Rectifier and Power Supply Applications of Diodes

Introduction

One of the commonest applications of diodes is the conversion of an alternating current into a unidirectional current. This conversion process occurs in the signal detector parts of radio based systems and in power supplies designed to convert power from an alternating current distribution system (such as the UK's 50Hz land based power distribution system or the 400Hz distribution systems found in marine and airborne applications) into a good quality direct current source for electronic circuitry. Signal based applications of rectifiers will be discussed first.

Detectors and clamps

Peak detectors are circuits that have an output voltage that is (ideally) equal to the peak value of an input signal. Diode clamp circuits clamp one extreme of a signal (ie, either its positive peak or its negative peak) to a defined potential.

Peak detectors

The circuit diagram of a peak detector is shown in figure 1. The key elements of source, diode and capacitor are connected in series and it is quite common to find a resistor in parallel with the capacitor. The source may be a transformer secondary as is common in radio circuits or an amplifier output as is more typical of instrumentation systems. The idea is simple; when V_i is positive, D conducts and C charges to (V_i - 0.7V) as current flows from V_i through D to C. For high values of R, ie, values that allow only a small fraction of the charge put into C to be lost in one periodic time of V_i, V_O remains more or less constant during a cycle time.

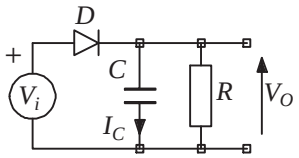

Figure 1

Peak detector circuit

The action of the peak detector circuit and the effect of R on its behaviour is illustrated in figures 2a and 2b. In figure 2a, the sinusoid is switched on at t = 0. Since the first half cycle is negative (ie, the wave is actually -sin ωt), the diode is reverse biassed for this half cycle and no

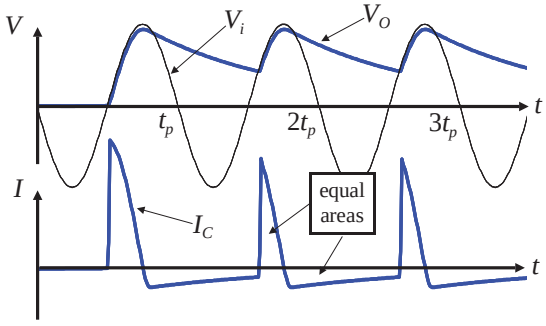

Figure 2a

Input voltage, output voltage and capacitor current in a peak detector circuit.

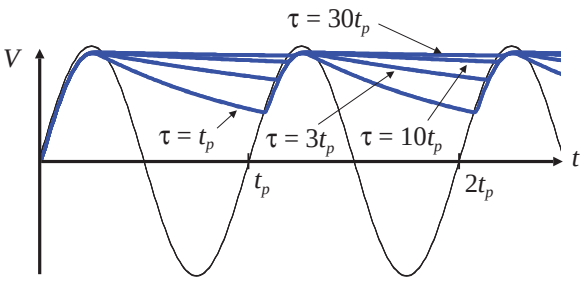

Figure 2b

Effect of RC on the output voltage of the peak detector

current flows. When $t = t_p/2 + \Delta t$, where Δt is the usually negligible time taken for the positive half cycle to rise to 0.7V, the diode begins to conduct and current flows from the source through C and R. The charge put into C is the amount needed to change its voltage from 0 to $V_{i\ \text{peak}}$ - 0.7V. The current rises quickly to a maximum and falls more slowly to zero as the source voltage approaches its positive peak (where the rate of change of source voltage equals zero).

The capacitor does not lose its charge straight away, instead it loses it slowly in the form of a current through R. This means that the source voltage, V_i, very soon becomes less than 0.7V higher than the capacitor voltage and consequently the diode ceases to conduct. Peak detectors are usually designed to ensure that the capacitor voltage does not fall too far in one source period so when the source voltage once again exceeds the capacitor voltage by 0.7V, only the charge lost through R need be replaced and the current pulse is correspondingly smaller in area than for the first charging event after switch on.

The effect of R on the rate at which the capacitor voltage falls is shown in figure 2b where there are waveshapes for four different values of R giving time constants of $RC = t_p$, $RC = 3t_p$, $RC = 10t_p$ and $RC = 30t_p$. If R were removed there would be no mechanism by which C could lose charge and the voltage across C would be $V_{i\ \text{peak max}}$ - 0.7V where $V_{i\ \text{peak max}}$ is the highest value of source voltage that had ever occurred.

Signal detectors

Peak detector circuits are often used to extract the information content from amplitude modulated radio signals. In this application the choice of RC is a compromise between a value large enough to give a small value of ripple at the carrier frequency and a value small enough to allow the capacitor voltage to follow the modulation envelope. Figures 3a to 3d illustrate this compromise in the context of a 100kHz carrier that is amplitude modulated by a 5kHz modulating signal. These conditions are approximately those relevant to the BBC Radio 4 long wave radio transmission at the upper limit of its signal bandwidth.

Notice that V_i, the modulated carrier, has an average value of zero whereas V_O has a non-zero average and a periodicity that is related mainly to the modulation envelope rather than the carrier. In figure 3a, $RC = 30t_p$, where t_p is the carrier frequency period. Although the carrier ripple is small, C cannot lose charge through R fast enough to allow its terminal voltage to follow the modulation shape. The remaining parts of figure 3 have RC products of $RC = 10t_p$, $RC = 3t_p$ and $RC = t_p$ for figures 3b, 3c and 3d respectively. Notice how the carrier ripple increases as V_O follows the modulation envelope more closely.

