it (the operating point). (See the chapter about "Linearity and Superposition" for more details about this point.)

12.4 Zener diodes

Up until now we've assumed that with large negative voltages, only a very small current (the saturation current) flows. In practice this is only true up to a certain negative voltage known as the *breakdown voltage*[8]. For voltages more negative than this breakdown voltage, diodes can suddenly start to conduct very large reverse currents.

Zener[9] diodes are diodes that are specifically engineered to have well-defined breakdown voltages, usually much lower than the breakdown voltage of a “normal” diode. For example, the 1N4148 small-signal diode has a typical breakdown voltage[10] of at least –100 V, whereas Zener diodes can be bought with a range of breakdown voltages starting at around –2.4 V.

---

[8] There are actually two physical effects which can cause this behaviour (quantum tunnelling and avalanche breakdown) but the result is the same for both: a sudden large increase in reverse current. Despite the name, breakdown doesn't damage the diodes (unless very large currents are allowed to flow), and diodes can operate quite happily in breakdown continuously.

[9] Named after Clarence Zener, who first described the breakdown property of semiconductors which leads to the Zener effect.

[10] Note that I've written this as a negative voltage to be consistent with the voltages specified elsewhere in this note. In most datasheets it's written as a positive number, even though it's the voltage across a reversed-biased diode. There's no confusion, since with a Zener diode everyone knows that the breakdown voltage occurs when the diode is reverse-biased.

This property can be very useful in producing voltage references (points in a circuit which remain at the same voltage independent of any other voltage in the circuit). For example, consider the following circuit, using a 3.3V Zener diode[11]:

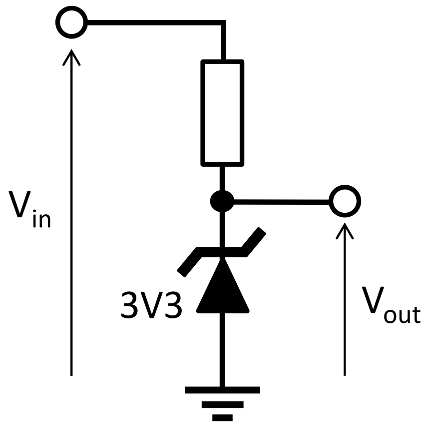

Figure 12.7 Zener voltage reference circuit

Provided the voltage at the top of the potential divider V_{in} is greater than 3.3 V, the voltage at the output (V_{out}) will ideally remain at approximately 3.3 V, irrespective of the voltage V_{in}[12].

12.4.1 Zener diodes in reverse breakdown

While there is a simple formula which gives an accurate prediction of the forward current through a standard diode (the Shockley equation), there is no simple formula that can be used to predict the reverse current flowing when the diode is in reverse breakdown.

One common approach to estimate the reverse current is to use a piecewise-linear approximation (which means approximating the current-voltage curve using a series of straight lines). The simplest way to do this is just to assume that in the breakdown region the dynamic resistance (i.e. the gradient of the curve) is constant, so that the graph of the I-V characteristic in reverse bias would look something like that shown in Figure 12.8.

The dynamic resistance (the inverse of the gradient of the I-V characteristic) of this diode in breakdown is:

$$R_D = \frac{\Delta V}{\Delta I} = \frac{-3.3-(-3.0)}{-10\times10^{-3}-\left(-0\times10^{-3}\right)} = \frac{-0.3}{-10\times10^{-3}} = 30\ \Omega \qquad (12.10)$$

Because the dynamic resistance is not zero (in other words the graph of current against voltage is not vertical), the reverse voltage across a Zener diodes in breakdown will depend on how much current is flowing through it. As a result, Zener diodes are specified as having a reverse breakdown voltage for one specific reverse current.

---

[11] Note the circuit symbol: like a diode, but with two small lines attached to the cathode line forming a sort of stylised ‘Z’.

