THE UNIJUNCTION TRANSISTOR

At present the principle use of this device appears to be as a trigger for the silicon controlled rectifier. However, it may in the future have a place in many other circuits. It consists of a bar of N type silicon having simple external connections at each end as shown in Fig. 1. If, for a moment, the effect of a third connection is ignored, this bar acts as an ordinary resistor. The third connection, called the emitter, is made to the opposite side of the bar, fairly near to base 1. This connection is to a P type region so that there is a rectifying junction at this point. Normally base 2 is biased positively with respect to base 1 to produce a small inter-base current. If the emitter current is zero, potential divider action causes a particular fraction of the inter base potential to be developed at the emitter. No emitter current flows provided that the emitter is negative with respect to this potential. If, however, it is made positive with respect to this the emitter base junction is forward biased and emitter current flows. Because of the construction and method of doping, this current is mainly comprised of holes moving from the emitter to base 1. The additional current carriers in the emitter-base 1 region reduces its resistivity and so leaves a greater portion of the forward bias originally applied between base 1 and the emitter to appear as forward bias across the PN junction. Thus the extra current leads to a <u>reduction</u> in the voltage required to produce it. This is a negative resistance characteristic. Once the resistance between base 1 and emitter is small enough it is dominated by the forward resistance of the PN junction and ceases to have a negative characteristic.

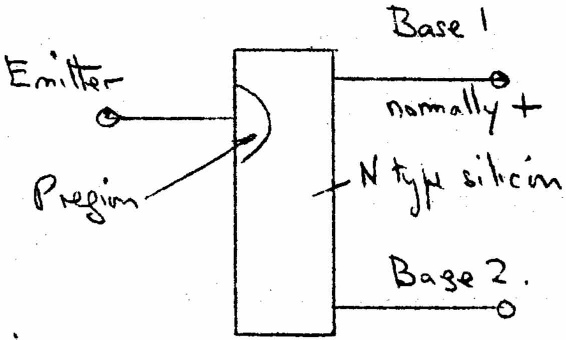


fig 1.

Symbol

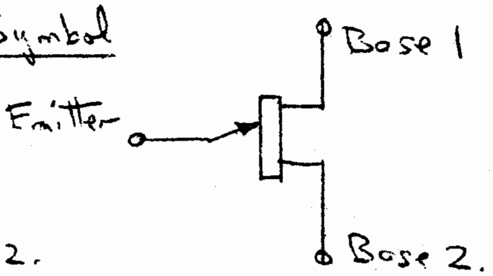


fig 2.

Fig. 3 shows the characteristics between base 1 and emitter for a number of values of inter-base voltages.

There are three important points about this characteristic.

1. The peak-point voltage at which the current switches on is a constant for a given inter-base voltage and device.

2. Until the peak point voltage is reached the input resistance is very high being equivalent to a reverse biased diode.

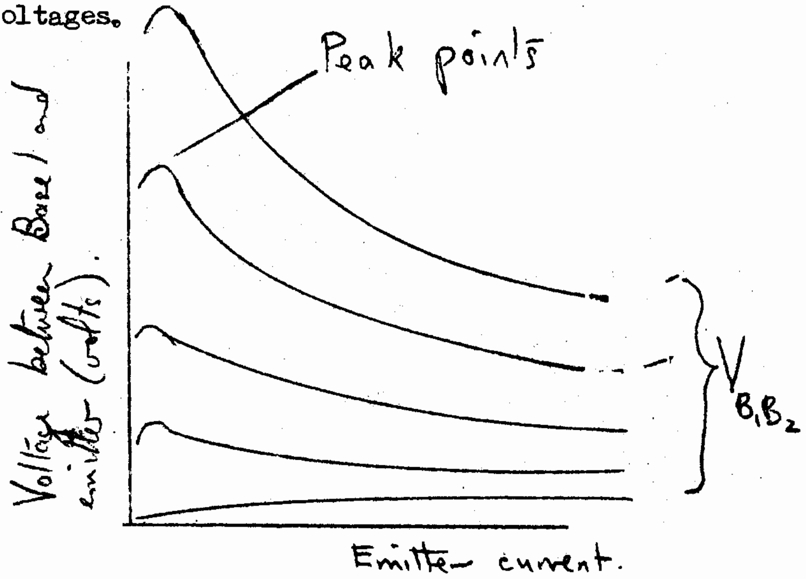


Fig. 3.

3. Once conduction occurs the input resistance drops to about 10Ω.

As might be expected the presence of an emitter current modifies the conditions between the two bases. Fig. 4 illustrates the effect of this.

The curve for $I_E = 0$ is approximately linear. The ratio of voltage and current of points on that curve gives an inter-base resistance of about 100kΩ. The curves above this are for various values of emitter current and correspond to lower values of d.c. resistance.

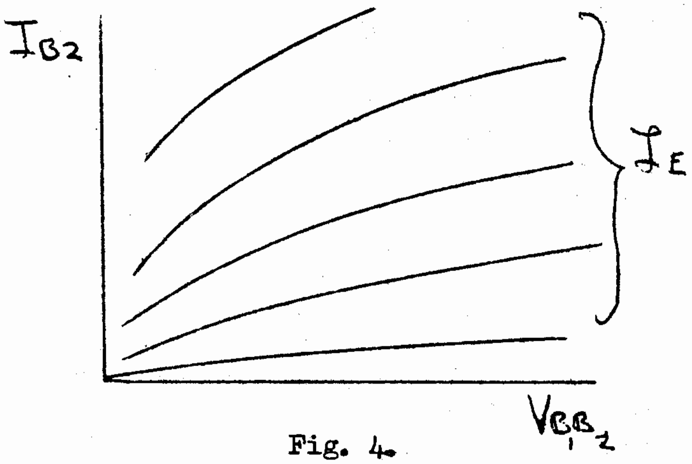


Fig. 4.

