A STUDY OF THE ESAKI TUNNEL DIODE AND DEVELOPMENT OF SOME

ASSOCIATED AMPLIFIER DESIGN RELATIONSHIPS

By

JOHN C. PAUL

Bachelor of Science

Oklahoma State University

Stillwater, Oklahoma

1959

Submitted to the Faculty of the Graduate School of
the Oklahoma State University
in partial fulfillment of the requirements
for the degree of
MASTER OF SCIENCE
May, 1960

A STUDY OF THE ESAKI TUNNEL DIODE AND DEVELOPMENT OF SOME

ASSOCIATED AMPLIFIER DESIGN RELATIONSHIPS

Thesis Approved:

_____
Thesis Advisor

_____

_____
Dean of the Graduate School

ii

PREFACE

This thesis reports the work performed by the author in compiling an analysis of the tunnel diode as a device and developing a suitable design procedure for utilizing the diode in an amplifier circuit. Because of the extreme non-linearity exhibited by the diode there are numerous possible applications. Some which have been explored include: logic elements, free running, astable, and bistable multivibrators, relaxation oscillators, sinusoidal oscillators, mixers and convertors, multi-function circuitry utilizing only one diode, and the topic discussed in the latter part of this thesis, straight amplification.

At the time of this writing, the tunnel diode is a relatively new device. For this reason it was felt that a complete theoretical explanation would enhance the understanding of any subsequent circuitry. Hence, the first portion of this thesis is mainly concerned with an explanation of the semiconductor characteristics of the diode. The latter portion is a developement of some design procedures and an experimental verification of them.

iii

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

In the fall of 1958, the phenomenon of negative resistance in very
narrow p-n junctions was first reported by the Japanese physicist L.
Esaki.[1] Since that time there has been considerable effort expended in
utilizing the negative resistance diode in circuitry. Experimental
amplifiers have been built and the results reported;[2] but to the author's
knowledge, only one source has given a detailed design procedure.[3] The
analysis reported was based solely on the Nyquist plot of the loop
impedance with the diode inserted and the resultant design equations
were first-order approximations. For this reason it was decided to
attempt an analysis from a different standpoint. The method developed
herein by the author incorporates a frequency response analysis with
both analytical and graphical solutions which allow an exact theoretical
design.

Since the diode is only a two terminal device, it might appear at
first glance that the problem of separating the input from the output
would be exceedingly difficult. However, if the device is inserted be-
tween the source and the load, amplification will be obtained although

[1] L. Esaki, "New Phenomenon in Narrow Ge P-N Junctions", Physical
Review, Vol. 109, p. 603, 1958.

[2] H. S. Sommers, Jr., "Tunnel Diodes as High Frequency Devices",
Proceedings, I.R.E., Vol. 47, p. 1201, 1959; and K. K. N. Chang,
"Low-Noise Tunnel Diode Amplifier", Proceedings, I.R.E., Vol. 47,
p. 1268, 1959.

[3] U. S. Davidsohn, Y. C. Hwang and G. B. Ober, "Characterization of
Tunnel Diodes and Circuit Stability Considerations", Electronic Design,
March 17, 1960.

the magnitude of the source and load impedance are still critical factors in the design.

Since there are essentially two types of sources, voltage and current, there are essentially two methods of insertion. Series insertion is used for voltage amplification and parallel insertion for current amplification. Only the series insertion amplifier was considered in this thesis.

Analysis of the series amplifier implies that two essential modes of amplification may be obtained. That is, the circuit may be adjusted to yield either selective or non-selective amplification. From the analysis it appeared that the non-selective amplifier was only a special case of the selective amplifier and for that reason the selective tuned amplifier was the only one considered in the analysis.

The primary reasons for the intense interest shown in the negative resistance or tunnel diode are its very high frequency response, extreme resistance to radiation and low noise figure. These characteristics result from the fact that the operation does not depend upon minority carriers. In order to illustrate more effectively how these advantages occur, the first portion of this thesis is devoted to an analysis of the tunnel diode as a semiconductor device.

The negative resistance characteristic depends upon quantum-mechanical tunneling of electrons through a potential which they do not appear to have the energy to surmount. Since electron tunneling theory is not widely understood, the second chapter is devoted to an explanation of how this effect is possible.

The third and fourth chapters are devoted to an explanation of the diode characteristics based upon conventional semiconductor theory and

Chapter V is the author's development of some design equations and pro-
cedures for utilizing the diode as an amplifier. Chapter VI reports
the results obtained from an experimental amplifier constructed from the
equations developed in Chapter V.

CHAPTER II

QUANTUM MECHANICAL DEVELOPMENT OF ELECTRON "TUNNELING"

Quantum mechanics is in essence a mathematical formulation to analyze
and predict particle behavior on an atomic scale. In the late nineteenth
century the older classical theory of particles began to deviate slightly
from the observed results of experiments and in certain situations gave
completely erroneous results.[1] It was felt at the time that perhaps these
deviations could still be fitted into the structure of classical mechanics
which up until this time had so adequately described all physical proc-
esses. However, time passed, and despite Herculean efforts of mathemati-
cal manipulation, the inconsistencies remained. Then a few farsighted
physicists among them, Heisenberg, Schrodinger, and Dirac, departed from
the older theory to formulate the new mathematical physics of quantum
mechanics.

The new mechanics, which came into full bloom in the late 1920's,
had as its underlying principle a postulate developed by Heisenberg known
as the "uncertainty principle", which fundamentally was a statement of
indeterminency. As applied to physical phenomena the uncertainty prin-
ciple simply says that certain related quantities (i.e., momentum and
position, energy and time) which describe the state of a particle are so
interrelated that a precise knowledge of one of the quantities involved
automatically implies uncertainty as to the magnitude of the other re-
lated quantity. Heisenberg proved that the order of magnitude of the

_____

[1]Banesh Hoffman, The Strange Story of the Quantum, Dover Publica-
tion, New York, 1959.

4

uncertainty was related to a physical quantity known as Plank's constant, 6.23×10^{-19} erg-sec. The order of magnitude of this uncertainty is so small that the effect of the indeterminency does not become noticable until the dimensions of the system approach the microscopic scale, but this is precisely where the difficulty lay in the application of classical mechanics to physical processes. The new theory filled the gap admirably and spurred developments in the field of atomic theory.

The method embraced by the new quantum theory as dictated by Heisenberg's uncertainty principle implied that, although precise formulation of the system state was no longer feasible, it was possible and, as a matter of fact, highly desirable to describe the uncertainty inherent in the system. This was accomplished by characterizing the physical quantities by their probabilities of existence.

If, for example, one wished to hypothesize an electron existing in space, he did not consider it as a particle existing at a certain point and moving with a definite velocity but instead formulated it as a probability wave. That is, the position of the electron would be characterized by a probability function throughout the region under consideration and the particle's momentum would be considered similarly. Now, in any mathematical treatment of the system, the probability waves would be used to represent the electron instead of the older method of considering the electron to be a particle at a fixed point with a fixed momentum.

From these first revolutionary concepts a highly intricate and complex theory has been developed from which it is possible for the modern physicist to accurately analyze and predict on an atomic scale. A complete exposition of that theory will not be considered here, but

only those aspects of it which will allow qualitative understanding of

the tunnel diode. With this in mind, a highly idealized example to

illustrate the foregoing discussion will now be considered.[2]

Consider the region in space around a hydrogen nucleus in which the

electrostatic attractive force between the nucleus and electron obeys the

Coloumb inverse square law of charged particles. Due to this attraction

the electron will have at any point in its orbit around the nucleus a

certain amount of potential energy due to its position. The electron

will also have kinetic energy determined by its velocity and the total

energy of the electron will be the sum of its kinetic and potential energy.

Since the system will be considered to be non-dissipative and will have

no external forces acting on it, the total electron energy must remain

constant. A diagram illustrating the energy of the system is shown in

Figure 2-1.

x (Position Relative to Center of Nucleus) ———→
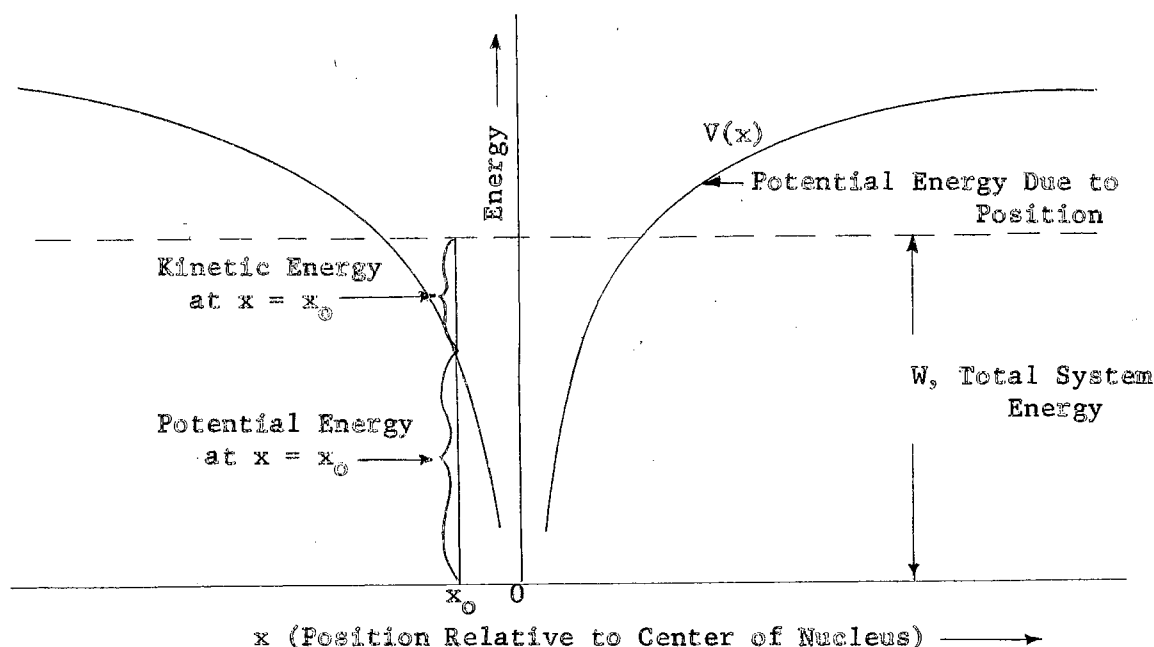
Figure 2-1. Electron Energy vs. Position in a Hydrogen Atom

[2]R. W. Gurney, Elementary Quantum Mechanics, Cambridge, 1934.

It is thus seen that the electron resides in a sort of potential "hole" the height of which is determined by the total energy of the system. According to the older theory of classical mechanics, there is no possible way for the electron to escape from the confines of this potential hole since that theory states

$$V \leq W \tag{2-1}$$

where

    V is the potential energy of the electron
    W is the total energy of the electron

In order to simplify the problem for quantum mechanical treatment without changing the essential results, the potential "hole" will be idealized into a potential "box" as illustrated in Figure 2-2. W_m is the amount of energy necessary to completely remove the electron from the influence of the nucleus.
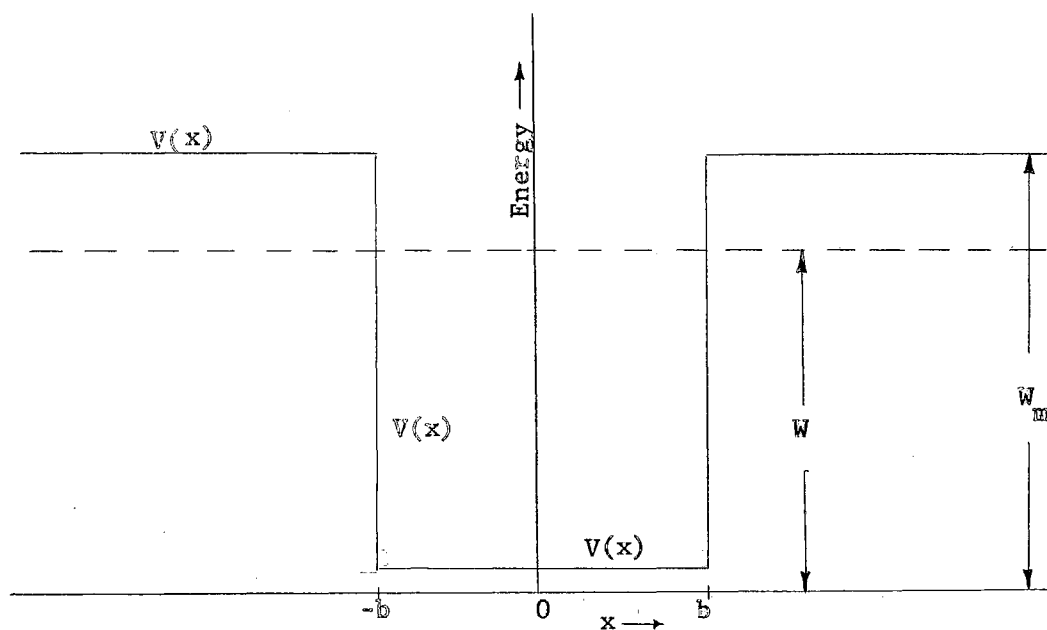
Figure 2-2. Idealized Electron Potential Energy vs. Position

In light of the earlier discussion it is evident that the electron does not exist at any particular location within the potential box, but rather there will exist some probability function describing the possibility of electron existence throughout the region. It is one of the triumphs of quantum mechanics that it exhibits a relatively simple method of obtaining this probability function through means of an equation developed by Schrodinger. In its time-free form for one dimension this equation is

$$\frac{d^2\Psi}{dx^2} + a(W - V(x))\Psi = 0 \tag{2-2}$$

where

$a = 8\pi^2 m/h^2$
m is the mass of the particle
h is Plank's constant
Ψ is the wave function
W is the total energy
V is the potential energy

The actual probability, $P(x)$, that an electron exists at any coordinate x is found from

$$P(x) = |\Psi(x)|^2 \tag{2-3}$$

providing $P(x)$ is normalized to unity. That is, the entire area under the probability curve must be set equal to unity as we are dealing with only one electron. Therefore

$$\int P(x)dx = \int |\Psi(x)|^2 dx = 1 \tag{2-4}$$

where the indicated integrations are taken over the whole of space.

The formal derivation, justification, and implications of the above equations are complex and abstract, and, furthermore, are not essential to the desired results. The interested reader is referred to the many excellent books on the subject of theoretical quantum mechanics.

The Schrodinger equation, when applied only to the interior of the potential box of the example yields the simple family of solutions

$$\Psi(x) = A \cos\sqrt{a(W-V(x))}\,x + B \sin\sqrt{a(W-V(x))}\,x \tag{2-5}$$

where

A and B are constants of integration determined by the boundary conditions

It was Schrodinger who first pointed out that the wave equation was valid outside the confines of the potential box. The question of how this situation arises is best justified by noting that it gives the necessary answer as dictated by experimental facts without being concerned with how the potential energy V of the system has seemingly managed to exceed its total energy. In any event, the solution of the wave equation for the case where (W-V) is negative yields for the region exterior to the potential box

$$\Psi(x) = C \exp\left(-\sqrt{a(V(x)-W)}\,x\right) + D \exp\left(\sqrt{a(V(x)-W)}\,x\right) \tag{2-6}$$

A necessary condition for the wave equation to represent the system is that the complete solution must tend to zero as x tends to infinity since it is not desirable to imply that the electron spends its entire existence exterior to the box. Under this restriction it is evident that the constant in the first term of Equation (2-6) must be zero for x less than -b and the constant in the second term of the equation must be zero for x greater than +b. The complete solution for all values of x is

$$\Psi(x) = A \cos\sqrt{a(W-V(x))}\,x + B \sin\sqrt{a(W-V(x))}\,x \qquad , -b \leq x \leq b$$

$$\Psi(x) = C \exp\left(-\sqrt{a(V(x)-W)}\,x\right) \qquad , b < x$$

$$\Psi(x) = D \exp\left(\sqrt{a(V(x)-W)}\,x\right) \qquad , -b > x \tag{2-7}$$

To preserve continuity the wave function and its first derivative
must be continuous across the boundaries of the problem. This restric-
tion is essential for fitting the complete solution together.

When the complete solution is fitted to the potential box it is seen
that at the boundaries of the box where V is varying extremely fast that
it would be possible to fit any number of solutions into the box were it
not for the boundary restrictions. Figure 2-3 (a) illustrates a number
of valid solutions to the wave equation, and Figure 2-3 (b) shows the
corresponding probability function. Figure 2-3 (c) shows what would be
the result if a shorter wavelength were to be used for the same boundary
condition on V as that for a longer wavelength.