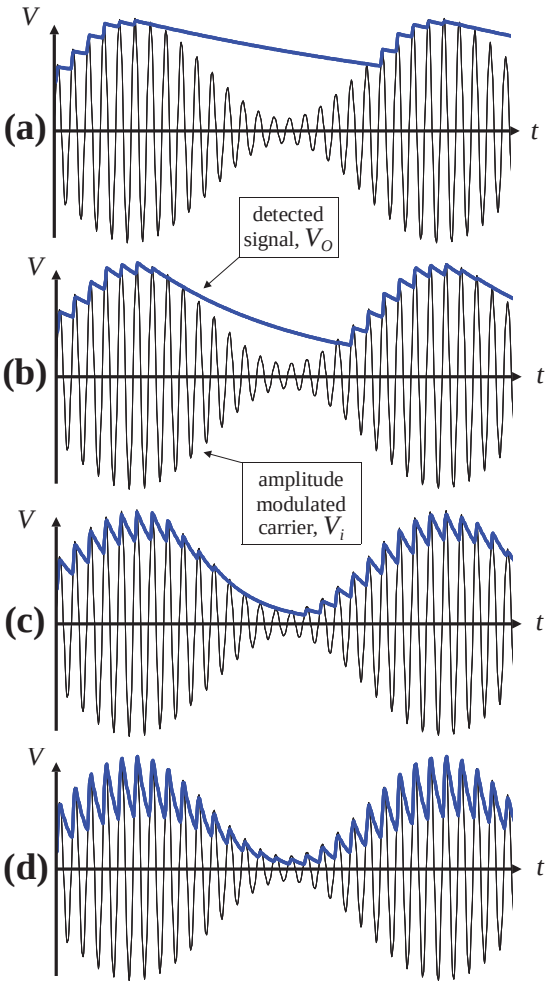

Figure 3

Effect of different time constants on a peak detector's performance.

Diode clamps

A diode clamp is a circuit that fixes or "clamps" either the positive or negative extreme of a waveform to a defined voltage level. There are some signals that transmit dc level information as part of the signal. The output of a signal detector is simply a waveshape consisting of a dc component and a superimposed signal waveshape. A clamping circuit discards the dc component and defines the dc level of either the positive or the negative signal peaks. Because clamping circuits resore the correct dc levels contained in signals, they are sometimes called "dc restoration circuits" and this term is used frequently in the context of analogue television video signals.

A basic diode clamp circuit is shown in figure 4. It contains the same components as the peak detector of figure 1 but they are put together in a different order. Figure 5 shows the input voltage output voltage and diode current that arise when the circuit of figure 4 is driven by a sinusoid switched on at $t = 0$. The action of the circuit is as follows.

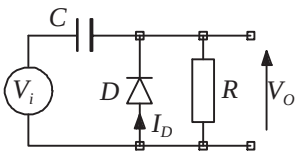

Figure 4
Diode clamp circuit

The first half cycle after switch on, $t = 0$ to $t = t_p/2$, is positive. The rising V_i will try and drive current through C towards the output. The diode will not conduct current in this direction so the only current that can flow in the circuit is through R. Since R is very large, the current is not large enough to affect sinificantly the charge stored in C. Thus there is a negligible change of charge in C, hence negligible *change* of voltage across it, and V_O follows V_i.

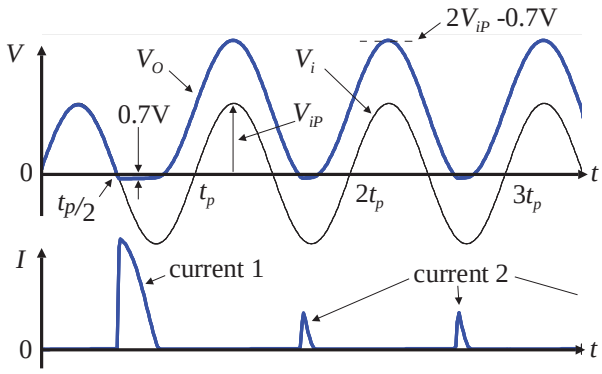

Figure 5
Voltage and current waveforms associated with the diode clamp of figure 4.

In the first half of the second half cycle, which begins at $t = t_p/2$, V_i continues to fall. V_O follows until it reaches -0.7V at which point D becomes forward biassed and begins to conduct (current 1 in figure 5). D conducts for the first half of the half cycle and thus holds the voltage on the output side of C at -0.7V. By the time V_i has reached its negative peak a charge of C ($V_iP - 0.7$) has been stored in C.

In the last quarter of the first cycle, V_i begins to increase again. This rising voltage once more tries to drive a current through C from the input side to the output side. As for the first half cycle, D will not support current flow in this direction so there is only a very small change of voltage across C because of the small current through R. Thus V_O follows the rise in V_i until the positive peak of V_i half a cycle later when $V_O = 2V_{iP} - 0.7$V and, because V_O is higher than -0.7V, continues to follow V_i as it falls back towards its negative peak one and three quarter cycles after the start.

When V_i gets close to its negative peak, V_O reaches -0.7V and a current flows once more through D into C. This current is replacing the charge lost by C because of the small current flowing through R for all the time V_O was above 0V. It is clear from figure 5 that this current (labelled as current 2) is much smaller than the current in the first half of the second half cycle (labelled as current 1). If R was infinitely large, current 2 would be zero because C would have lost no charge.

Notes:

1 V_i can be superimposed on any dc voltage without affecting the circuit behaviour. The practical limit to any such dc voltage is the votage rating of the capacitor C. In figure 5 V_i is drawn with an average value of zero because it is useful to have the waveshapes close to each other.