The following typical application of the device will illustrate its use.

1. Pulse Generator See Fig. 5. Assume that T_1 is conducting so that its emitter-base potential is very small. C_1 therefore changes towards +10V through R_3, causing the emitter potential of T_2 to become more positive. Initially the impedance between the emitter of T_2 and earth will be very high. The charging of C_1 will however, eventually cause the emitter of T_2 to reach the peak point when the emitter-earth potential falls abruptly to the order of a few tens of ohms. The right hand plate of C_1 then falls almost to earth potential and by capacitive induction the left hand plate falls by the same amount to a potential which is negative w.r. to earth so that T_1 is cut off. C_1 then discharges through R_2 until the discharge current falls to a small value at which T_2 again opens and, by capacitive induction, switches T_1 back on so that the cycle is repeated.

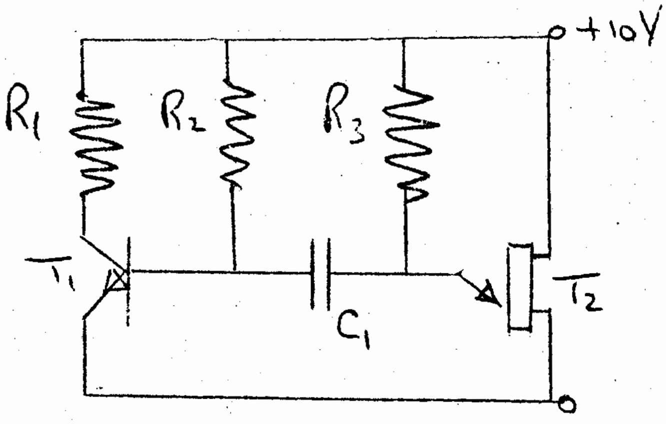


fig 5

THE FIELD EFFECT TRANSISTOR

This device is named in this way because its output current is controlled by the use of an electric field. It is also sometimes called the Unipolar Transistor.

Two main types of this device will be described;

(a) The junction field effect transistor - at present the only one which is regularly available.

(b) The insulated gate field effect transistor - also known as the metal-oxide-semiconductor-transistor (MOST). This type has recently been developed to the point at which one manufacturer has made some available to professional designers for their evaluation. NOW IN USE AVAILABLE & IMPORTANT.

The Junction Field Effect Transistor

This usually consists of a bar of N-type silicon having connections to each end which are called respectively the source and drain (an alternative terminology uses the terms Cathode and Anode). This bar is surrounded, by a PN junction about half way along its length. Alternatively two PN junctions may be situated on either side of the bar at about the same point along its length. The P region which forms this junction is called the Gate (the alternative terminology makes this the Grid). The device is shown diagrammatically in Fig. 1, which also shows its normally accepted circuit symbol. In the absence of a potential between gate and source the material between the drain and source acts as an ordinary resistor and conducts quite easily. The drain is the most positive of the two end terminals. Normally the device is symmetrical so that either end may be so used. If now the gate is made negative with respect to the source, the PN junction is reverse biased and negligable gate current flows. A depletion layer is then projected into the N type material which passes between the junctions and which is called the CHANNEL. This is shown in Fig. 2.

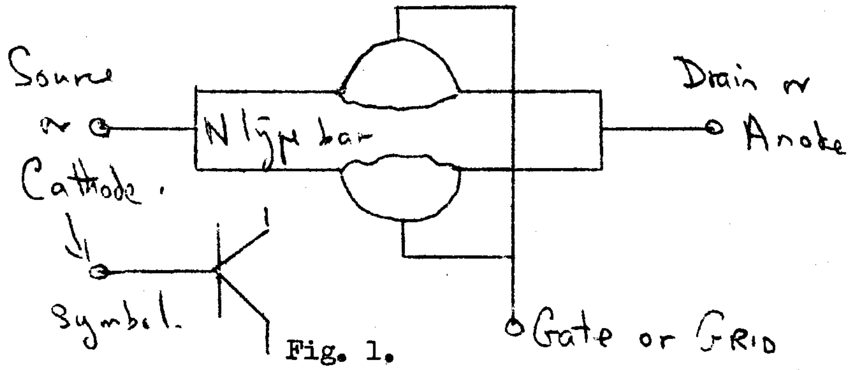


Fig. 1.

The depletion layer is widest towards the drain because the reverse bias is largest at that end. The extension of the depletion layer reduces the area of the channel and so limits the flow of current between the source and drain. When the reverse bias is sufficient the area of the channel is reduced to zero and an increase of V_{SD} then has no effect in increasing the drain current. The drain current flowing when this occurs is called the "pinch off current". The reverse bias to produce this effect is dependent upon both V_{GS} and V_{SD} so that pinch off can be produced by either increasing the drain potential positively, or the gate potential negatively, with respect to the source. The characteristics of the device which result from this action are very similar in shape to those of a pentode valve and are shown in Fig. 3. The input impedance of this device is due to a reverse biased PN junction and is therefore very high (in some cases up to 100MΩ). However, because this impedance is largely due to the depletion layer capacitance it is quite frequency dependent. The d.c. input resistance is also dependent upon temperature as the d.c. gate current is due to thermally generated minority carriers. The device may have a g_m of the order of a few mA/V. At present it is quite expensive but costs are expected to fall as production rises.