[12] Not quite true if you think about it: the resistor has to be sufficiently small so that the voltage dropped across the resistor when the Zener diode switches on doesn’t reduce the voltage across the diode to below the threshold voltage, at which point the Zener diode would just switch off again. This is usually straightforward to achieve with a suitable choice of resistor.

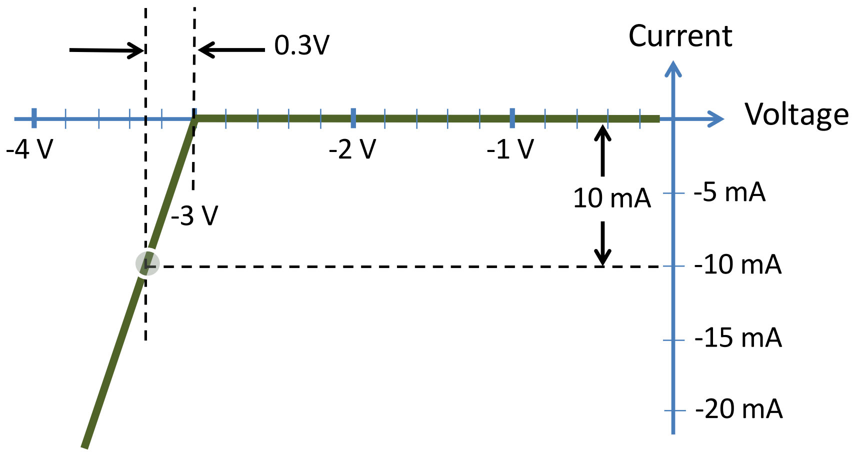

Figure 12.8 Piecewise linear approximation of a diode in breakdown

Using this piecewise-linear approximation to the characteristic, knowing the reverse breakdown voltage at a particular reverse current and the dynamic resistance in breakdown is enough to predict the behaviour of the diode at any value of reverse current.

For example, the diode with the characteristic shown in Figure 12.8 could be specified as having a dynamic resistance of 0.3 / 0.01 = 30 ohms, and a breakdown voltage of 3.3 V at 10 mA of reverse current. (Alternatively, it could be specified as having a breakdown voltage of 3.15 V at 5 mA of reverse current, which would amount to the same thing: both points lie on the line.)

12.5 Calculating diode voltages in circuits

Consider the diode circuit shown in Figure 12.9 below, in which you want to know the output voltage for a given input voltage. In theory, Shockley and Ohm's equations are all that is required to calculate the voltage at the output of this potential divider, however attempting to do this turns out to be harder than you might at first think.

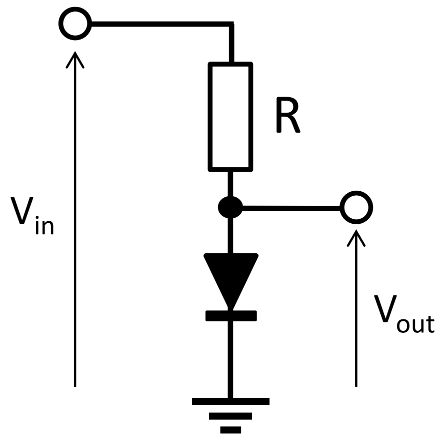

Figure 12.9 Potential divider with forward-biased diode

The analysis might proceed as follows: first, we can consider the potential divider using the standard potential divider formula and the static resistance of the diode:

$$V_{out} = V_{in} \frac{R_S}{R + R_S} \tag{12.11}$$

and the static resistance of the diode is the ratio of the voltage across the diode to the current through it, which is given by Shockley's equation as:

$$R_S = \frac{V_{out}}{I_{diode}} = \frac{V_{out}}{I_S\left(\exp\left(\frac{eV_{out}}{nkT}\right) - 1\right)} \tag{12.12}$$

and eliminating R_S between these equations and re-arranging gives:

$$V_{out} + I_S\left(\exp\left(\frac{eV_{out}}{nkT}\right) - 1\right)R = V_{in} \tag{12.13}$$

Now what? It turns out to be impossible to manipulate this equation into the form V_{out} = something. So how do you work out V_{out}?

