PRIMER ON
Semiconductor Devices

by JIM ROWE

2 — The Junction Diode

The simplest semiconductor device most of us are likely to meet is the junction diode. Here's a quick look at how it works, and the kinds of jobs it can be used for in practical circuits.

A junction diode is produced by forming a P-type region and an N-type region directly alongside one another inside a semiconductor crystal. This can be done by either growing a layer of one type epitaxially on the top of a wafer formed from the opposite type, or by reversing the nett doping of a selected region on the top of the wafer by diffusing or ion implanting it with atoms of the opposite kind of dopant. The resulting kinds of structure are shown in Fig.1.

Either way, you end up with P-type and N-type regions inside the crystal, directly alongside one another. And when you do this, interesting things happen at the "junction" area which separates the two.

As you may recall, N-type material has a surplus of electrons in its conduction band, and conducts a current using these as its *carriers*. At normal temperatures these surplus electrons move around the crystal in a random fashion due to thermal (heat) energy.

On the other hand P-type material has a shortage of electrons in its valence band, and the resulting electron "holes" act as this material's current carriers. At normal temperatures these also tend to move around randomly.

What happens at the P-N junction is that electrons tend to wander over from the N-type region into the P side, and holes tend to wander the other way from the P-type region into the N side. When this happens the electrons tend to fill the holes, cancelling each other out as far as nett charge is concerned. So a kind of "no-carrier's land" tends to develop in the vicinity of the junction, with neither kind of current carrier able to survive there for more than a very short time before meeting its opposite number and being neutralised.

Because this region around the junction becomes depleted of free carriers, it is known as the *depletion layer*. And the effect of the depletion layer is to set up a potential barrier between the P-type and N-type regions, because of the lack of carriers. In effect, the N-type region has lost some of its electrons and becomes positively charged, while the P-type region has lost some of its holes (i.e., gained electrons), and becomes negatively charged — see Fig.2.

The wider the depletion layer spreads away from the actual junction, the larger this potential barrier grows. Eventually it stops growing, because the potential barrier repels carriers back into their own regions. So for a given temperature, the depletion layer width remains fairly constant — providing no external voltage is applied. At normal room temperatures it produces a potential barrier of about 300mV for a junction in germanium, and about 600mV for a junction in silicon.

What happens when we apply an external voltage? It depends on the polarity. If we connect it "in reverse", so that the N-type region is made more positive than the P-type region, all we do is boost the depletion layer's own potential barrier. The layer simply widens, and the number of carriers able to "scale its wall" becomes even smaller.

So *reverse biasing* the junction doesn't achieve a great deal. Very little current flows at all — a few microamps or less (until the reverse voltage is increased to a point where the depletion layer can widen no further, as we discuss later).

On the other hand if we apply an external voltage in the "forward" direction, with the P-type region made more positive than the N-type region, things are rather different. Here the external voltage is opposite in polarity to the depletion layer's potential barrier, so as soon as the two become comparable, the effect of the barrier disappears.

So in this case the current increases rapidly, as soon as the external forward bias exceeds the barrier potential of 300mV (germanium) or 600mV (silicon).

In other words, a semiconductor P-N junction tends to behave in a very similar fashion to one of the old thermionic valve *diodes*: it conducts very easily in one direction (with forward bias), but very poorly in the other (with reverse bias).

This makes the semiconductor P-N junction diode generally quite suitable for most of the circuit functions that previously called for valve diodes: rectifying AC, detecting AM signals, preventing circuit voltages from rising or falling below specified levels, and so on. In fact they're often rather better at these jobs than the older valve diodes, because they generally have a much lower resistance when conducting in the forward direction. This makes them closer to a perfect "one-way" diode.