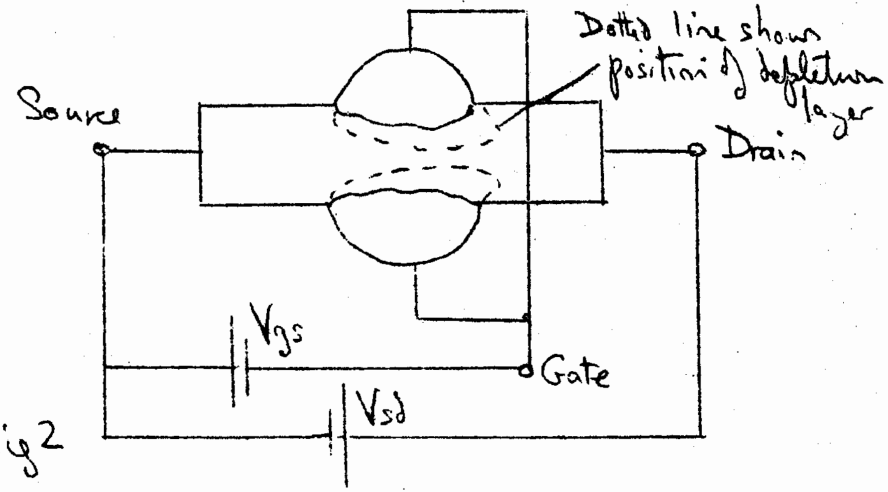


fig2

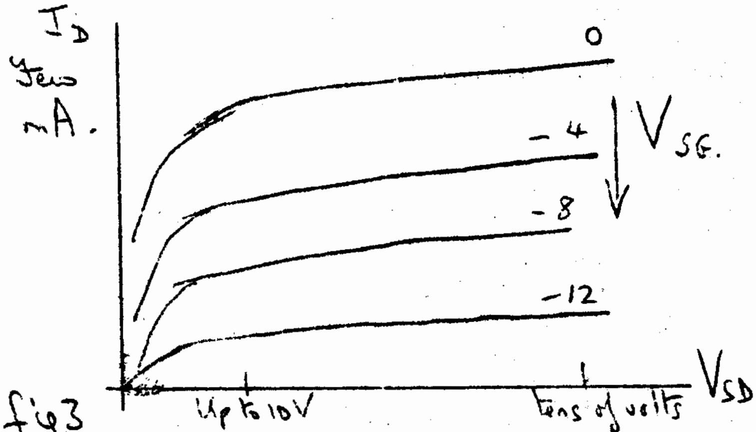


fig3

Fig.4 shows the construction of one typical field effect transistor of this type.

(a) Slice of P type silicon forming the base of a number of devices. Thin N layer added by epitaxy.

(b) P regions now introduced by masked diffusion to separate the devices from each other.

(c) P type gate regions now added by a second diffusion process.

(d) Devices separated and appropriate external connections added.

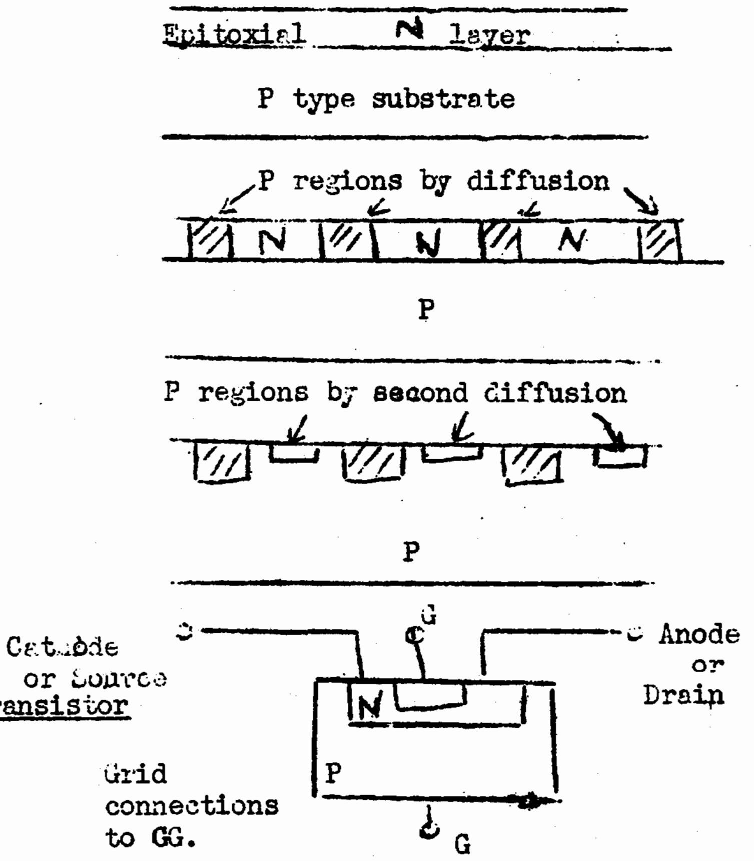


Fig.4.