12.5.1 Solving non-linear circuits using iteration

When faced with this sort of problem, one common approach is to use iteration: guess a sensible value for V_{out}, then work out what the current through the resistor and diode would be, and if they are not equal then adjust the guess.

For example, here we could start with an intelligent guess: that V_{out} will be about 0.6 V. For that output voltage, the current through the diode would be:

$$I_{diode} = I_S\left(\exp\left(\frac{0.6e}{nkT}\right) - 1\right) \tag{12.14}$$

and the current through the resistor would be:

$$I_{resistor} = \left(\frac{V_{in} - V_{out}}{R}\right) \tag{12.15}$$

If this guess about the value of V_{out} was correct, then these two currents should be equal. If the diode current is greater than the resistor current, then to reach the operating point of this circuit the diode current would need to be reduced and the resistor current increased, which suggests that the output voltage is actually slightly less than 0.6 V. On the other hand, if the diode current is less than the resistor current, that suggests increasing the estimate of the output voltage so that the diode current increases and the resistor current decreases.

By making successive guesses of the output in this way, the actual output voltage can be approximated more and more closely with each guess.

The question remains: if we find that the estimate of 600 mV is too small, how much should we increase it as our next guess? One approach that can find the correct solution quickly is as follows:

1. Guess an initial estimate of the voltage across the diode
2. Work out the current through the resistor given the estimate current estimate of the voltage across the diode

3. Using this current, work out what the voltage across the diode would be, thus providing a new estimate of the voltage across the diode
4. Go to step 2

Repeating this process can lead to an accurate answer with only a few iterations.

For example, suppose that in the example above the input voltage is 3.3 V, the resistor was 1000 ohms, and the diode is a typical 1N4148 with a saturation current of 3 nA and an ideality factor of 2.

1. Our initial guess will be 0.6 V.
2. This gives a current through the resistor of:

$$I = \frac{(3.3 - 0.6)}{1000} = 2.7 \text{ mA} \quad (12.16)$$

3. This current through the diode suggests the output voltage should be:

$$0.0027 = 3 \times 10^{-9} \left(\exp\left(\frac{V_{out}}{0.05} \right) - 1 \right) \quad (12.17)$$

$$V_{out} = 0.05 \times \ln\left(\frac{0.0027}{3 \times 10^{-9}} + 1 \right) = 686 \text{ mV} \quad (12.18)$$

2. Back to stage two, with the new estimate, which suggest a current through the resistor of:

$$I = \frac{(3.3 - 0.686)}{1000} = 2.614 \text{ mA} \quad (12.19)$$

3. And this new current estimate suggests the output voltage should be:

$$V_{out} = 0.05 \times \ln\left(\frac{0.002614}{3 \times 10^{-9}} + 1 \right) = 684 \text{ mV} \quad (12.20)$$

This is only 2 mV (or around 0.3%) different from the last estimate. If that is within an acceptable level of accuracy (it would be for most purposes) we can stop here, and quote the output voltage as being approximately 684 mV.

Going through the iteration process more times would give a more accurate answer, but is rarely necessary.

12.5.2 Solving Zener-breakdown circuits using iteration

The same approach can be taken to solve for circuits involving the reverse breakdown voltages and currents in Zener diodes. For example, consider once again the circuit shown in Figure 12.7 using a 1000 ohm resistor. If this Zener diode followed the characteristic shown in Figure 12.8 we could determine a more accurate result for the output of this potential divider.