2 In practice C must be able to lose charge slowly so that the circuit can adapt to sources with slowly changing signal amplitudes. This means that R must be finite. It is usually the case though that R is sufficiently large to make charge loss per cycle a small fraction of the charge injected by current 1.

3 If the diode is turned round, the positive peaks of V_i are clamped to +0.7V and the output goes negative from there. The action is a mirror image of the action described here.

Peak to peak detector (voltage doubler)

A peak to peak detector is essentially a diode clamp followed by a peak detector. The diode clamp takes a sinusoid and clamps its negative peaks to -0.7V and the peak detector peak detects the positive peaks to give an output voltage approximately twice what it would have been without the clamp.

A peak to peak detector circuit is shown in figure 6. C_1 and D_1 form the diode clamp and D_2 and C_2 form the peak detector. R is effectively the circuit load. The circuit is used both in signal applications and in low power power supply applications.

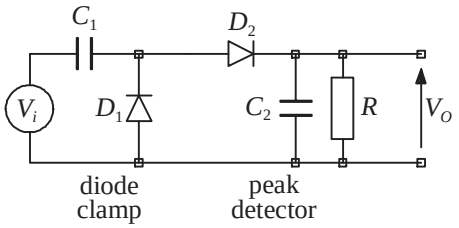

Figure 6

A peak to peak detector circuit.

The circuit is really one of a class of circuits called "diode pump" circuits or "charge pump" circuits. V_i is a driving force that is more or less equivalent to the reciprocating action needed to operate a pump such as a bicycle pump. When V_i is low, charge flows through D_1 into C_1 as described in the diode clamp explanation and when V_i goes high, that charge is tipped through D_2 into C_2. C_2 is essentially a charge reservoir from which charge continually drains away in the form of a current through R. If C_2 is a to be an effective charge reservoir, the constant leakage of charge through R should cause only a very small fall in voltage across it. Another way of putting this condition is RC_2 must be very much greater than the period of V_i. The output voltage is then determined by the equilibrium condition, charge pumped from C_1 to C_2 per second equals the charge lost through R per second.

Assuming that C_2 is an effective reservoir, C_1 affects the equilibrium output voltage and the time it takes the output to reach equilibrium. A C_1 that is small compared to C_2 will only be able on each cycle to transfer a small fraction of the charge needed to fill C_2. Thus charge will build up in C_2 over many input cycles (hence V_O will build up over many cycles) and because the rate at which charge is entering C_2 is small, V_O is sensitive to small current drains such as that through R because of the need to maintain an equilibrium between the charge flowing into and that flowing out of C_2. A C_1 that is large compared to C_2 can pump a large charge per cycle into C_2 so V_O rises rapidly and is relatively insensitive to R.

The voltage multiplier circuit

The peak to peak detector circuit can be extended to obtain higher output voltages using the circuit of figure 7. Let V_i be a sinusoid with a peak value V_{iP} big enough for the 0.7V drop associated with the diodes to be neglected.

C_1, D_1, C_2 and D_2 form a peak to peak detector that generates a voltage across C_2 of approximately $2V_{iP}$ as described in the previous section.

C_3, D_3, C_4 and D_4 form another peak to peak detector. In this second circuit, D_3 clamps the negative peak of the signal at node **B** to the top of C_2 - ie to a voltage of $2V_{iP}$. The positive peak of the voltage at node **B** is $2V_{iP}$ above its clamped negative peak giving a peak detected voltage across C_4 of $2V_{iP}$.

C_5, D_5, C_6 and D_6 form yet another peak to peak detector. This time the signal at node **C** is clamped to the voltage at the top of C_4, ie, $4V_{iP}$. As a result, the positive peak at node **C** is at a voltage of $6V_{iP}$ and this is peak detected by D_6 and C_6 to give $2V_{iP}$ across C_6.

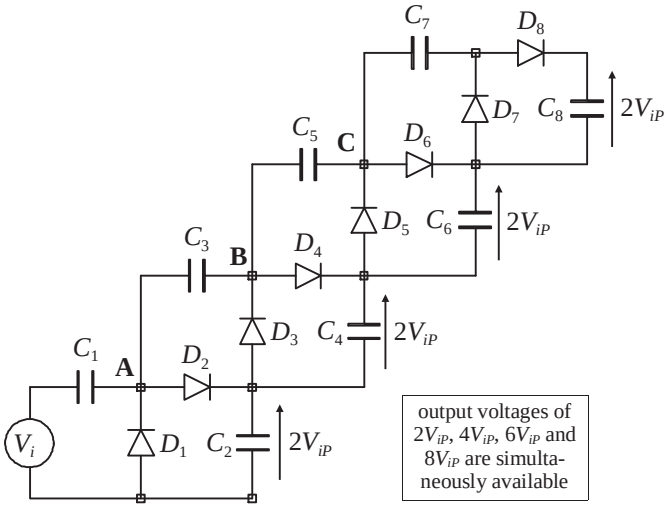

Figure 7

A votage multiplier circuit made from a cascade of peak to peak detectors

C_7, D_7, C_8 and D_8 form yet another peak to peak detector. The action is similar to the previous cases and the result is a voltage of $2V_{iP}$ across C_8. Thus there is a voltage of $8V_{iP}$ between the top of C_8 and the bottom of C_2; V_{iP} has been multiplied.