The insulated gate field effect transistor

Fig.5 shows a diagram of a device of this type. When a positive potential is applied to make the gate positive with respect to the substrate a negative charge is induced in the part of the substrate adjacent to it so forming an N type channel connecting the source and drain. This channel prevents the formation of a reverse biased PN junction between the substrate and the drain so that when this electrode is made positive with

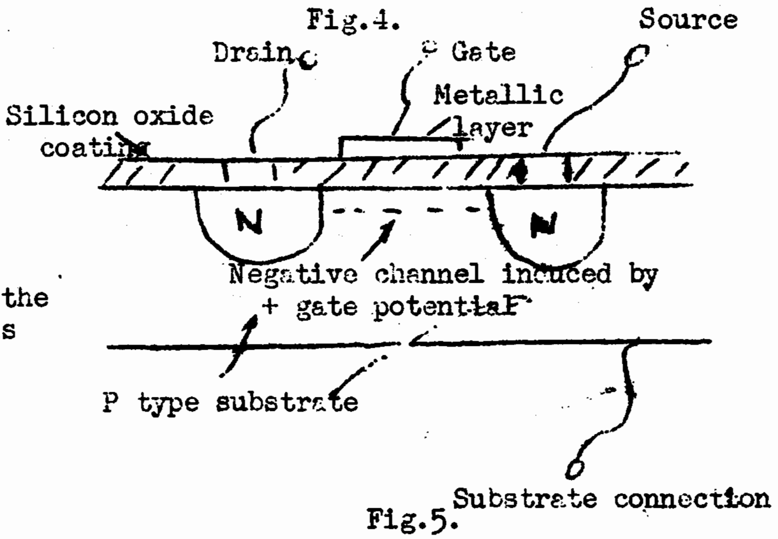


Fig.5.

respect to the source a current flows. The magnitude of this current is directly controllable by the intensity and thickness of the induced channel which is itself dependent upon the gate potential. The characteristic of this device is again similar to the pentode, an example being shown in Fig. 6.

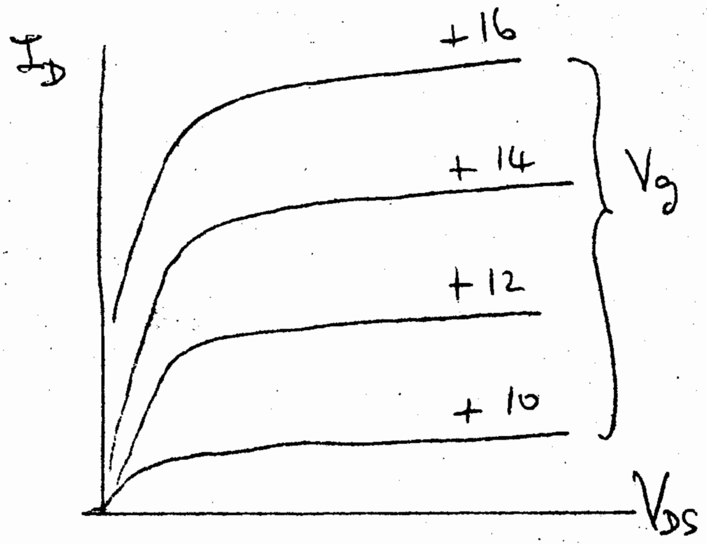


Fig. 6.

The gate voltage which is required to prevent the reverse PN junction forming is called the threshold voltage and has a value which is typically up to about 8 volts as shown. Below the threshold voltage no current flows. The device is symmetrical - that is either N region may be biased to act as the drain. The input impedance is very high, a value of 1,000,000MΩ is quoted at d.c. The input capacitance is typically about 4 p.F and the g_m may be about 1mA/V. One great advantage is that they are not temperature dependent over their normal operating range. At present they are not generally available and are very expensive. This state of affairs is unlikely to continue although at present there are one or two difficulties yet to be overcome. The oxide layer is liable to break down permanently even with quite small gate voltages and the electrical characteristics are rather inclined to drift. Particular advantages of all field effect transistors appears to be their high input impedance and their ability to produce very low noise figures.

THE TUNNEL DIODE

The action of the tunnel diode cannot be fully explained by the rather simple theory which has been used hitherto. A complete explanation of its properties is not attempted in this paper but it is hoped that the following will give a reasonably sufficient understanding of this device.

Fig. 1(a) illustrates the characteristic of an ordinary semiconductor diode. In this the breakdown or zener point on the reverse characteristic is usually the result of avalanche conduction but it can be due to breakdown due to a high electric field across the junction (strictly only the latter should be called zener breakdown although normal usage seldom makes this distinction. As the impurity concentration in the P and N regions is increased, the junction barrier becomes narrower, and consequently, for a given applied voltage, the gradient of the electric field across it becomes greater. Thus, increasing the degree of doping first of all ensures that breakdown is due to the electric field, and then, as the degree of doping is further increased, causes the zener point to occur at progressively lower voltages. Eventually breakdown occurs virtually at the origin as in Fig. 1(b).

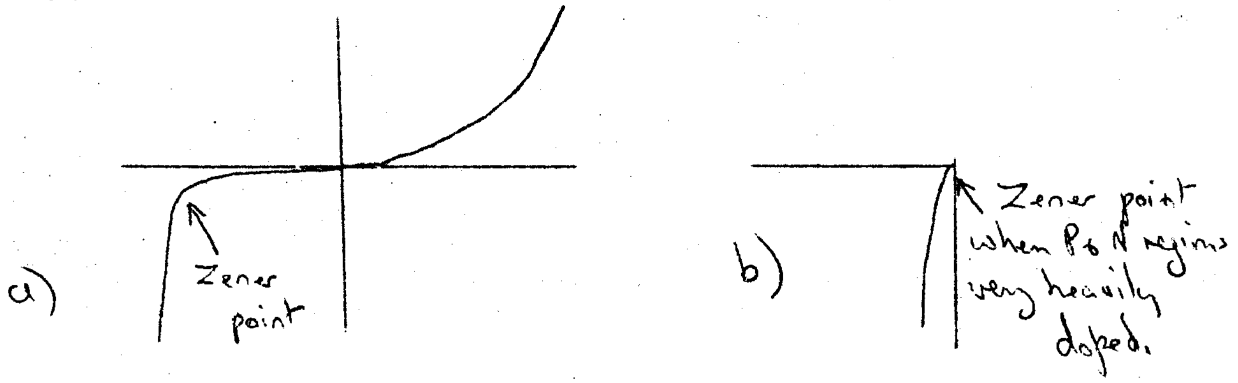


Fig. 1.