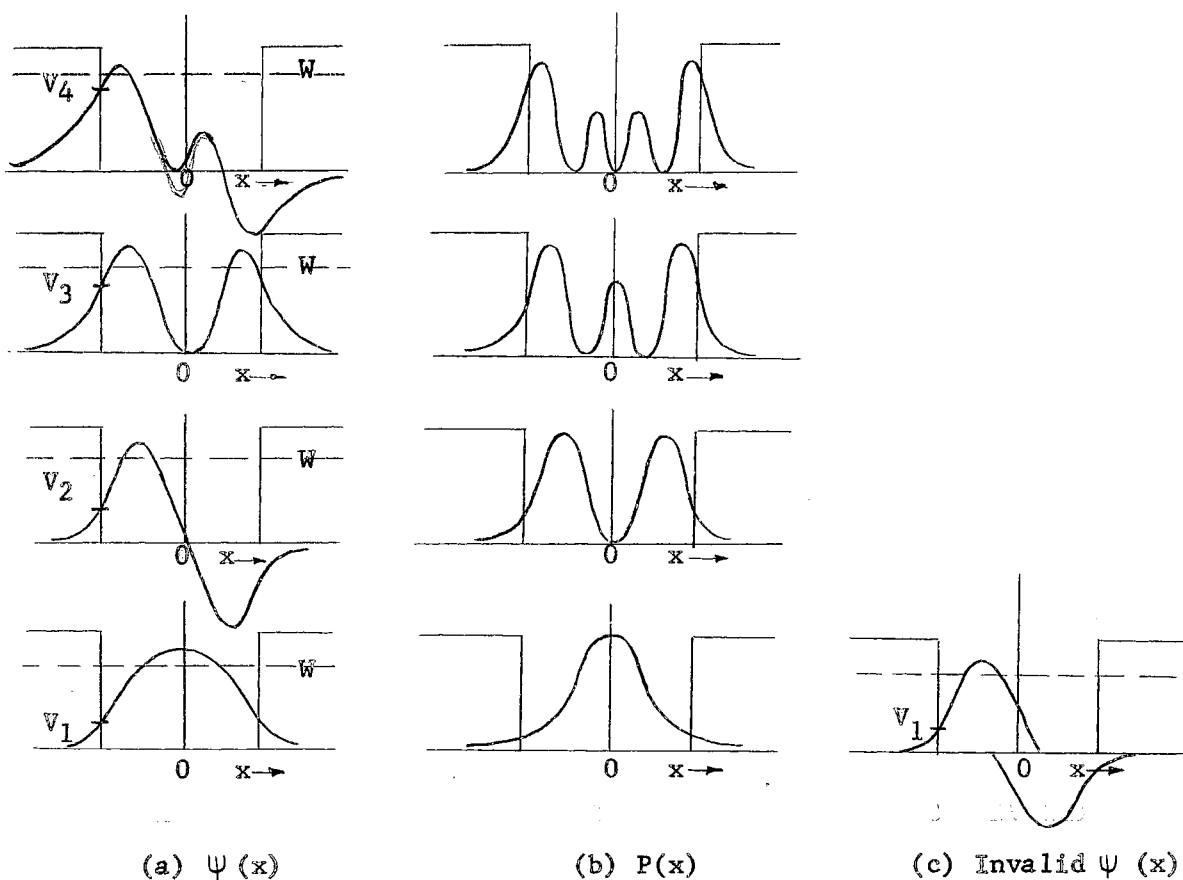(a) $\psi(x)$ (b) P(x)   (c) Invalid $\psi(x)$

Figure 2-3. Schrodinger's Wave Equation Solutions

It is interesting to observe that these solutions yield a finite existence probability outside the confines of the potential box where, according to classical mechanics, the electron could not possibly exist. It is also worthy of note that in the valid solutions of the wave equation, the V_i, which are an indication of the electron potential energy, are discrete in value with $V_{i+1} > V_i$ and that the V_i approach W as a limit. This is known as quantization of electron energy and is the mathematical justification for the evidence which indicates that the electron exists in its orbit at definite discrete energy levels. As a matter of fact, when the V_i from the solution of the Schrodinger wave equation are compared with the line spectra emitted by the hydrogen atom, there is an exact numerical correspondence.

Each boundary energy level is characterized by a quantum number, and is known as a quantum state in one dimension. When the problem is extended to three dimensions, it is found that there exist two other quantum numbers associated with the other two variables which describe the system and, in addition, there is a quantum number associated with the spin on the electron which is $\pm 1/2$ according to whether the spin is "parallel", or "anti-parallel".[3] It would be well at this time to introduce another principle of theoretical physics which will again be referred to at a later time. It is known as the "Pauli exclusion principle" and states that no two electrons may occupy the same quantum state simultaneously.

Now let the problem of the electron in the potential box be extended to that of two adjacent potential boxes with an electron initially in the box on the left in Figure 2-4.

[3]Ibid.
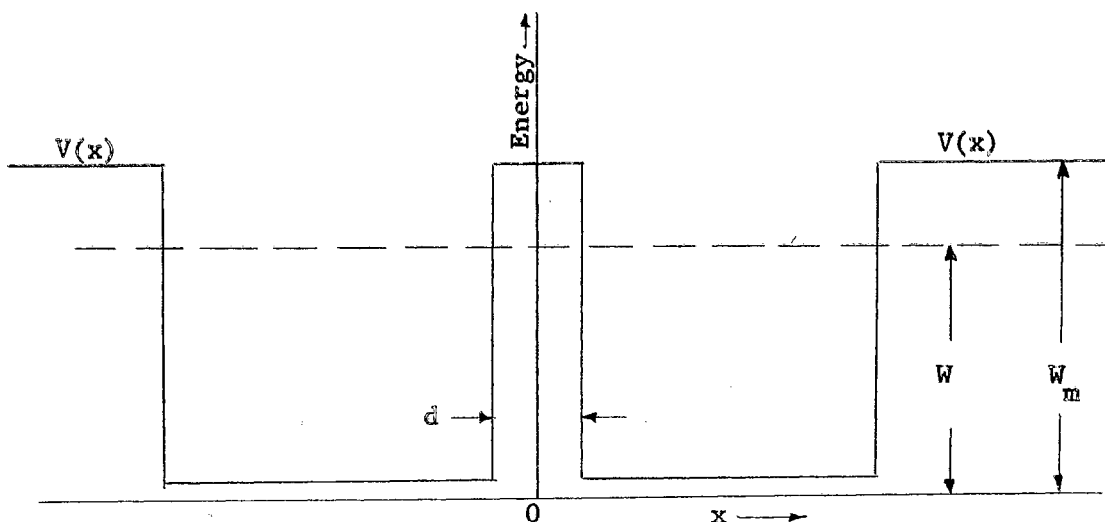
Figure 2-4. Idealized Adjacent Potential Boxes

The method of solution for the wave function indicating the electron position is essentially the same as that given for the previous problem. The Schrodinger wave equation in one dimension is solved subject to the boundary restrictions with the principal difference being in utilizing the increasing expotential in the region of the "barrier" d. This variation is admissible as the complete solution still tends to zero as x increases without bound. Two of the possible valid solutions are illustrated in Figure 2-5, where both of the solutions correspond to the "ground state" or longest wavelength solutions. Note that the term wavelength refers to only that portion of the solution which lies within one of the potential boxes. The only difference in the solutions is that the increasing expotential in (b) is the negative of the increasing expotential in (a). Although it might first appear that both solutions correspond to the same energy state, it is found, when fitting the solution to the boundary conditions, that it is necessary to make the solution of (a) of slightly longer wavelength than (b). This is shown qualitatively in Figure 2-5.

This slightly longer wavelength, in fact, corresponds to a slightly

lower energy state than that for the solution with the shorter wavelength.
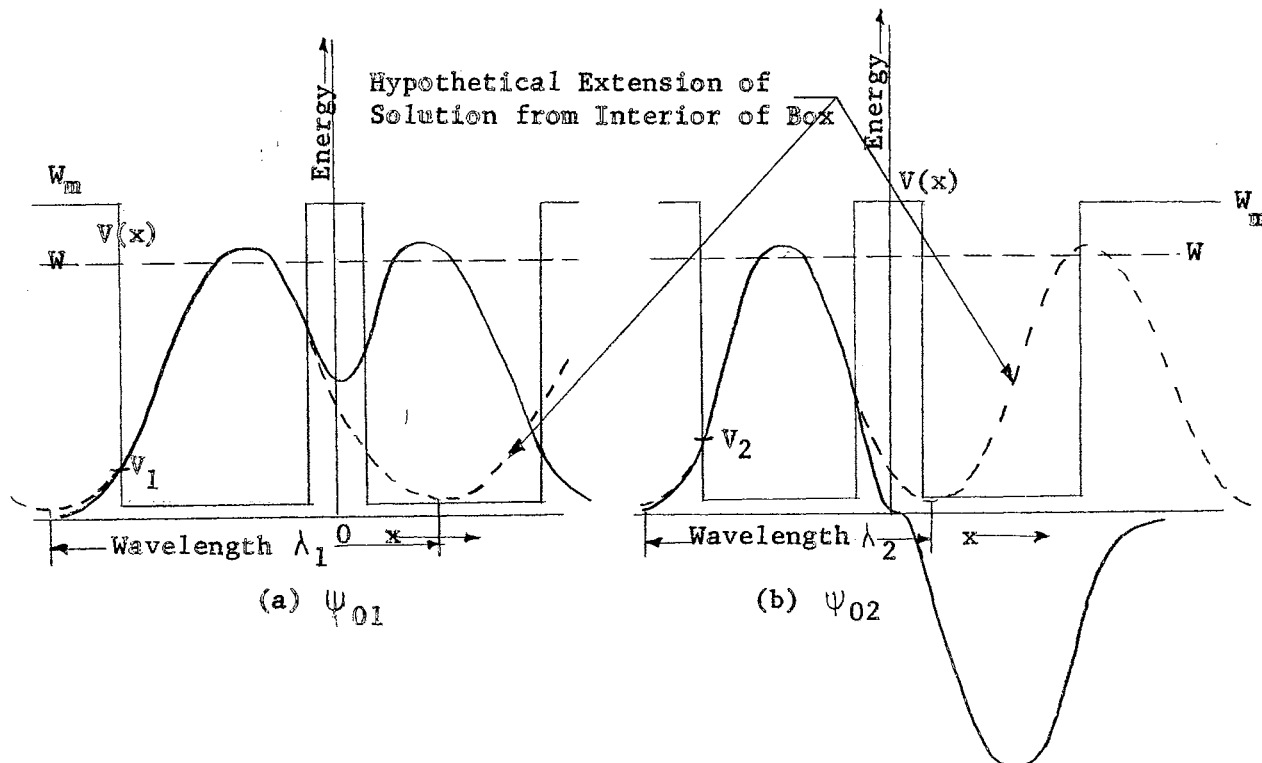
Figure 2-5. "Ground State" Wave Functions for Adjacent Potential Boxes

To summarize the essential result contained in Figure 2-5, the pre-

sence of the second potential box has caused a splitting of the initial

energy level into twin energy states.

This result is immediately extendible to the case of N potential

boxes close together as might be present in a crystalline structure. Now

the initial "ground state" solution of the crystal would be split into an

energy band of N discrete energy levels. Now by considering the other

two variables as well as the "spin" which are needed for a 3-dimensional

description of the problem, the energy band would contain 4N levels which

are known as quantum levels since a quantum number is associated with each

of the 4 variables. This energy band concept is extremely useful in discussing semiconduction processes as will be done in later chapters.

A final point to be made in relation to Figure 2-5, is that an analytic development shows the energy difference (V_2-V_1) of the two solutions is an inverse function of the barrier width.[4] From this it may be seen that the band splitting effect does not become appreciable until the barrier width becomes relatively small as in the case of a crystalline material.

All of the previous discussion, although correct under the assumption made, must now be modified somewhat to become more in accord with the situation as it actually exists. In the normalizing process on ψ , the probability of the electron's existence was normalized to unity to signify the existence of a single electron. However, this process was only done for one energy level. As may be seen from observation of Figure 2-3, there are a number of possible solutions of the Schrodinger wave equation, and since the electron only exists in one of the probability states, the preceeding results must be modified to allow the electron to exist in only one of these states at a particular instant. Proceeding according to the methods of quantum mechanics, the concept of probability is again introduced in the sense that the probability of any particular solution being correct must be evaluated.

If $\psi_1 \ldots \ldots \psi_n$ denote the different wave functions obtained from the Schrodinger wave equation, then it is possible to assign a certain "weight" value to each of the ψ_i, corresponding to the probability that the electron actually has this wave function as its solution state. This

---

[4]Ibid.

"weight" will be denoted by a_i where a_i is the weight associated with the wave function ψ_i. These a_i are merely factors which indicate a relative ignorance as to which of the specific states the electron actually occupies.

Since the electron must exist in one and only one state, then a normalization is applied to normalize the a_i to unity as follows

$$|a_1|^2 + \cdots + |a_n|^2 = 1 \tag{2-8}$$

Utilizing the appropriate "weight" factors, the complete time-free wave function in one dimension becomes

$$\psi(x) = a_1 \psi_1(x) + \cdots + a_n \psi_n(x) \tag{2-9}$$

The use of this representation for $\psi(x)$ yields an answer to a previous difficulty which was not discussed at the time it arose. To be specific, the problem of the electron in one of two adjacent potential boxes will be discussed again.

It is reasonable to suppose that if the electron is initially in the box on the left, it is very likely that when the system is examined again after a very short interval of time that the electron will still be in the region of the box on the left. However, in examination of Figure 2-5, it is seen that this is not the case under the conditions set forth for the solution. The reason that the solutions of Figure 2-5, which imply an equal probability of finding the electron in either box, are incorrect is that the assumption was implicitly made that the ψ pattern belonging to a single energy level would adequately represent the system. However, as was just shown, this is not correct. Instead it is necessary to proceed according to Equation (2-9), and assign the proper "weight" to all possible solutions before the probability is computed. For simplicity in illustration, the ψ function will be taken as

$$\psi(x) = a_1 \psi_1(x) + a_2 \psi_2(x) \tag{2-10}$$

where

a_1 is the weight associated with solution ψ_1
a_2 is the weight associated with solution ψ_2
ψ_1 and ψ_2 are the solutions shown in Figure 2-5

with the assumption being made that all higher solutions are negligible.

Now if the weight of either solution is identical ($a_1 = a_2$) or expressed according to the uncertainty principle, if knowledge of the electron energy state is very gross, then it would be likely to suspect that knowledge of the electron position would be correspondingly refined and a fairly accurate presentation of the position probability should be the result. The position probability is found as follows

$$P(x) = |\psi(x)|^2 = |a_1\psi_1 + a_2\psi_2|^2 = |\psi_1 + \psi_2|^2 \tag{2-11}$$

The operation indicated in Equation (2-11) is that of adding ordinates of the two solutions shown in Figure 2-5, and squaring the magnitude of the resultant. The result obtained is that shown in Figure 2-6, and it is observed that this is more in agreement with the situation as it might be expected to exist. That is, if the electron is initially in the box on the left, the position probability should indicate it on the left.
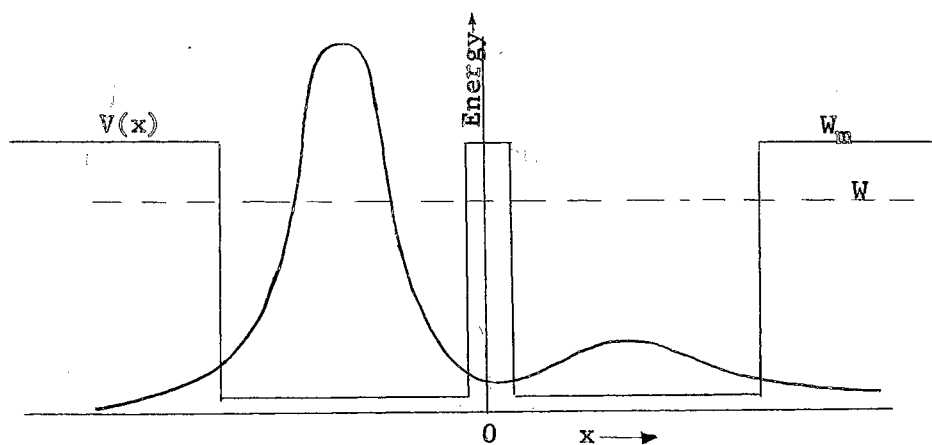

Figure 2-6. "Weighted" Probability Function

The next logical development and the last feature of quantum mechanical theory needed to qualitatively discuss the tunnel diode is that of the time varying wave function so that it may be ascertained in what manner the probability wave associated with the electron varies with time.

