PARAMETRIC AMPLIFIERS USING TUNNEL DIODES AS NEGATIVE RESISTANCE ELEMENT

Submitted to : Prof.Dr. Mustafa Santur

Submitted by : M.Erhan KARABEKİR (Mehmet)

Graduate EE ,1964-1965

SCHOOL OF ENGINEERING
ROBERT COLLEGE

THIS THESIS IS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

TABLE OF CONTENTS

ACKNOWLEDGEMENT

I would like to express my deep gratitude to Prof. Dr. Mustafa Santur for his invaluable advice and instructions regarding the preparation of this thesis.

The recent interest in amplifiers which derive their gain from nonlinear circuit elements such as variable capacitance ,inductance, and resistance components stems chiefly from the development of low-loss variable-capacitance diodes or varactors and negative resistance diodes .There are two reasons for this interest .One reason is the fact that such amplifiers have low noise and the other is that the diodes are expected to have extremely long life .Either one of these properties is adequate justification for the excitement ...currently rumpant throughout the world concerning the exploitation of this new type of amplifier ,but,with two good reasons readily apparent, this excitement is doubled .

Mystery seemed to invade the thoughts of people when the scientists announced this new type of amplifier which was called a variety of names ,such as Parametric Amplifier,Reactance amplifier, and MAVAR (Modulator Amplifier by Variable Reactance) .Some of this mystery could have been avoided had the modern men known or mentioned that the underlying principle whereby electrical amplification was effected was an old principle .This principle may be broadly stated thus : The energy of an oscillating system may be increased by supplying energy at a frequency which differs from the fundamental frequency of the oscillator .One mechanical illustration of this principle is the simple pendulum .The child in the swing learns that he can " pump " up the amplitude of his oscillation of the swing by lowering his center of gravity on the down swing and raising it on the up swing .He thus pumps at twice the frequency of the swing .

Faraday,Melde and Lord Rayleigh have published observations and calculations concerning this principle .Quoting Lord Rayleigh " Faraday , ... with great ingenuity and success (upon examining) the crispations upon the surface of water which oscillates vertically arrived at the conclusion experimentally that there were two complete vibrations of the support for each complete vibration of the liquid. Crispations (may be) observed upon the surface of the liquid in a large wine glass or finger glass which is caused to vibrate in the usual manner by carrying the moistened finger round the circumference . All that is essential to the production of crispations is that the body of liquid with a free surface be constrained to execute a vertical vibration .Faraday's assertion that the waves have a period double that of the support has been disputed,but it may be verified in various ways ." The following example of the principle,reported by Melde in 1859,is,however , readily observed and understood . Quoting again from Lord Rayleigh " Perhaps the best known example is that form of Melde's experiment in which a fine string is maintained in transverse vibration by connecting one of its extrme-ties with a vibrating prong of a massive tuning fork,the direction of motion of the point of attachment being parallel to the length of the string .Under these circumstances .... the string may settle down into a permanent and vigorous vibration whose period is the double of that of the fork ." Lord Rayleigh analyzed and experimented with this and other similar mechanical phenomena in 1887 .This leads to analogous experiments with electrical cicuits .

The elctrical principle is readily understood by the following simple explanation .Suppose that we have a capacitor formed by two metal plates separated by air .Assume that a charge exists on the capacitor .The plates will be attracted to each other because of equal and opposite charges so that to separate the plates requires work .Upon separating the plates ,say to twice the original distance the capacitance will be reduced to 1 / 2 its original value .The elctrostatic energy,however,has been doubled ,since it is propor-tional to the square of the voltage and directly proportional to the capacitance .The energy required to separate the plates now appears as electrostatic energy in the capacitor .

Now suppose that the capacitor is combined with an inductor to form an oscillating circuit .The voltage on the capacitor will reach a maximum value twice each cycle .Now ,if on each half cycle,the capacitance is decreased when the voltage is maximum and increased when the voltage is zero,net energy will be imparted to the oscillations since no electrical energy is used to restore the capacitor to its original value when the voltage is zero .

Similarly ,it is apparent that net energy could be imparted to the circuit had the inductance been varied in the appropriate phase. This electrical principle was expanded to include frequencies other than the two-to-one ratio and the resulting device was used successfully in radio telecommunication between Berlin and Vienn prior to World War I .Next came new devices called Magnetic amplifiers .The objective then was to modulate a continous wave arc transmitter by means of a nonlinear inductance or saturable reactance .

Alexanderson,in the discussion ,also suggested amplification of incoming signals by cascaded stages of up-conversion ,rectification, and up-conversion ,etc .Currently we recognize the RF version of the above device as a type of parametric amplifier ,reactance amplifier or MAVAR .Alexanderson presented curves to show that negative-resistan effects could exist .Quoting again from his 1916 paper :Under some conditions " instability and generation of self-excited oscillations can exist." " This is a condition that must be avoided for telephone control ;whereas it may have useful applications for other purposes."

In the 1920's and 30's interest developed in subharmonic oscillations in electrical circuits containing a variable reactance .These "parametric " oscillations could exist at any one of f/n frequencies, where n is the subharmonic fraction of the fundamental frequency .

About thirty years later, interest developed in capacitance reactance modulators at microwave frequencies. The failure of reciprocity in some crystal converters observed in the middle 1940's and the peculiar behavior of welded contact Ge diodes were interpreted to mean that the contact capacity varied with bias.

However, in none of the devices suggested or manufactured till that time (1954), was a very-low-loss variable capacitance diode available and hence the gain limited and the noise figure was not especially good.

In 1954 a project was sponsored by the U.S. Signal Corps to develop semiconductor devices. A technique was found for making low-loss units. This technique of making low-loss diode varactors or varicaps advanced rapidly and interest in these units began to expand.

In the meantime, Suhl discovered that the variable reactance in the microwave range was obtained in ferrite materials when properly excited by a pumping frequncy. He proposed using this effect to obtain parametric amplification and discussed suitable materials.

Early in 1958, the low-noise properties predicted by theory were verified experimentally at the Bell Laboratories at 6000 Mc. Later distributed parametric amplifiers were devised. This was followed by traveling wave parametric amplifiers, modulated electron beams. A noise figure as low as 1.4 db was obtained.

The development of the vacuum tube in Alexanderson's time curtailed the interest in RF parametric transducers. Now the development of parametric transducers has stimulated the interest in vacuum tubes as parametric devices.

General Properties of Non-Linear Elements

A reactance may be defined as a circuit element that stores and transfers electromagnetic energy. If the stored energy is prodominantly in the electric field, the reactance is said to be capacitive; if the stored energy is predominantly in the magnetic field, the reactance is said to be inductive. In many applications, even at microwave frequencies, it is most convenient to speak in terms of voltages and currents rather than electric or magnetic field. The various functional representations for capacitive, inductive, and resistive elements are given below.

$q = f(v)$ Capacitive $\qquad \dot{q} = f(v)$ Resistive

$\phi = f(i)$ Inductive

A reactance is said to be linear if the particular functional relationship happens to be a linear one. E.g.: a linear capacitance implies a linear relation between charge and voltage.

$$q = Cv$$

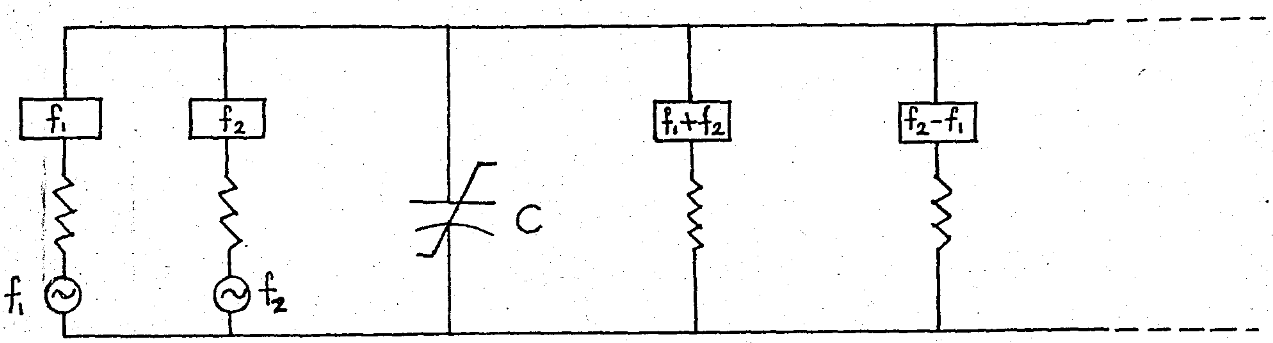
If these relations are not linear, the reactance is said to be non-linear. For example, if the relation between charge and voltage happens to be $q = av^2$, the element is said to be a non-linear capacitance. In this case, it is convenient to define capacitance as the partial derivative of charge with respect to voltage.

$$C = \frac{\partial q}{\partial v}$$

If the relation between charge and voltage is linear, but the coefficient C happens to be a function of time, the resulting capacitance is said to be linear, but time-varying. The analysis of time-varying capacitance circuit is facilitated by the use of the principle of superposition and other powerful methods. But the analysis of non-linear reactive circuits is not as easy.

Non-linear reactances are capable of creating a transfer of power from one frequency to another frequency. This behavior of non-linear elements was studied by Manley and Rowe.

The circuit model used by Manley and Rowe in their original derivation of power flow relations for nonlinear reactances is shown below :



We have two voltage generators at frequencies f_1 and f_2 together with associated series resistances and bandpass filters, placed across a nonlinear capacitor. These filters are designed to reject power at all frequencies other than their respective signal frequencies. In addition to the two signal generators, an infinite array of load resistors and bandpass filters are also connected to the nonlinear capacitor. These filters are tuned to the various sum and difference frequencies which will arise because of the nonlinear reactance. The sign convention will be used that power flowing into the nonlinear capacitance is positive (e.g. the power coming from the two signal generators), while power flowing from the capacitance (, the power flowing into the load resistances) will be negative in sign .

As stated before, a capacitive reactance may be defined as a circuit element for which a functional relation between charge and voltage can be written : $q = f(v)$
or the inverse function will do just as well in defining capacitance that is, we may also write $v = h(q)$

The only restriction placed on h(q) is that it be single-valued, i.e., free from hysteresis .

Next it is of interest to investigate the time-dependence of the charge q in the circuit model shown above. With impressed frequencies of and , there will certainly be variations in charge across the nonlinear capacitor at frequencies f_1 and f_2. In addition, because of the mixing action in the nonlinear capacitor, there will be generated charge variations at all the possible sum and difference frequencies $f_{mn} = mf_1 + nf_2$, where m and n take on all integral values. As in usual circuit analysis, it is convenient here to use the exponential representation of sinusoids so that differentiation with respect to time reduces to multiplication by $j\omega$. However, since we will here be interested in products involving current and voltage, we must exercise care in dealing with complex quantities. It will be most convenient to use only real quantities, expressing them in complex form. For example,

$$A \cos \omega t = \frac{A}{2} \left(e^{j\omega t} + e^{-j\omega t}\right)$$

Using this, we can express the total charge flowing into the nonlinear capacitor in a double Fourier Series in exponential form.

$$1^\circ \quad q = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} Q_{mn}\, e^{j(m\omega_1 t + n\omega_2 t)}$$

q must be real, so we impose the condition $Q_{mn} = Q^*_{-m,-n}$. We may find the total current "i" flowing through the nonlinear capacitor by taking the total derivative of 1° with respect to time.

$$2^\circ \quad i = \frac{dq}{dt} = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} j(m\omega_1 + n\omega_2)\, Q_{mn}\, e^{j(m\omega_1 + n\omega_2)t} = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} I_{mn}\, e^{j(m\omega_1 + n\omega_2)t}$$

where $j(m\omega_1 + n\omega_2)Q_{mn} = I_{mn}$ and $I_{mn} = I^*_{m,-n}$ since " i " has to be real

The voltage may be expressed as a function of q, and consequently it may also be expressed as a function of $m\omega_1 t + n\omega_2 t$. That is, we may also represent "v" in a double Fourier Series, just as we did for q.

$$v = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} V_{m,n}\, e^{j(m\omega_1 + n\omega_2)t} \quad , V_{m,n} = V^*_{-m,-n} \qquad 3^\circ$$

The Fourier coefficients can be evaluated by the usual integral :

$$V_{m,n} = \frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} v\, e^{-j(m\omega_1 + n\omega_2)t}\, d(\omega_1 t)\, d(\omega_2 t) \qquad 4^\circ$$

We are after an expression involving power; therefore we need to form products involving I_{mn} and $V_{m,n}$. As a first step toward this goal, let us multiply 4° by $jmQ^*_{m,n}$ and sum over m and n.

$$5^\circ \quad \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} jm Q^*_{mn} V_{mn} = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} jm Q^*_{mn} \frac{1}{4\pi^2} \int_0^{2\pi}\int_0^{2\pi} v e^{-j(m\omega_1+n\omega_2)t} d(\omega_1 t) d(\omega_2 t)$$

If we interchange the order of summation and integration, we obtain

$$6^\circ \quad \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} jm Q^*_{m,n} V_{mn} = \frac{1}{4\pi^2} \int_0^{2\pi}\int_0^{2\pi} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} jm Q^*_{mn} e^{-j(m\omega_1+n\omega_2)t} d(\omega_1 t) d(\omega_2 t)$$

The double sum inside the integral is reminiscent of the expression for charge. If we perform the following differentiation

$$7^\circ \quad \frac{\partial q}{\partial(\omega_1 t)} = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} jm Q_{mn} e^{j(m\omega_1+n\omega_2)t}$$

Since q is real, we may take the complex conjugate of $\frac{\partial q}{\partial(\omega_1 t)}$ without changing its value.

$$8^\circ \quad \frac{\partial q}{\partial(\omega_1 t)} = -\sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} jm Q^*_{mn} e^{-j(m\omega_1+n\omega_2)t}$$

Similarly the left side of is strongly reminiscent of the products of current and voltage. Remembering that $I_{mn} = j(m\omega_1 + n\omega_2) Q_{mn}$ we may express the left side of 6° as

$$9^\circ \quad \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} jm Q^*_{mn} V_{mn} = -\sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{m V_{m,n} I_{mn}}{m\omega_1 + n\omega_2}$$

Hence we may re-write 6° in the following form :

$$10^\circ \quad \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{m V_{m,n} I^*_{mn}}{mf_1 + nf_2} = \frac{1}{2\pi} \int_0^{2\pi}\int_0^{2\pi} v \frac{\partial q}{\partial(\omega_1 t)} d(\omega_1 t) d(\omega_2 t)$$

Now . Hence we can write the integral over $d(\omega_1 t)$ as

$$11^\circ \quad \int_{\omega_1 t=0}^{2\pi} v \frac{\partial q}{\partial(\omega_1 t)} d(\omega_1 t) = \int_{\omega_1 t=0}^{2\pi} v \frac{dq}{d(\omega_1 t)} d(\omega_1 t) = \int_{q(\omega_1 t=0)}^{q(\omega_1 t=2\pi)} v(q) dq$$

Therefore can be simplified even further to obtain

$$13^\circ \quad \left.\frac{\partial q}{\partial(\omega_1 t)} = \frac{dq}{d(\omega_1 t)}\right]_{\omega_2 t = \text{constant}}$$

$$12^\circ \quad \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{m V_{mn} I^*_{mn}}{mf_1+nf_2} = \frac{1}{2\pi}\int_{\omega_2 t=0}^{2\pi} d(\omega_2 t) \int_{q(0,\omega_2 t)}^{q(2\pi,\omega_2 t)} v(q) dq$$

With the mathematical manipulation now almost complete, it is time to stop and consider how the products of current and voltage as expressed by 12° are related to the power flow in the nonlinear capacitor. We know that active power is obtained by taking the real part of VI^*. Since Eq. 12° consists of such products, the real part of Eq. 12° will give us the desired information regarding power flow in a nonlinear reactance. Let us break the summation m into two parts, one summation from o to ∞, and one summation from 0 to $-\infty$

14° $$\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\frac{mV_{mn}I^*_{mn}}{mf_1+nf_2}=\sum_{m=0}^{\infty}\sum_{n=-\infty}^{\infty}\frac{mV_{mn}I^*_{mn}}{mf_1+nf_2}+\sum_{m=0}^{-\infty}\sum_{n=-\infty}^{\infty}\frac{mV_{mn}I^*_{mn}}{mf_1+nf_2}$$

Recalling $I_{mn}=I^*_{-m-n}$ $V_{mn}=V^*_{-m-n}$

we can effectively change the summation over the negative values of m into a summation over the positive values of m by simply replacing V_{mn} and I_{mn} by their complex conjugates. Hence 14° becomes

15° $$\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\frac{mV_{mn}I^*_{mn}}{mf_1+nf_2}=\sum_{m=0}^{\infty}\sum_{n=-\infty}^{\infty}\frac{m[V_{mn}I^*_{mn}+V^*_{mn}I_{mn}]}{mf_1+nf_2}$$

The right side of 15° is seen to consist of a quantity plus its complex conjugate, which is equal to twice the real part of that quantity. So Eq. 15° can be written as

16° $$\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\frac{mV_{mn}I^*_{mn}}{mf_1+nf_2}=\sum_{m=0}^{\infty}\sum_{n=-\infty}^{\infty}\frac{m\,2\,\mathrm{Re}[V_{mn}I^*_{mn}]}{mf_1+nf_2}=\sum_{m=0}^{\infty}\sum_{n=-\infty}^{\infty}\frac{mP_{mn}}{mf_1+nf_2}$$

where P_{mn} is the power flowing at frequency mf_1+nf_2. Here the magnitudes of V and I have been defined so that $P_{mn}=2\,\mathrm{Re}[V_{mn}I^*_{mn}]$

All that now remains is the evaluation of the double integral in Eq. 12°. Since q is periodic in both $\omega_1 t$ and $\omega_2 t$ with period the integrals over the range 0 to 2π will be identically equal to zero, as $v(q)$ is singl-valued. The desired result for the power flow in a non-linear reactance is therefore

17° $$\sum_{m=0}^{\infty}\sum_{n=-\infty}^{\infty}\frac{mP_{mn}}{mf_1+nf_2}=0$$

In exactly the same manner, a second result can be obtained involving the index n instead of m.