If the circuit is stretched out and redrawn it appears in the form of figure 8 and this is how it is usually presented in circuit diagrams. The circuit is exactly the same as that of figure 7

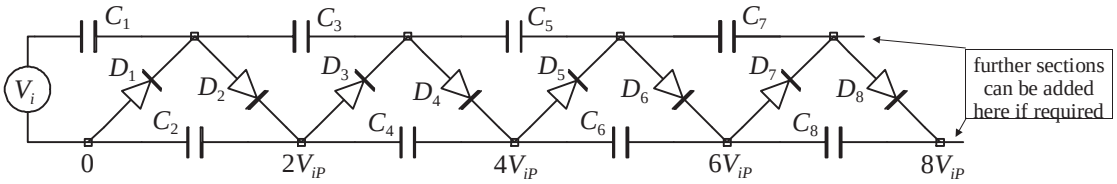

Figure 8

An alternative way of drawing the voltage multiplier circuit of figure 7

This circuit is used commonly in colour cathode ray tube based displays where it is used to multiply pulses of approximately 6kV in amplitude by a factor of around 4 to obtain the 24kV or so required to accelerate the electron beam towards the screen. It is also used in integrated form to multiply 5V pulses up to around 100V to drive the small plasma displays often found on CD and cassette tape players. It has been used in experimental physics research to obtain extremely high voltages - of the order of MV - for electric discharge studies and in such applications, multiplications of several hundred may be used. All these applications require low currents - typically a few tens of μA. The circuit is not suitable for high currents.

Rectifier circuits

The term "rectifier" is usually used in the context of the conversion of ac voltage into a unipolar voltage that is usually of the form of a dc component with a superimposed ac component. The dc component is always the *average* value of the rectifier output voltage and the superimposed ac component, which is rarely sinusoidal, is called the ripple voltage. In what follows, the ripple voltage will always be the peak to peak value of the superimposed ac component.

Rectifier circuits are often divided into two categories "half wave" and "full wave". In fact, full wave circuits can be looked at as two or more half wave circuits connected together and that is the approach used here. Other descriptive terms applied to rectifier circuits are "single phase" and "three phase" and these terms relate to the nature of the ac power supply - most high power industrial circuits (kilowatts and above) will be supplied by a three phase ac power source while low power industrial applications are more likely to be supplied by a single phase source. Domestic dwellings are supplied by a single phase source.

Single phase half wave rectifier

Figure 9 shows a single phase half wave rectifier circuit. The transformer can be considered as an ideal voltage source producing an rms voltage V_S across its secondary. (In fact, real transformers are not ideal in rectifier circuits as we will see later. All the calculations used in this module will, however stick to the ideal assumption.) The primary voltage V_P is derived from the local power system, eg 230V 50Hz for the UK, 115V 60Hz for the USA and 115V 400Hz for marine and airborne applications.

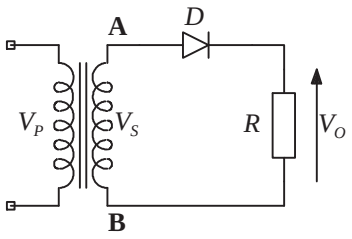

Figure 9

A half wave rectifier circuit.

It is important to remember that the voltage ratings of transformers designed for use with normal ac power distribution systems are given as rms voltages. Since the waveshape is sinusoidal, the peak value of voltage is $V_{\text{rms}}\sqrt{2}$ so in the case of figure 9, the voltage of node **A** with respect to node **B** will have a maximum value of $V_S\sqrt{2}$ and a minimum value of $-V_S\sqrt{2}$. The diode will conduct whenever node **A** is more than 0.7V positive with respect to node **B** and thus positive half cycles are transmitted to *R* but negative half cycles are blocked. The output voltage, V_O, from the circuit of figure 9 is shown in figure 10. Note that if the direction of *D* in figure 9 was reversed, the polarity of V_O would be reversed because *D* would conduct when node **A** was negative with respect to node **B**.

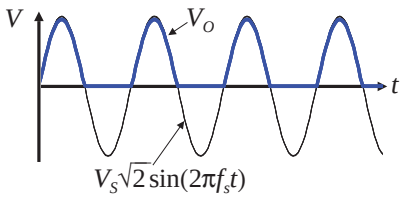

Figure 10

The output waveshape from the circuit of figure 9

There is a limited range of power supply applications for a waveshape such as that of V_O in figure 10. Most electronic equipment requires a relatively smooth supply voltage that approximates to the dc one might expect from a battery. V_O in figure 10 is clearly unipolar (ie, all positive in this case) and has a positive average (or dc) value but the magnitude of the superimposed ripple is too great. In particular the value of V_O is zero for the duration of alternate half cycles. For the circuit to be useful as a dc power source for electronic circuitry, the output needs to be "smoothed" or "filtered" in order to reduce significantly the amplitude of the superimposed ripple voltage.

Half wave rectifier with smoothing

There are various ways in which smoothing can be achieved - all rely on energy storage devices that are capable of filling in the gaps and smoothing out the peaks of the unfiltered waveform. The simplest of these is a capacitor connected as shown in figure 11. This smoothing process is called "capacitor input" filtering but in essence the circuit is a peak detector designed for relatively large voltages and currents and C stores energy which it uses to fill in the gaps in the waveform of figure 10.

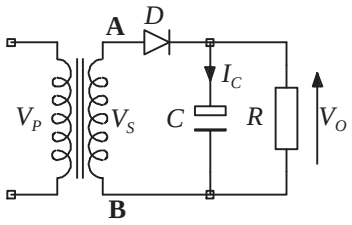

Figure 11

A half wave rectifier circuit with "capacitor input" filtering.