From all previous discussion it would be likely to presuppose that the time solution would exhibit a wave-like nature and, indeed, such is the case. The time varying Ψ function is found to be expressible mathematically as[5]

$$\Psi(x,t) = \Psi(x) \exp(-i 2\pi \gamma t) \tag{2-12}$$

where

γ = V/h and is the frequency of wave oscillation
V is the particle energy level
h is Plank's constant

Thus each Ψ pattern associated with an allowed energy level V_i, oscillates at its own particular frequency of V_i/h. The probability P(x) is

$$P(x,t) = |\Psi(x,t)|^2 = \Psi \Psi^* \tag{2-13}$$

where

Ψ^* denotes the complex conjugate of Ψ

The Ψ pattern representing the electron is thus oscillating in time and over some interval will represent all variations in the "weight" of the time-free Schrodinger wave solution for a given energy level.

We may now apply this time varying function to the problem previously discussed of the electron in one of the two adjacent potential boxes. Again giving identical "weight" to either of the solutions shown in Figure 2-5 it is seen by utilizing Equation (2-12), that the time varying wave function is

---

[5]Ibid.

$$\Psi(x,t) = a_1 \left(\Psi_1 \exp(-i2\pi\gamma_1 t) + \Psi_2 \exp(-i2\pi\gamma_2 t) \right) \qquad (2\text{-}14)$$

where

γ_1 and γ_2 are the frequencies associated with the time varying solutions[2]

The time varying probability $P(x,t)$ is

$$P(x,t) = \Psi\Psi^* \qquad (2\text{-}15)$$

substituting Equation (2-14) in Equation (2-15) yields

$$P(x,t) = \left[|\Psi_1|^2 + |\Psi_2|^2 - 2\mathrm{Im}\left(\Psi_1 \Psi_2^* \exp(-i2\pi(\gamma_2 - \gamma_1)t) \right) \right] \qquad (2\text{-}16)$$

which contains the important result that it is the difference or beat in

the two frequencies ($\gamma_2 - \gamma_1$) which determines the probability variation.

Now, ($\gamma_2 - \gamma_1$) is proportional to ($V_2 - V_1$), the difference in energy levels

of the two solutions, and, from previous discussion, proportional to an

inverse function of the barrier width. The probability variation is

actually found to be proportional to $\exp(-kd)$ where k is a constant and d

is the barrier width.[6]

Since the frequencies γ_1 and γ_2 are nearly equal, the total prob-

ability comprised of Ψ_1 and Ψ_2 starts with Ψ_1 and Ψ_2 in phase with

each other. As time progresses, the synchronism gets worse and worse

with the value of Ψ in one box growing at the expense of the Ψ in the

other box until finally the vibrations are completely out of phase with

each other. The situation is now the reverse of that indicated in Figure

2-6, with the larger probability wave in the box on the right. Note the

important result that the electron has traversed the region separating

the two potential boxes, a region that, according to the older concepts

----

[6]Ibid.

of classical mechanics, would completely contain the electron in the box on the left. Since the electron did not possess the required energy to climb the potential "hill", it might be said to have "tunneled through" and for this reason the phenomenon is known as the "tunnel effect". It is important to observe that, if "tunneling" is a desired effect, one requirement would be that of a small barrier width.

This at last is the end towards which the previous discussion has been pointed because it is precisely this tunneling effect which is predominant at low values of forward bias in the tunnel diode and is the major explanation for its unusual characteristics.

Now that the phenomena of tunneling can at least be qualitatively visualized it will be profitable to return to more conventional methods for the remainder of the theoretical discussion, keeping in mind that the tunneling process may be drawn upon as an explanation for an appropriate effect should the need arise.

CHAPTER III

PREPARATORY SEMICONDUCTOR THEORY

It has been found that the elements of Group IV of the periodic

table exhibit to a marked degree the property of crystal formation as

a consequence of being tetravalent or having four electrons in the outer

shell or orbit. The atoms of the elements in this group tend to share

electrons with other atoms to form covalent bonds so that the atoms are

formed into a tetrahedron crystalline struction which is illustrated

diagrammatically in Figure 3-1.[1] The dashed lines indicate a covalent

bond.
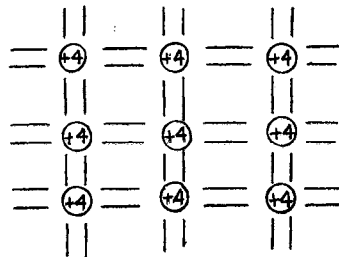
Figure 3-1. Group IV Diagrammatic Crystalline Struction

The situation as seen by any single valence electron in the crystal-

line struction may be found by applying the discussion contained in

Chapter II. The electron exists in a potential box determined by the

atomic core and the other three valence electrons. It is, however, also

influenced by the relatively close presence of a number of other potential

---

[1]Aldert Von Der Ziel, Solid State Physical Electronics, Prentice
Hall Publishing Company, Englewood Cliffs, New Jersey, 1957.

boxes due to the other atoms throughout the crystal so that any single
energy level corresponding to a solution of Schrodinger's equation is
split into an energy band containing 4N discrete quantum levels each
differing slightly in value from all others. Now assuming the material
is at $0^{\circ}K$, with no external fields applied, all of the electrons will
exist in the "ground state" solution and, since there are 4N electrons
to fill the 4N quantum states, all states in the valence energy band are
filled since the Pauli exclusion principle must hold. The density of the
4N available quantum energy states as a function of energy may be cal-
culated from the methods of quantum mechanics as[2]

$$N(E) = C(E - E_o)^{\frac{1}{2}} \tag{3-1}$$

where

> N(E) is the energy state density, i.e., number of states per unit
> volume
> E_o is the energy reference level
> C is a constant depending on the crystal

Depending upon which elements constitute the atoms in the crystal
structure, there may or may not be a distinct energy "gap" between the
highest energy level in the valence energy band and the energy level
necessary to make the electron available for conduction through the
crystal. This energy difference between the top of the valence band and
the "conduction" band is known as the forbidden gap.[3] It is the width
of this forbidden gap which largely determines the external electrical
characteristics of the crystal.

---

[2] W. Shockley, Electrons and Holes in Semiconductors, D. Van Norstad
Publishing Company, New York, 1956.

[3] E. Wolfendale (Editor), The Junction Transistor and Its Applications,
New York, 1958.

The crystalline structure will tend to reject all applied energies which are not of sufficient magnitude to raise an electron from the valence band into the conduction band. Those materials with a large forbidden gap consequently have high electrical resistivities and are good insulators. Those in which the top of the valence band overlaps the conduction band have low electrical resistivity and are good conductors. The materials between these two extremes have moderate values of resistivity and are known as semiconductors. Relative energy band diagrams for these three cases at $0^\circ K$ are illustrated in Figure 3-2.
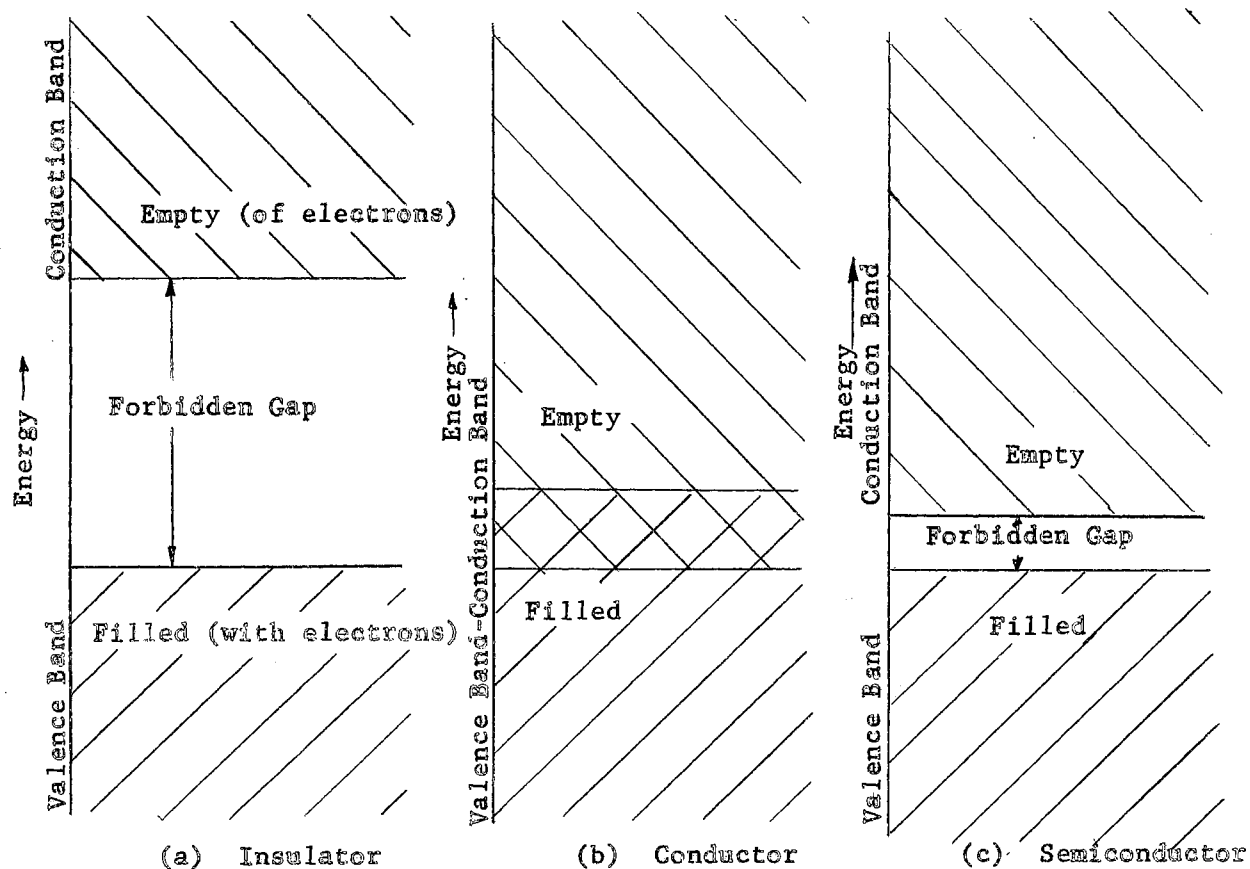
Figure 3-2. Energy Band Diagrams

Consider the semiconducting crystal energy band diagram illustrated in Figure 3-2 (c). As the temperature of the crystal is raised, a few of the higher energy electrons in the valence band will accept enough thermal energy to jump the forbidden gap and become available for conduction. As the electrons leave the covalent bond, they leave a vacancy or "hole" in the valence band which acts in many respects like a charge carrier of opposite sign since electrons from adjacent covalent bonds may fill the vacancy and thus in effect move the hole through the crystal. The energy band diagram of the semiconductor as it might appear at room temperature is shown in Figure 3-3 which illustrates that the energy distribution of the electrons has been altered from that shown in Figure 3-2 (c) in that some electrons are now in the conduction band, and some holes have appeared in the valence band.

By considering the bands as a continum rather than a discrete set, two physicists, Fermi and Dirac, working independently derived the energy distribution of the electrons as a function of temperature and crystalline material. The Fermi-Dirac distribution function for the number of electrons with energies between E and E + dE is[4]

$$n(E) = \frac{N(E)\, dE}{1 + exp[(E - E_f)/kT]} \qquad (3\text{-}2)$$

where

    n(E) is the number of electrons per unit volume with energy E
    N(E) is the number of energy states per unit volume with energy E
    (Equation 3-1)
    k is Boltzman's constant
    T is the absolute temperature in degrees Kelvin
    E_f is the Fermi energy level

The quantity, E_f, or the Fermi energy level may be thought of as that energy where the probability of a quantum state being occupied is 50

[4]Ibid.

percent. In Figure 3-2 (c) the Fermi level would fall in the exact center

of the forbidden gap since all states in the valence band are occupied

and all states in the conduction band are empty.

If certain selected impurities from Group III or V, which contain

an excess hole and electron respectively compared to the elements of
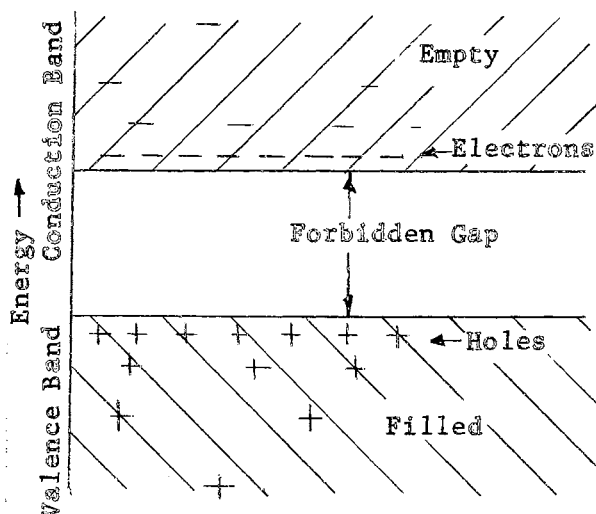
Figure 3-3. Carrier Distribution at Room Temperature

Group IV, are introduced into the crystalline structure during the crystal

growing process, these elements will displace a normal atom from the

crystal structure as illustrated in Figure 3-4.[5] Now the crystal structure

as a whole will contain a number of loosely bound electrons or holes which

are readily available for conduction.

(a) p-type Crystal   (b) n-type Crystal

Figure 3-4. Impurity Crystals

[5]Aldert Von Der Ziel, op. cit.

The impurities from Group III are selected so that the energy required to break an existing covalent bond and allow an electron to fill the hole, is only slightly above the energy level at the top of the valence band. Those from Group V are selected so that the energy levels of their loosely bound electrons are only slightly less than the energy level at the bottom of the conduction band. The energy band diagrams for these impure crystals at $0^{\circ}K$ are shown in Figure 3-5.
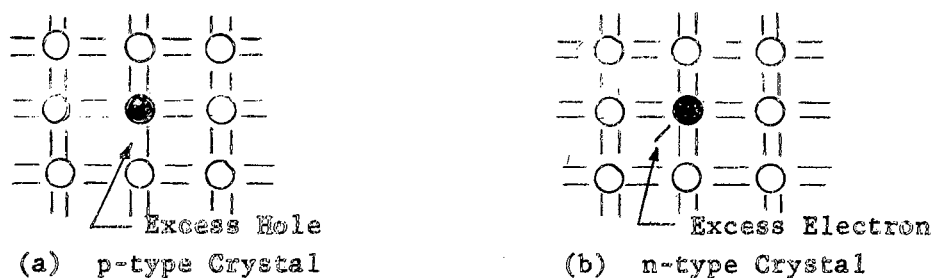

(a) p-type Crystal   (b) n-type Crystal
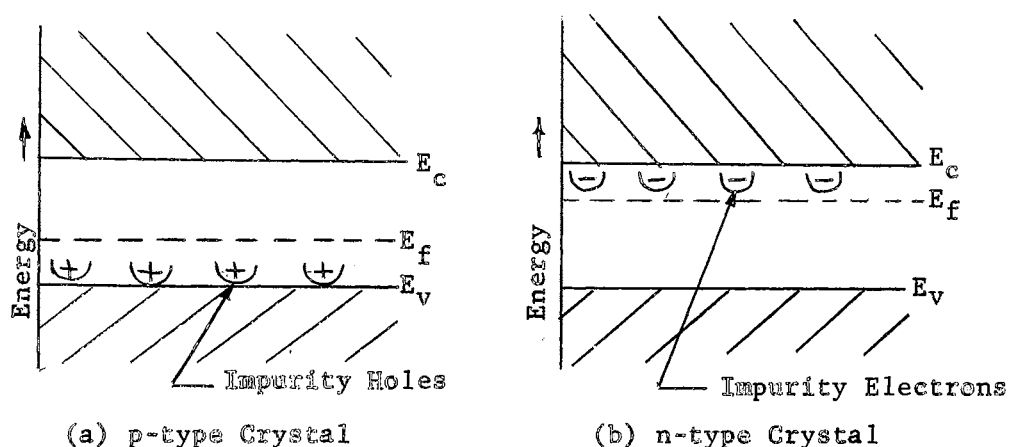
Figure 3-5. Impurity Crystal Band Diagram

It is now apparent that much less energy than for the pure crystal must be accepted by the impure crystalline structure in order to make some charge carriers available for conduction purposes. The impurity concentration is very small (approximately 1 part to 10^{8}) but, due to the large number of atoms present (approximately 10^{18} atoms/cm^{3}) there are still a large number of excess carriers existing so that relatively large currents may be maintained through the crystal. Note that the Fermi level of the p-type has been lowered and that of the n-type has been raised since the energy for 50 percent occupation probability is now different.