18° $$\sum_{n=0}^{\infty}\sum_{m=-\infty}^{\infty}\frac{nP_{mn}}{mf_1+nf_2}=0$$

These last two equations are called Manley-Rowe relations, which relate the power flow in a non-linear reactance to frequencies present.

We will illustrate the usefulness of the Manley-Rowe power relations by a specific example : Let the signal generator at f_1 represent a signal source, and the generator at f_2 the so-called pump source. Consider the case where power flow is allowed at a frequency f_3 which is the sum of f_1 and f_2 . Then , the former of the Manley-Rowe relations reduces to

$$\frac{P_1}{f_1} + \frac{P_3}{f_3} = 0$$

and the latter reduces to

$$\frac{P_2}{f_2} + \frac{P_3}{f_3} = 0$$

Since we are supplying energy to the non-linear reactance at frequency f_1 , P_1 is positive . Then the first of the above equations tells us that the power P_3 is negative ; that is , P_3 flows from the reactance and to our resistive termination at f_3 . We can define a gain which is the ratio of the power delivered by the reactance at frequency f_3 to that absorbed by the reactance at frequency f_1 .

This gain is seen to be $G_{1-3} = \dfrac{f_3}{f_1}$

This equation gives the maximum gain that can be attained by such a device .

Many non-linear media have been observed to obey Manley-Rowe power relations, which unfortunately do not apply to a non-linear resistor because they were derived for a non-linear inductance or capacitance . Therefore, they must be shown to be valid for a medium before application to the medium in question .

Methods of Analysis :

In this section we will take up the case of a nonlinear capacitor and present a method of approach to the solution of a circuit employing this capacitor although the tunnel diode is the basic component to be considered in our discussion .But since both the varactor and tunnel diode have different characteristics, the former being a variable-capacitance diode while the latter exhibits an actual negative resistance ,the inclusion of the following discussion may seem irrelevant to the subject .Nevertheless both the tunnel diode and varactor, as the variable capacitance diode is called, fall under the same category ,namely nonlinear amplifying circuit components ,so a discussion of the methods of analysis will illustrate how such problems associated with nonlinear elements can be handled .

The analysis of parametric devices often involves the analysis of nonlinear circuits .While there is, in general, no method by which exact solutions can be obtained ,various approximate methods give results which are quite satisfactory in most applications . Perhaps the most well-known method is the so-called small-signal approximation .In this approximation, it is assumed that the signal voltage is small compared to the pump voltage which is supposed to provide the exciting source for varying the junction capacitance so as to obtain the necessary nonlinear characteristics .In other words, the signal level is small relative to the local oscillator. Another method, which will be shown to be equivalent to the small-signal approach ,is that of replacing the nonlinear reactance with a time-varying linear reactance .A third method, which for the lack of a better name will be called the large signal approach, is to expand the nonlinear characteristic in a Taylor series about a d-c point and consider only the first few terms . As in the small-signal method , the signal level need not be restricted in amplitude.

Small-signal method : In the analysis of parametric devices, we are interested in computing the mixing effects that occur when voltages at two or more different frequencies are impressed on a nonlinear reactance . As a simple example, let us consider the case when two voltages are present, one at a frequency $\omega = 2\pi f$, and one at a frequency 2ω . One quantity of interest is the resultant current at frequency ω , for by dividing this current by the assumed voltage at ω we can obtain an equivalent linear admittance which the remainder of the circuit sees at ω . Once this admittance is determined , the significant properties of the parametric device can be obtained by linear analysis .

The small-signal analysis applies when one RF voltage (the signal voltage) is small compared to another voltage (the pump voltage or the local oscillator voltage to control the nonlinear capacitance) . In the study of low-noise amplifiers, for example, the small-signal approach is certainly justified since we are interested in signal voltages which are typically more than I00 db below the level of the pump . Let the two voltages be

$$v_1 = V_1 \cos \omega t \quad , \quad v_2 = V_2 \cos 2\omega t$$

where $V_1 << V_2$. Because $V_1 << V_2$, we can expand the charge on the capacitor in a Taylor series about v_2 and consider only the first two terms .

$$1^\circ \quad q(v) = q(v_2) + \frac{dq}{dv}(v_2)\, v_1$$

For convenience, let us define a capacitance by

$$2^\circ \quad C(v_2) = \frac{dq}{dv}(v_2)$$

The current through the capacitor is given by the time derivation of the charge .

3° $$i = \frac{dq}{dt} = \frac{d}{dt} q(v_2) + \frac{d}{dt}\left[C(v_2) V_1 \cos \omega t\right]$$

Since $C(v_2)$ is periodic with a fundamental frequency of 2ω, we can also expand it in a Fourier Series. Assuming that we choose our reference to make $C(v_2)$ an even function, we obtain

4° $$C(v_2) = \sum_{n=0}^{\infty} C_n \cos 2\omega n t$$

We can interpret 4° as a time-varying linear capacitance in this approximation, for we see that the second term of Eq. 3° is of the form

$$\frac{d}{dt}\left[C(t)\, v(t)\right]$$

which is the result obtained for a linear but time-varying capacitance. The C_n coefficients can be interpreted as the magnitude of each harmonic of this time-varying capacitance. That is, C_0 gives the magnitude of the constant capacitance, C_1 the magnitude of the capacitance variation at 2ω, etc.

Using Eq. 3° and 4°, the current component at the frequency ω can be computed to be

5° $$i(\omega) = \frac{d}{dt}\left[(C_1 \cos 2\omega t)(V_1 \cos \omega t)\right] = \frac{d}{dt}\left[\frac{C_1 V_1}{2} \cos \omega t\right] = -\frac{\omega C_1 V_1}{2} \sin \omega t$$

From the foregoing analysis it is seen that insofar as small-signal effects at the signal frequency are involved, a circuit containing a nonlinear reactance can be replaced by a time-varying <u>linear</u> reactance.

Large-signal Method

The preceding small-signal approach cannot be used to predict saturation effects since the signal level must remain small relative to the local oscillator level or the pump. However, if we choose to expand our relation for charge about a d-c point rather than a-c point, we need not restrict the magnitude of one RF voltage relative to another. We do need to specify, however, that the a-c excursions be relatively small so that again only a few terms in the Taylor series expansion will be necessary to adequately represent the circuit. Expanding the charge about a bias voltage V_0

6° $$q(v) \approx q(V_0) + \frac{dq}{dv}(V_0)\left[v_1+v_2\right] + \frac{1}{2}\frac{d^2q}{dv^2}(V_0)\left[v_1+v_2\right]^2$$

Here no restriction is placed on the magnitude of the signal voltage relative to the pump voltage. Let

7° $$\frac{dq}{dv}(V_0) = C_0 \qquad v_1 = V_1 \cos \omega t$$
$$\frac{d^2q}{dv^2}(V_0) = C_1' \qquad v_2 = V_2 \cos 2\omega t$$

The current component due to the nonlinear mixing is

8° $$i = \frac{d}{dt}\left[\frac{1}{2} C_1' \left(v_1 \cos \omega t + v_2 \cos 2\omega t\right)^2\right]$$

The component of this at frequency ω is

9° $$i(\omega) = \frac{d}{dt}\left[\frac{C_1' V_1 V_2}{2} \cos \omega t\right] = -\frac{\omega_1 C_1' V_2}{2} V_1 \sin \omega t$$

Comparing this result with that obtained by the small-signal analysis, we see that at frequency ω the results are identical if $C_1' V_2 = C_1$. C_1' may be interpreted as the slope of the capacitance versus voltage curve and therefore $C_1' V_2$ may be interpreted as the amplitude of the capacitance change at the local oscillator frequency.

In addition to yielding information about mixing effects at frequency ω, this large signal approach also yields similar information at the pump frequncy 2ω. It is this additional information, not obtainable from the small-signal appraoch, which is useful in predicting the effects of signal voltage on the remainder of the circuit. Thus saturation effects may be considered.

The behavior of varactors or nonlinear capacitance diodes was discussed previously. It was shown how these devices could be utilized so as to effect transfer of power at one frequency to another frequency and also how such elements could be employed for amplification. It should be pointed out here that varactors are not the only low-noise device that we have today. There is a new type of semiconductor diode which performs essentially the same job. It is the so-called Tunnel diode or Esaki diode named after its discoverer. We remarked before that the primary reason for the great interest in these new devices was their low-noise performance and compactness in size. Although tunnel diodes cannot compete with varactors in low-noise performance, they are nevertheless very useful for very high frequencies. Tunnel diodes easily lend themselves to amplification in the microwave region as well as amplification in the usual high frequency range. Therefore it is relevant to discuss the fundamental properties of tunnel diodes in order to pave the way for further discussion on their uses as amplifying devices.

Fundamentally, the tunnel diode is a two-terminal semiconductor device that displays an ac negative resistance over a portion of its current-voltage curve. This phenomenon occurs if the tunnel diode is biased in the forward direction. It is the ac negative resistance property of the tunnel diode that is exploited in a number of circuit arrangements to provide amplification, oscillation, switching, and memory functions extensively used in electronics.

The usefulness of the tunnel diode is due to the negative part of its characteristic curve. Negative resistance acts like a source of power, like a generator, results in signal power being increased, just as a positive resistance results in signal power being dissipated or reduced.

The important feature of the tunnel diode that distinguishes it from ordinary semiconductor diodes is its very high degree of doping. The most marked effect of heavy doping of semiconductor materials is the effect of the quantity of doping on the width of the transition region. The heavier the doping, the narrower the transition region. If the two types of semiconductor materials contain different amounts of impurities, the portion of the transition region in the more heavily doped material is narrower than that in the lightly doped material. The reason for these conditions is that the barrier is formed by a given number of donor ions on one side of the barrier and an equal number of acceptor ions on the other side of the barrier. The heavier the impurity concentration in the semiconductor material, the smaller the depth of penetration on that side of the junction necessary to establish a given number of uncompensated ions.

Let us now consider the effect of heavy doping on the location of Fermi levels in both the p-type and n-type materials :

The Fermi level in the n-type material is given by the expression

$$E_F = E_c - kT \ln \frac{N_c}{N_d} \qquad \text{where} \quad N_c = \left[\frac{2\pi m_e kT}{h^2}\right]^{3/2} \times 2$$

N_d = number of donor atoms
E_c = conduction energy level

Also the Fermi level in the p-type material is given by the expression

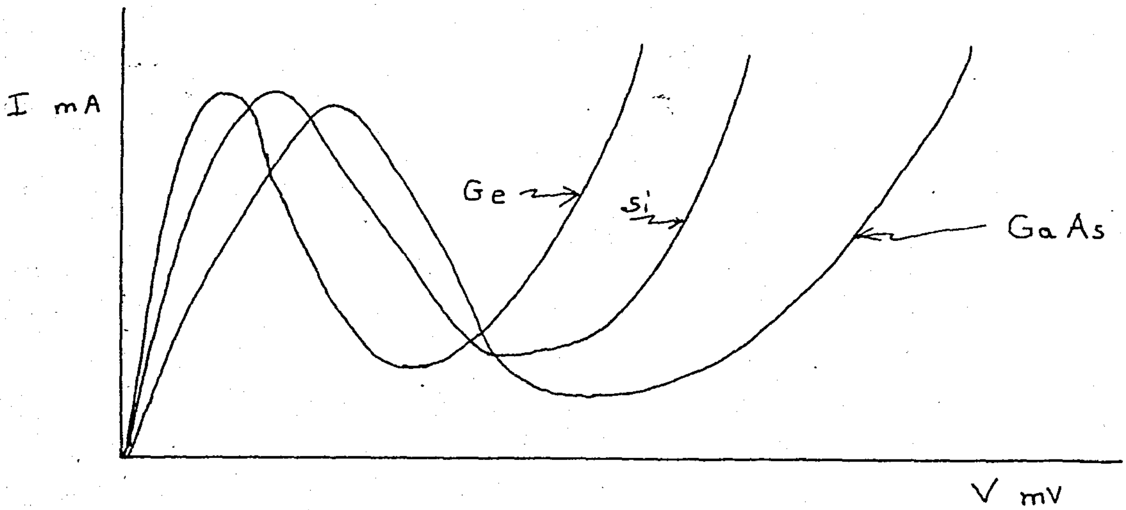
$$E_F = E_v + kT \ln \frac{N_v}{N_a} \qquad \text{where} \quad N_v = 2\left[\frac{2\pi m_h kT}{h^2}\right]^{3/2}$$

N_a = number of acceptor atoms
E_v = valence energy level

If the degree of doping is very high, then $N_d > N_c$, and the term $\ln \frac{N_c}{N_d}$ becomes negative. The result is that the Fermi level lies above E_c for the n-type material. Applying the same reasoning to the p-type material we see that the Fermi level lies slightly below E_v.
It is this characteristic of the junction formed by the abovementioned p- and n- type materials that constitutes the basis for discussion

of the characteristics of the tunnel diode .In tunnel diodes ,the degree of doping is IO to IOO times larger than in usual semiconductors. A discussion of tunnel diode materials seems appropriate at this point. Most tunnel diodes studies were based around the well-known materials, Germanium and Silicon .Almost simultaneously with this work,however, it was recognized that intermetallic compounds offered many attractive features as tunnel diode materials .This is because tunneling and tunnel diodes are not dependent upon the minority carrier lifetime, which has never been very high in the intermetallics ,and because such proprties as low effective masses and high mobilities ,which are characteristic of many of the intermetallic compounds ,are of direct benefit to tunnel diode operation .

Among the various intermetallic compounds that have exhibited tunneling action or from which succesful tunnel diodes have been fabricated ,GaAs is one of the most attractive .First of all,the wide energy gap of GaAs (I.35 - I.45 eV) makes possible tunnel diodes with a voltage swing significantly larger than those made of Ge or Si. The theoretical decrease in tunneling current due to a wider band gap is more than offset by the low effective masses of GaAs relative to those of Ge or Si .Also the larger mobilities of carriers in GaAs ,again relative to Ge and Si , make possible a reduction in diode's series resistance .Finally the lower dielectric constant (II.I) of GaAs is of some value in reducing the junction capacitance .
The general appearance of a GaAs tunnel diode v-I characteristic as compared to those of Ge and Si is shown in the figure below :



Characteristics of Ge ,Si , and GaAs tunnel diodes

Good electrical properties,high resistance to radiation damage, larger voltage swings,greater stability with temperature make GaAs tunnel diodes competitive .

Let us now briefly review the fundamental principle underlying the operation of the tunnel diode .

First ,we should note that the transition region in a tunnel diode is appreciably narrow because this is essential to the discussion. For the n-type material,the Fermi level lies above the bottom edge of the conduction band .At normal temperatures ,those energy levels slightly below the Fermi level will be filled ,moreover there will be some electrons occupying energy levels above the Fermi level .There also exist some vacant energy levels above the Fermi level . Considering now , the p-type material : The Fermi level lies a little below the top edge of the valence band .There are some unoccupied energy level above the Fermi level and within the valence band; the energy levels below the Fermi level are filled with electrons to a considerable extent;there are a few vacant energy levels below the Fermi level .

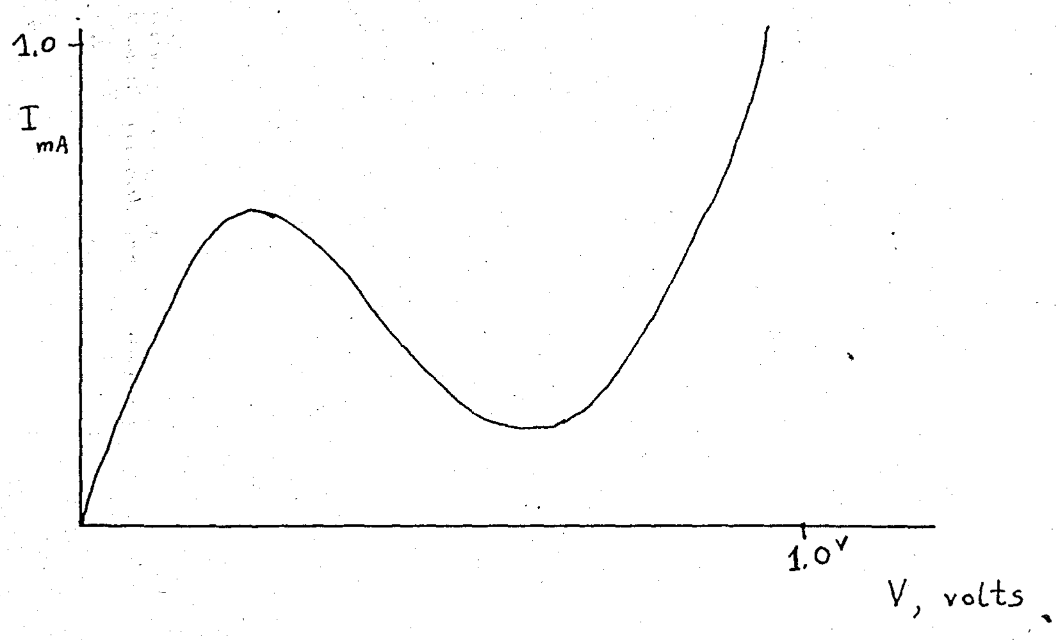
If a p-n junction is formed by such highly doped semiconductors, the Fermi levels will align .This is the case at zero bias .We see from the following figure that the valence band of the p-type material is opposite the conduction band of the n-type material. As is known ,a potential barrier is created at the junction which prevents or retards the passage of majority carriers across the junction so that an equilibrium between the majority carriers and the minority carriers is established .But ,in the case of the junction under consideration now,the electrons in the conduction band of the n-type material have two choices : either to gain sufficient energy to overcome the potential barrier at the junction and cross into the conduction band of the p-type material ;or to go through the forbidden zone right into the valence band of the p-type material. The latter phenomenon is known as electron tunneling :which can take place if there is a vacant energy level in the valence band of the p-type material at the same energy level as that of a free electron in the n-type region .In this case ,quantum mechanics predicts that the free electron in the n-type region may appear at the same enrgy level in the valence band of the p-type region on the other side of the potential barrier .This process has nothing to do with diffusion .Thus this tunneling effect forms the basis of operation of tunnel diodes .