The waveshapes of V_S, V_O and I_C are shown in figure 12. These graphs were simulated using a real diode model but the transformer and capacitor are ideal. C is charged in the vicinity of the peak of every positive half cycle and provides current for the load inbetween the positive peaks. The charge lost by C inbetween peaks must equal the charge gained by C at the peaks if the average dc output voltage remains constant.

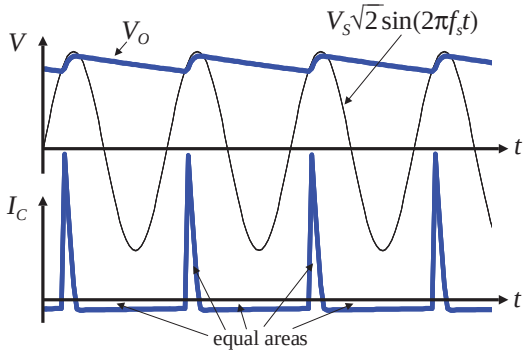

Figure 12

The waveshapes associated with the circuit of figure 11. All voltages are measured with respect to node B.

Note:

(i) The ripple is a very good approximation to a triangle - this fact will be useful when constructing a simplified model

(ii) The current waveform has large amplitude charging pulses of short duration. The pulse amplitude is some fifteen times the load current (which is equal to the magnitude of the negative capacitor current). Pulses like this are a nuisance for a number of reasons - their rms value to average value ratio is high (implies excessive losses due to I^2R heating in the circuit), the utilities are supplying all the energy used at one point on a cycle (sets up impulsive mechanical loads in the utility machinery) and rapid rates of change of current with time can cause radiative interference problems.

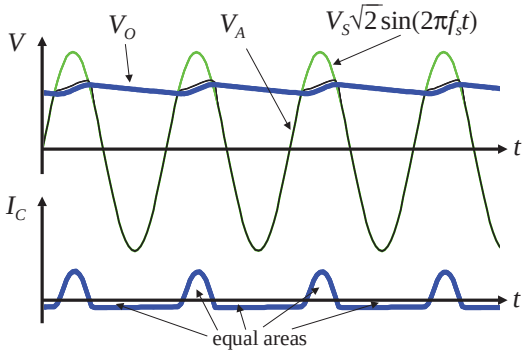

Figure 13

The waveshapes associated with the circuit of figure 11 when a realistic transformer impedance is included. All voltages are measured with respect to node B

In reality, the transformer and, indeed, the main supply have a series impedance that helps to mitigate the problems associated with charging C, at the expense of output voltage. Figure 13 shows the same waveforms as figure 12 with a series impedance appropriate for a 6VA transformer added between transformer and diode. In the simulation, the circuit is delivering less than 1W to the load and this is well within the transformer's capability. $V_S\sqrt{2}\sin(2\pi f_s t)$ is the ideal transformer secondary voltage and V_A is the voltage actually appearing at the diode anode.

Notice that the series impedance means that a large impulsive charging current is no longer

possible. The waveform is effectively clipped at a value significantly below its peak and this has two main consequences; firstly the output voltage is somewhat lower than in the ideal case of figure 12 and secondly because charge cannot get into *C* at the rate shown in figure 12, the charging pulses are broader and smaller in amplitude than (albeit of similar area to) those in figure 12. ***Notice, though, that the ripple voltage is still triangular in shape.***

Manufacturers do not give figures for the impedance of their transformers and even if they did, the calculation of the effects of that impedance would be very difficult. The message here is that the output voltage from a simple power supply with capacitor input filtering is poorly defined and in practice will be lower than ideal estimates. Consequently the use of simplified models for design and prediction of behaviour can easily be justified.

Choosing a capacitor to meet a ripple specification

The model usually used in order to describe the behaviour of a circuit such as that of figure 11 is based on the triangular nature of the ripple as shown in figure 14. In figure 14, the solid line represents the triangular ripple such as that in figures 12 and 13. It is common to assume, as a further simplification, that the capacitor discharges throughout the charging cycle and charges instantaneously at the peak of each charging cycle. Instantaneous charging implies infinitely high charging current pulses and is thus an inappropriate model for any consideration of diode or transformer current. Assumptions usually used are:

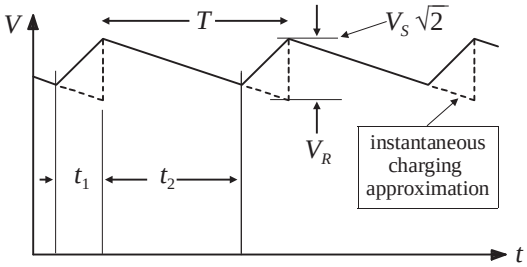

Figure 14

A simplified model of the output voltage from a capacitor input filtered rectifier circuit.

(i) The transformer and power source are ideal (ie, zero seies impedance)

(ii) Diode forward voltage drop is negligible

(iii) Load current is constant

(iv) Discharge occurs for the whole interval between charging peaks

Using the instantaneous charging model, the voltage across $_C$ reduces at a constant rate over the interval T as load current I_L is drawn from C. Thus

$$I_L = C\,\frac{dV}{dt} = C\,\frac{V_R}{T} \qquad \textbf{(1)}$$

since I_L is constant. Since this is a half wave rectifier

$$T = \frac{1}{f_s} \qquad \textbf{(2)}$$

where f_s is the supply frequency and if the load is a resistance R, as in figure 11,

$$I_L = \frac{V_{O\,peak}}{R}\,. \qquad \textbf{(3)}$$

Note that the V_O used in equation **(3)** is the peak value, $V_S\sqrt{2}$. This will give the largest possible value of I_L so will always overestimate C for a given ripple requirement. Equations **(1)**, **(2)** and **(3)** can be combined and manipulated in a number of ways to get the answer required depending on the information given.