The effect of increasing the degree of doping in this way, so far as the forward characteristic is concerned, can be most surprising. It has already been shown that the barrier at the PN junction is very thin in this type of diode. Forward bias does not remove this barrier, but reduces it by making it thinner, in fact we are now thinking of a barrier which in the tunnel diode is probably thinner than 150A^{o}, (i.e. 150 x 10^{-10} metres). Now the elementary description of diode action previously given implies that electrons are rather like minute, negatively charged balls having an energy of motion which varies with temperature. The problem of causing electrons to cross the PN junction barrier could therefore be solved in either of the following ways.

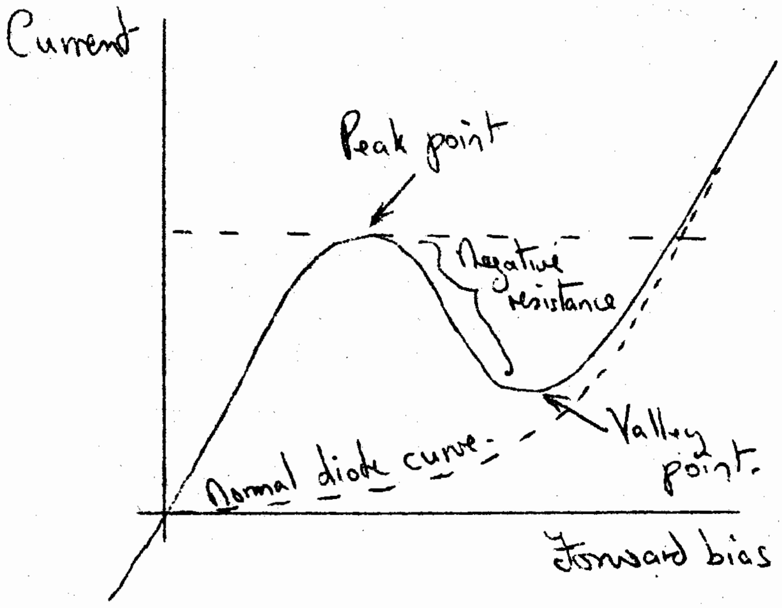
(a) Give the electrons enough energy to jump across the barrier by raising the temperature of the material.

(b) Lower the barrier to be jumped by applying a forward bias across the junction.

This simple picture has proved to be sufficient for a reasonable understanding of the previous work but it fails in this case. For the tunnel diode we must visualise an electron which is in the N region, but close to the junction, as being separated from the P region by an extremely thin but high barrier. Ideas have previously suggested that the electron must leap over this barrier to enter the P region. Now it must be realised that electrons are not miniature cannon balls after all, in fact when it comes down to it we may not really be at all clear as to what they are, even though a great deal is known of what they do and how they may be controlled.

Let us therefore use our elementary ideas but at the same time realise that this amounts to stretching an insufficient theory to fit a rather complicated action.

How can the electron know that it cannot penetrate the junction? Simply by banging into it and being bounced back. This however, must make a difference to the junction. The energy of the collision must penetrate into the surface before the stresses which cause the reflection of the electron can build up. In the case of the tunnel diode the barrier is so thin that we might visualise the effect of this energy as sometimes even becoming apparent on the opposite side of the barrier. In fact if the conditions are right this energy might even dislodge an electron from that side into the P region. The energy would then be absorbed by this action so that the original impingeing electron on the N side would remain to take its place. It is as if the electron had tunnelled through the barrier although in fact it is energy which has done this. One could almost visualise this action in terms of matching impedances, whether or not an electron is dislodged into the P region depending on whether the impedance match at this surface is correct for the absorption of energy. This match depends upon the magnitude of the forward bias. There is a particular small forward bias which produces the optimum condition for this effect to occur. This means that a forward current, due to tunnelling, can flow before the PN barrier has been reduced sufficiently by forward bias to permit a current by the more normal processes. The consequency is that an additional current, due to tunnelling, is superimposed on the forward characteristic which we have come to expect of the normal semi-conductor diode. The resultant tunnel diode characteristic is shown in Fig. 2.



The region between the peak and valley points represents a negative resistance, that is a condition in which an <u>increase</u> of voltage across the device causes a <u>reduction</u> in the current through it. Typically the peak and valley points occur at about 55mW and 350mV respectively with a germanium device (there are no silicon tunnel diodes at present). One important feature of this characteristic is the accuracy with which the peak point current is defined. Most semi-conductor specifications are to a tolerance of about 50% or more but the peak point current can already be specified to an accuracy of 2·5% and an accuracy of 1% is anticipated. This is important because it enables the tunnel diode to be used as a switch which is capable of sensing and switching to small and accurately defined current pulses. Usually the peak current is of the order of a few, to a few tens of mA but one device with a peak current of over 100A is available although it is extremely expensive. Another great advantage of the device is that its action is relatively insensitive to temperature changes. It can operate over a temperature range from almost 0°k to hundreds of $^{\circ}$C with little alteration in its characteristics. This makes it very suitable for use in high stability oscillators. Noise figures and the maximum frequency of operation of amplifiers using this device can be superior to anything which a transistor can produce.