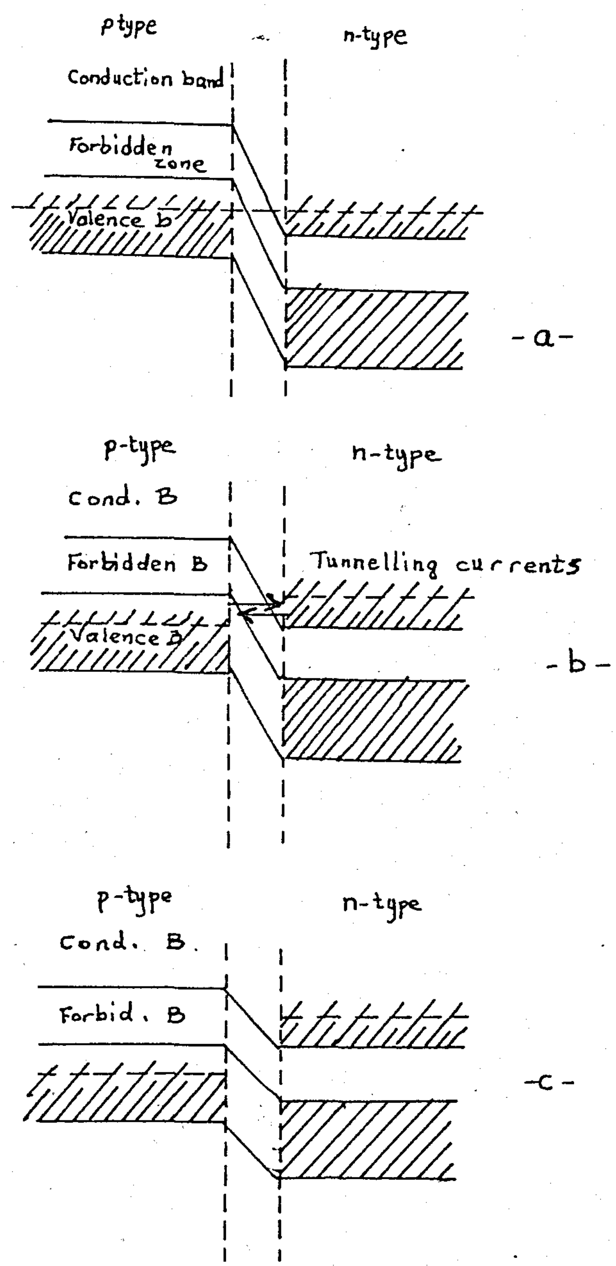
As a slight forward bias is applied,the filled levels in the n-type material take a position opposite those vacant energy levels in the p-type material ,hence electrons pass from the n-side to the p-side by tunneling ;likewise some electrons in the valence band of the p-side may pass from the p-side to the n-side ,thus constituting a reverse tunneling current .As the forward bias is increased more and more filled energy levels in the n-side come opposite those vacant energy states in the p-side so that the forward tunneling current increases while the reverse tunneling current decreases slightly .When all of the filled energy levels in the n-side are opposite all the vacant energy levels in the p-side ,the current reaches a maximum .As forward bias is increased more and more , the filled energy levels in the n-side are no longer opposite all the vacant energy levels in the p-side ,so the tunneling current begins to decrease .This drop in current goes on until the filled energy levels in the n-side are opposite the forbidden zone :therefore tunneling current stops completely .

Superimposed upon the tunneling current is the normal diode current .After the tunneling current stops ,the junction begins to function as an ordinary p-n junction .

If the tunnel diode is back-biased ,the number of minority carriers increases due to additional bonds being disrupted .Also the reverse tunneling current increases so there is a large reverse current .

A typical characteristic of a tunnel diode is shown below :





Energy Level Diagrams for a Tunnel Diode

a-Zero bias

b-Small forward bias

c-Larger forward bias

Esaki developed two equations for the two tunneling currents as follows :

$$I_{c\rightarrow v} = A\int_{E_c}^{E_v} f_c(E)\rho_c(E)\, Z_{c\rightarrow v}\left[1 - f_v(E)\right]\rho_v(E)\,dE$$

$$I_{v\rightarrow c} = A\int_{E_c}^{E_v} f_v(E)\,\rho_v(E)\, Z_{v\rightarrow c}\left[1 - f_c(E)\right]\rho_c(E)\,dE$$

where $Z_{c\rightarrow v}$ and $Z_{v\rightarrow c}$ are the probabilities of penetrating the gap (these could be assumed to be approximately equal) ; $f_c(E)$ and $f_v(E)$ are the Fermi-Dirac distribution functions ,namely ,the probabilities that a quantum state is occupied in the conduction and valence bands,respectively ; $\rho_c(E)$ and $\rho_v(E)$ are the energy level densities in the conduction and valence bands ,respectively .

When the junction is slightly biased positively and negatively, the observed current I will be given by

$$I = I_{c\rightarrow v} - I_{v\rightarrow c} = A\int_{E_c}^{E_v}\left[f_c(E) - f_v(E)\right] Z\ \rho_c(E)\,\rho_v(E)\,dE$$

From this equation, if Z may be considered to be almost constant in the small voltage range involved, we could fairly well calculate the I-V curve at a certain temperature ,indicating the dynatron-type characteristic in the forward direction .

At the very beginning of our discussion of tunnel diodes, it was pointed out that it was the degree of doping that gave the tunnel diode its primary importance. In other words, the characteristics of tunnel diodes are entirely dependent upon the degree of doping. Therefore it is pertinent here to discuss the effect(s) of non-symmetrical doping of tunnel diodes. For the final choice as to the kind of tunnel diodes in different applications, such as amplifiers, oscillators, and switching circuits, etc ..., rests upon certain parameters of tunnel diodes. So we will consider shortly the effect of such doping.

We will use the original model emplyed by Esaki, and assume that the tunneling probability is constant. The model will not be restricted to equal degeneracy energies on each side of the junction. The model is calculated for $0°K$. Comparison with graphical solutions at $300°K$ indicates that the resulting voltages derived do not change by more than a few percent.

At any given voltage, the tunnel current I_t may be found by evaluating the following integral given by Esaki :

$$I_t = A\int_{E_c}^{E_v} \left[f_c(E) - f_v(E) \right] \rho_c(E)\, \rho_v(E)\, dE$$

where the various terms are defined by Esaki. At $0°K$ the above integral may be expressed exactly by equations containing transcendental and quadratic terms. Two degeneracy voltages, V_{dv} and V_{dc} are defined as

$$V_{dv} = \frac{E_v - E_f}{e} \quad \text{and} \quad V_{dc} = \frac{E_f - E_c}{e}$$

~~wsh~~ where E_f is the fermi energy and E_v and E_c are the band edges of the p-type and n-type sides of the junction, respectively. It has been shown that a forward bias equal to $V_{dv} + V_{dc}$ causes tunnel current to cut off .We shall then define the tunnel cut-off as

$$V_{co} = V_{dv} + V_{dc}$$

A degeneracy ratio α is defined as

$$\alpha = \begin{cases} \dfrac{V_{dv}}{V_{dc}} & \text{for } V_{dv} > V_{dc} \\ \dfrac{V_{dc}}{V_{dv}} & \text{for } V_{dc} > V_{dv} \end{cases}$$

The general properties of the tunnel diode characteristic can be shown to be functions of V_{co} and α .

The peak voltage : The position of the peak voltage is determined by

$$\left.\frac{dI_t}{dv}\right]_{V_p} = 0 \qquad \text{letting} \quad V_p = \beta V_d$$

where V_d is the smaller of the two degeneracy voltages .The solutions of theequation defining the peak voltage give the following relation

$$\sin^{-1}\left[\frac{\beta-(\alpha-1)}{(1+\alpha)-\beta}\right] + \sin^{-1}\left[\frac{\beta+(\alpha-1)}{(1+\alpha)-\beta}\right] = \frac{4}{(1+\alpha)-\beta}\left\{[\alpha-\beta]^{\frac{1}{2}} + [\alpha(1-\beta)]^{\frac{1}{2}}\right\}$$

Resulting solutions of $\beta(\alpha)$ are given in Fig . 1 .It can be seen that in the limit of large α $V_p \cong \frac{1}{1+\alpha} V_{co}$

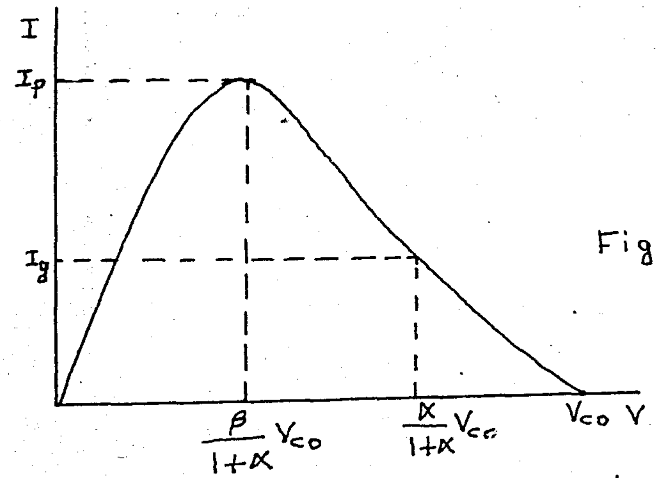
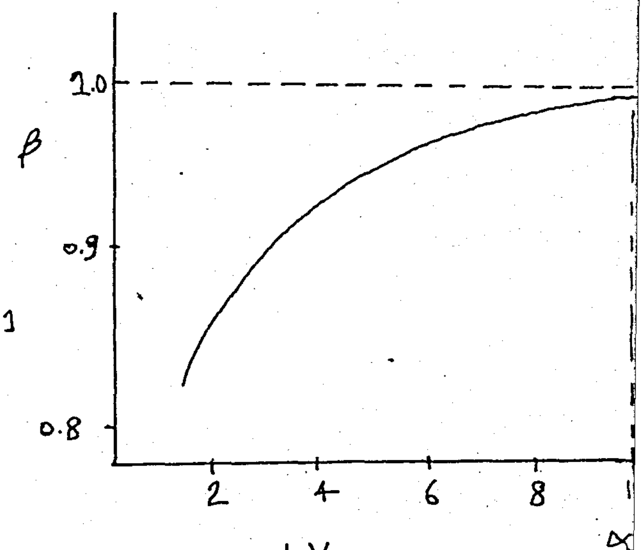


Fig.1



Variation of V_p and V_g with α and V_{co}

Maximum negative conductance G_m :

The condition of maximum negative conductance is :

$$\left.\frac{d^2 I_t}{d v^2}\right]_{V_g} = 0$$

This was shown to occur at $V_g = \frac{\alpha}{1+\alpha} V_{co}$

The current at the peak negative conductance was

$$I_g = \frac{A\ Z\ \pi\ V_{co}^2}{8} \left(\frac{1}{1+\alpha}\right)^2$$

with a resulting peak negative conductance

$$G_m = \frac{A\ Z\ \pi\ V_{co}}{4} \left(\frac{1}{1+\alpha}\right)$$

where A and Z were defined previously .The peak current was found to be related to I_g as follows :

$$\frac{I_p}{I_g} = \frac{4}{\pi}\left(\sqrt{\alpha-\beta} + \alpha\sqrt{\alpha(1-\beta)}\right)$$

It should be noted that, although the voltages derived remain nearly constant with temperature, the currents can vary by as much as 2 to I between $0°K$ and room temperature .

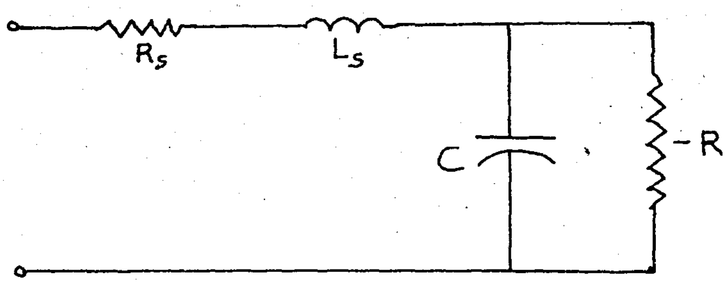
We have shown so far that the properties of the tunnel diode are quite dependent upon the parameters α and V_{co} .

We ,see ,from the circuit point of view that, given a constant V_{co} ,the impedance level would rise and the current level fall at G_m with increasing α .Both of the above results indicate that diodes with a large α would be more suitable for use in amplifiers, and conversely, diodes with $\alpha \approx 1$ would be more suitable for oscillators and switches .

It has also been shown that the peak current of a tunnel diode is also dependent upon the temperature .The peak current temperature coefficient may be positive or negative depending upon other factors .

Our discussion has so far been concerned with the tunnel diode itself, its properties ,construction etc . From now on we will enter into a discussion of circuit principles for tunnel diode applications. As is always known,the basic step in the study of circuit aspects is the development of an equivalent circuit to facilitate the analysis of circuits .So we will first construct an equivalent circuit for the tunnel diode and then discuss various amplification possibilities. The tunnel diode itself is represented by a negative resistance $-R$ shunted by a capacitance which is the junction capacitance . The series resistance R_s includes the lead resistances as well as the losses in the diode.The series inductance L_s also represents the inductance introduced by parasitic elements and leads .In the ideal diode these parasitic series resistance and inductance should not exist .Since they cannot be done away with,they should be minimized .As an example,for the application of tunnel diodes in the microwave region it is important that the series inductance of the diode is well below one nanohenry .There are various methods to minimize this parasitic series inductance .References are given in the appendix .

The equivalent circuit for a tunnel diode is shown below :

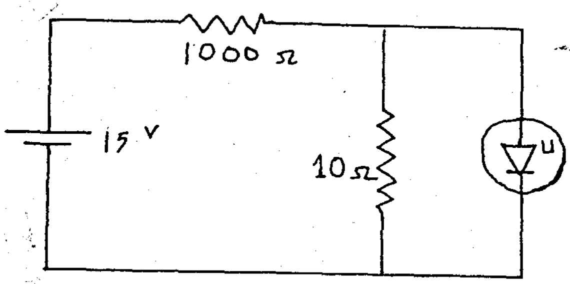


We have previously pointed out that it is the negative resistance region of the tunnel diode characteristic that is of primary importance .If the tunnel has no d-c bias applied to it ,the input voltage swing may or may not extend into the negative resistance region even if it does extend into the negative resistance region, the tunnel diode will not be useful as far as amplification and oscillation are concerned .The diode will exhibit a negative resistance during some portion of the input voltage cycle while it will have a positive resistance during some other interval ,depending upon the swing of the input voltage .Such an operation is very much undesirable if the tunnel diode is to be used either as an amplifier or an oscillator. Therefore a need arises for biasing the tunnel diode in the negative resistance region so as to employ the negative resistance .If the tunnel diode is to be used for amplification purposes, the operation should be limited to the negative resistance region .This imposes a limitation on the amplitude of the input signal .

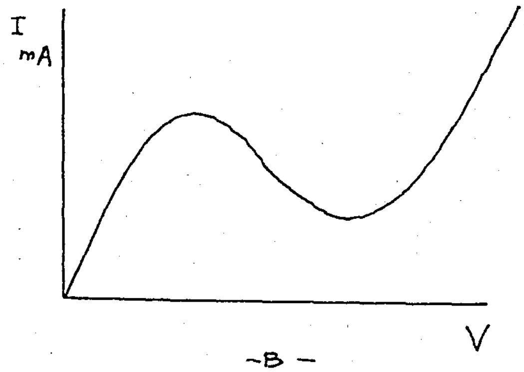
After recognizing the need for biasing the tunnel diode, we will briefly consider the biasing methods for tunnel diodes .

There are two main categories of d-c operation of the tunnel diode, and each is affected by the total series resistance in the circuit .In one category, the tunnel diode operates in the switching mode, and in the other category ,the tunnel diode operates as an oscillator or amplifier .