The key is to understand the model and how equation **(1)** relates to it.

Single phase full wave rectifier circuits

In terms of their circuit shapes, single phase full wave circuits can be thought of as combinations of half wave rectifier circuits.

One of the oldest full wave circuit shapes is shown in figure 15. In this circuit V_{S1} and V_{S2} are 180^o out of phase - ie, when node **A** is at $V_S\sqrt{2}$, node **B** is at $-V_S\sqrt{2}$ and vice versa. V_{S1} and D_1 form one half wave rectifier and V_{S2} and D_2 form another. The outputs of these two half wave rectifiers are combined before being applied to the load. Because of the 180^o phase shift between V_{S1} and V_{S2}, V_{S2} and D_2 are active when V_{S1} and D_1 are not and there is an output across R on every half cycle as shown in figure 16.

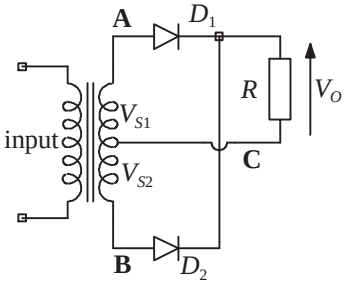

Figure 15

A two diode full wave rectifier circuit.

If the diodes in figure 15 were reversed, V_O would be negative - ie an upside down version of V_O in figure 16. In fact, there is no reason why reversed versions of D_1 and D_2 should not be connected to nodes **A** and **B** of figure 15 in order to obtain a simultaneous positive and negative output with respect to node **C**. Such a circuit, shown in figure 17, consists of four half wave rectifiers connected together to form a full wave rectifier. The circuit is more commonly drawn as shown in figure 18 and in this form is called a "full wave bridge rectifier".

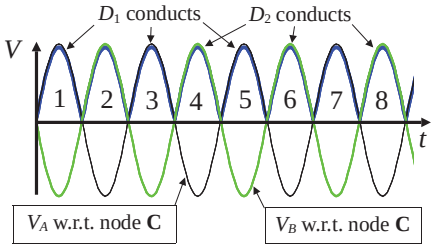

Figure 16

The output from a full wave rectifier. D_1 conducts on half cycles labelled 1, 3, 5, 7, etc and D_2 conducts on half cycles 2, 4, 6, 8, etc.

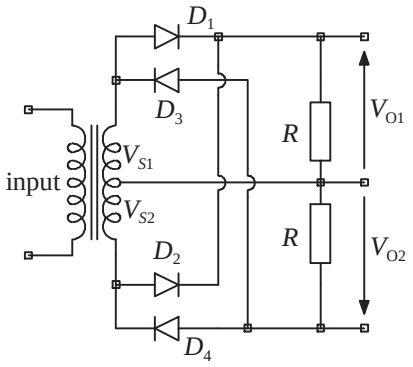

Figure 17

Four half wave rectifiers arranged to give a dual output full wave rectifier circuit

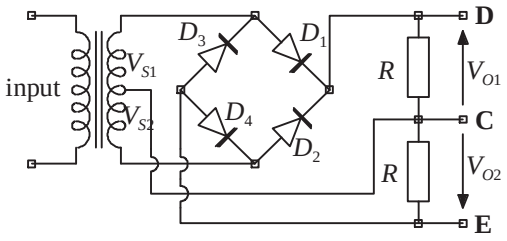

Figure 18

Figure 17 redrawn in a recognisable full wave bridge circuit shape.

In figures 17 and 18, diodes D_1 and D_2, the top R and V_{O1} are exactly the same as in figure 15.

The outputs from figures 17 and 18 can be used in various ways. If node **C** is the reference point, node **D** is a positive output and node **E** is a negative output. If node **E** is the reference point, both nodes **C** and **D** are positive with **D** having twice the magnitude of node **C**. If node **D** is the reference point, both nodes **C** and **E** are negative with node **E** having twice the magnitude

of node **C**. The two most common applications are firstly a positive and negative output with respect to node **C** (often called a "centre zero" power supply and used extensively for audio amplifiers) and secondly a single output of node **D** with respect to node **E** (as would normally be found in a low cost car battery charger). In the latter case, the connection between the centre of the transformer secondary and the mid-point of the *R*s is often omitted.

When deciding the output voltage from circuits like figure 17 and 18 it is important to take note of the way the transformer seconary voltage is specified. Transformers with a centre tap on their secondary winding are often specified as, say, 12-0-12 and in the context of figure 18 this would mean that $V_{S1} = V_{S2} = 12$Vrms. Very occasionally a secondary like this might be described as "24V centre tapped".

Smoothing a full wave rectifier

Smoothing a full wave rectifier output is very similar to smoothing a half wave circuit. The main difference is

the half wave rectifier charges the smoothing capacitor ***once*** per input cycle.

the full wave rectifier charges the smoothing capacitor ***twice*** per input cycle.

In all other respects the behaviour is the same and the model of figure 14, together with equations **(1)**, **(2)** and **(3)** that follow from it can be used to choose a capacitor to meet ripple requirements or to estimate ripple given the circuit values and operating conditions. The ripple frequency from a single phase full wave rectifier is a twice the input frequency.

Three phase full wave rectifiers (not examinable)

Most power systems that handle more than a few kW are three phase. A three phase power system has three live conductors (instead of the single live conductor in most domestic systems) that each carry power at the same frequency, but displaced in phase from one another by 120^o. There are many reasons why three phase power systems are attractive and some of them are connected with the fact that when one phase is going through zero, the other two are not, so power is continuously available from somewhere. (There are many other advantages that are irrelevant here.)