At room temperature, practically all the impurities have been ionized due to acquisition of thermal energy and their charges are available for conduction purposes in addition to those charges from the valence band which have jumped the forbidden gap. Due to the large difference in relative energy levels, the charges in the conducting bands are almost exclusively determined by the impurity. That is, in the n-type material for example, there are so many unbound electrons that holes are combined with one of these electrons almost as soon as the hole is formed. Thus, to a very good approximation, it may be said that the charge carriers in the p-material are holes and the charge carriers in the n-material are electrons.

The P-N Junction

Now if by some means, a p-type crystal could be grown or attached to an n-type crystal such that the transition from one region to the other is abrupt, the total system would fall into equilibrium in the following manner. At room temperature, the excess electrons in the n-material are in much greater concentration than the excess electrons in the p-material. Hence, they tend to diffuse across the junction much as molecules of a gas tend to diffuse from a region of high pressure to a lower pressure. These electrons which traverse the junction are in a region of large hole concentration and hence tend to combine with the p-type impurities to form ionized atoms. A similar situation holds for the excess holes in the p-material. As these ionizations occur, an electric field is built up across the junction due to the different charges on the immobile ions. This process tends to continue until the

transition region has been swept free of excess charge carriers and an equilibrium condition is reached. Eventually the tendency of the excess charges to flow by diffusion is just matched by the force exerted on the excess charges by the electric field since the field is in such a direction as to oppose the flow by diffusion. The situation is illustrated in Figure 3-6.
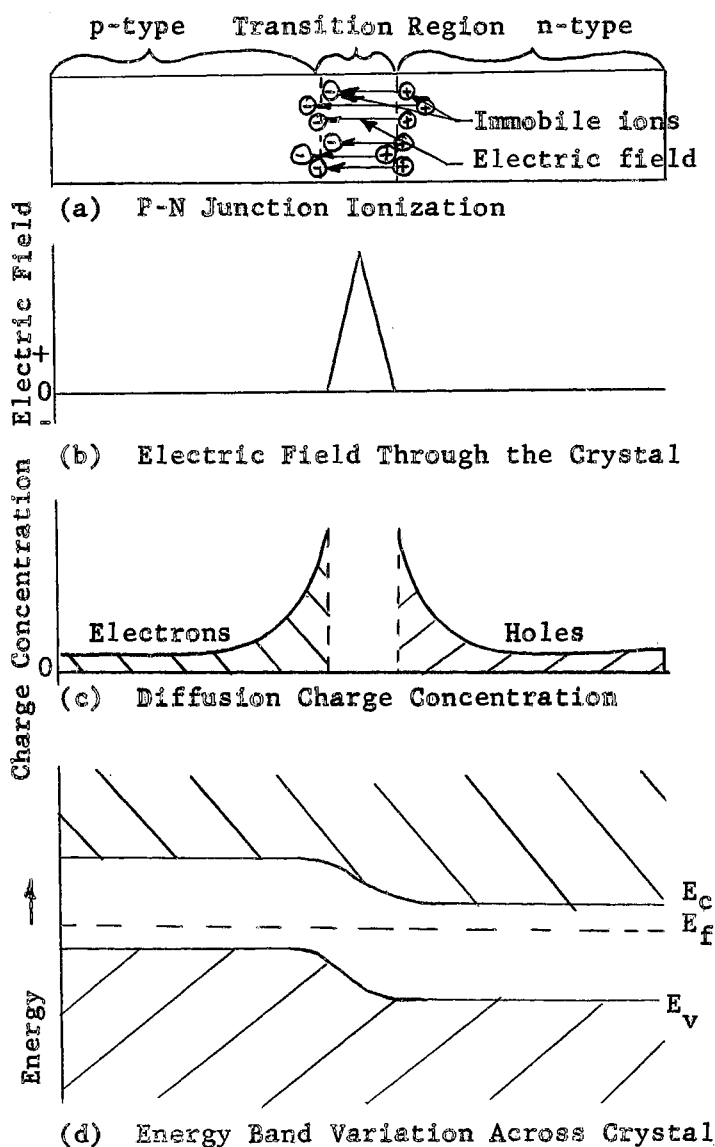

(a) P-N Junction Ionization

(b) Electric Field Through the Crystal

(c) Diffusion Charge Concentration

(d) Energy Band Variation Across Crystal

Figure 3-6. Pertinant Variables Across a P-N Junction

Even in the equilibrium condition, there will be some holes in the p-side which attain the necessary energy to surmount the potential barrier at the junction. Due to thermal agitation, holes are also being generated in the junction region which tend to flow away from the junction in the direction of the electric field and opposite to the high energy holes. Since at zero external applied voltage no net current may flow, these two currents must be equal and opposite. A similar analysis holds for electrons and when the analytic expressions are solved for the concentration variation, they yield the results of Figure 3-6 (c) which indicates an expotential fall-off in concentration as the carriers move into the region where they are in the minority. Physically, this is due to the large tendency for ionization recombination which is present. In essence this is a storage or capacitive effect and greatly limits the frequency response of the p-n junction since it is this minority current which is predominant when the external bias is in the forward direction.

Figure 3-6 (d) is worthy of comment at this time. When the junction is formed and the electric field is stabilized across it the energy states on the p-side are raised, the energy states on the n-side are lowered, and equilibrium is reached when the Fermi levels of both sides of the crystal attain the same value.[6] The result of this energy band shifting across the junction as shown in Figure 3-6 (d) may be thought of in this way. The electrons in the lower level energy states on the n-side are situated directly opposite forbidden regions at the same energy level so their energy states must be raised if they are to pass to the p-type material. The same is true for the holes in the p-type although the situation is not so immediately obvious.

_____

[6]Wolfendale, op. cit.

It may now be observed that the amount which the bands are shifted relative to each other is a function of the impurity concentrations of both the P and N types of semiconductors since in the discussion of Figure 3-5, it was indicated that the shifting of the Fermi level from the mid-point of the forbidden region depended on how many impurity ions were present. Thus, for large impurity concentrations, there is a large difference in initial Fermi levels which yields a large band shift when the sections are joined. In fact if the impurity concentration is so high that the impurity energy levels form a fairly wide band of their own, these bands may overlap the available energy bands in the pure semiconductor and actually move the Fermi level into the overlapping region. This is one condition which must be satisfied in the fabrication of the tunnel diode.

If an external bias is applied to the junction, the bands are again shifted relative to each other depending upon the direction of the applied field. If the field is in such a direction to oppose the field already existing across the junction, the bands will tend to become aligned horizontally, thus making available more energy states for the minority carriers and the external current will increase greatly. If the bias is reversed the junction barrier height will be increased and the majority of the small current which flows will be due to the thermal generation of carriers in the junction region.

CHAPTER IV

SEMICONDUCTOR THEORY OF THE TUNNEL DIODE

Qualitative Theory

If the reverse bias on an ordinary p-n junction is increased, an

eventual point will be reached where the V-I characteristic of the diode

exhibits an unusual feature. The diode starts conducting current heavily

and the current changes by an extremely large amount for very little

increase in applied voltage. This V-I characteristic is indicated in

Figure 4-1. The theoretical origin of this phenomena may, in at least

some cases, be ascribed to the electron tunneling process described in

Chapter II. It will be of value to examine Figure 4-1 to obtain a quali-

tative explanation of the breakdown. The numbered regions correspond to

the various energy levels shown in Figure 4-2.
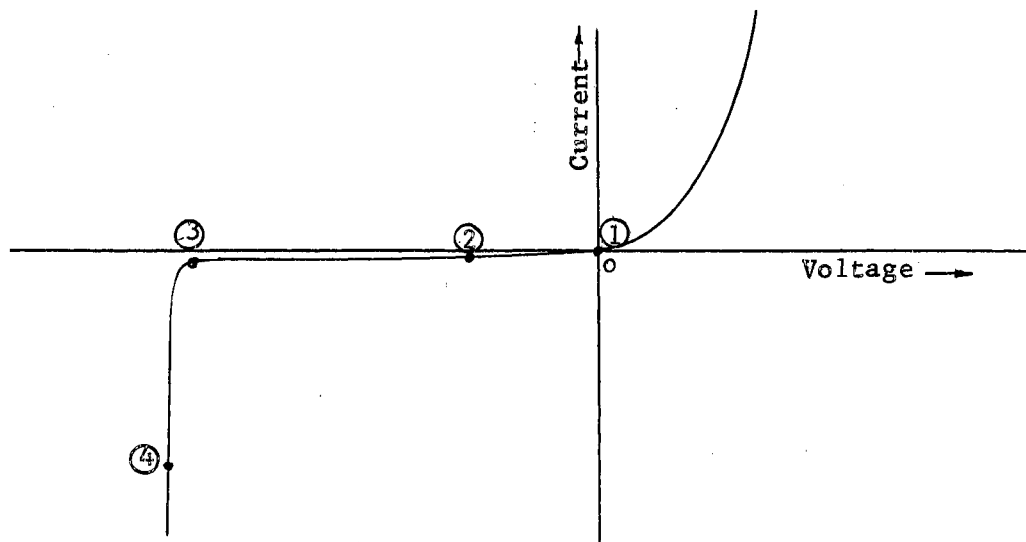
Figure 4-1. V-I Characteristics of a Lightly "Doped" P-N Junction

30
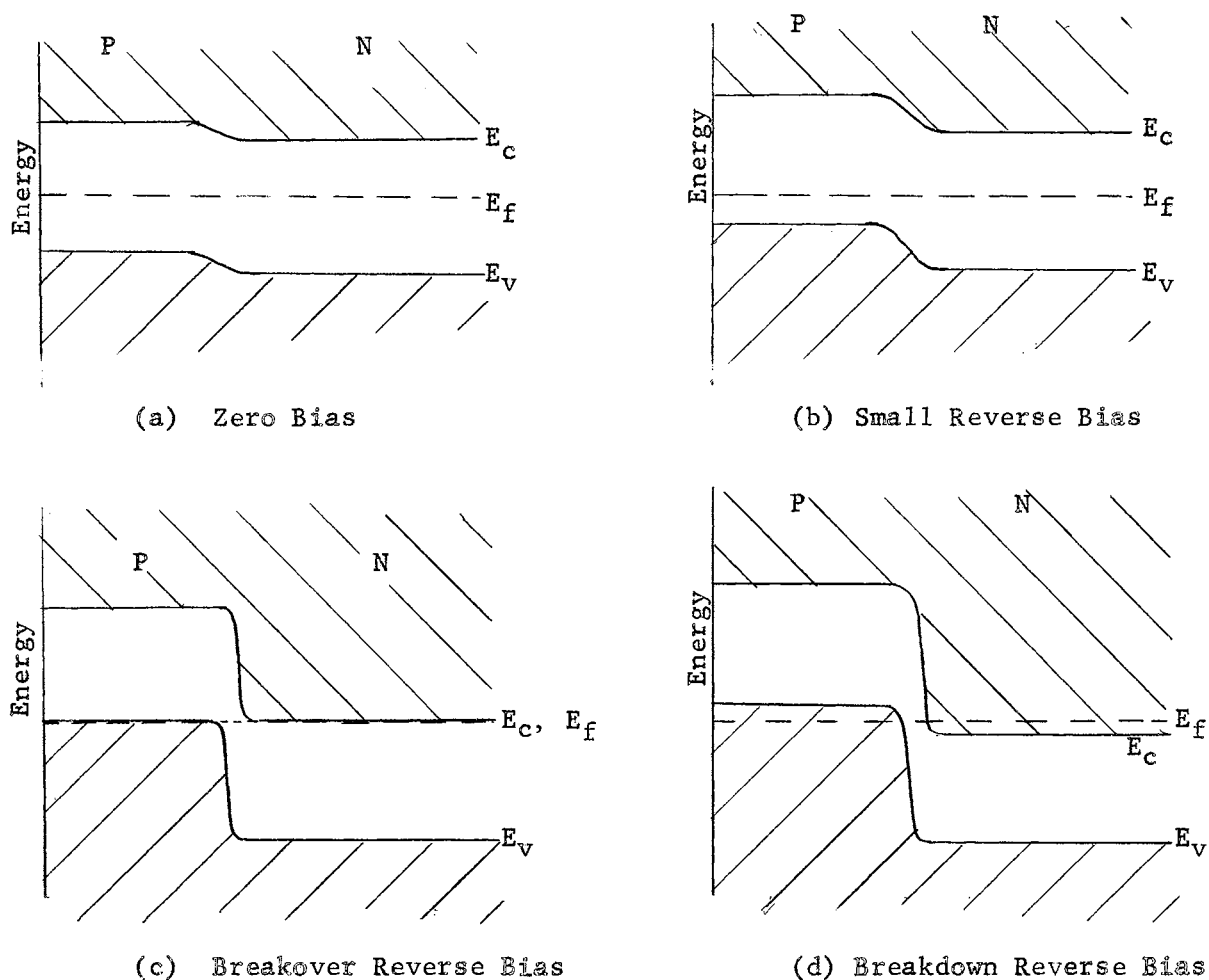
Figure 4-2. Relative Energy Band Diagrams

Nothing of unusual interest is shown in Figure 4-2 (a). The diode
is at zero bias and no net current flows across the junction until
(Figure 4-2 (d)) the reverse bias is applied which causes the small current
due to electron-hole generation in the junction to flow. As the bias is
increased, the breakover voltage will eventually be reached (Figure 4-2 (c))
and large currents will start to flow. The explanation of how this occurs
is fundamental to an understanding of tunnel diode action and is quali-
tatively fairly simple to understand.

In Figure 4-2 (c) the highest energy electrons in the valence band of the p-side of the junction are just approaching the empty energy states at the same energy level on the n-side of the junction. As a consequence, the only impediment to prevent their flowing to the n-side is the junction barrier. In Chapter II it was shown that a definite probability existed for an electron to tunnel through a potential barrier to an empty state if the energy level on both sides of the barrier were equal. (See Equation 2-16.) This is exactly what occurs in this case. The electrons from the valence band of the p-side tunnel through the junction barrier to the empty available energy states in the conduction band of the n-side and constitute a current flow.

Since the impurity concentration and thermal agitation have raised some electrons to the conduction band in the n-side and also made available states in the valence band of the p-side, the electrons from the conduction band of the n-side also tunnel through the barrier to the available states in the valence band of the n-side but being much smaller in number the two will still add algebraically to give a large net current flow across the junction.

As the impurity concentration or "doping" of the intrinsic material is increased it is found that the breakover voltage is lower and lower in magnitude as shown in Figure 4-3.[1] This may be understood by recalling previous discussion where it was indicated that the value of the zero bias junction field became larger and larger for increased doping. Since the initial field is larger, then obviously less bias is required in

[1]Bernard Sklar, "The Tunnel Diode - It's Action and Properties," Electronics, November, 1959.

order to align the empty states in the n-side conduction band to the

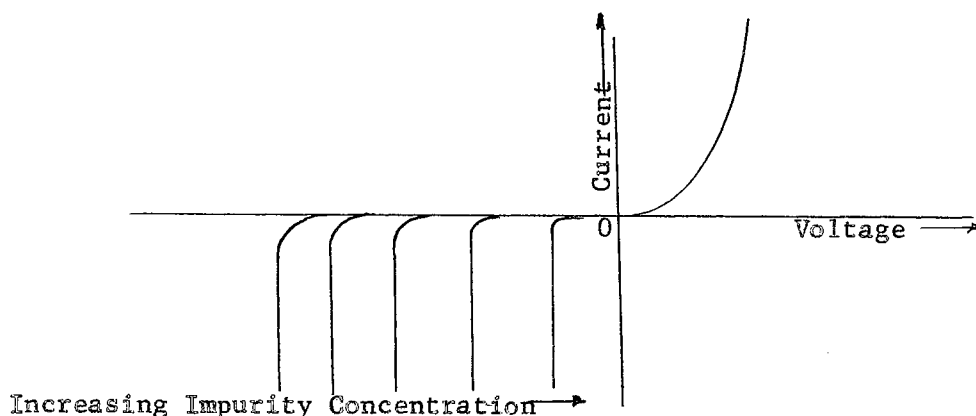filled states in the p-type valence band.

Figure 4-3. Effect of Increased "Doping"

If the impurity concentration were made high enough, the junction

would be in a breakdown condition at zero bias and would continue in such

a condition until the forward bias was of such a value that the valence

and conduction levels on the opposite sides of the junction were separated

by an energy gap. This high impurity concentration is one requirement for

tunnel diode operation along with the requirement that the junction barrier

be extremely thin to facilitate the tunneling.