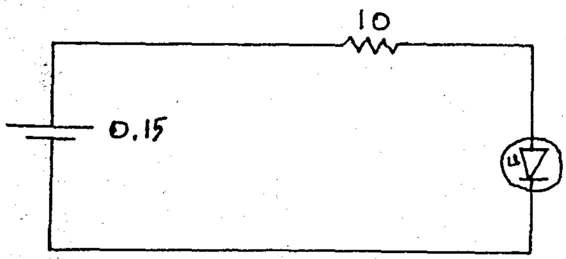
The second category of d-c operation biases the tunnel diode in the negative resistance region ,as shown in the figure below :



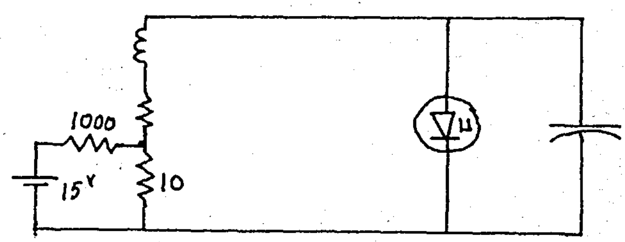
-A-



-B-



-C-



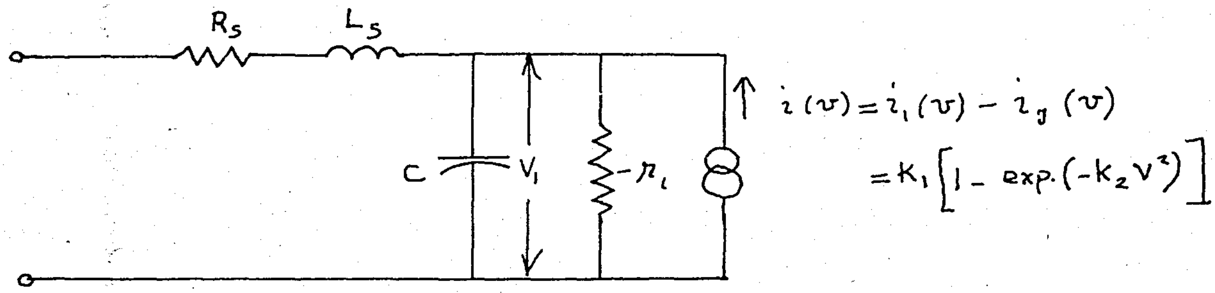
-D-

For d-c load line analysis ,it is convenient to replace A by the circuit shown in C . In D ,a typical amplifier circuit with the bias circuit included is shown .For reasons to be explained later the sum of the d-c resistance of the coil and the bias supply in D must be less than the magnitude of the negative resistance of the tunnel diode .It is also necessay that the total series a-c resistance be positive .

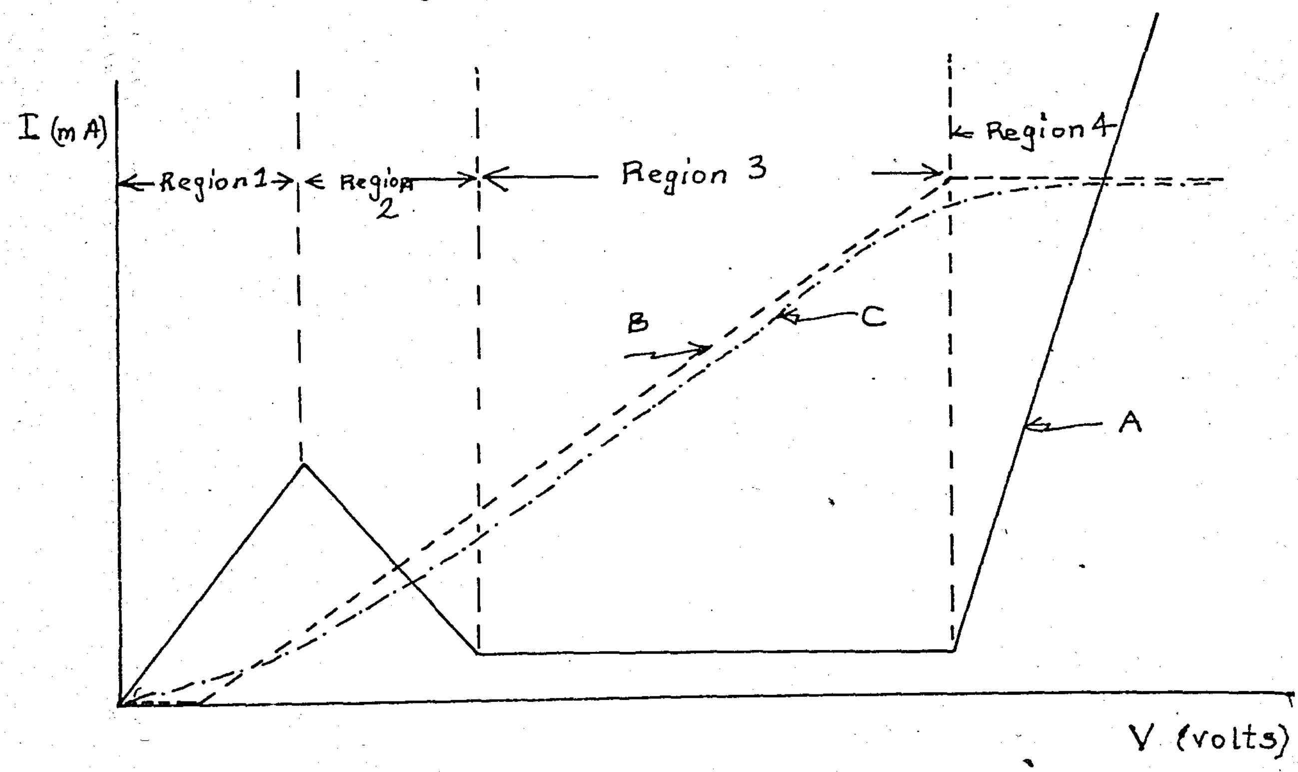
Since we are at a stage of developing an equivalent circuit for the tunnel diode and biasing it ,it is appropriate to add somemore to this discussion .

Most of the equivalent circuit models such as the one we gave in the forgoing pages have been small-signal types.The small signal operation of this semiconductor device can be correctly characterized by its incremental resistance at a fixed bias point with associated reactive elements.However, the large-signal departures from this bias point which typify bistable swituhing and relaxation oscillations make the model inadequate .So a large signal equivalent circuit for the tunnel diode is needed .Such a circuit is shown below. This circuit characterizes both the static and dynamic responses of the device for all models of operation .The network inclosed in the dashed box simulates the static V-I characteristic of the tunnel diode .

The validity of the small-signal tunnel diode model has been extended over a large-signal range by incorporating the static terminal characteristics of the device into the model shown .The lead inductance L_s ,the series body resistance R_s and the barrier capacitance C are assumed to remain constant over the operating range of interest .



Large signal tunnel diode equivalent circuit



Curve A is an approximation to tunnel diode characteristic.

Curve B is a broken-line approximation to the current from the $i(v)$ generator.

Curve C is analytic approximation to the current from the $i(v)$ generator.

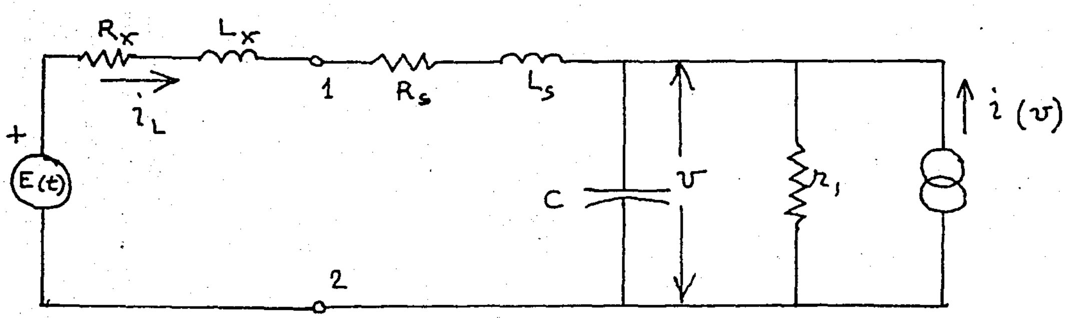
The output from the $i(v)$ generator is defined by either

$$i(v) = i_1(v) - i_j(v)$$

or by an analytic approximation to curve B.

$i_1(v)$ is defined by the linear equation relating v and i in region 1 of curve A. i_j is defined by the linear equations relating v and i in each of the j linear regions (j is 1,2,3,4) of curve A. The resistance r_1 is set equal in value to the inverse slope of region 1 of curve A.

The following pair of equations can be written on the tunnel diode circuit shown below, with which all principal modes of operation such as switching and oscillating can be realized.



Tunnel diode circuit contains external Rx and Lx elements.

$$C\frac{dv}{dt} + \frac{v}{r_1} - k\left[1 - \exp(-k_2 v^2)\right] - i_L = 0$$

$$L\frac{di_L}{dt} + R\, i_L = E(t) - v$$

$R = R_x + R_s$

$L = L_x + L_s$

$E(t)$ = applied voltage

Combining the two above equations produces a nonlinear differential equation whose solution yields the junction voltage ,V , i.e

$$\frac{d^2v}{dt^2} + \frac{1}{C}\left[\left(\frac{1}{r_1}+\frac{RC}{L}\right) - 2\,k_1k_2\exp(-k_2v^2)\right]\frac{dv}{dt}$$

$$+\ \omega_o^2\left[\left(1+\frac{R}{r_1}\right)v - Rk_1\left[1-\exp(-k_2v^2)\right]\right] = E(t)\,\omega_o^2$$

where $\omega_o^2 = \frac{1}{LC}$ .This equation does not lend itself to a closed form solution .It falls into the class of nonlinear equations of the type

$$\frac{d^2v}{dt^2} + f(v)\frac{dv}{dt} + g(v) = h(t)$$

After some investigation a graphical solution for this form was found and techniques for its application were worked out .The solution process begins with an integration of the combined equation with respect to time which yields

$$\frac{dv}{dt} + F(v) + \int_0^t g(v)\,dt = H(t) + K$$

where $F(v) = \frac{1}{C}\int_0^v \left[\left(\frac{1}{r_1}+\frac{RC}{L}\right) - 2\,k_1k_2v\exp(-k_2v^2)\right]dv$

$$g(v) = \omega_o^2\left[\left(1+\frac{R}{r_1}\right)v - Rk_1\left[1-\exp(-k_2v^2)\right]\right]$$

$$H(t) = \omega_o^2\int_0^t E(t)\,dt$$

K = integration constant

The last equation is converted into incremental form

$$\frac{\Delta v_n}{\Delta t} + F(v_n) + S_n(t) = H(n\Delta t) - K$$

where $S_n(t) = \int_0^t g\left[v(t)\right] dt$

A fixed value of Δt is chosen and the angle $\alpha = \arctan \frac{1}{\Delta t}$ is determined. It may be necessary to multiply the equation in incremental form in order to produce useful α's. is corrected to accommodate the departure of the abscissa-to-ordinate scaling of $F(v)$ v.s v from a one-to-one numerical relation to graphical solution as

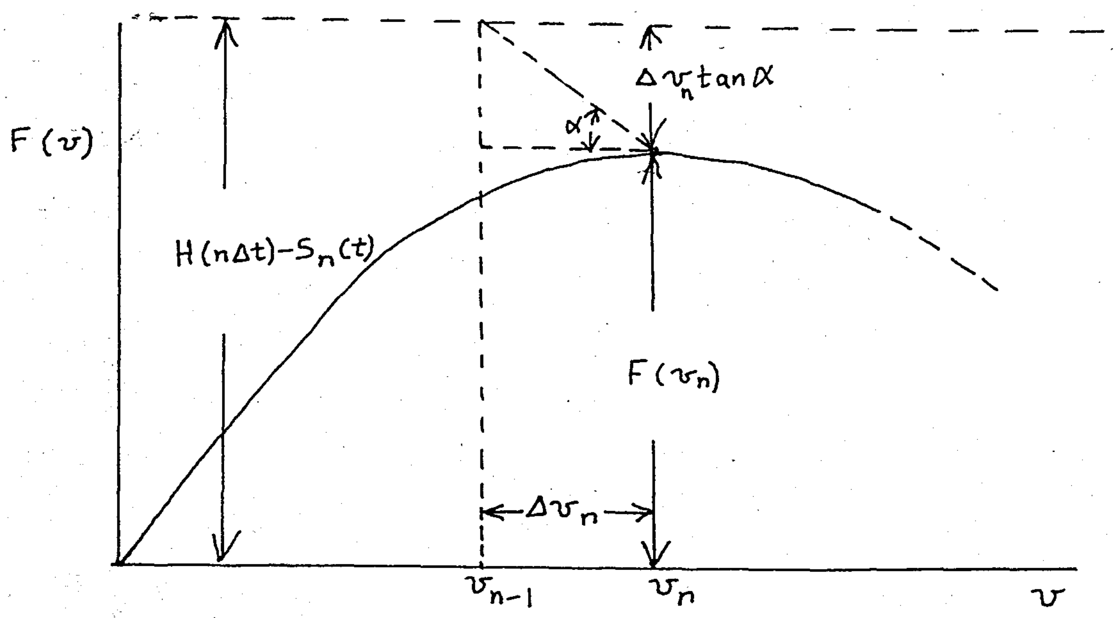
$$\Delta v \tan\alpha + F(v_n) = H(n\Delta t) + K - S_n(t)$$

The angle α is constructed physically on a template. Vertical lines of length $H(n\Delta t) - S_n(t) + K$ are constructed in the $F(v)$ v.s v plane as the solution proceeds. The successive Δv_n's are determined by the construction technique shown in the typical segment in the figure . $S_n(t)$ is determined at each step of the process by applying the trapozoidal rule to the values of $g(v_n)$ and $g(v_{n+1})$. This will be further detailed. Values of v as a function of time such as $v(n\Delta t)$ and $v\left[(n+1)\Delta t\right]$ are determined by the successive constructions in the $F(v)$ vs v plane. Then the values of $g\left[v(n\Delta t)\right]$ and $g\left[v(n+1)\Delta t\right]$ are directly obtained from the $g(v)$ vs v plot. As the construction proceeds each quantity

$$S_n(t) = \left[\frac{g(v\, n\Delta t) + g(v(n+1)\Delta t}{2}\right]\Delta t$$

is added on to the preceding total value of $S_n(t)$ and this new total is subtracted from the next applied $H(n\Delta t)$. This gives the length of the vertical construction line, $H(n\Delta t) - S_n(t)$

As each construction is made, the time required for each Δv_n to occur is known from the relationship of Δt to the angle α employed in the construction.



Typical construction increment in the $F(v)$ v.s v plane

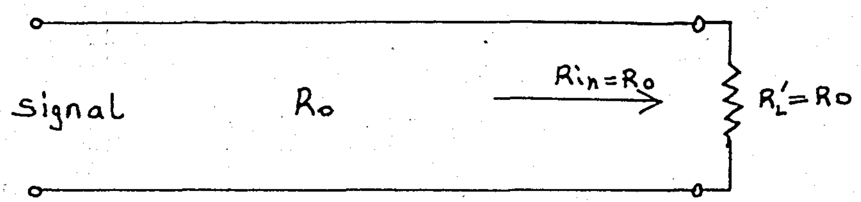
When the $H(n\Delta t) - S_n(t)$ line falls below the $F(v)$ curve the construction revereses direction and proceeds towards the origin. The angle α may be changed to secure maximum resolution and solution speed .

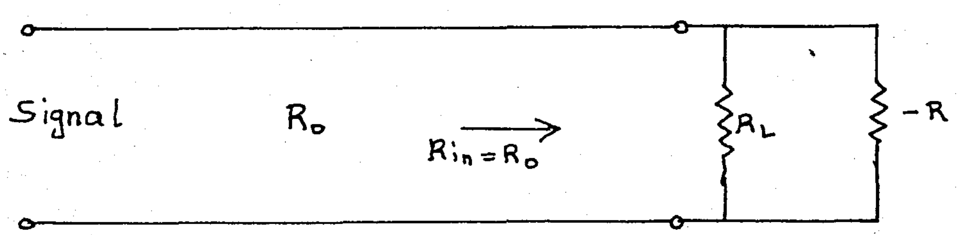
The discussion thus far illustrates pretty well the method of analysis of large-signal operation of the tunnel diode .The results obtained from the analysis given above agree fairly well with the actual ones .

In our discussion of amplifiers using tunnel diodes as their basic element large signal behavior is not considered .It is assumed that the input signal and the d-c bias are such that the device operates in the negative resistance region .