Figure 19 shows the circuit diagram of a three phase full wave rectifier and figure 20 shows its unsmoothed output. If one compares figure 20 with the unsmoothed full wave output of

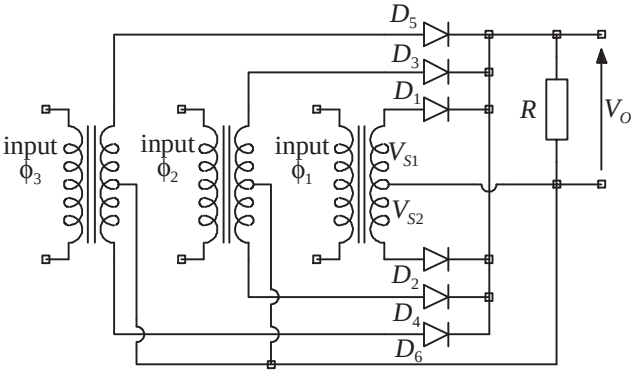

Figure 19
A three phase full wave rectifier circuit. Each transformer is the same as that in figure 15.

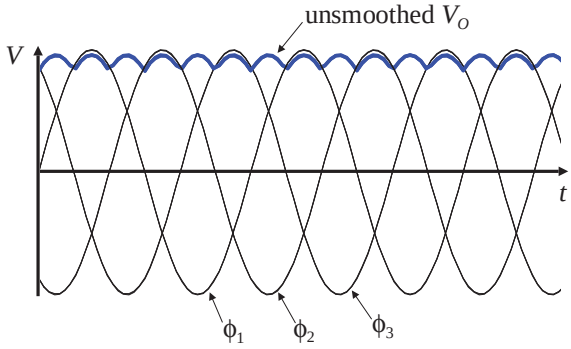

Figure 20
Output from a three phase full wave rectifier. Note that even with no smoothing the ripple is small.

figure 16, a couple of key points are immediately obvious:

(i) There is a large dc component of output voltage, $V_{DC} = 0.955V_P$, in the unsmoothed output waveform.

(ii) The peak to peak ripple voltage of the unsmoothed output waveform is $0.133V_P$ (compared to a value of V_P for single phase full wave and half wave rectifier circuits).

(iii) The fundamental frequency of the ripple is six times the input frequency (compared to the input frequency and twice the input frequency for single phase half wave and full wave rectifiers respectively).

These three points mean that for many applications smoothing is not necessary but when it is necessary, it is easier to achieve than in single phase circuits both because of the intrinsically smaller ripple voltage and because of the higher ripple frequency.

Stabilisation and regulation

The output from a rectifier and smoothing circuit is rarely of sufficient quality to supply an electronic circuit directly. There are many reasons for this

- The effects of finite supply and transformer impedance make it difficult accurately to predict the dc component in the output waveform (as illustrated in figure 13)
- The permitted range of supply voltage (in the UK, a nominal 230V with an upper limit of 253V (a + 10% rise) and a lower limit of 116V (a - 6% fall) that allows the utilities to manage load to ensure a constant supply frequency).
- The ripple voltage - very small ripple voltages demand very large smoothing capacitors. (bulky and expensive)
- The dependency of output dc voltage and ripple voltage on load current.

These problems have historically been divided into the two categories; **stabilisation** - the process of making the output independent of changes in supply voltage, and **regulation** - the process of making the output independent of load current changes. Although in the early days of electronics, the methods used to achieve good regulation were different from those used to achieve good stabilisation, good regulation and stabilisation are now simultaneously achieved by circuits that are usually called regulator circuits.

Regulator circuits

There are two types of regulator circuit; series regulators and shunt regulators. These can be further subdivided depending on their mode of operation but here the interest is in the basic forms shown in figures 21a and 21b. In both cases the unregulated dc input must be bigger than the required output voltage by some defined margin - and this must be true at the lowest instantaneous value of input, ie, the instant of the ripple negative peak.

The series regulator is connected in series with the load and works in the same way as a water tap or gas regulator by restricting the flow of current from input to output in such a way as to maintain a constant V_O across the load.

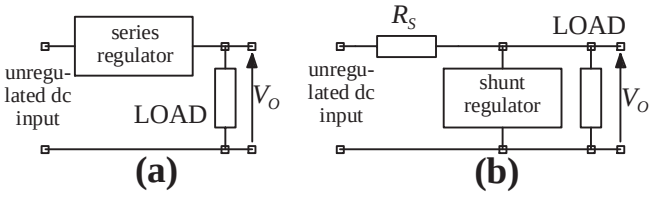

Figure 21

Organisation of a series regulator (figure 21a) and a shunt regulator (figure 21b).

The shunt regulator is in parallel with the load and works in conjunction with a series resistor R_S. It works by altering the current it draws, and thus the voltage drop across R_S, in such a way as to maintain a constant V_O across the load.

It is the shunt regulator that is of interest in this module.

Zener diode regulators

The simplest form of shunt regulator exploits the characteristic of a special type of silicon p-n junction diode called a "Zener" diode - named after its inventor. In a forward bias direction the Zener diode behaves in the same way as any other p-n junction diode. Its special properties lie in its reverse bias region, in particular, its reverse breakdown region. Once reverse breakdown in a p-n junction occurs, large increases in reverse current cause only small increases in reverse voltage and this property is exploited by Zener diodes which are special in that they are designed to break down at a particular well defined voltage. Devices are available with reverse breakdown voltages in the range 3V to 300V and the breakdown voltage is specified to an accuracy of typically ±5%.