Again, the first step is to guess an approximate value for the voltage across the diode, and in this case we could reasonably take as a first guess the specified breakdown voltage of 3.3 V. This produces a first estimate of the current through the resistor of:

$$I = \frac{5.0 - 3.3}{1000} = 1.7 \text{ mA} \tag{12.21}$$

The equation of the straight-line approximation to the I-V characteristic of this diode in breakdown is:

$$I = \frac{(V + 3.3)}{30} \tag{12.22}$$

so for a current of 1.7 mA, the second estimate of the reverse voltage across this diode would be:

$$\begin{aligned} 1.7 \times 10^{-3} &= \frac{(V + 3.3)}{30} \\ V &= 30 \times 1.7 \times 10^{-3} - 3.3 = -3.249 \text{ V} \end{aligned} \tag{12.23}$$

producing an estimated output voltage of 3.249 V. If this isn't accurate enough, then another round of the iteration would reveal a better approximation of the current as:

$$I = \frac{5.0 - 3.249}{1000} = 1.751 \text{ mA} \tag{12.24}$$

which produces a better approximation of the voltage across the Zener diode of:

$$V = 30 \times 1.751 \times 10^{-3} - 3.3 = -3.2475 \text{ V} \tag{12.25}$$

which suggests that the result after the first iteration step was very accurate: the second iteration only changes the value by 4 mV.

12.5.3 Solving Zener-breakdown circuits without iteration

While the iteration process works, one benefit of using a simple piecewise linear approximation for the breakdown characteristic is that you can solve for Zener diodes in breakdown directly.

For example, from the specification of the Zener in breakdown, we know that:

$$I = \frac{(V + 3.3)}{30} \tag{12.26}$$

and since V_{out} is the reverse voltage across the Zener diode, and I is the current through the resistor and therefore must be equal to (V_{in} - V_{out}) / 1k, we could convert equation (12.26) into:

$$\frac{V_{in} - V_{out}}{1000} = \frac{(-V_{out} + 3.3)}{30} \tag{12.27}$$

and this can be solved directly, to produce:

$$V_{out} = \frac{\frac{3.3}{30} - \frac{V_{in}}{1000}}{\frac{1}{30} - \frac{1}{1000}} \tag{12.28}$$

For an input voltage of 5 V this evaluates to:

$$V_{out} = \frac{\frac{3.3}{30} - \frac{5}{1000}}{\frac{1}{30} - \frac{1}{1000}} = \frac{0.1050}{0.03233} = 3.247 \text{ V} \tag{12.29}$$

This direct way of getting to the answer is only possible because we've chosen a simple piecewise linear approximation to the behaviour of the diode. Needless to say, real diodes don't behave exactly like this, but this choice does allow a relatively simple calculation to give an approximate answer, and approximate is often all that is required.

12.6 Light emitting diodes

Another very common form of diode is the light emitting diode (or LED[13]). These tend to have much lower saturation currents than other forms of diode (and hence require larger forward voltages for the same forward current), but otherwise behave in a very similar way to other types of diodes. Except of course they light up when you pass forward current through them.

12.7 Schottky diodes

Also worth a mention are the Schottky diodes; these operate using a slightly different principle from most other semiconductor diodes. The result is a diode with a very fast switching speed (it can turn on and off very quickly), and a very low forward voltage drop (often less than half that of a typical silicon diode). The disadvantage is a low breakdown voltage; they are also slightly more expensive than the most common silicon switching diodes.

12.8 Summary: the most important things to know

- When current is flowing through a typical silicon diode, the voltage across the diode is usually between 0.6 and 0.7 volts.
    - Other types of diodes can have lower forward voltage drops (e.g. Schottky or germanium) or higher forward voltage drops (e.g. LEDs).
- Diodes are characterised in terms of their saturation current (the maximum current that flows in reverse bias, at least until the breakdown voltage is reached) and the ideality factor.
    - The saturation current is a strong function of temperature.
- Ideal diodes can be assumed to obey the Shockley equation: $I = I_0\left(\exp\left(\frac{eV}{nkT}\right) - 1\right)$
- Solving circuits using the Shockley equation requires iteration.

---

[13] Usually pronounced 'el-ee-dee', rather than 'led'.