Amplification through tunnel diodes :

The tunnel diode has been a promiseful device for amplification purposes .We have repeatedly indicated that a negative resistance may be used for amplification .Let us now attempt to illustrate how a tunnel diode can be used as an amplifier .Consider a signal from a source (such as a line) at characteristic impedance R_o as shown in the figure .If we are to obtain maximum power from this signal we must match the line to a load which is also R_o ,and any deviation from this will results in a reduction of power .It is possible to increase the effective value of a resistance,as seen across its terminals ,by placing a negative resistance in parallel with it (just as a positive resistance would reduce the value) Thus we may reduce the value of the load resistance R_L in the figure ,while still presenting a load of R_o to the line ,by placing a negative resistance in parallel with R_L .As the load presented to the line is unchanged, the signal voltage will be unchanged .However the value of the actual load resistance is decreased,and the current ,and therefore the signal power in the load is increased .





$$R_o = \frac{-RR_L}{R_L - R}$$

if $R_L < R$

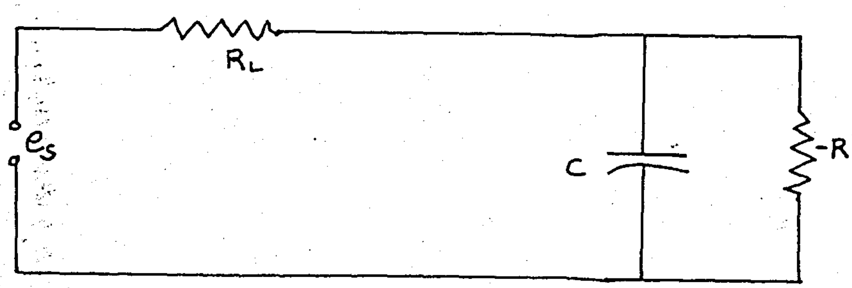
$R_o > 0$

The example discussed above illustrates the use of tunnel diode as a power amplifier only ;for the voltage across the load remains the same since the load presented to the line is the same in both cases. There is ,however ,a different form of circuit through which voltage amplification is possible .

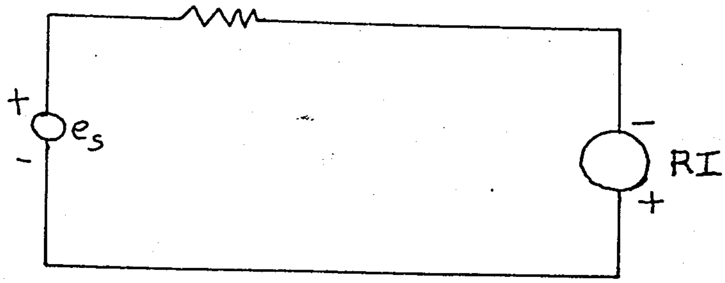
Therefore there are, in general, two basic types of tunnel diode amplifiers : the series amplifier ,and the parallel amplifier. We will discuss both types .

A- <u>Series Amplifier</u> :

In this type ,the load ,the tunnel diode ,and the signal source are all connected in series .Such an arrangement is shown below :



If we neglect the junction capacitance ,we may still devise another equivalent circuit for the tunnel diode ,which illustrates best why a negative resistance acts as a generator which <u>aids</u> the applied voltage .For this purpose we draw upon the Compensation Theorem which states that any resistance can be replaced by a generator whose voltag is equal to product of the current through the resistance and the resistance .If we deal with a positive resistance,this generator has its polarity opposing the applied voltage .In the case of negative resistance the polarity is such as to increase the applied voltage. This form of equivalent circuit is also shown below .



The equation below can be written for the former circuit :

$$e_s = (R_L - R)\, i \quad \therefore \quad i = \frac{e_s}{R_L - R}$$

The output voltage is equal to

$$e_o = i\, R_L = \frac{e_s R_L}{R_L - R}$$

From these the voltage gain is

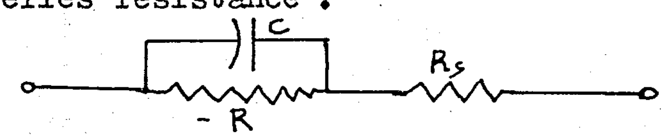
$$A_v = \frac{R_L}{R_L - R}$$

Likewise the power gain can be determined to be

$$G = \frac{R_L^2}{(R_L - R)^2}$$

In the above equations we neglected the junction capacitance .If the capacitance is taken into account,the equations will have to be modified .In fact the effective negative resistance of the tunnel diode depends upon the shunting effect of the capacitive reactance of the junction capacitance .

The above consideration leads us to suspect that there may be some limit to the frequency of operation of the tunnel diode. Let us depict the tunnel diode equivalent circuit including the series resistance .



The two-terminal impedance is given by the expression :

$$Z = R_s - \frac{R}{1+(\omega RC)^2}$$

For amplification $\quad R_s < \dfrac{R}{1+(\omega RC)^2}$

When the real part is zero :

$$R_s = \frac{R}{1+(\omega RC)^2} \quad \therefore \quad f_c = \frac{1}{2\pi RC}\sqrt{\frac{R}{R_s}-1}$$

is called "the resistive cutoff frequency :" This sets a limit upon the frequency of operation .

B- Parallel Amplifier :

Amplification can also be achieved by placing a tunnel diode in parallel with a load resistor .We discussed this before .

If the signal source does not have any internal impedance ,then there is no voltage amplification in a parallel arrangement .If the signal source has an internal resistance ,the equations for voltage and current gains will be modified .In this case the gain formulas are given below :

Current gain $\quad \dfrac{G_T}{G_T - G}$

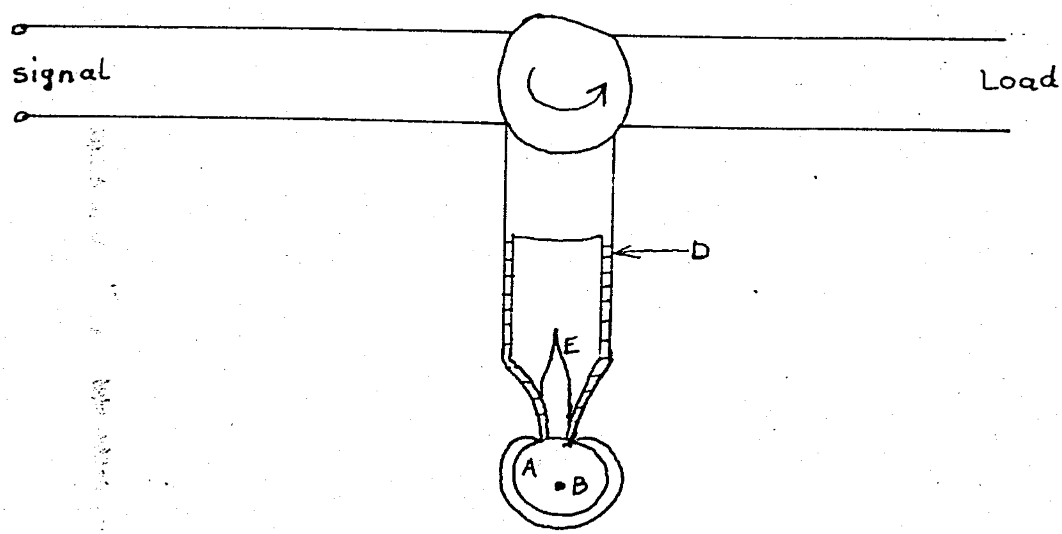
Voltage gain $\quad \dfrac{|-R|}{|-R| - R_T}$

Power gain $\quad \dfrac{|-R|^2}{(|-R| - R_T)^2}$

R_T = total resistance including the source resistance

R = tunnel diode resistance

We will briefly look at the microwave version of a tunnel diode amplifier .The diagram below shows a simple tunnel diode microwave amplifier :



The amplifier consists of single-port cavity A, containing the Esaki diode B ,which is coupled via an iris to a short section of a waveguide beyond cutoff C ,separating it from the main prepagating waveguide D .The incident input signal and the reflected (and amplified) output signal are separated by a low-loss circulator ,which is a device which does not absorb energy but simply directs it in the desired direction .

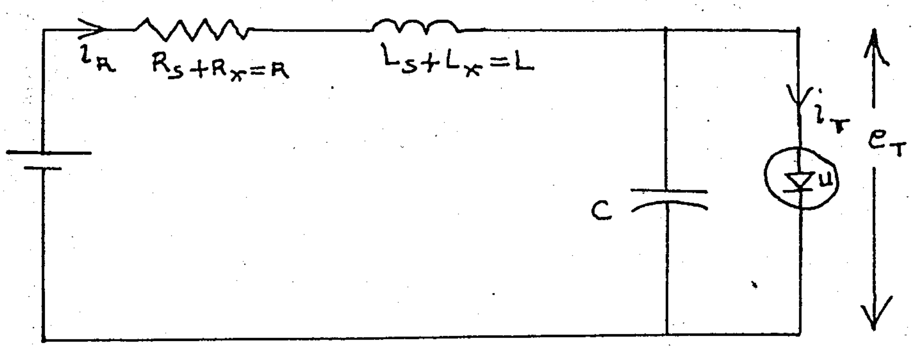
Sufficient loading must be provided to prevent oscillation from taking place .This is achieved by a dielectric plug E ,which, when completely inserted into the cutoff section, enables the latter to propagate .The depth of insertion is thus used to vary the external coupling of the cavity .It can be shown that, for high gain, the voltage gain-bandwidth product is given by

$$\sqrt{G}\,\Delta f \cong \frac{1 - \frac{Q_d}{Q_c}}{\pi RC}$$

where $\frac{Q_d}{Q_c}$ is the ratio of the power dissipated inside the cavity to that generated by the negative resistance of the diode .R and C are the negative resistance and the capacitance of the diode. G is the power gain, Δf is the bandwidth .

So far amplification using tunnel diodes was discussed ,but without any regard to whether or not such circuits would be stable .By definition stable operation is that mode of operation in which all amplitudes remain finite.If ,somehow,the current through the circuit keeps on increasing indefinitely,then the circuit is said to be unstable .Therefore it is essential to choose the parameter values in the circuit in such a way that the circuit constructed will be stable .So we recognize the need for an analysis of stability ; otherwise trial-and error methods will be too lengthy to resort to .

In practice ,there is a variety of circuits incorporating tunnel diodes .If an attempt is made to analyze all of them ,too much time will be wasted .In the literature there various articles dealing with the stability problem of various types of tunnel diode amplifier . However ,we preferred to analyze the basic tunnel diode circuit for stability,thus indicating the method of approach to such a problem . This basic circuit will consist of a tunnel diode with its series resistance and inductance together with external series resistance and inductance .In the analysis we also take into account the nonlinearity in the negative conductance region .



Let the tunnel diode nonlinear voltage-current characteristic be represented by the relation

$$i_T = \phi\ (e_t)$$

Then the circuit equations are

$$L\frac{di_R}{dt}+Ri_R+e_t=E$$

$$C\frac{de_t}{dt}+\phi(e_t)=i_R$$

Singular or equilibrium points occur where simulthaneously

$$E=Ri_{R_s}+e_{t_s}$$
$$i_{R_s}=\phi(e_{t_s})=i_{T_s}$$

The singular points could ,of course ,have been found by drawing the load line on the (e_t , i_T) characteristic .The intersection of the load line with the characteristic determines ths singular points.

In the neighborhood of a singularity ,we can make the substitutions

$$i_R=i_{R_s}+i$$
$$e_t=e_{ts}+e$$

where i and e are small changes ; subscript "s" refers to quiescent values .

Expanding $\phi(e_t)$ by Taylor series ,we have

$$\phi(e_t)=\phi(e_{ts})+\phi'(e_{t_s})e+\cdots\cdot$$

Neglecting the higher terms

$$\phi(e_t)=\phi(e_{ts})+ge$$

where $g=\phi'(e_{ts})$ is the variational conductance at the singularity .

Hence the circuit equations reduce to

$$\frac{di}{dt}=\frac{1}{L}\left[-Ri-e\right] \qquad \frac{de}{dt}=\frac{1}{C}\left[i-ge\right]$$

$$\frac{di}{de}=\frac{1/L}{1/C}\left[\frac{-e-Ri}{-ge+i}\right]$$

We know that,for a differential equation

$$\frac{dy}{dx} = \frac{cx+dy}{ax+by}$$

the characteristic eqution is

$$\lambda^2 - (a+d)\lambda + (ad-bc) = 0$$

In the present case $a = -\frac{g}{C}$, $c = -\frac{1}{L}$, $b = \frac{1}{C}$, $d = -\frac{R}{L} = -\frac{1}{GL}$

Therefore the characteristic equation is

$$\lambda^2 + \left(\frac{g}{C} + \frac{1}{GL}\right)\lambda + \frac{1}{LC}\left(1 + \frac{g}{G}\right) = 0$$

and the characteristic roots are

$$(\lambda_1, \lambda_2) = -\frac{C+GLg}{2LCG} \mp \frac{1}{2LCG}\left[C^2 + G^2L^2g^2 - 2GLCg - 4G^2LC\right]^{1/2}$$

$$= \frac{1}{2LCG}\left[A \mp \sqrt{\Delta}\right]$$

where $A = -(C + GLg$

and $\Delta = C^2 + G^2L^2g^2 - 2GLCg - 4G^2LC$

The nature of the roots depends on the four parameters L, G, C , and g . We can study the effect of these parameters on the stability by considering three two-dimensional cross sections (G,g), (L,g), (C,g) . Of course, the parameters L, C and G can have only positive values, whereas g can have both positive and negative values as determined by the characteristic .

I- The (G,g) plane :

From the last equation , we have

$$\Delta = (C - LGg)^2 - (2G.\sqrt{LC})^2$$

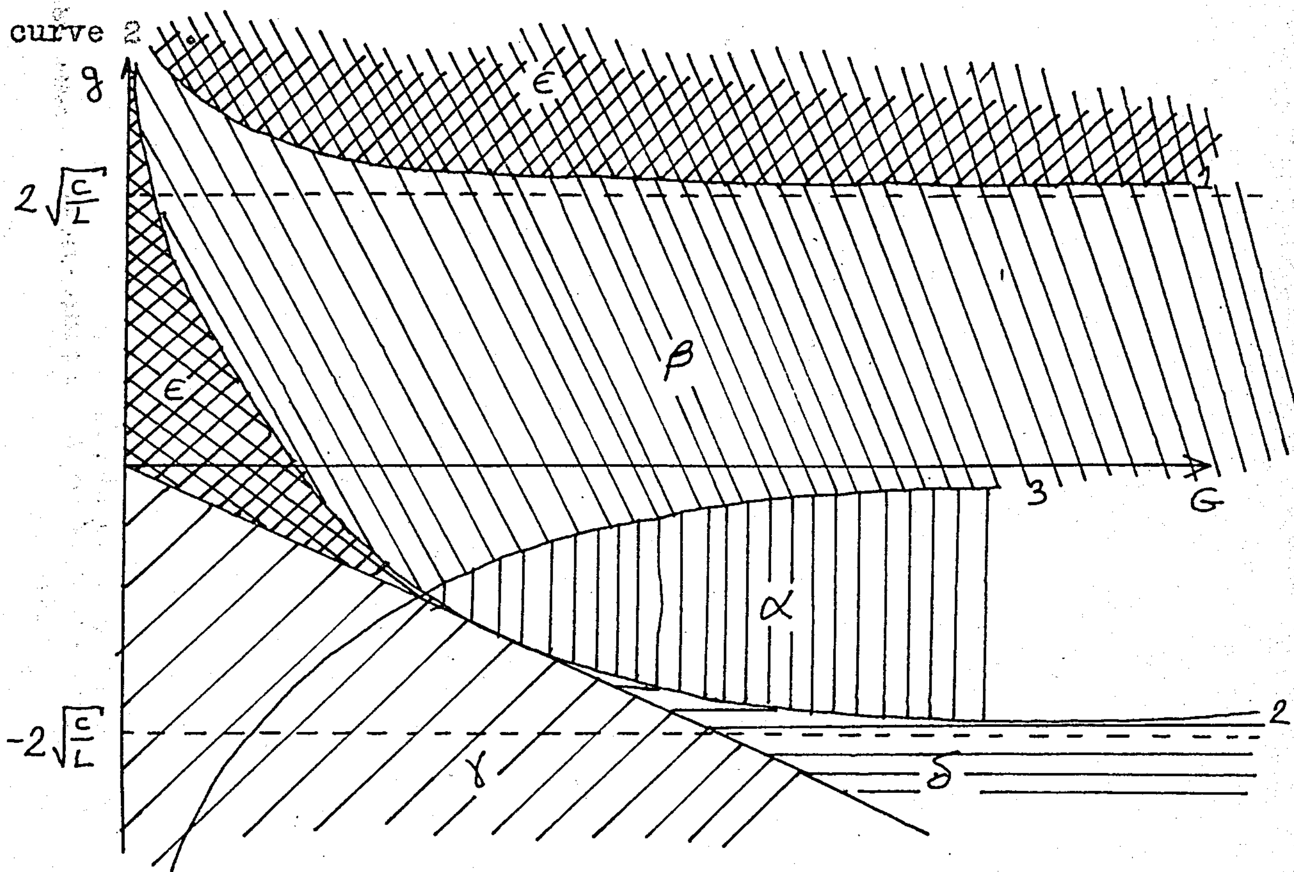
The condition for complex roots is

$$(c - LGg)^2 - (2G\sqrt{LC})^2 < 0$$

Consider the two equations

$$c - LGg + 2G\sqrt{LC} = 0$$
$$c - LGg - 2G\sqrt{LC} = 0$$

These two equations represent hyperbolas in the (G,g)plane .These are shown by curves I and 2 in the figure below .The g axis is an asymptote for both hyperbolas .Also,the line $g = 2\sqrt{\frac{c}{L}}$,is an asymptote to curve I,whereas the line $g = -2\sqrt{\frac{c}{L}}$ is an asymptote to curve 2



Hence the region between the hyperbolas correspond to the region where the complex roots of the solution for $\lambda_{1,2}$ are situated .Since the complex roots correspond to focal points,we may deduce that the focal points are situated in the region between the two hyperbolas. These focal points are unstable if

$$C + GLg < 0$$

and stable if

$$C + GLg > 0 \quad \therefore \quad \frac{1}{G} > \frac{Lg}{C}$$

The equation $C + GLg = 0$ represents a hyperbola with the g and G axes as asymptotes ,which is shown as curve 3 in the above figure .