Uses of the Tunnel Diode

Oscillators

If a single current pulse is used to cause a parallel tuned circuit to "ring" it will do so at a frequency which is determined by its components and with an exponential decay due to resistive losses. An ideal tuned circuit would oscillate continuously as it would have no losses. The tunnel diode can be used to produce this effect. To achieve this, it is placed in parallel with the tuned circuit and provided that its negative resistance is smaller than the dynamic resistance of the circuit $(R_D = \frac{L}{CR})$ the net resistance of the combination remains negative and the circuit oscillates continuously. The point is illustrated in Fig. 3.

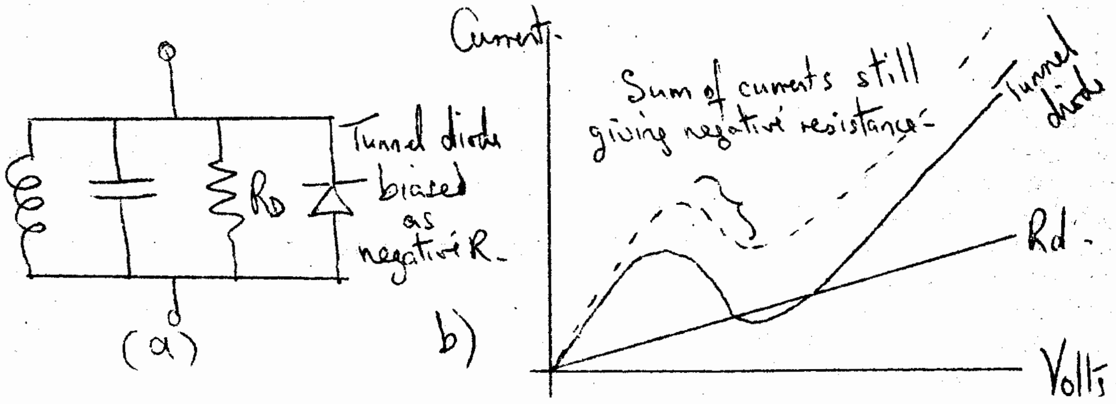


Fig. 3.

In Fig. 3(b) curves relating the current through, to the voltage across, both R_d and the tunnel diode have been drawn. Since these devices always have the same voltage across them the total current taken by the combination is the sum of these curves. This is shown by the dotted curve. Provided that this resultant curve retains a negative slope to which the tunnel diode is biased oscillation occurs.

This action requires that the tunnel diode shall be unambiguously biased to its negative resistance region. This can only be done by using a fairly steep d.c. load line as otherwise there will be more than one intersection between this and the characteristic and the precise operating point will be uncertain.

This is illustrated in Fig. 4.

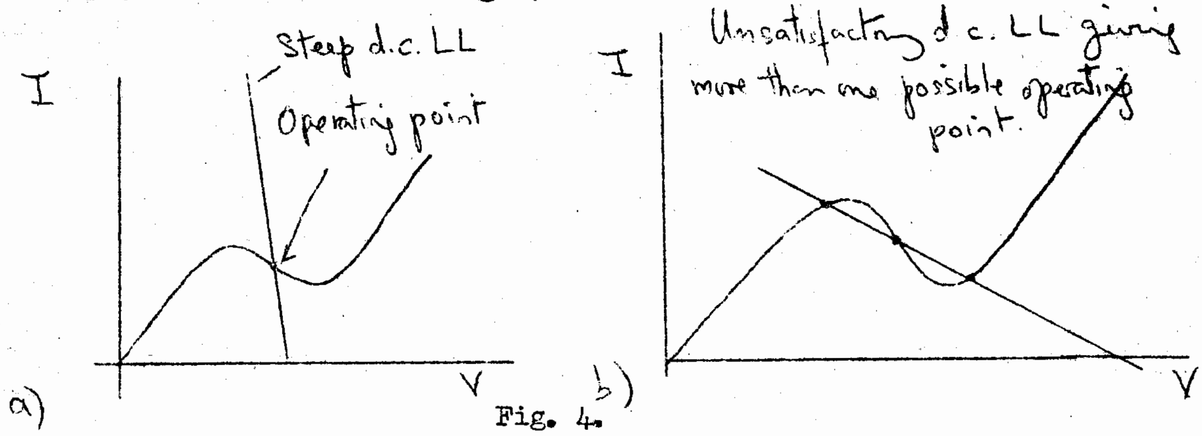


Fig. 4.

Fig. 5 shows the circuit of a typical tunnel diode oscillator. The d.c. load line for this circuit is due to $R_1//R_2$ (normally R_2 is quite small). These components are arranged to bias the diode to its negative resistance region. R_2 is decoupled by C_1 so that the diode is effectively in parallel with the tuned circuit (C_2, L) so far as a.c. is concerned. There is of course, capacitance across the diode junction which has a small but almost constant effect on the resonant frequency of the circuit. Sometimes R_2 is left undecoupled to damp out the possibility of a parasitic oscillation in the series mode between the junction capacitance and lead inductance. This kind of thing is always a danger in a negative resistance circuit of this kind.