A typical V-I characteristic for the tunnel diode is shown in Figure

4-4 where the numbered regions correspond to the band level diagrams of

Figure 4-5.[2]

At (1), the diode is in breakdown condition and the tunneling currents

are equal and opposite. These currents are produced by the quantum

---

[2]"Tunnel Diodes", Technical Information Sheet, General Electric
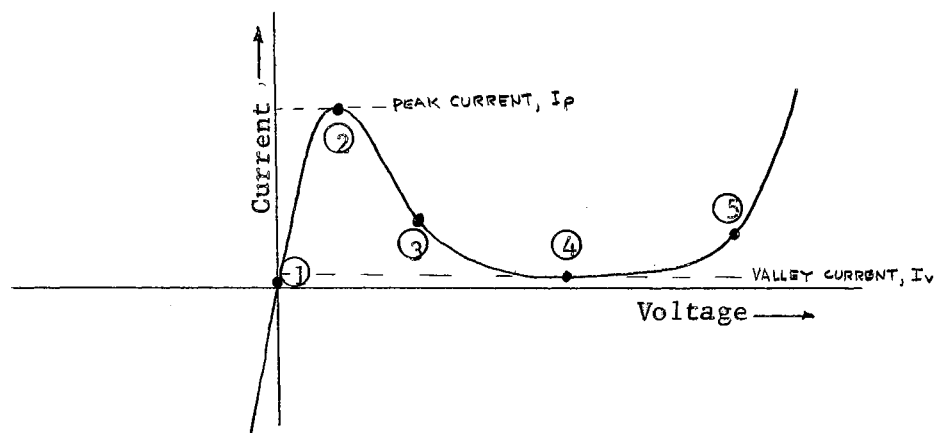Research Laboratory, New York, 1959.

Figure 4-4. Tunnel Diode V-I Characteristics

mechanical probabilities describing electron position in adjacent potential boxes which were discussed in Chapter II. As a small positive bias is applied, the electrons tunneling from the conduction to the valence band greatly outnumber the electrons tunneling from the valence to the conduction band since there are many more empty available states in the valence band and the electron current will increase to a positive maximum.

Now as the bias is raised higher (3), some of the electrons in the conduction band are directly opposite the forbidden gap and since there are no available states for the electrons so described to tunnel to, the total tunneling currents starts to fall with an increasing bias. Thus, the V-I characteristic exhibits a decrease in current for an increase in voltage? current. This is the negative resistance region for the diode.

As the bias is further increased (4),(5), the minority current starts to flow in the forward direction as the potential barrier is reduced enough to allow the high energy excess carriers to surmount the potential barrier and contribute to the conduction.

The principle feature which is advantageous in this type of diode, is that the current carriers are not minority carriers in the negative
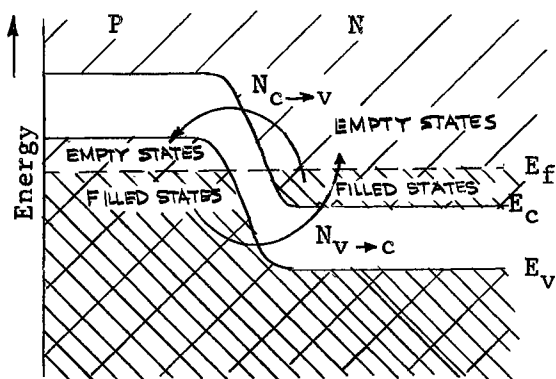
Region (1) Zero Bias

$n_{v \to c}$ electrons tunneling from valence to conduction band.

$n_{c \to v}$ electrons tunneling from conduction to valence band.
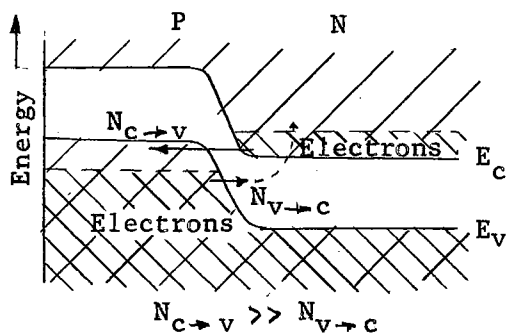
Region (2) Small Positive Bias

Region (3) Positive Bias Increasing

Region (4) Higher Bias, Tunnel Current Falls to Zero, Minority Current Starts

Region (5) Normal Diode Conduction

Figure 4-5. Energy Band Variation of the Tunnel Diode for Positive Bias

resistance region and hence the high frequency response of the junction

should be greatly improved over the normal junction diode. Also the

resistance of the junction to the effects of high energy radiation should

be improved. Such is found to be the case in the tunnel diode.

Quantitative Theory

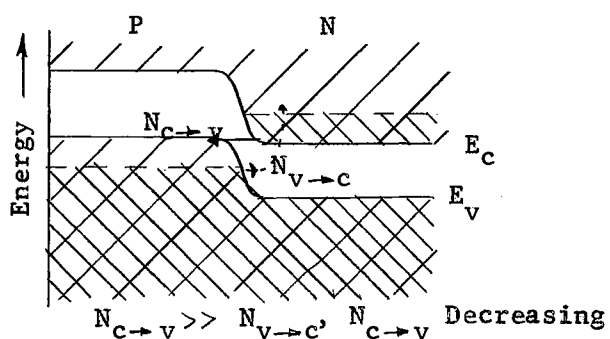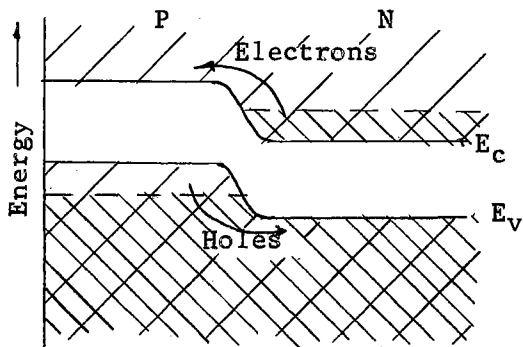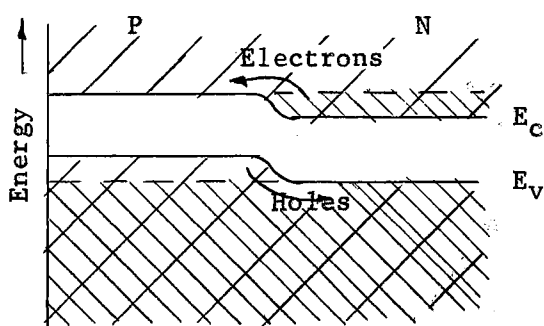The first important factor which affects the current flow across

the junction is the probability that any electron which strikes the

potential barrier will tunnel through. It was seen in Chapter II that

this probability would depend expotentially upon the barrier thickness.

However, in the actual case, the potential barrier is not linear as was

assumed in Chapter II. Under these circumstances, the Schrodinger

equation has non-constant coefficients and the exact solution is exceed-

ingly difficult. However, an approximate development indicates[3]

$$\psi(x) = \exp\left[-\int_0^b \sqrt{a(W-V(x))}\, dx\right] \tag{4-1}$$

Now the tunneling probability is

$$T = \psi\psi^* = \exp\left[-2\int_0^b \sqrt{a(W-V(x))}\, dx\right] \tag{4-2}$$

where

   T is the tunneling probability
   a is $\dfrac{8\pi^2 M_{eff}}{h^2}$
   M_{eff} is the effective mass of electron in crystal
   b is the barrier thickness

To a good approximation, the potential barrier is triangular in

shape as in Figure 4-6, and the electric field across the junction is

constant.

_____

[3] I. A. Lesk, N. Holonyak, Jr., U. S. Davidsohn, and M. W. Aarons, "Germanium and Silicon Tunnel Diodes - Design, Operation, and Application", 1959 Wescon Convention Record.
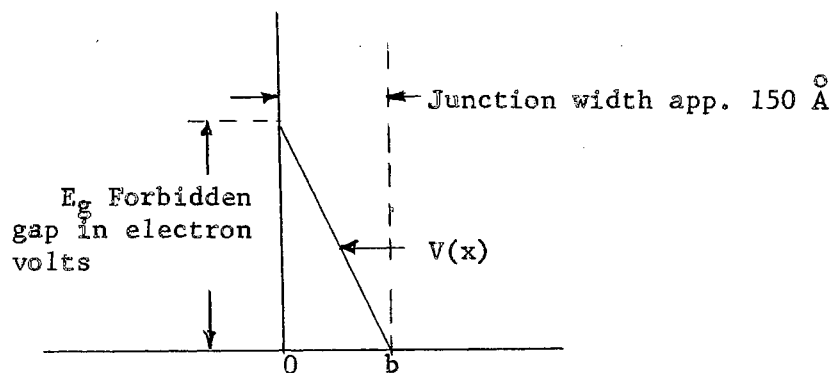
Figure 4-6. Potential Across the Junction

Under these conditions the solution of Equation (4-2) becomes

$$T = exp\left[-\frac{8\pi\sqrt{2m_{eff}}\,E_g^{\frac{3}{2}}}{3\hbar e\,\mathcal{E}}\right] \tag{4-3}$$

where

$\mathcal{E}$ is the electric field in the barrier
e is the charge on the electron

To obtain the tunneling rate, T is multiplied by the number of collisions per second of electrons against the potential barrier as follows.

From Newton's law, the time rate of change of momentum is equal to the applied forces. The quantity under the square root in Equation (4-2) is inversely proportional to the wavelength of the ψ function and through the de Broglie theorm relating the wavelength and momentum of a moving electron T becomes proportional to the electron momentum. Thus,

$$T = exp\left[-2\int_0^b \sqrt{a(W-V(x))}\,dx\right] = exp\left[-\int_0^b \frac{2\pi}{\lambda(x)}\,dx\right] \tag{4-4}$$

where

$\lambda(x)$ is the wavelength of the ψ function, $\lambda(x) = \frac{2\pi}{\sqrt{a(W-V(x))}}$

According to the de Broglie theorm

$$\text{Momentum} = p(x) = \frac{\hbar}{\lambda(x)} \tag{4-5}$$

Therefore

$$T = \exp\left[-\frac{4\pi}{\hbar}\int_0^b p(x)\,dx\right] \tag{4-6}$$

Now since

$$\Sigma \text{ applied forces} = \text{mass x accelaration} = \frac{d(p(x))}{dt} = e\mathcal{E} \tag{4-7}$$

the change in momentum with respect to time becomes

$$\frac{\Delta p}{\Delta t} = e\mathcal{E} = \hbar\frac{\Delta\left(\frac{1}{\lambda}\right)}{\Delta t} \tag{4-8}$$

or

$$\frac{1}{\Delta\lambda} = \frac{e\mathcal{E}}{\hbar}\Delta t \tag{4-9}$$

Now if an electron striking the barrier does not tunnel through, then its momentum must reverse in the barrier and then reverse again before the electron strikes the barrier again. This means that the wavelength associated with the electron must take on all allowable values from the upper band edge to the lower band edge and back again to its original value. It was seen in Chapter II that the band width of solutions was proportional to the spacing between potential boxes. In a crystal, this spacing corresponds to the atomic spacing and may be signified by a constant called the lattice spacing A.

It is found that the change in wavelength is exactly equal to this lattice spacing constant or

$$\Delta\lambda = A \tag{4-10}$$

Using this in Equation (4-9), and solving for the time between collisions gives

$$\Delta t = \frac{\hbar}{e\mathcal{E}\Delta\lambda} = \frac{\hbar}{e\mathcal{E}A} \tag{4-11}$$

Dividing Equation (4-11) into Equation (4-3) yields the tunneling probability per second or tunneling rate Z as

$$Z = \frac{T}{\Delta t} = \frac{Ae\mathcal{E}}{\hbar}\exp\left[-\frac{8\pi}{3e\mathcal{E}\hbar}\sqrt{2m_{EFF}}\,E_q^{\frac{3}{2}}\right] \tag{4-12}$$

When this equation is plotted against $\mathcal{E}$, with normal values for the other factors, it exhibits an extremely sharp rise in tunneling rate at around $10^6/cm.$[4] Again this illustrates the need for a narrow junction of around 100 $\overset{o}{A}$ since the value of E_g for most semiconducting materials is usually low and a high tunneling rate is desirable.

The remainder of this discussion is similar to that followed by Esaki,[5] but it will prove more illustrative to refer it to Figure 4-7.

Following the normal rules of probability, the probability of two events occuring simultaneously is equal to the product of the individual probabilities. Thus the probability or "tunnel" current of Energy dE flowing in the unbiased p-n junction from the conduction band in the n-region to the valence band in the p-region is equal to the number of
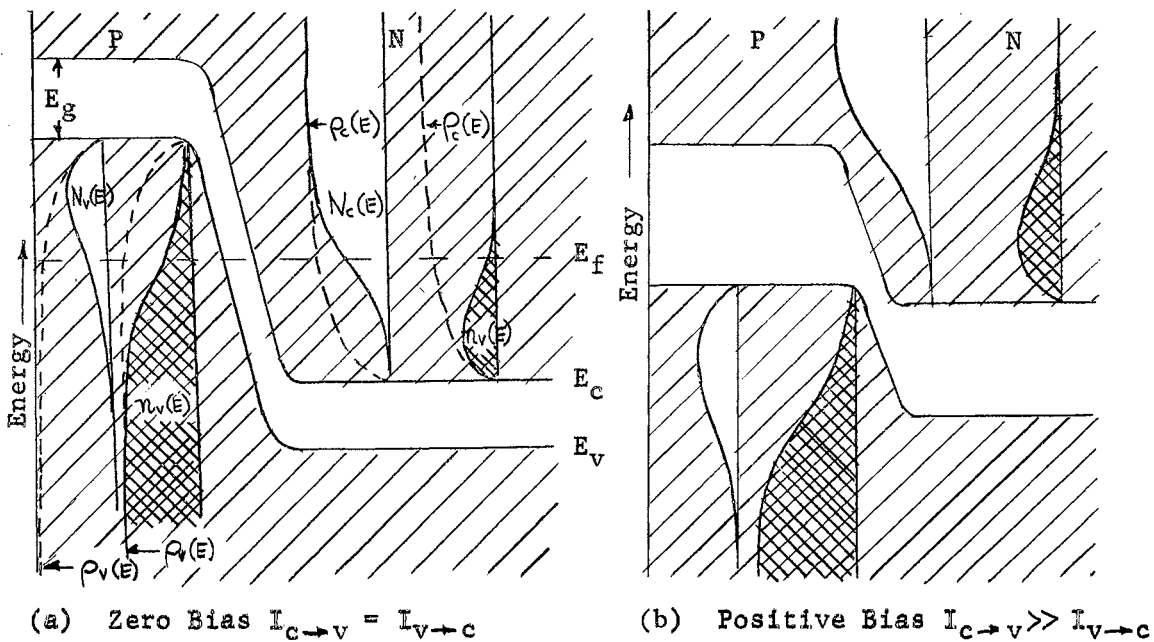

(a) Zero Bias $I_{c \to v} = I_{v \to c}$ (b) Positive Bias $I_{c \to v} \gg I_{v \to c}$

Figure 4-7. Density Functions to Determine Tunnel Current

---

[4]Ibid.

[5]Esaki, op. cit.

electrons in the conduction band times the number of available states in the valence band (unoccupied states) times the tunneling rate from the conduction band to an idential energy level in the valence band. The total current $I_{c \to v}$ is obtained by integrating over the range of overlapping energy states.

Quantitatively, the number of electrons in the conduction band in n-region is determined by the product of the Fermi-Dirac probability distribution function

$$f_c(E) = \frac{1}{1 + \exp\left(\frac{E - E_f}{kT}\right)} \tag{4-13}$$

where

$f_c(E)$ is the Fermi-Dirac distribution function

and the density of available states that the distribution function applies to. It is assumed that the available state density is given by Equation (3-1).

$$\rho_c(E) = C\left(E - E_c\right)^{\frac{1}{2}} \tag{3-1}$$

where

ρ_c is the density of available states of conduction band in n-region
E_c is the bottom of conduction band in n-region

Thus the number of electrons $\eta_c(E)$ available is

$$\eta_c(E) = f_c(E) \cdot \rho_c(E) \tag{4-14}$$

This function is shown qualitatively in Figure 4-7.

Next, the number of holes in the conduction band in the n-region is found by[6]

$$N_c(E) = 1 - \eta_c(E) \tag{4-15}$$

---

[6]Wolfendale, op. cit.

where

$N_c(E)$ is the number of holes in the conduction band in n-region

The electron and hole density in the valence band are found similarly

yielding

$$N_V(E) = \left(1 - f_V(E)\right) \cdot \rho_V(E) \tag{4-16}$$

where

$N_v(E)$ is the number of holes in the valence band
E_v is the top of valence band in the n-region
ρ_V is $C(E_{v-E})^{1/2}$ in the p-region

and

$$\eta_V(E) = 1 - N_V(E) \tag{4-17}$$

where

$\eta_v(E)$ is the number of electrons in the valence band in the p-region

All these functions are illustrated in Figure 4-7.