A Zener diode characteristic is shown in figure 22 together with the symbol for this diode variant. The direction of current and voltage indicated are consistent with a forward bias being regarded as positive. The diode is always used in its reverse biassed direction so in a circuit environment, its reverse bias current is usually taken as positive. Once the diode has broken down in a reverse direction, an increase in reverse bias, ΔV will lead to an increase in reverse current ΔI. The ratio $\Delta V/\Delta I$ is known as the "**Zener slope resistance**" and is usually given the symbol r_Z (the lower case r indicates an inremental rather than a static resistance). The slope of the reverse breakdown characteristic is much steeper than drawn in figure 22 and typical r_Z for a typical low power device is 5Ω to 10Ω.

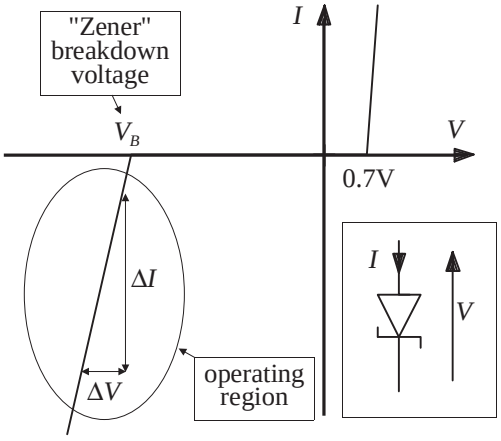

Figure 22

The I-V characteristic of a Zener diode. The diode symbol is shown in the inset diagram. The arrow head in the symbol has the same significance as for any other p-n junction

A Zener diode regulator circuit is shown in figure 23. V_i is an unregulated dc supply that may come from a battery or may come from a rectifier circuit and have a waveshape such as that in figure 14. The circuit output voltage will be close to V_Z, the so called "Zener voltage" (which is the same as V_B, the diode breakdown voltage), providing that the diode is biassed in its operating region (shown in figure 22). If the diode is to be biassed in its operating region, $I_Z > 0$ at all times. The operation is as follows.

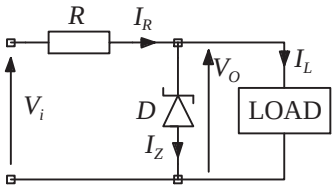

Figure 23

A Zener diode regulator circuit

First let I_L be constant but let V_i change. A reduction in V_i of ΔV_i will cause a change of voltage across R of ΔV_i and hence a reduction in I_R of $\Delta V_i/R$. This will tend to reduce V_O slightly which will in turn reduce I_Z. Only a small change in V_O is required to cause a large change in I_Z because of the steepness of the Zener diode characteristic in the operating region. Thus, the reduction in I_R is compensated by an almost equal reduction in I_Z leaving V_O and I_L more or less

unchanged.

If V_i remains constant but I_L changes, the action is very similar. An increase in I_L would also tend to lower V_O. Again a small drop on V_O causes a relatively large drop in I_Z so effectively the extra I_L has come from a reduction in I_Z and I_R is essentially unchanged.

design of a zener diode regulator

From a design point of view the circuit must be set up so that it can work as described above. At the cathode node of the diode

$$I_R = I_L + I_Z \qquad \textbf{(4)}$$

I_R can be written in terms of V_i, V_O and R, $I_R = \dfrac{V_i - V_O}{R}$ **(5)**

I_L will be defined either by V_O and the resistance of the load or, if the load is not a resistance, by a specified value or range of values that the load may draw.

I_Z is the variable that must be set by the designer to keep the diode in its operating region under all input voltage and load current conditions. In many cases Zener diode manufacturers will specify a minimum value of I_Z, $I_{Z\,MIN}$, that must be maintained for proper device operation.

The design proceeds by considering the conditions most likely to violate the minimum I_Z condition that must be satisfied. A moment's thought should lead to the conclusion that the circumstances most likely to threaten the minimum I_Z condition are when V_i is at its smallest, $V_{i\,MIN}$, and when I_L is at its largest, $I_{L\,MAX}$. Combining **(4)** and **(5)** and the appropriate worst case conditions,

$$\frac{V_{i\,MIN} - V_O}{R} = I_{L\,MAX} + I_{Z\,MIN} \qquad \textbf{(6)}$$

from which a suitable value of R can be found. Note that equation **(6)** is using the conditions most likely to make I_Z too small. If a value of R larger than that suggested by equation **(6)** were used, there would be insufficient I_R to satisfy the worst case demands of $I_{L\,MAX}$ and $I_{Z\,MIN}$. Thus, the R calculated in equation **(6)** is the largest value that can be used.

ripple considerations

To calculate the effect of the Zener diode regulator on ripple, a ripple equivalent circuit is used. If the circuit has been properly designed, the diode will be operating as intended and the incremental resistance of the reverse current region in conjuction with R will give the relationship between the input and output ripple voltages. The ripple equivalent circuit is shown in figure 24 where V_R is the ripple component of the input voltage and v_r is the output ripple. The relationship between input and output ripple is given by the potential division between R and r_Z,

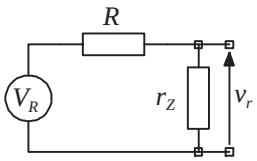

Figure 24

The ripple equivalent circuit of a Zener diode regulator

$$v_r = V_R \frac{r_Z}{R + r_Z} \qquad \textbf{(7)}$$

The load has been ignored in equation **(7)** but any load resistance will appear in parallel with r_Z and give a v_r smaller than that predicted.