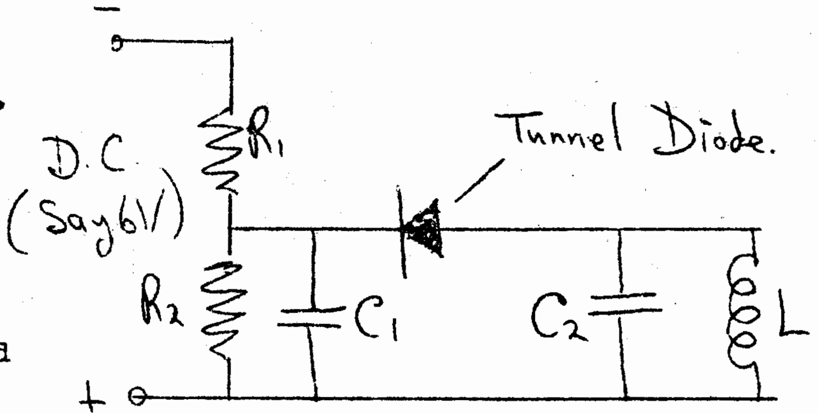


fig 5

Amplifier

Readers will already be familiar with the concept of "insertion loss". This expresses the effect on load power of introducing an additional network between source and load. Insertion loss is the ratio (expressed in dB) of the load powers before and after the additional network is introduced. The loss is due to the components of the new network and possibly to alterations in matching conditions. The tunnel diode can be introduced between source and load in the same way and because of its negative resistance characteristic can give rise to an "insertion gain". As an example consider the arrangements of Fig. 6.

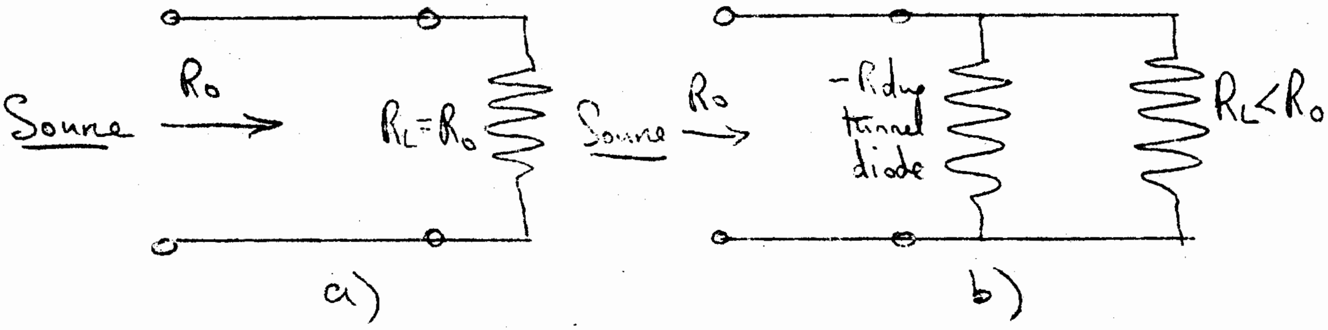


Fig. 6.

In Fig. 6(a) a matched source feeds a load R_o any deviation from the conditions of optimum matching leading to a reduction in the load power. In Fig. 6(b) the load resistor has been made smaller but has also been shunted by a negative resistance due to a tunnel diode. The resistance seen by the source is then given by $\frac{-R \cdot R_L}{-R+R_L}$. Now provided that the negative resistance is greater in magnitude than R_L the combination will give a positive resistance which can in fact be made equal to R_o so that correct matching is retained. As the matching condition is unaltered by this arrangement similar source voltages in Figs. 6(a) and (b) will give similar load voltages. Load power is $\frac{V^2}{R_L}$ and since R_L is smaller in the circuit of Fig. 6(b) the load power will be greatest in this case. The tunnel diode has had the effect of increasing the power delivered to the load.

A tuned amplifier operating in this way is similar to the oscillator circuit of Fig. 5 but is loaded so that oscillation does not quite occur. See Fig. 7.

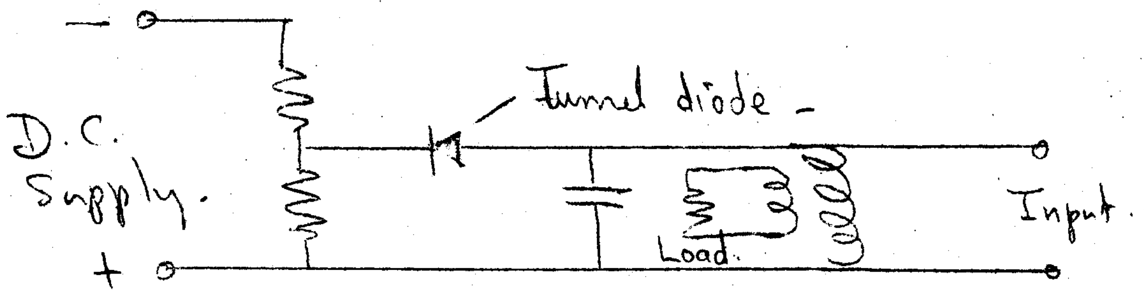


Fig. 7.

Alternatively the diode may be placed in series with the load as in Fig. 8.