Hence the region between curves I and 2 ,but above curve 3, corresponds to stable focal points and that below curve 3 to unstable focal points .

Therefore,we have,in region α ,that self-excitation of oscillation is possible since the oscillations increase on the divergent spiral starting from an unstable focal point .Also ,in region β ,self-excitation is not possible .

When $C + GLg > 0$,the real part of the expression for is negative ,so the circuit is stable .

Saddle points arise if

$$G + g < 0 \quad \therefore \quad \text{Stable when } G + g > 0 \therefore \frac{1}{g} > \frac{1}{G}$$

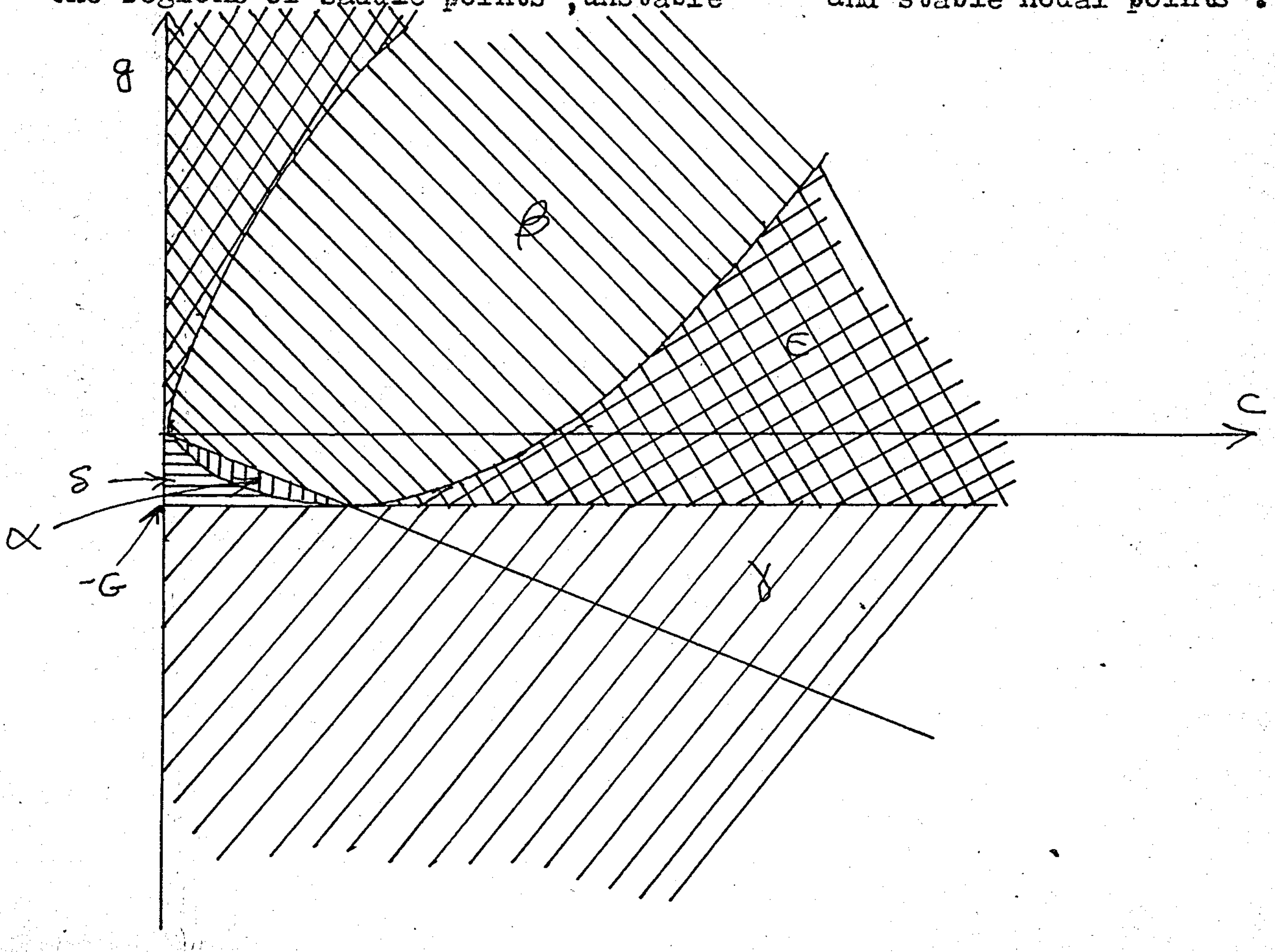
Hence,the region below the straight line $G + g = 0$ corresponds to the region of saddle points ,and is a region of instability . The co-ordinates of the points of intersection of curves I and 2 are $\left(\sqrt{\frac{C}{L}}, -\sqrt{\frac{C}{L}}\right)$ and the straight line passes through this point .Since the intermediate region separating regions of saddle points and focal points is a region of nodal points .The regions δ and ϵ ,respectively,correspond to the regions of unstable and stable nodal points .Thus the point X is a point of separation for the different types of singularities .

2- The (C,g) Plane

Let us consider the equation $\Delta = 0$,or

$$C^2 + (GLg)^2 - 2GCLg - 4LCG^2 = 0$$

This is a parabola in the (C,g) plane,passing through (0,0) .This parabola has a vertical tangent at the origin and a horizontal tangent (G^2L, $-G$) .Thus ,the focal points are situated inside this parabola .As before the stable and unstable focal points may be separated by drawing the curve $C + LGg = 0$,which is a straight line through the origin with a slope of $-\frac{1}{GL}$ .The regions α and β ,respectively,correspond to the regions of unstable and stable focal points .Thus ,self-excited oscillations are possible only for the negative values of g and for not too large values of C . The region of saddle points is obtained by drawing the line $G + g = 0$ As before, the regions γ , δ , and ϵ ,respectively ,correspond to the regions of saddle points ,unstable and stable nodal points .



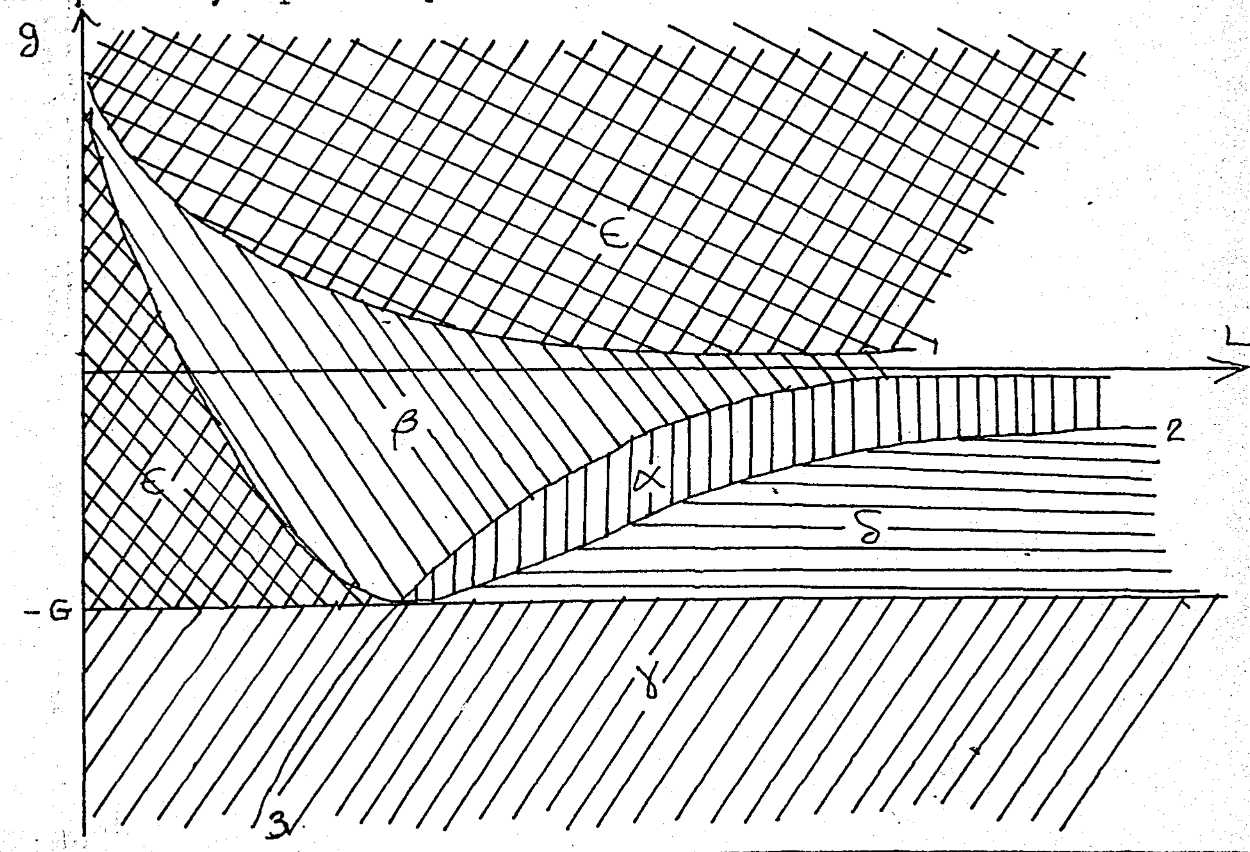
3- The (L,g) Plane :

Again considering the equation $\Delta = 0$,we have

$$C - GLg + 2G\sqrt{C}\,.\sqrt{L} = 0$$

$$C - GLg - 2G\sqrt{C}\sqrt{L} = 0$$

These two equations are plotted in the diagram below ,shown as curves I and 2 .The g and L axes are asymptotes to curve I .Curve 2 has the g axis as an asymptote and a horizontal tangent at $\left(\frac{C}{G^2}, -G\right)$. To find the regions of unstable and stable focal points ,we draw the curve $C + GLg$,which is a hyperbola with the g and L axes as asymptotes ,and passing through the point x ,where curve 2 has a horizontal tangent .The regions α and β are the regions of unstable and stable focal points .The region of saddle points is obtained by drawing the line $G + g = 0$ .The regions γ , δ and ϵ represent ,as before ,the regions of saddle points ,unstable and stable nodal points ,respectively .



Conclusion :

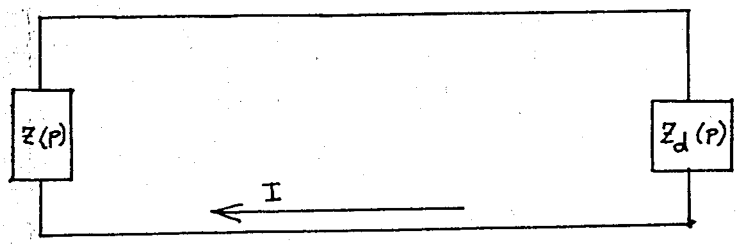
We may summarize the stability criteria as follows :

$$\frac{1}{g} > \frac{1}{G} > \frac{Lg}{C}$$

There is also the question of resistive cutoff frequency and the self resonant frequency .For stable operation it is also necessary that the resistive cutoff frequency be below the self resonant frequency ;otherwise the circuit will oscillate and will become unstable .

The "plane" analysis given thus far may be extended further to investigate the stability of various points on the tunnel diode characteristic ,which are the intersection of a load line with the characteristic curve .

There could be a different approach to the stability investigation of a tunnel diode .But the answers that are obtained agree with those we deduced in this section .So we will state one more thing in connection with the stability of tunnel diode :

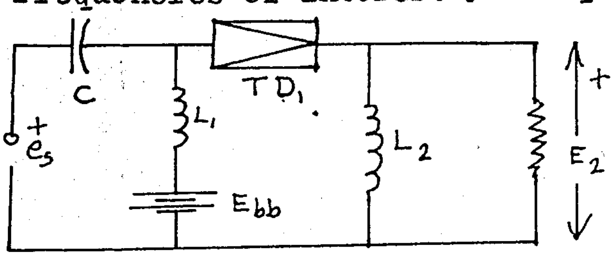


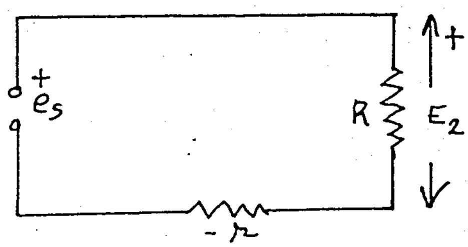
Considering the above circuit , $Z_d(p)$ the two-terminal device is potentially stable if there exists at least one finite positive real function $Z(p)$ such that the equation $Z(p) + Z_d(p) = 0$ has no solutions in the closed-right half s-plane (or the "p" plane) Re $p \geq 0$ .Then $Z(p)$ is said to stabilize $Z_d(p)$.No oscillations of ever increasing amplitudes are possible in this case .

We have considered the stability of tunnel diode only in the

the basic circuit. In general, depending upon the form of the circuit the stability conditions will differ; nevertheless the approach is the same for all, namely that there are no solutions whose real parts are positive.

Previously various gain formulas were derived for voltage, current and power gain. Consider a simple tunnel diode amplifier circuit whose voltage gain we know; if we are to ignore the stability conditions, we see that the voltage gain can be increased as large as we please by decreasing the value of the external resistance or by increasing the value of tunnel diode negative resistance so as to make the difference between them minimum so that we obtain a very large voltage gain. But this is not so in practice because at some point the circuit may become tottally unstable. We gather then that the amplification that can be obtained through a single tunnel diode stage has an upper limit if the circuit stability is not to be disturbed. But if the maximum allowable gain per stage is less than the required over-all gain, then several amplifier stages must be cascaded. Let us take the basic circuit below: here we assume that and are such that they act as open circuits at all frequencies of interest. The equivalent circuit is also shown.

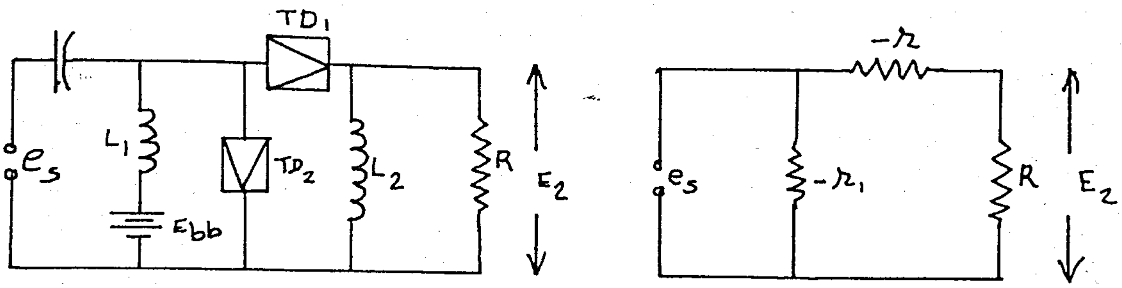




The voltage gain of this circuit is given by $K = \frac{E_2}{e_s} = \frac{R}{R-\rho}$

Let us now consider another circuit in which we added another tunnel diode with a different negative conductance. Its equivalent circuit is also shown. The input resistance of this circuit is given by

We assume that we have tunnel diodes with required negative resistances.



The input resistance of the circuit depicted above is given by

$$R_{in} = \frac{-r_1 (R-r)}{R-r-r_1}$$

while the input resistance of the former circuit is $R-r$

If we choose $-r_1$ so that $R_{in}=R$,then solving for $-r_1$

we obtain $$-r_1 = \frac{R(R-r)}{-r}$$

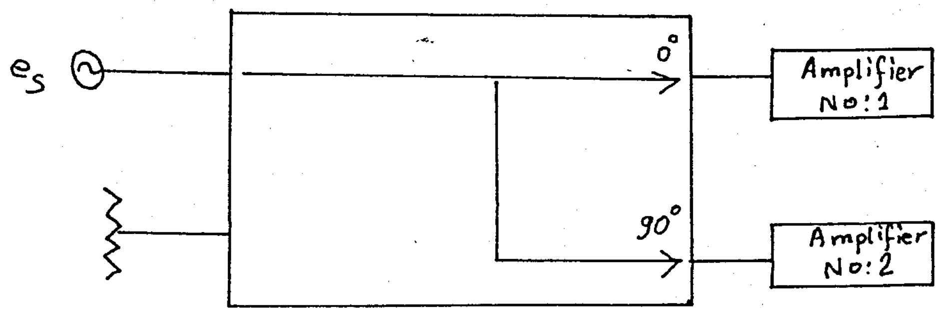
The input resistance of this circuit is now equal to the load resistance ,thus this circuit may be used as the load resistance of a similar circuit .When several of such stages are cascaded the voltage gain of each will be $\frac{R}{R-r}$ .The over-all gain of this circuit will be the product of the individual stage gains .

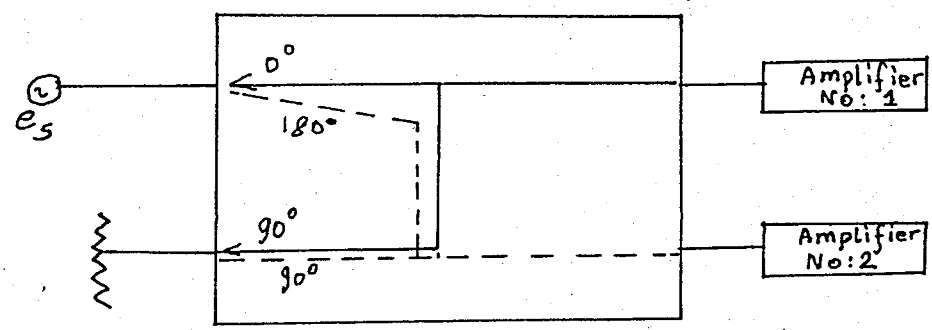
The resulting circuit resembles a transmission line made up of negative resistance elements ,and produces a gain rather than a loss .