Of course they're also much more compact than valve diodes — another point in their favour. Even a diode capable of conducting 6 amps typically has a body size of less than 10mm in diameter and say 15mm long. Small-

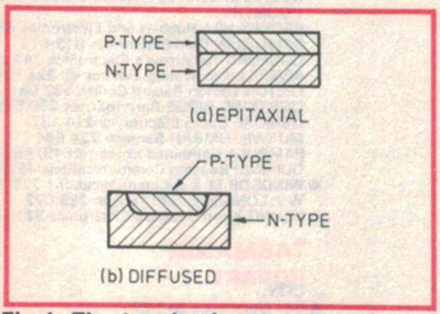

Fig.1: The two basic ways to make a P-N junction diode, epitaxy and diffusion.

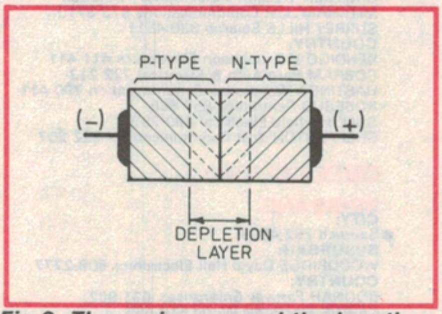

Fig.2: The region around the junction becomes depleted of carriers, setting up a potential barrier.

signal detector diodes can be very much smaller, say 2.5mm in diameter and 4mm long.

Apart from rectification and detection, P-N junction diodes have other uses as well. For example while the internal depletion layer prevents the junction from conducting any significant DC when reverse biased, at the same time it has significant *capacitance*. And because the width of the depletion layer varies according to the applied reverse bias, the capacitance varies as well. In fact it varies inversely with voltage, in a parabolic fashion; for high reverse bias it is small, and for low reverse bias it can become quite large. So a P-N junction diode can be used as a voltage-controlled variable capacitor.

Although virtually all silicon junction diodes can be used for this (germanium diodes have rather too much reverse bias leakage current), diodes made specifically for rectification/detection are generally not ideal. Instead special diodes are made to exploit the effect, and these are known as *varicaps* or *varactors*. They are used for tuning radio and TV receivers, for control of oscillator frequency in AFC (automatic frequency control) circuits, and for more esoteric things like frequency multiplication.

Another circuit application for P-N diodes makes use of the fact that when the reverse bias applied to a diode is increased beyond a certain point, its current suddenly increases from the normal very low figure. This is due to it "breaking down" in one way or another (a number of different mechanisms can occur).

When it does enter this "breakdown" mode, the diode's voltage drop tends to remain almost constant, despite the rapid rise in current. In other words, its *dynamic resistance* becomes quite low. At the same time, the diode need not be damaged, providing the current flow and power dissipation are prevented from rising to damaging levels. So in this reverse voltage breakdown mode, the P-N junction can be used to stabilise or *regulate* circuit voltages.

Although most silicon diodes can be used in this way, again it is usually better to use diodes that are specially made to exploit the effect. These *zener* or regulator diodes are designed to have rather lower dynamic resistance in the breakdown mode, and are also designed to conduct more current in this mode without damage. They are also made to break down at a range of convenient voltages, so that you can select a diode to suit the voltage you want to regulate: 6V, 15V, 27V or whatever.

Yet another use for what is still essentially a simple P-N junction diode is to detect light and/or convert it into an EMF to provide power. This relies on the fact that when photons of light energy fall on the P-N junction, they "knock" electrons out of the crystal structure and create a free electron and a hole. The extra carriers produced cause the junction's depletion layer to widen, increasing the potential difference between the P-type and the N-type regions. So in effect, the light energy is converted into electrical energy — which can be used to power circuit if desired.

Again most P-N diodes can be used for this, providing they are in a package which allows light to reach the actual diode "chip". But *solar cells* are even better, with very thin junctions spread over a large area to allow lots of light to enter the depletion layer and be converted.

Silicon is generally used for solar cells, because it generates 600mV per diode cell instead of 300mV. The larger the cell area, the higher the current (and hence power) it can deliver.