Proceeding as indicated previously the tunneling current $I_{c \rightarrow v}$ is

found by

$$I_{c \rightarrow v} = A \int_{E_c}^{E_v} \eta_c(E) \cdot N_V(E) \cdot Z_{c \rightarrow v} \, dE \tag{4-18}$$

where

A is the junction area

A similar expression holds for current from the valence band to the

conduction band.

$$I_{v \rightarrow c} = A \int_{E_c}^{E_v} \eta_V(E) \cdot N_c(E) \cdot Z_{v \rightarrow c} \cdot dE \tag{4-19}$$

Over the bias range indicated, it is a reasonable assumption that

$Z_{c \rightarrow v}$ equals $Z_{v \rightarrow c}$ and it is thus found that the total net current across

the junction is

$$I(V) = I_{NET} = B \int_{E_c}^{E_v} \left\{ f_c(E) - f_V(E) \right\} \cdot Z \cdot (E - E_c)^{\frac{1}{2}} (E_v - E)^{\frac{1}{2}} \, dE \tag{4-20}$$

where

   V is the applied bias
   B is a constant dependent on the crystal

   This is the final result of this section, an expression for the

current as a function of the applied voltage. Equation (4-20) is a

particularly difficult integral to evaluate analytically but Esaki[7] gives

the calculated V-I characteristic obtained from Equation (4-20), and it

compares favorably with that which is obtained experimentally.

---

[7]Esaki, op. cit.

CHAPTER V

AMPLIFIER ANALYSIS

Consider the region between I_p and I_v in Figure 4-4. The slope of the V-I characteristic is negative between these limits implying that the diode may act as an energy source to yield amplification if the proper circuit conditions are maintained. To examine these considerations in more detail consider the circuit in Figure 5-1. The diode is assumed to be biased positively and the bias source is isolated from the a-c circuit.

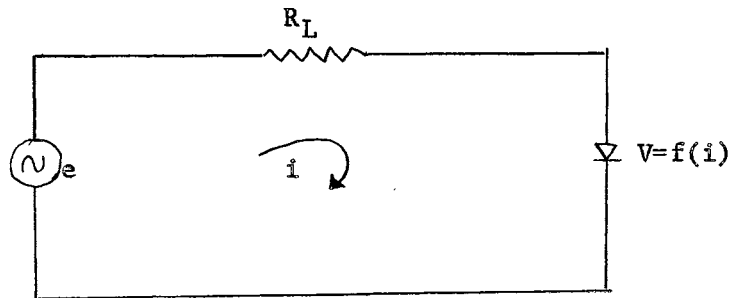Figure 5-1. Simple Amplifier Circuit

The loop voltage equation may be written as

$$e = i R_L + f(i) \tag{5-1}$$

where f(i) defines the V-I characteristic of the diode

Solving for f(i),

$$f(i) = e - i R_L \tag{5-2}$$

This is the familiar load line equation employed so often in vacuum tube circuit analysis with an exception in that the voltage e is not necessarily a constant but may vary with time.

43

The instantaneous diode resistance may now be defined as

$$r_d = \frac{\partial f(i)}{\partial i} \tag{5-3}$$

and it may be seen from Figure 4-4 that r_d goes through a negative

minimum at the inflection point where the characteristic changes from

concave downward to concave upward.

When Equation (5-2) is solved graphically the condition follows in

Figure 5-2 that unless

$$R_L \leq r_{d \, MIN} \tag{5-4}$$

the load line has the possibility of intersecting the characteristic at

three points instead of one.
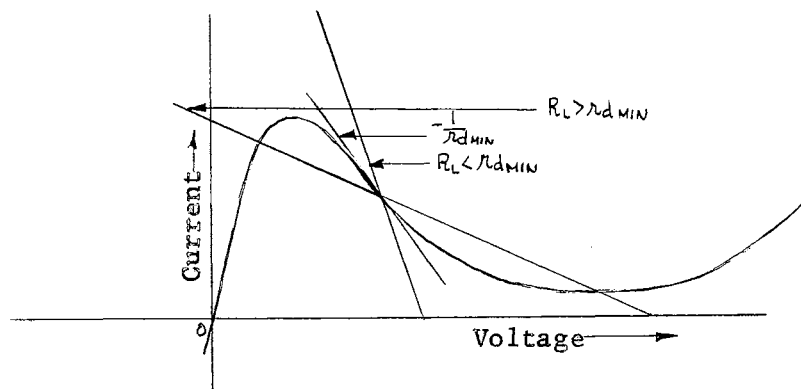
Figure 5-2. Graphical Interpretation of Stability Conditions

This multiple intersection implies an unstable configuration which must

be avoided in the amplifier if relaxation oscillations are to be avoided.

Assuming Equation (5-4) is satisfied, let e in Equation (5-2) be

$$e = E_m \sin \omega t \tag{5-5}$$

Now there is a simple graphical interpretation which indicates how

amplification is possible. As e changes, the load line retains the same

slope but shifts horizontally with the voltage e intersecting the V-I

characteristic at different points.

The time varying load line along with the output current and voltage of
the device may be found graphically by following the arrows in Figure 5-3.
The output voltage is larger than the input voltage indicating that
amplification has been obtained.

Figure 5-3. Graphical Solution for Diode Waveforms

The actual equivalent circuit of the diode is more complex than that
of a simple negative resistance and has been shown to be the circuit in
Figure 5-4.[1]

_____

[1]Sommers, op. cit.
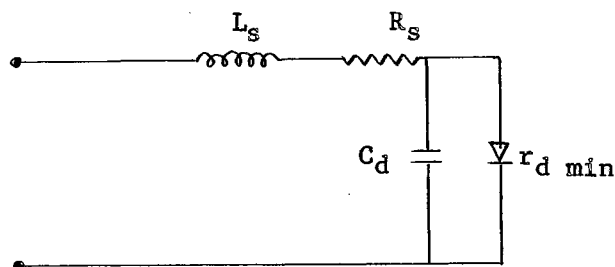
Figure 5-4. Tunnel Diode Equivalent Circuit

where

L_s is the lead and body inductance of the diode
R_s is the bulk resistance of the diode
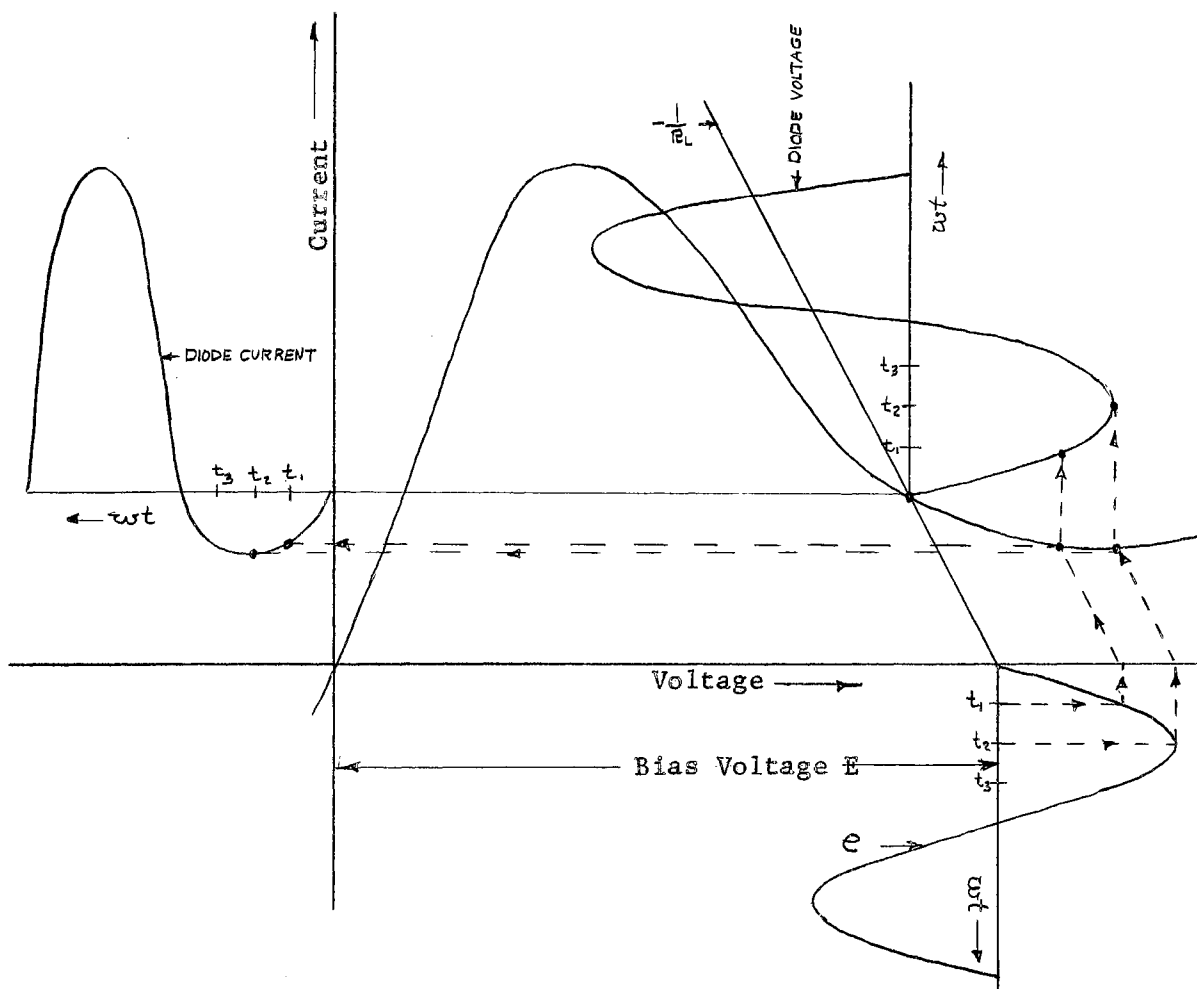C_d is the function capacitance
$r_{d\ min}$ is the slope of the V-I characteristic of the inflection point

In order to reduce the distortion which is apparent in the output

waveforms of Figure 4-3, the amplification mode usually used is that of

a tuned circuit. The essential circuit can be reduced to one similar to

Figure 5-4 where L_s, R_s C_d, and $r_{d\ min}$ are replaced by L, R_t, C, and -R

where -R is considered to be linear for small signal analysis,

$$L = L_s + L_1 \tag{5-5a}$$

$$R_t = R_s + R_1 \tag{5-5b}$$

$$C = C_d + C_c \tag{5-5c}$$

where

L_1 is the external circuit inductance
R_1 is the external circuit resistance
C_c is the case and stray capacitance

This amplifier circuit has been analyzed from the Nyquist plot of

its loop impedance[2] and design procedures have been obtained. However,

it will be of interest to examine the circuit from a different point of

view so that more precise design equations may be developed.

---

[2] U. S. Davidsohn, et. al., op. cit.
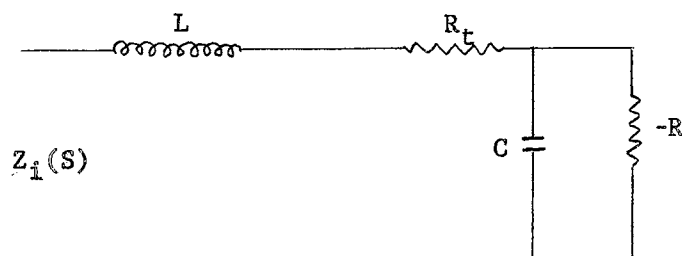
Figure 5-5. Amplifier Equivalent Circuit

The input impedance of Figure 5-5 is

$$Z_i(S) = \frac{S^2 LRC + (RR_tC - L)S + (R - R_t)}{SRC - 1} \tag{5-6}$$

where

S is the ~~Laplacian operator.~~ Laplace operator or Laplacian variable

Amplifier stability, which is a prime consideration, will be preserved if the zeros of $Z_i(S)$, which are the poles of the response, do not lie in the right hand half of the S plane.[3]

The roots of Equation (5-6) are

$$S = -\frac{(RR_tC - L)}{2LRC} \pm \sqrt{\left(\frac{RR_tC - L}{2LRC}\right)^2 - \left(\frac{R - R_t}{LRC}\right)} \tag{5-7}$$

The circuit will be unconditionally stable if

$$(1) \qquad R_t > \frac{L}{RC} \tag{5-8}$$

and

$$(2) \qquad R_t < R \tag{5-9}$$

Now defining

$$\delta \omega_A = \frac{1}{2}\left(\frac{R_t}{L} - \frac{1}{RC}\right) \tag{5-10}$$

$$\delta = \frac{1}{2}\left(\frac{R_t}{L} - \frac{1}{RC}\right)\sqrt{\frac{LRC}{R - R_t}} \tag{5-11}$$

$$\omega_A = \sqrt{\frac{R - R_t}{LRC}} \tag{5-12}$$

Equation (5-7) becomes

$$S = \omega_A\left(-\delta \pm \sqrt{\delta^2 - 1}\right) \tag{5-13}$$

---

[3] J. L. Brown and Peter M. Schultheiss, <u>Introduction to the Design of Servomechanisms</u>, John Wiley and Sons, Inc., New York, 1958.

Obviously two cases exist (note δ must be positive and ω_a must be real for stability to be maintained) depending on whether the radical is real or imaginary. If the radical is real ($\delta > 1$) the amplifier will have a non-selective response and if the radical is imaginary ($\delta < 1$) the amplifier response will be selective. It is the latter case which will be considered most carefully in this analysis.

Equation (5-13) is now substituted into Equation (5-6) so that the numerator corresponds to an accepted convention for the description of a normalized quadratic response. The numerator of Equation (5-6) will then be

$$\left[\left(\frac{S}{\omega_A} \right)^2 + 2\delta \left(\frac{S}{\omega_A} \right) + 1 \right] \left[R - R_t \right] \tag{5-14}$$

Now the input impedance as a function of frequency becomes

$$Z_i(\omega) = \frac{\left[\left(\frac{j\omega}{\omega_A} \right)^2 + 2\delta \left(\frac{j\omega}{\omega_A} \right) + 1 \right] \left[R - R_t \right]}{j\omega RC - 1} \tag{5-15}$$

A convenient power gain definition for the amplifier circuit is "insertion power gain". This is defined as the ratio of the power in the load with the amplifier inserted (P_{L1})

$$P_{L_1} = \left| i_4 \right|^2 Re\, Z_L \tag{5-16}$$

to the power in the load with the amplifier removed (P_{L0}).

$$P_{L_0} = \left| i_{0_1} \right|^2 Re\, Z_L \tag{5-17}$$

Therefore, G the power gain is

$$G = \left| \frac{P_{L_1}}{P_{L_0}} \right| = \left| \frac{i_4}{i_0} \right|^2 \tag{5-18}$$

It will be convenient to consider the generator and load to be purely resistive. Under these conditions

$$G = \left| \frac{\frac{e_q}{Z_{L_1}}}{\frac{e_q}{Z_0}} \right|^2 = \left| \frac{R_L + R_q}{Z_i} \right|^2 \tag{5-19}$$

With Equation (5-15) substituted for Z_i in Equation (5-19) the power gain is

$$G = \left| \frac{R_L + R_q}{R - R_t} \right|^2 \cdot \left| \frac{j\omega RC - 1}{\left(\frac{j\omega}{\omega_A}\right)^2 + 2\delta\left(\frac{j\omega}{\omega_A}\right) + 1} \right|^2 \tag{5-20}$$

Expressed in db

$$G_{db} = 10 \log_{10} \left| \frac{R_L + R_q}{R - R_t} \right|^2 + 10 \log_{10} \left| -1 + j\omega RC \right|^2$$
$$- 10 \log_{10} \left| \left(\frac{j\omega}{\omega_A}\right)^2 + 2\delta\left(\frac{j\omega}{\omega_A}\right) + 1 \right|^2 \tag{5-21}$$

Note that there are three possible terms which might yield power gain. These terms will be defined as

$$G_c = \left| \frac{R_L + R_q}{R - R_t} \right|^2 \tag{5-22}$$

$$G_d = \left| -1 + j\omega RC \right|^2 \tag{5-23}$$

$$G_R = \frac{1}{\left| \left(\frac{j\omega}{\omega_A}\right)^2 + 2\delta\left(\frac{j\omega}{\omega_A}\right) + 1 \right|^2} \tag{5-24}$$

Equation (5-24) has been plotted in db, and is shown in Figure 5-6 along with Equations (5-22) and (5-23) which are of nominal value.[4]

Although Figure 5-6 has been derived by conventional methods there is a difference from conventional passive response in that the G_c term exhibits a power gain which is not present in the plot of the same circuit as Figure 5-5 with a passive resistance for the diode.

The G_c term has some interesting significance and therefore its power gain contribution as a function of the $\frac{R_t}{R}$ ratio will be shown.