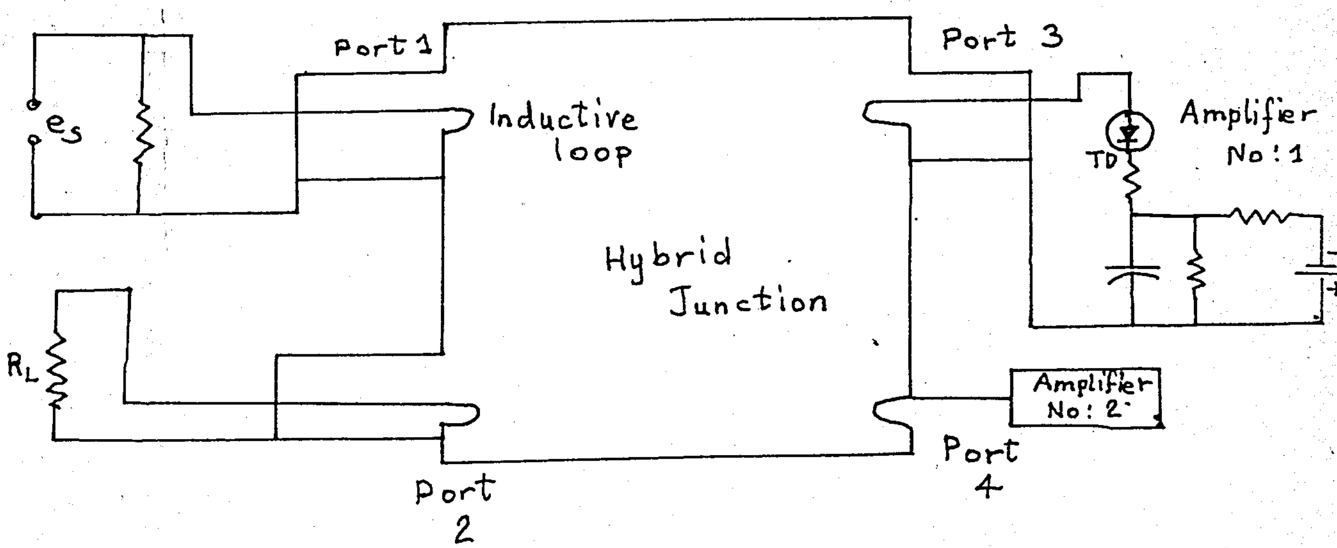
Other methods of cascading tunnel diode amplifier stages are available ,using isolators and circulator ,etc...

If tunnel diode is to be used for amplification in the microwave range ,other coupling arrangements for cascading individual stages have also been devised .

One example would be Hybrid coupling .We will discuss the operation of such a coupler .


e_s
0°
90°
Amplifier No:1
Amplifier No:2


e_s
0°
180°
90°
90°
Amplifier No: 1
Amplifier No:2


Port 1
Port 3
e_s
Inductive loop
Amplifier No:1
TD
Hybrid Junction
R_L
Amplifier No: 2
Port 4
Port 2

We will describe the process of amplification and coupling as below : equal amounts of energy arrive from the source at amplifiers No:I and No:2 ,o and 90 out of phase .The signal is amplified in each amplifier and transferred back to the hybrid junction .The field from amplifier 1 remains unchanged in phase . The field from amplifier No:2 undergoes another 90 deg. phase shift so that it arrives at the source 180 deg.out of phase with that from amplifier No:I .These two fields being in phase opposition cancel ,and no energy is transmitted to the source from the amplifiers .The field from amplifier No:I is shifted in phase by 90 deg. ;the field from amplifier No:2 does not suffer any phase shift,since it had already undergone a 90 deg.phase shift .The result is that the two fields from the two amplifiers arrive at the load in phase ,so they add. The amplified signal is thus transferred to the load .

A simple circuit using such an arrangement is also shown .

There are other methods of cascading tunnel diode amplifiers either for use in the microwave range or for use in the normal HF region ; but they are already available in the literature so they will not be repeated .(References will be found in the appendix .)

Noise Considerations :

So much attention has been focused on the noise performance of amplifying devices of all kinds ,for instance vacuum tube amplifier, the transistor amplifier ; amplifiers used for microwave amplification. Why do we attach so much importance to noise ? This is because the noise level may be such that the actual signal may be "lost" in the noise signal and it is only the noise signal that is detected .We must be sure that our amplifiying devices do not add noise by themselves and that the noise level is considerably below the signal .

Sources of noise in tunnel diodes :

First ,there is the so-called Johnson noise or the thermal noise generated by the ohmic losses in the spreading and contact resistance of the diode and the surrounding microwave structure (if a microwave amplifier is under study) .Secondly there is the shot noise accompanying the tunneling process .The shot-noise is very familiar in an ordinary vacuum tube ,it results from changes in the rates of amission of electrons ,thus the d-c component does not remain constant ,but it shows some fluctuations about a mean point. The shot noise in a tunnel diode is believed to be the result of carriers deflected from the barrier when unlike charges of equal energy levels are not on opposite sides of the barrier .

In a tunnel diode there are two currents in opposite directions in the tunneling region ,if both are assumed to be uncorrelated ,then they both contribute to shot noise and although the net tunneling current is the difference between the two tunneling current ,the net current contributing full shot noise is the sum of the two .

The two tunneling currents can be represented by the following expressions :

$$I_e = A \int_E F(E,V)\, f(E)\left[1 - f(E+qV)\right] dE$$

$$I_z = A \int_E F(E,V)\, f(E+qV)\left[1 - f(E)\right] dE$$

All terms in the above equations have already been defined . The net electron current flowing across the junction is the difference between I_E and I_z .The current I_E is caused by electrons tunneling from the conduction band of the n-type material into the valence band of the p-type material . I_E is known as the Esaki current and I_z as the Zener current .

The net current is given by

$$I = I_E - I_z$$

For calculations of the mean square noise current ,we shall assume that both currents are completely uncorrelated .Then both contribute to shot noise as pointed out before .The total or equivalent current contributing shot noise is the sum of I_E and I_z .

$$I_{eq} = |I_e| + |I_z|$$

We will try to relate the individual tunnel components to the total current I ,we note that the Fermi factors in the tunneling integrals are related as follows :

$$f(E+qV)\left[1-f(E)\right] = f(E)\left[1-f(E+qV)\right] \exp\left(-\frac{qV}{kT}\right)$$

This is the key device .Since V is not an integration variable, then

$$I_z = I_E \exp\left(-\frac{qV}{kT}\right)$$

hence

$$I_E = \frac{I}{1-\exp\left(-\frac{qV}{kT}\right)} \qquad I_z = \frac{I}{\exp\left(\frac{qV}{kT}\right)-1}$$

If the two tunnel currents exhibit full shot noise ,and if the noise currentsare uncorrelated ,then the mean square noise current in a frequency interval δf is expressed as

$$i^2 = 2q\, I_n\, \delta f$$

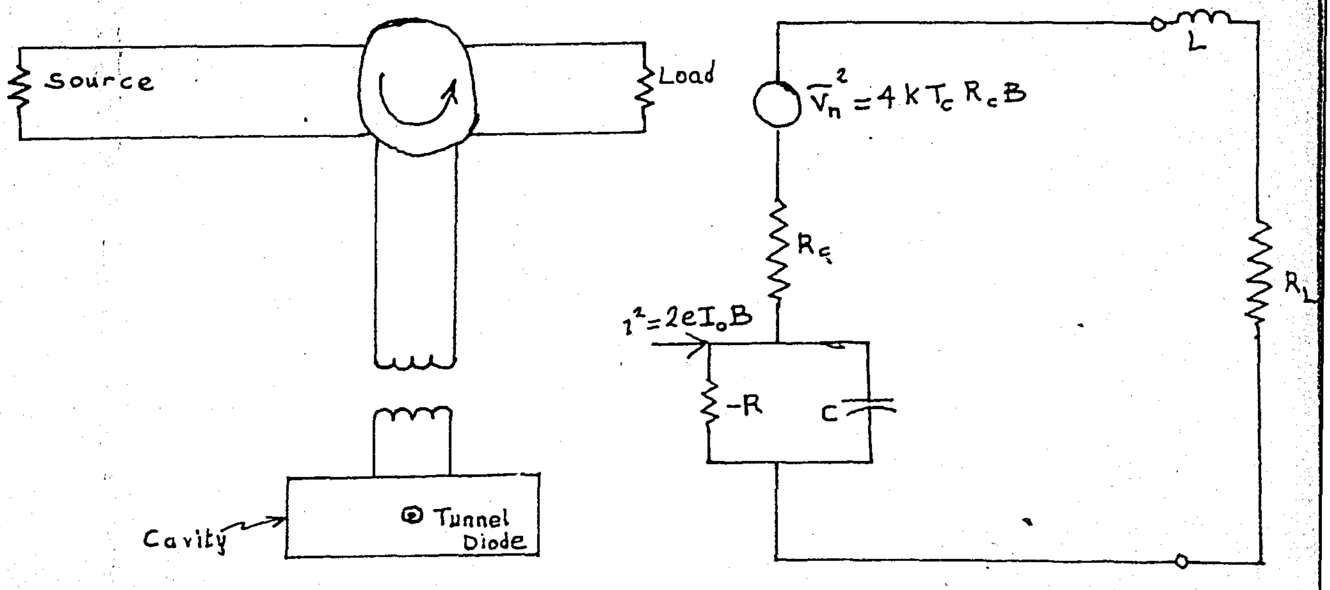
The equivalent diode noise current is equal to the sum of the two tunnel current and is given by $I_n = I \coth \frac{qV}{2kT}$

If the diode I-V characteristic is given in the tunneling region, then the equivalent noise current can be obtained from it with the help of the above equation .

At zero bias it is important to observe that the equation giving the noise current reduces correctly to the expression for thermal or Johnson noise , $\overline{i}^2 = 4kTg\,\delta f$.

Let us now analyze the noise performance from the circuit point of view .

The amplifier used as the model is shown in the figure below : (also shown is the equivalent circuit with noise generators)

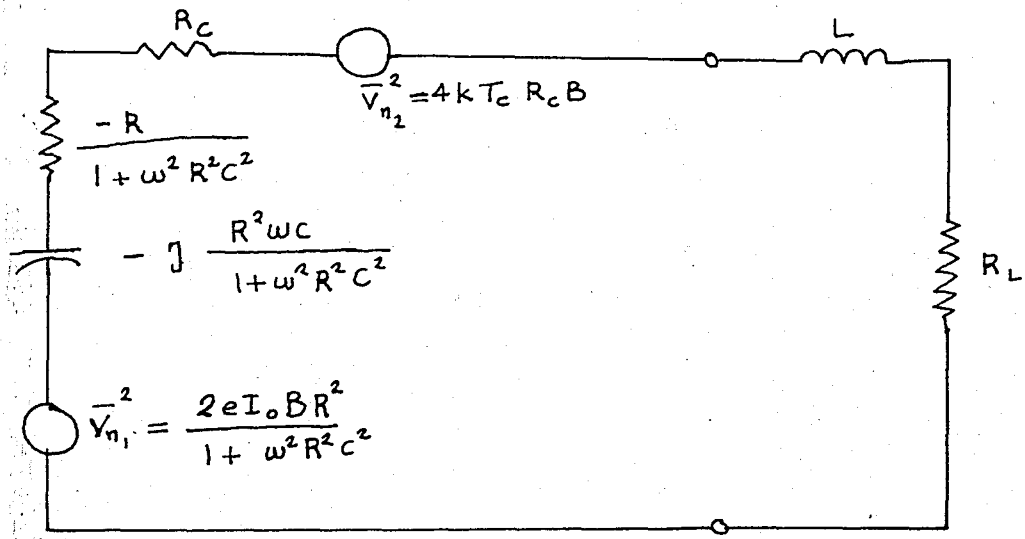


$v^2 = 4kT_c R_c B$ is provided by the Johnson noise generator where T_c is the ambient temperature of the cavity .L stands for the total inductance (parasitic and intentional) .

The shot noise is provided by the current generator

$$i^2 = 2eI_o B$$

Let us now transform all parallel elements into series equivalents .The circuit in this case is shown below :



Series resonance in this circuit occurs at

$$\omega_r = \omega_o\sqrt{1-\frac{1}{\omega_o^2 R^2 C^2}} \qquad \omega_o^2 = \frac{1}{LC}$$

A necessary and sufficient condition for positive gain is that the real part of the total series resistance ,as seen looking back into the amplifier terminals ,be ~~positive~~ negative ; or

$$\frac{R}{1+\omega^2 R^2 C^2} > R_c$$

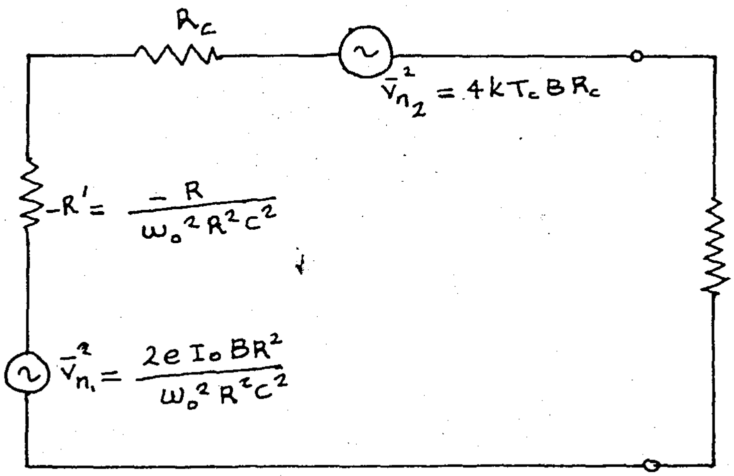
The resistive cutoff frequency is given by

$$f_{co} = \frac{\sqrt{\frac{R}{R_c}-1}}{2\pi RC}$$

Our main interest is in amplifier operating at resonance,we evaluate the elements of the above circuit for $\omega=\omega_r$ .The effective negative resistance is represented by the series resistance

$$-R' = \frac{-R}{\omega_o^2 R^2 C^2}$$

The equivalent circuit at resonance is as shownbelow :



The power gain of the amplifier is defined as the ratio of the power delivered to the load (the reflected power) to the power available from the source .It is equal to

$$G = \left[\frac{R_L - (R_c - R')}{R_L + (R_c - R')} \right]^2$$

Our aim is to derive an expression for the noise figure of the amplifier .But before going into details ,let us make a few definitions :

The noise figure at a specified input frequency is the ratio of the total noise power per unit bandwidth available at the output, to that portion of this power engendered at the input frequency by the input termination at the standard noise temperature of 290 K, which is the room temperature .That is to say, the noise figure is the ratio of total noise power output to noise input at the standard temperature. Therefore,

Noise figure $= F = \dfrac{N_o}{G k T_o B}$ $\qquad T_o = 290°K$

We will make one more definition:

The effective input noise temperature of a network is defined as the temperature of the input termination which results in output noise power per unit bandwidth double that which would occur if the input termination were at absolute zero.

It can be shown that the noise figure is also given by

$$F = 1 + \frac{T_e}{290} \qquad T_e = \text{effective temperature, } K°$$

Let us now return to our discussion.

We could also define the effective input noise temperature as the increase in the source temperature which is required to keep the noise power output constant if the amplifier, hypothetically were rendered noiseless. It is found by equating that part of the total noise output originating within the amplifier, N_a to $kT_e GB$ i.e.,

From the equivalent circuit and the gain definition:

$$k T_e B \frac{(R_L - R_c + R')^2}{(R_L + R_c - R')^2} = 4 k T_e B \frac{R_c R_L}{(R_L + R_c - R')^2} + \frac{2 e I_o B R R' R_L}{(R_L + R_c - R')^2}$$

Solving for T_e yields

$$T_e = \frac{4 T_c R_c R_L}{(R_L - R_c + R')^2} + \frac{2 e I_o R R' R_L}{k (R_L - R_c + R')^2}$$

which can be transformed to

$$T_e = \frac{(\sqrt{G} + 1)^2}{G} \left[T_c \left(\frac{R_c}{R_L} + \frac{e I_o R}{2k} \quad \frac{R'}{R_L} \right] \right.$$

After having determined the effective temperature in terms of circuit parameters, the noise figure of the tunnel diode amplifier stage can be found from the expression

$$F = 1 + \frac{T_e}{290}$$

The dependence of T_e is contained in the factor $(\sqrt{G}+1)^2/G$ which decreases asymptotically toward unity with increasing gain so that high gain operation is imperative .

The shot noise contribution enters through the term I_oR . It can be minimized by optimum biasing of the diode in such a way that the product I_oR is a minimum .In practice ,it was found impractical to operate the diode for any considerable distance beyond the mid-range of the negative resistance ,because the increasing value of R ,due to the curvature of the I-V curve,more than offsets the decrease in I_o.We will discuss how a minimum product can be obtained for optimum performance .

The I_oR product is inherently smaller for semiconductors with small energy gaps (InSb ,0.17 V ; Ge - 0.72 V) which should therefore be preferred for low noise applications .One futher point must be made : the reduction of the time constant ,RC of tunnel diode has a beneficial effect on the noise temperature of the amplifier as can be seen from the expressions.