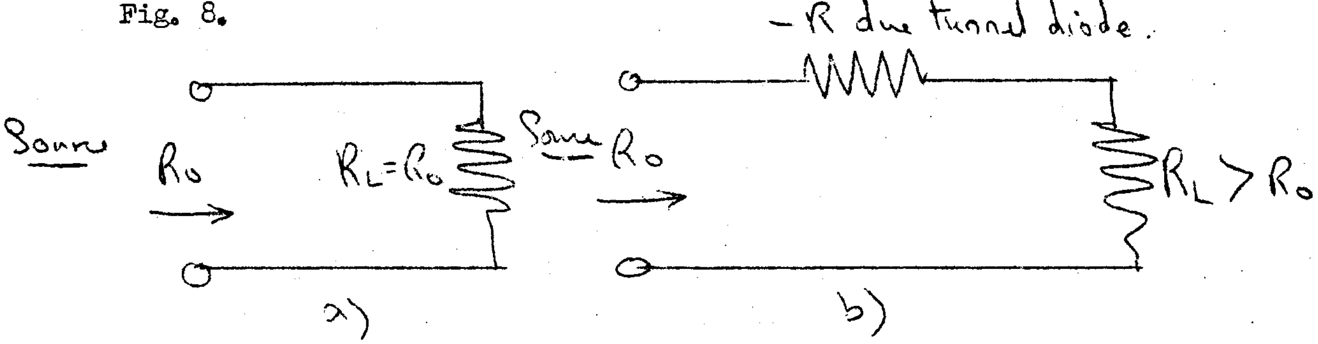


Fig. 8.

Again optimum matching is retained and the load current is the same for both circuits. The load power is I^2R_L and this is greater in the circuit containing the tunnel diode as R_L is greatest in this case. A circuit using this principle is shown in Fig. 9. The series amplifier comprises the diode and the components L and R_L. D.C. biasing is due to R_1, R_2 but their effects are stood off so far as a.c. is concerned, by the radio frequency choke. The circuit is series resonant between the capacitance of the diode and L so that it is frequency selective.

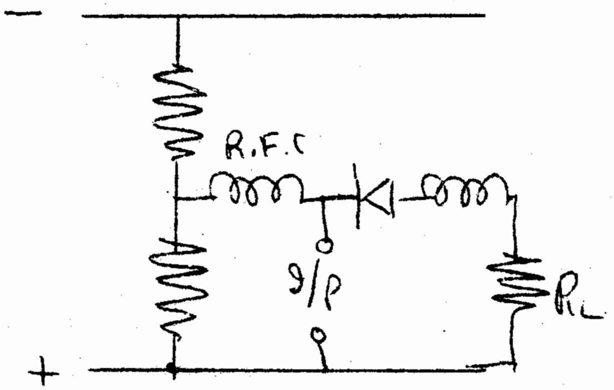


Fig. 9.

The Backward Diode

This device is very similar to the tunnel diode. It may be thought of as a germanium diode with a very low "forward" voltage drop (it conducts with only a few mV across it) which is only useable with very low signal voltages. Its "forward" characteristic is due to zener conduction from the origin as was shown in Fig. 1(b) so that "forward" bias for this is the reverse of all other diodes. The "reverse" characteristic is the low conduction region of the ordinary diode. The device is made so that the tunnelling effect is minimised but quite often a small peak due to this is noticeable. Fig. 10 shows the characteristic of this device superimposed on the dotted characteristic of an ordinary diode.

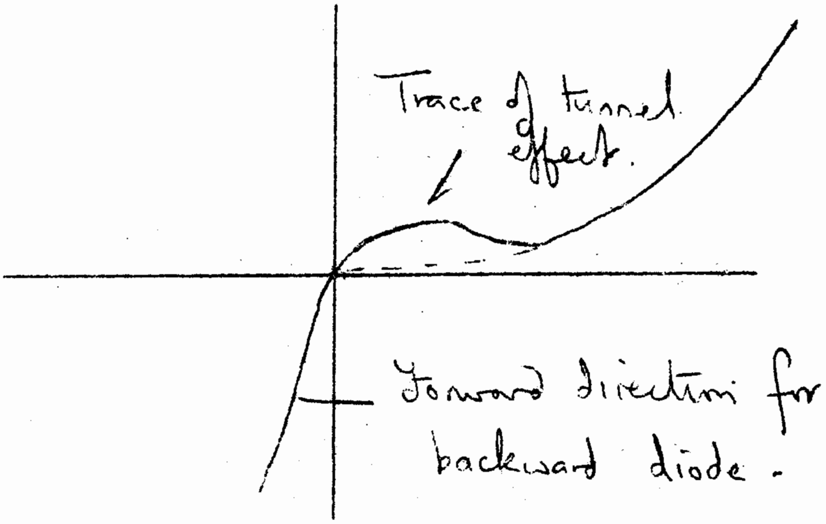


Fig. 10.

SEMI-CONDUCTOR INTEGRATED CIRCUITS

Such a circuit is one in which normally "individual" components (transistors, diodes, resistors and capacitors) are fabricated and suitably interconnected in a single semiconductor crystal. Resistances up to about 100kΩ and capacitances up to 0.1μF/sq.cm are practical possibilities at present. It is not yet possible to produce inductors or transformers by the same techniques which are principally those of diffusion and epitaxy. Briefly, the process begins by growing an oxide coating onto the surface of a single silicon crystal slice. This coating is then used as a masking medium, selective regions being removed by etching to permit the properties of certain areas of the slice to be modified by diffusion, or new regions to be grown by epitaxy. The process may be quite tedious but can be applied to many units at one time as with the most modern transistor manufacturing techniques. The circuit is built up, with interconnections, in layers (about three layers seems to be the present limit). The tolerances on the circuits so produced are at present fairly wide but this is not always important (with certain switching applications for example). Manufacturers can already purchase as stock items, simple units such as a multi-vibrator or a d.c. coupled amplifier with feedback. Such circuits need be no larger than, say, an OC71 transistor. They may have four connecting wires, one common, and three which are respectively the supply, input and output terminals. Present limits of the technique are that it must be applied to large numbers for economic production, that tolerances are wide and that the problems of heat dissipation and high frequency operation have yet to be satisfactorily overcome. In spite of these factors it may be confidently assumed that the future will bring big developments in this field.

JHB/RAS
13.11.64.