First

$$R_t \approx R_L + R_q \tag{5-25}$$

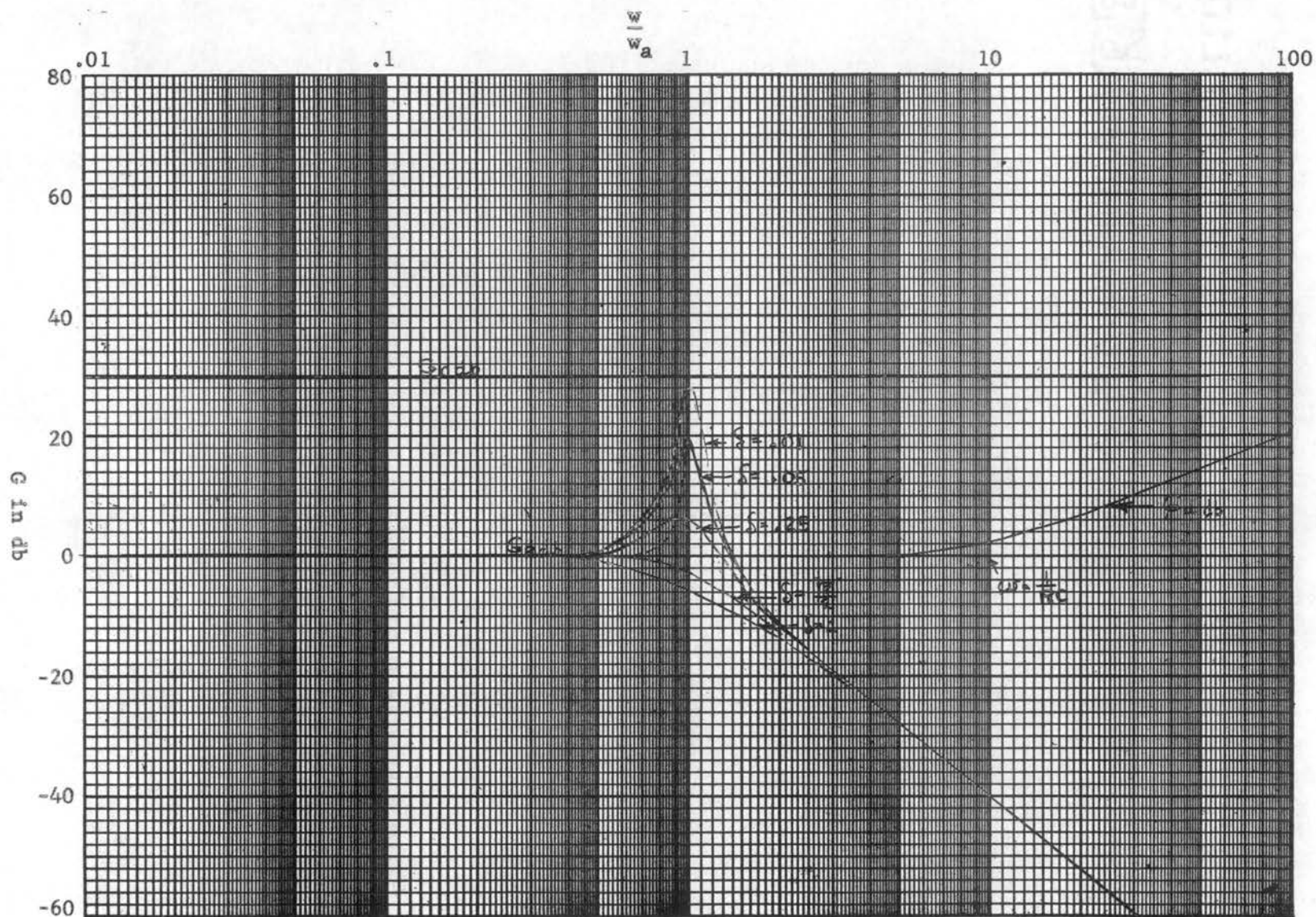since R_s is small (approximately 2 ohms).

---

[4] Ibid.

Figure 5-6. Theoretical Amplifier Gain Terms

Therefore

$$G_c = \left| \frac{R_t}{R - R_t} \right|^2 = \left| \frac{A}{1-A} \right|^2 \tag{5-26}$$

Where

$$A = \frac{R_t}{R} \tag{5-27}$$

In db

$$G_c db = 10 \log_{10} \left| \frac{A}{1-A} \right|^2 \tag{5-28}$$

Equation (5-28) is shown in Figure 5-7. Note the significant point where A = 1/2. G_c is unity for this value.

Now the gain due to the resonant term at the resonant frequency will be derived. First

$$\frac{1}{G_R} = \left| \left(\frac{j\omega}{\omega_A} \right)^2 + 2\delta \left(\frac{j\omega}{\omega_A} \right) + 1 \right|^2$$

$$= \left[\left(1 - \left(\frac{\omega}{\omega_A} \right)^2 \right)^2 + 4\delta^2 \left(\frac{\omega}{\omega_A} \right)^2 \right]$$

$$= \left[1 + (4\delta^2 - 2) \left(\frac{\omega}{\omega_A} \right)^2 + \left(\frac{\omega}{\omega_A} \right)^4 \right] \tag{5-29}$$

This function exhibits a minimum value where

$$\frac{d\left(\frac{1}{G} \right)}{d\left(\frac{\omega}{\omega_A} \right)} = (8\delta^2 - 4)\left(\frac{\omega}{\omega_A} \right) + 4\left(\frac{\omega}{\omega_A} \right)^3 = 0 \tag{5-30}$$

This occurs at

$$\frac{\omega}{\omega_A} = \frac{\omega_0}{\omega_A} = \sqrt{1 - 2\delta^2} \tag{5-31}$$

for $\delta < \frac{\sqrt{2}}{2}$, AND

$$\frac{1}{G_{RO}} = 4\delta^2 (1 - \delta^2) \tag{5-32}$$

The subscript zero refers to the resonant frequency. For $\delta > \frac{\sqrt{2}}{2}$ the function exhibits no resonant peak whatsoever.

Figure 5-8 shows the variation in G_{ro}db as a function of δ. Observe that above unity power gain the relationship is almost linear.
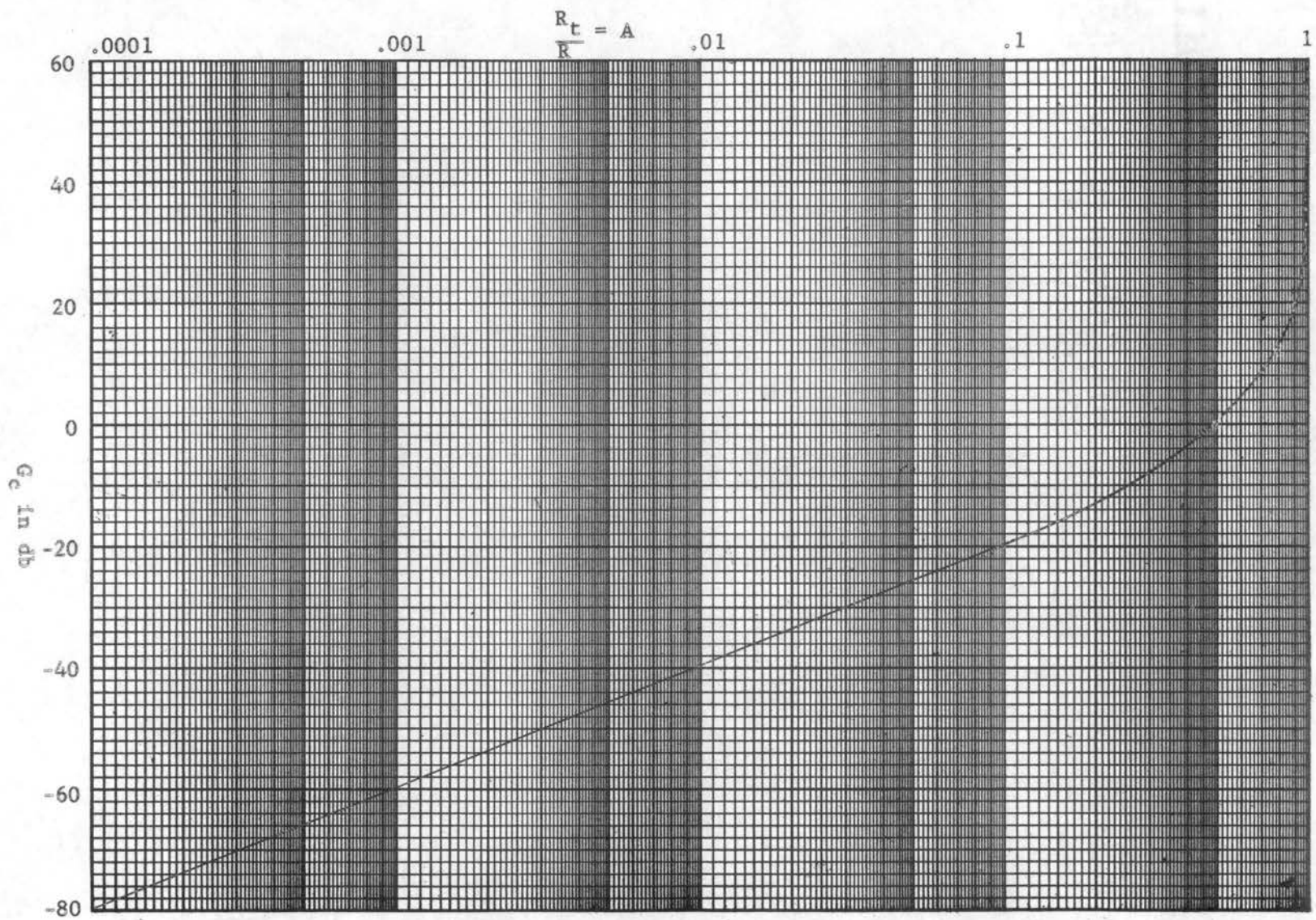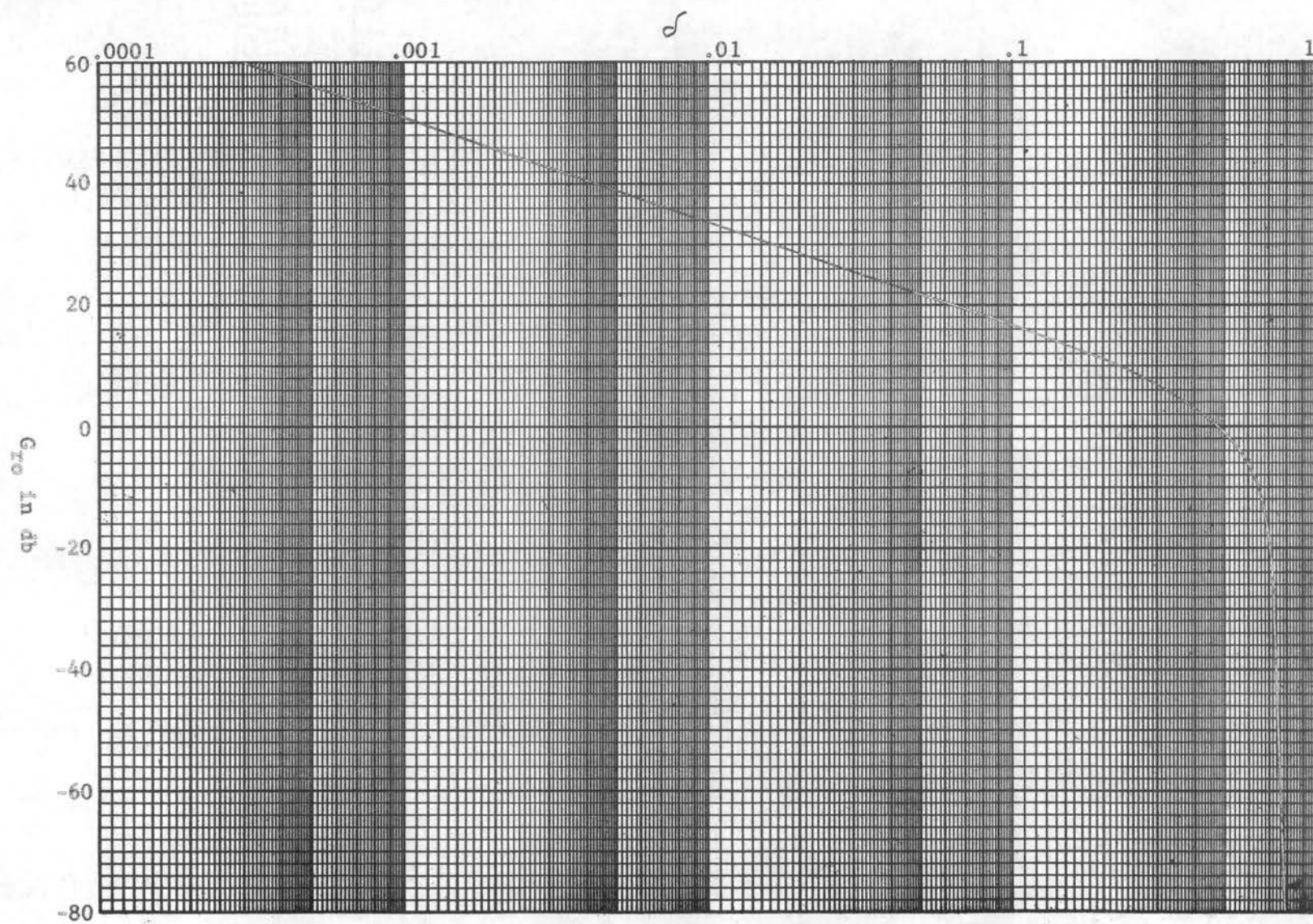
Figure 5-7. Constant Gain Term Variation With A

Figure 5-8. Resonant Gain Term Variation With δ

Before proceeding with the last power gain term G_d, it will be advantageous to investigate the gain bandwidth relationship of the circuit including the two power gain terms already discussed.

First the bandwidth (B.W.) must be found. Let ω_x be the frequency at which the current through the load resistance which is associated with G_r is $\frac{1}{\sqrt{2}}$ times its maximum value or

$$\frac{2}{G_{RO}} = 8\delta^2(1-\delta^2) \tag{5-33}$$

Substituting in Equation (5-29)

$$8\delta^2(1-\delta^2) = 1 + (4\delta^2-2)\left(\frac{\omega_x}{\omega_A}\right)^2 + \left(\frac{\omega_x}{\omega_A}\right)^4 \tag{5-34}$$

The solution of Equation (5-34) is

$$\left(\frac{\omega_x}{\omega_A}\right)^2 = (1-2\delta^2) \pm 2\delta\sqrt{1-\delta^2} \tag{5-35}$$

Now using Equation (5-31) and Equation (5-32), Equation (5-35) becomes

$$\omega_x^2 = \omega_o^2 \pm \frac{\omega_A^2}{G_{RO}} \tag{5-36}$$

In terms of δ the fractional bandwidth is

$$\left(\frac{B.W.}{\omega_o}\right) = \sqrt{1 + \frac{2\delta\sqrt{1-\delta^2}}{1-2\delta^2}} - \sqrt{1 - \frac{2\delta\sqrt{1-\delta^2}}{1-2\delta^2}} \tag{5-37}$$

But Equation (5-37) may also be expressed as

$$\left(\frac{B.W.}{\omega_o}\right) = \left[\frac{2\omega_o}{2\omega_o + \Delta\omega_1 - \Delta\omega_2}\right]\left[\frac{1}{1-2\delta^2}\right]\left[\frac{1}{G_{RO}^{\frac{1}{2}}}\right] \tag{5-38}$$

where $\Delta\omega_1 - \Delta\omega_2$ is the difference between the half power frequencies on either side of W_o.

If

$$\Delta\omega_1 \approx \Delta\omega_2 \tag{5-39}$$

then Equation (5-38) becomes

$$\left(\frac{B.W.}{\omega_o}\right) \approx \frac{1}{(1-2\delta^2)G_{RO}^{\frac{1}{2}}} \tag{5-40}$$

It is important to note from Equation (5-35) that unless $\delta < \frac{1}{2}\frac{\sqrt{2}}{2}$ the bandwidth does not exist since the gain never drops to one-half its maximum value for frequencies below resonance.