In short , RC products must be small ,also energy gaps must be low ,meaning small I_oR product .

Having discussed how to obtain good noise performance, or rather the factors giving rise to good noise performance, let us now attempt to describe some means of actually realizing low noise figures. As we said earlier in connection with conditions for low noise figure, the I_0R product must be a minimum if the noise figure is to be improved. So we will now analyze this case.

The noise figure will now be expressed in terms of the standard temperature, so the expression will be slightly different. Also a new definition is made. This is the excess noise factor which is defined as the noise factor F minus unity,:

$$F-1=\left[\frac{G_n}{G_g}+\frac{G_L}{G_g}+\frac{G_e}{G_g}\right]\frac{T}{T_0}$$

where G_g is the source conductance at the standard temperature T_0, G_n is the loss conductance which contributes thermal noise, G_L is the load conductance, and G_e is the equivalent shot noise conductance of the diode at the ambient temperature.

If $R\left(=\frac{1}{G}\right)$ is the negative resistance of the diode corresponding to the operating diode current I, and G_p is the power gain, it follows that

$$A\,(F-1)=\left(G\,|R|+G_L\,|R|+\frac{eI}{2kT}\,|R|\right)$$

where $$A=\left(|R|\,G_g\right)\frac{T_0}{T}=\left(1+\frac{G_L}{G_g}+\frac{G_n}{G_g}-\frac{2}{\sqrt{G_p}}\sqrt{\frac{G_L}{G_g}}\right)^{-1}\frac{T_0}{T}$$

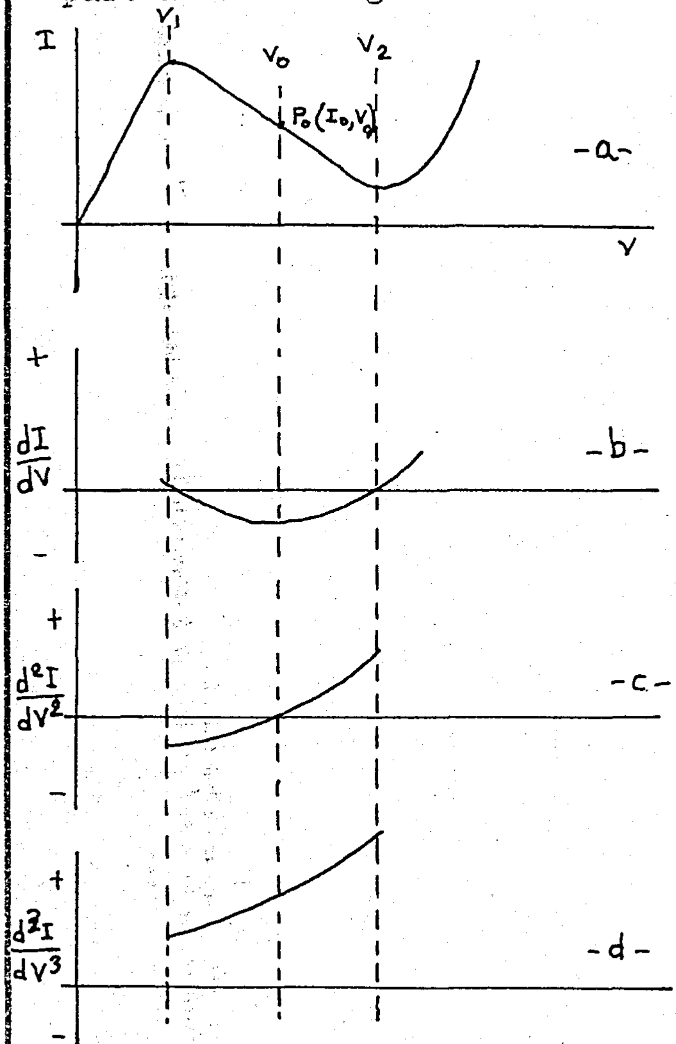
$$G_p=\frac{4\,G_gG_L}{\left(G_n+G_g+G_L-|G|\right)^2}$$

Here $A(F-1)$, the product of the excess noise and a constant dependin on gain, is a measure of the noise-and-gain behavior of the tunnel diode amplifier. A small value of $A(F-1)$ means small noise factor and large power gain G_p.

Our objective is to show analytically that with a given

I-V characteristic of a tunnel diode, a minimum value of exists at a unique point on the characteristic . Another aim is to suggest an ideal I-V characteristic which can be a guide for the design of tunnel diode amplifiers with low values of .

The I-V characteristic of a typical tunnel diode is shown in part a of the figure below :



The negative-slope portion of the curve lies between V_1 and V_2 , where the slopes are zero . An inflection point (V_o , I_o) exists between these two points . The negative-slope portion of the curve can be represented by the function

$$I - I_o = f(V - V_o)$$

Then its derivative with respect to V is

$$I' = \frac{dI}{dV} = f'(V - V_o) = \frac{1}{R}$$

where R , the reciprocal of the slope ,is the negative resistance of the diode .Consider first the case in which the thermal noise term $G_n|R|$ and the load noise term $G_L|R|$ are ignored .The IR product $(A(F-1)2KT/e)$

is

$$IR = \frac{f(V-V_o)+I_o}{f'(V-V_o)}$$

IR is maximum or a minimum at the point $P_m(V_m, I_m)$,where

$$\frac{d(IR)}{d(V-V_o)} = 0$$

that is ,

$$\frac{\left[f'(V_m-V_o)\right]^2 - \left\{f(V_m-V_o)+I_o\right\} f''(V-V_o)}{\left[f'(V_m-V_o)\right]^2} = 0$$

since $V_1 < V_m < V_2$

by part -a- of the above figure , $f'(V_m-V_o) \neq 0$

the last equation thus becomes

$$\left[f'(V_m-V_o)\right]^2 = \left\{f(V_m-V_o)+I_o\right\} f''(V_m-V_o)$$

It follows that

$$(IR)_{max\ or\ min} = \frac{f'(V_m-V_o)}{f''(V_m-V_o)}$$

We have to show whether the IR product is a minimum .

The second derivative of the expression for IR at the point V_m, I_m is

$$\left.\frac{d^2(IR)}{dV^2}\right]_{\substack{V=V_m\\ I=I_m}} = \frac{f'(V_m-V_o)f''(V_m-V_o) - \left\{f(V_m-V_o)+I_o\right\} f'''(V_m-V_o)}{-\left[f'(V_m-V_o)\right]^2}$$

Since $f(V_m-V_o)+I_o$ is always positive , $f''(V_m-V_o)$ is positive too .Thus by the graph of $f''(V-V_o)$, V_m must be greater than V_o ; that is , the maximum or minimum IR product occurs beyond the inflection point I_o, V_o .

Also $f'(V_m - V_o) < 0$

$f'''(V_m - V_o) > 0$

In fact , for the I-V characteristic of the tunnel diode, it can be shown graphically that the numerator of the second derivative expression is always negative and hence the second derivative is always positive . Therefore the absolute value of IR is a minimum and a minimum value of A (F -I) does exist .

If the thermal noise term $G_n|R|$ and the load noise term $G_L|R|$ are included in the original expression for IR , then the point of optimum noise factor V_m, I_m shifts towards the inflection point V_o, I_o . The actual minimum value of $A(F-1)$ at this new point is the same as that obtained by the present analysis provided that I_o is replaced by $I_o + (G_n + G_L)^2 \, 2kT/e$.

Example

Let us assume that

$$I = I_o - G_o V_o \sin \frac{V - V_o}{V_o}$$

where G_o is the slope at the inflection point . This is a crude but simple approximation for the characteristics of a tunnel diode. (See reference list) . It follows that

$$I' = - G_o \cos \frac{V - V_o}{V_o}$$

$$I'' = \frac{G_o}{V_o} \sin \frac{V - V_o}{V_o}$$

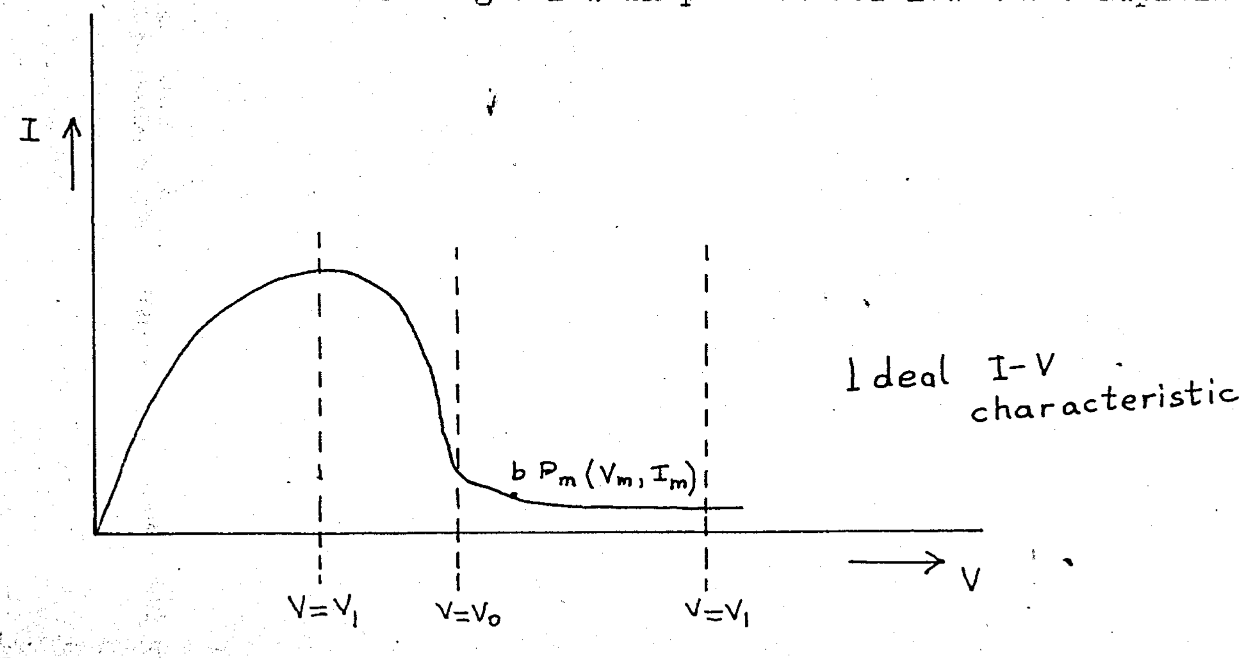
the voltage which yields a minimum IR is

$$V_m = \left\{ \left[\sin^{-1} \left(\frac{V_o}{I_o} G_o \right) \right] + 1 \right\} V_o$$

$$(IR)_{min} = \frac{I_o}{G_o}\sqrt{1-\left(\frac{V_o}{I_o}G_o\right)^2}$$

where $\left(V_o G_o / I_o\right)^2 < 1$. Accordingly IR can be made small by making $(V_o)^2 (G_o)^2 / I_o^2$ approach unity . Also , the minimum IR product is smallest for small values of $f'(V_m - V_o)$ as can be seen from the expression for $(IR)_{min}$:, and large values of $f''(V_m-V_o)$. This, of course , requires a small value of the minimizing current I_m according to the equation giving the condition for minimum or maximum IR product . An I-V characteristic which has a steep drop followed by a slow decay toward a very small current minimum, is, for example, an ideal curve to fulfill these requirements for a low minimum IR product . The additional advantage offered by the slow decay is that a region of wide choice of stable operating points of low slopes is provided .

However , low operating slopes result in high operating negative resistances which can be realized only by small capacitances at high operating frequencies . This is probably the restriction for obtaining a low IR product for low noise amplification.



Thus far we discussed many aspects of tunnel diodes and tunnel diode circuits.

A- The construction and properties of tunnel diodes

B-Amplification principles

C-Stability criteria

D-Cascading

E-Noise performance

It should addedat this point that amplifiers are not the only field of application of tunnel diodes .In fact there is a variey of uses to which tunnel diodes may be put .So it is our aim to list briefly some more application of tunnel diodes .

I- Oscillators

2-Pulse and Switching circuits

3- Modulation ,demodulation ,heterodyne detection

5-Communications

6-Tunnel diodes may possibly replace variable capacitance diodes for some low- noise applications .

Although the tunnel diode cannot compete with variable capacitance diode for low-noise applications at very high frequencies;new methods of construction ,new techniques make it possible to manufacture tunnel diodes with very low noise figures. But the simplicity of construction of tunnel diode amplifiers is a point in its favor .There are many other advantages of tunnel diodes ,that it is not possible to list all of them .

Low noise figures make tunnel diodes especially useful at microwave frequencies .In fact these are really the useful range of frequencies .

Tunnel diodes have increasingly poor noise figures at low frequencies .Vacuum tube and transistor amplifiers already give satisfactory noise performance at low frequencies .So it is not preferrable to employ tunnel diodes for low-frequency applications.

CONCLUDING REMARKS :

We were able to touch upon a limited field of tunnel diode applications ,performance and characteristics .However,the material included discusses the basic aspects of tunnel diodes and tunnel diode circuits .There are many possible circuit configurations both in the microwave region and low-frequency range .All of them are special circuits; it would be meaningless to attempt to analyze all of them one by one ; if the basic principles are available for further study ,the sole aim of discussion is achieved.

Our discussion is by no means complete ; but it is hoped that our attempts at the discussion of basic principles will be of help to a reader in understanding the subject since no comprehensive text is available .Many references will be found in the Appendix .However ,none of them presents a complete picture of tunnel diodes and applications .We hoped to present a unified discussion on tunnel diodes and tunnel diode applications .It should be born in mind that this is only an introduction to the subject and that further survey has to be conducted if one is to gain a thorough knowledge of this subject .

Respectfully submitted

E. Karabekir

M. Erhan Karabekir

Graduate EE ,1964-1965

Robert College

APPENDIX

BIBLIOGRAPHY

I- Lawrence A. Blackwell & Kenneth L. Kotzebue, Semiconductor - Diode Parametric Amplifiers , Prentice - Hall , Inc., Englewood Cliffs, N.J. , pp. 3 - 9 ; I6 - I7 . I96I .

2- Sylvester P. Gentile , Basic Theory and Application of Tunnel Diodes , D. Van Nostrand Company , Inc., Princeton , N.J. pp. 57 - 69 ; 76 - 83 ; I20 - I24 ; I43 - I44 .

3- New Phenomenon in Narrow p-n junctions , Physical Review , Vol. I09 , pp.603 - 4 , Jan . I958 .

4- History of Parametric Transducers , PROC. IRE , Vol 48 , p. 848 , May I960 .

5- An Introduction to the Tunnel Diode as a circuit Element , The Journal of the British Institution of Radio Engineers , Vol 22 , No : I . pp.75 - 9 . July I96I .

6- Tunnel Diode Large -Signal Equivalent Circuit Study and the Solutions of its Nonlinear Differential Equations , Journal of Research of the National Bureau of Standards . Engineering and Instrumentation , Vol 66 C , No : I . January - March I962 .

7-Tunnel Diodes as amplifying devices , Electronics , Vol 36 , June 2I , I963 , p. 68 .

8-Negative - Resistance Amplifier Design , Electronics , Vol 33 , No : 22 , pp . II0 - II2 , May 27 , I960 .

9-Biasing Methods for Tunnel Diodes , Electronics , Vol 33 , pp . 82 - 84 , June 3 , I960 .

I0- Tunnel Diodes as High - Frequency Devices , PROC IRE ,Vol 47 , p . I20I , July 1959 .

II- The Effect of Non-Symmetrical Doping on Tunnel Diodes , PROC IRE , Vol 49 , pp. I435 - 36 ,Sept . I96I .

I2- Temperature Dependence of the Peak Current of Ge Tunnel Diodes, PROC IRE , Vol 49 , pp . I428 - 9 , Sept. I96I .

I3- Stability Criteria for a Tunnel Diode Amplifier , Proc IRE , Vol 49 , p. I937 , Dec . I96I .

I4- Stability Criteria for Tunnel Diodes , PROC IRE , Vol 49 , pp . I206 - 7 ; July I96I .

I5- Stability Considerations for a Tunnel Diode Circuit , Journal of Franklin Inststute , Vol 274 , pp . 444 - 5I ,Dec.I962

I6- A Technique for Cascading Tunnel Diode Amplifiers , PROC IRE , Vol 49 , pp . 373 - 375 , Jan. I96I .

I7- A Technique for Cascading Tunnel Diode Amplifiers , PROC IRE , Vol 48 , p . II56 ,June I960 .

I8- A Unidirectional Amplifier with Esaki Diodes , IRE TRANSACTIONS ON ELECTRON DEVICES ,March I96I, pp.I63 - I69

I9- The Equivalent Noise Current of Esaki Diodes , PROC IRE , Vol 49 , pp . I080 -I ; June I96I .

20- A Noise Investigation of Tunnel -Diode Microwave Amplifier , PROC IRE , Vol 49 , pp . 739 - 743 ,April I96I .

2I- Noise Performance of Tunnel Diode Amplifiers , PROC IRE , Vol 48 ,pp. I478 - 79 , August I960 .

22- Shot Noise in Tunnel Diode Amplifiers , PROC IRE , Vol 48 , pp. I4I8 - 23 ,August ,I960 .

23- The Optimum Noise Performance of Tunnel Diode Amplifiers, PROC IRE , VOL 48 , pp . I07 - I08 ,Jan . I960 .