The fractional gain-bandwidth product is defined as $\sqrt{G}\left(\frac{BW}{\omega_0}\right)$ Therefore from Equation (5-38)

$$\sqrt{G_{RO}}\left(\frac{B.W.}{\omega_0}\right) = \left[\frac{2\omega_0}{2\omega_0 + \Delta\omega_1 - \Delta\omega_2}\right]\left[\frac{1}{1-2\delta^2}\right] \tag{5-41}$$

Again, if $\Delta\omega_1$ is approximately equal to $\Delta\omega_2$, then Equation (5-41) becomes

$$\sqrt{G_{RO}}\left(\frac{B.W.}{\omega_0}\right) \approx \frac{1}{1-2\delta^2} \tag{5-42}$$

Now the constant gain term G_c may be included in the gain B.W. product although its contribution is solely in terms of gain.

$$\sqrt{G_1}\left(\frac{B.W.}{\omega_0}\right) = \sqrt{G_c G_{RO}}\left(\frac{B.W.}{\omega_0}\right) = \left[\frac{A}{1-A}\right]\left[\frac{2\omega_0}{2\omega_0 + \Delta\omega_1 - \Delta\omega_2}\right]\left[\frac{1}{1-2\delta^2}\right] \tag{5-43}$$

and if Equation (5-39) is satisfied

$$\sqrt{G_1}\left(\frac{B.W.}{\omega_0}\right) \approx \left[\frac{A}{1-A}\right]\left[\frac{1}{1-2\delta^2}\right] \tag{5-44}$$

Observe that as $A \to 1$, very large gain bandwidth figures result, especially for non-selective application.

To obtain analytical relationships suitable for design purposes, it is necessary to know what quantities would determine the response curve. In most tuned response applications it is reasonable to assume that the resonant frequency (W_0) and B. W. would be specified. Under this assumption it is possible to relate all the design variables to W_0 and B.W.

From Equation (5-42)

$$\sqrt{G_{RO}}\left(\frac{B.W.}{\omega_0}\right) \approx \frac{1}{1-2\delta^2} \tag{5-42}$$

but

$$\sqrt{G_{RO}} = \frac{1}{2\delta\sqrt{1-\delta^2}} \tag{5-32}$$

therefore

$$\left(\frac{B.W.}{\omega_0}\right) = K = 2\delta\left[\frac{(1-\delta^2)^{\frac{1}{2}}}{1-2\delta^2}\right] \tag{5-45}$$

The factor in the brackets of Equation (5-45) is very nearly unity for small δ and hence

$$\delta \approx \frac{K}{2} \tag{5-46}$$

If there is doubt as to the approximation involved in Equation (5-46), Equation (5-37) has been plotted exactly in Figure 5-9 and for every K a B may be found such that

$$\delta = BK \tag{5-47}$$

where

B is 1/2 for small δ

Since δ has already been defined in terms of the circuit, there is now one design relationship involving the four circuit variables.

Also

$$\omega_A = \frac{\omega_0}{\sqrt{1-2\delta^2}} = \sqrt{\frac{1-A}{LC}} \tag{5-48}$$

Therefore

$$\frac{1}{LC} = \left(\frac{1}{1-A}\right)\left(\frac{\omega_0^2}{1-2\delta^2}\right) \tag{5-49}$$

The desired design equations are now

(1) $\qquad \frac{R_t}{R} = A \tag{5-50}$

where A is yet to be determined

(2) $\qquad \delta = BK = \frac{1}{2}\left(\frac{R_t}{L} - \frac{1}{RC}\right)\left(\frac{1}{\omega_A}\right) \tag{5-51}$

(3) $\qquad \frac{1}{LC} = \left(\frac{1}{1-A}\right)\left(\frac{\omega_0^2}{1-2\delta^2}\right) \tag{5-52}$

Solving the three previous equations simultaneously yields

$$\frac{1}{RC} = \frac{\omega_0}{\sqrt{1-2\delta^2}}\left[\sqrt{\delta^2 + \frac{A}{1-A}} - \delta\right] \tag{5-53}$$

Therefore

$$C = \frac{\sqrt{1-2\delta^2}}{R\omega_0}\left[\frac{1}{\sqrt{\delta^2 + \frac{A}{1-A}} - \delta}\right] \tag{5-54}$$
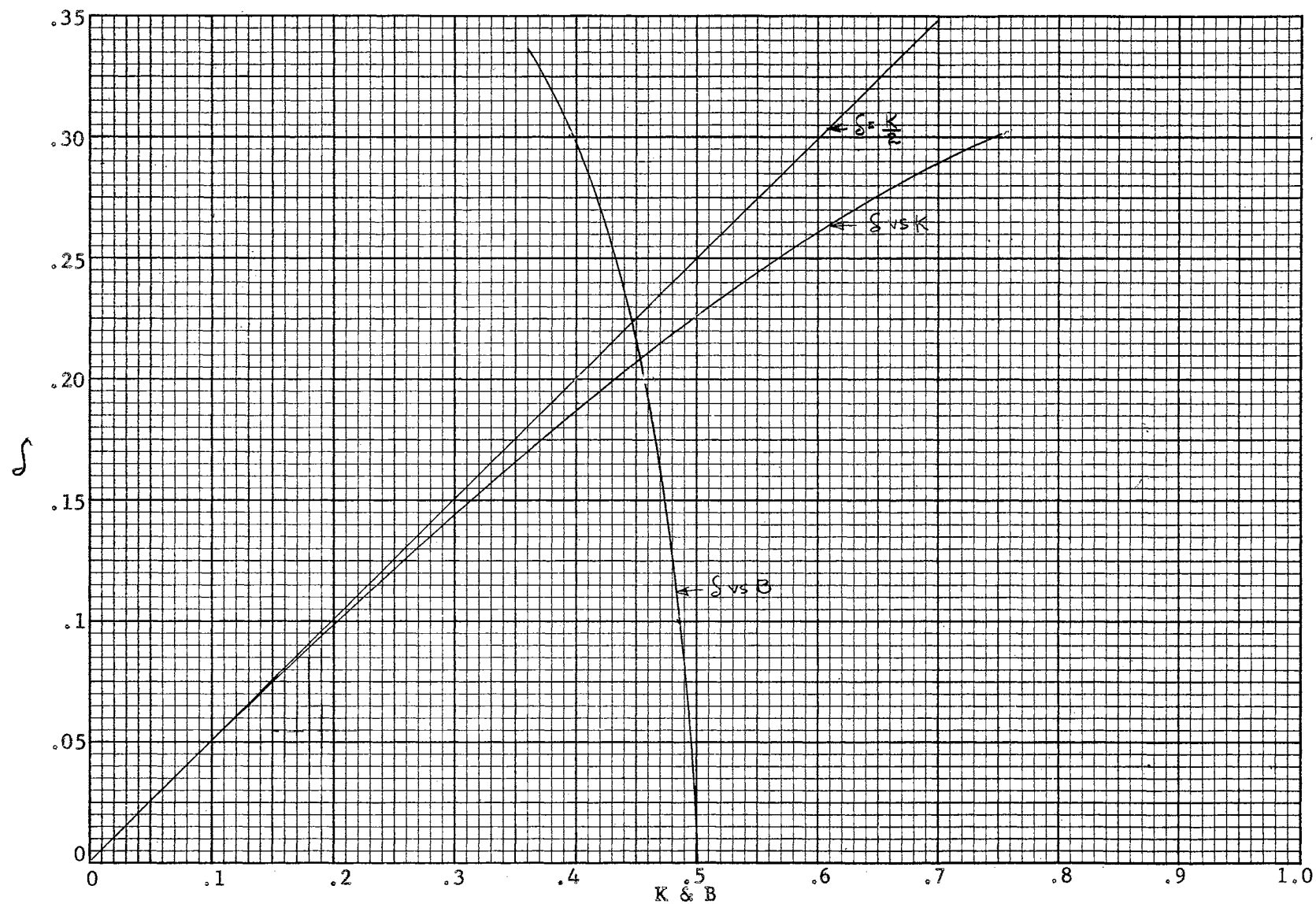
Figure 5-9. Relationships Between δ and the Fractional Bandwidth K

Note that if C is also specified, as would be the case if the diode capacitance alone were to be used in the tuned circuit, then A is fixed.

From Equation (5-54) and Equation (5-52) L is now found as

$$L = \frac{(1-A)(1-2\varsigma^2)}{C\,\omega_0^2} = \frac{R}{\omega_0}(1-A)(1-2\varsigma^2)^{\frac{1}{2}}\left(\sqrt{\varsigma^2 + \frac{A}{1-A}} - \varsigma\right) \tag{5-55}$$

From Figure 5-6 it is seen that $\frac{1}{RC}$ in Equation (5-53) is the "corner frequency" of the G_d term and thus the only way in which G_d can make a significant contribution to the gain at resonance is by letting A be very much less than unity.

Now let

$$\frac{1}{RC} = \omega_c \tag{5-56}$$

Where W_c is the corner frequency of the G_d term. From Equation (5-53)

$$\frac{\omega_c}{\omega_A} = \sqrt{\varsigma^2 + \frac{A}{1-A}} - \varsigma \tag{5-57}$$

This expression is plotted in Figure 5-10 which gives the corner frequency to be used in a graphical analysis. From observation of Figure 5-10 it may be seen that unless excessive attenuation is introduced in G_c, the principal effect of G_d is a slight widening of the bandwidth and a small gain contribution. These effects must be accounted for since W_0 and B.W. are the two given design parameters.

Squaring both sides of Equation (5-57) gives

$$\left(\frac{\omega_c}{\omega_A}\right)^2 = \frac{A}{1-A} + 2\left[\varsigma^2 - \varsigma\sqrt{\varsigma^2 + \frac{A}{1-A}}\right] \tag{5-58}$$

but

$$G_D = \left|-1 + j\omega RC\right|^2 = 1 + \left(\frac{\omega}{\omega_c}\right)^2 \tag{5-59}$$

Substituting Equation (5-58) into Equation (5-59)

$$G_D = 1 + \left(\frac{\omega}{\omega_A}\right)^2 \left[\frac{1}{\frac{A}{1-A} + 2\left(\varsigma^2 - \varsigma\sqrt{\varsigma^2 + \frac{A}{1-A}}\right)}\right] \tag{5-60}$$
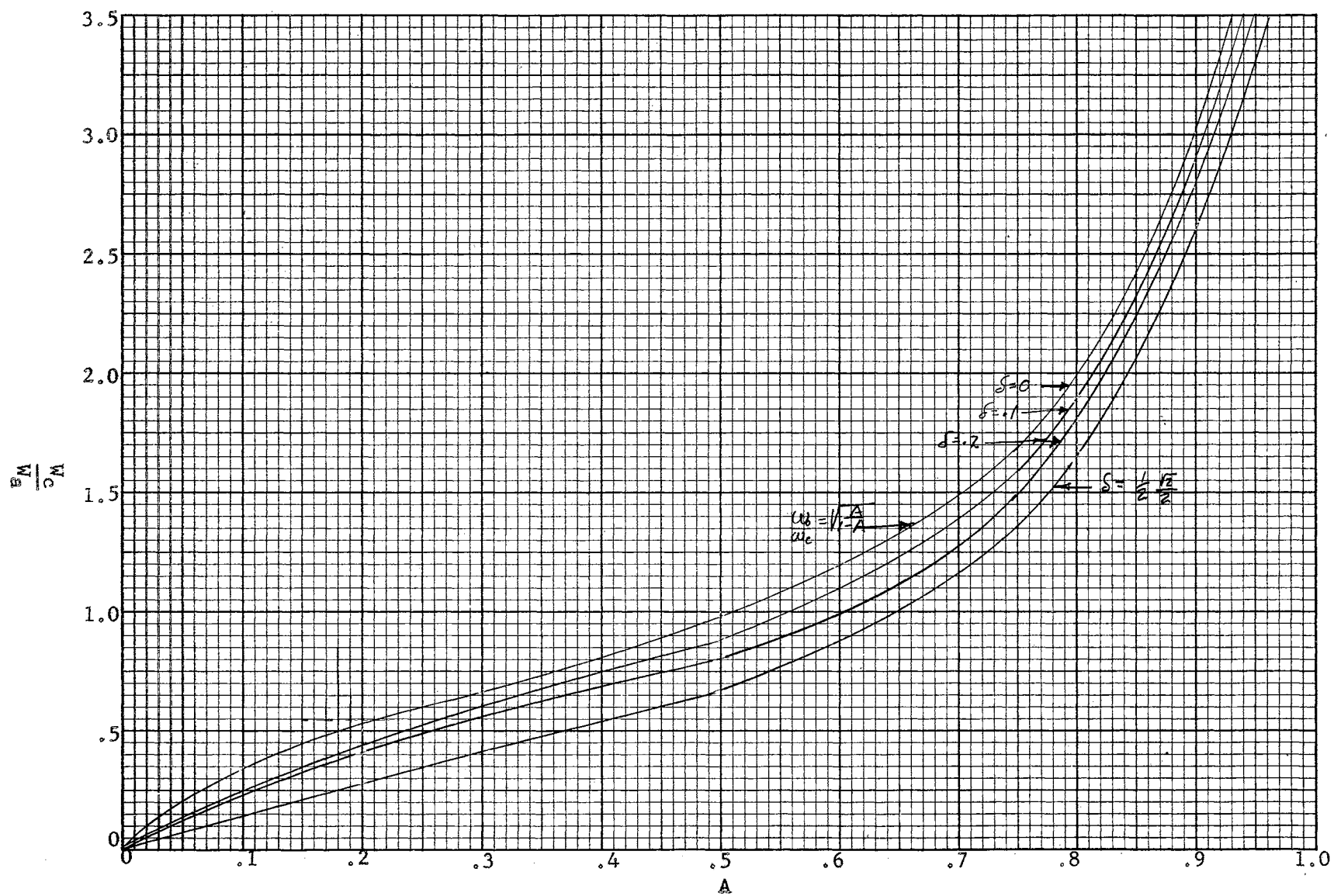
Figure 5-10. Normalized Variation in G_d Corner Frequency as a Function of A With δ a Parameter

At $W = W_o$, G_d simplifies to

$$G_{DO} = \frac{\frac{1}{1-A} - 2S\sqrt{S^2 + \frac{A}{1-A}}}{\frac{A}{1-A} + 2\left(S^2 - S\sqrt{S^2 + \frac{A}{1-A}}\right)} \tag{5-61}$$

And for small S this expression reduces to

$$G_{DO} \approx \frac{1 - 2S\sqrt{A(1-A)}}{A - 2S\sqrt{A(1-A)}} \tag{5-62}$$

which for moderately large values of $A(A > .1)$ reduces still further to

$$G_{DO} \approx \frac{1}{A} \tag{5-63}$$

If $A = 1/2$, $G_d \approx 2$ and from Figure 5-10 the corner frequency is approximately equal to W_o if S is small. Under these conditions it is seen that the essential contribution of G_d occurs after the response has fallen to unity where it alters the fall-off in response from - 12db per octave to -6db per octave. In general it might be noted that G_d goes up when G_c goes down and vice versa. Now it is possible to state that G_d causes no essential widening of the B.W. and hence the previous discussion is still appropriate for design.

For all reasonable tuned amplifier design the foregoing relationships may be summarized.

The total gain bandwidth expression becomes

$$\left(G_{co}G_{Ro}G_{do}\right)^{\frac{1}{2}}\left(\frac{B.W.}{W_o}\right) \approx \frac{A}{1-A}\left(\frac{1 - 2S\sqrt{A(1-A)}}{A - 2S\sqrt{A(1-A)}}\right)\left(\frac{1}{1-2S^2}\right) \approx \frac{1}{(1-A)(1-2S^2)} \tag{5-64}$$

From the desired G_c gain

$$\frac{R_t}{R} = A \tag{5-50}$$

S is determined by the fractional bandwidth

$$S = BK \tag{5-47}$$

C is now determined

$$C = \frac{1}{R\omega_A\left(\sqrt{S^2 + \frac{A}{1-A}} - S\right)} \tag{5-54}$$

and now L is completely specified

$$L = \frac{(1-A)}{\omega_A^2 C} \qquad (5\text{-}55)$$

Two further design considerations deserve discussion at this point. First the value of A must not be so large that the d-c biasing is near instability since the amplifier must be stable in the d-c circuit as well as the a-c circuit. This means that unless moderate values of A are used the biasing problem will become difficult.

Also it may be that for the W_o selected, C_d may not lie in the allowed range dictated by Equation (5-54) when the d-c biasing considerations discussed above are adhered to. One method of remedying this difficulty is by paralleling the diode with a capacitance to bring Equation (5-54) within the limits set by d-c stability.

While this is the simplest method of utilizing the diode as an amplifier at lower frequencies (less than 50 mc) it also introduces a stability problem. There is now a minor loop in the amplifier configuration which is potentially unstable and suitable measures must be taken to avoid oscillation. The solution is to insert a series resistance with the parallel added capacitance. The open loop minor circuit is now as shown in Figure 5-11 where R_p is the added resistance and terminals xx' represent the insertation point of the parallel capacitance.

This is obviously the same essential circuit configuration as that observed at the terminals of the amplifier proper and hence the same stability conditions must be met, that is

$$\frac{L_s}{RC} < R_t' < R \qquad (5\text{-}55)$$

Where

$R_t' = R_p + R_s$
L_s is the series lead inductance around the loop including the capacitor leads
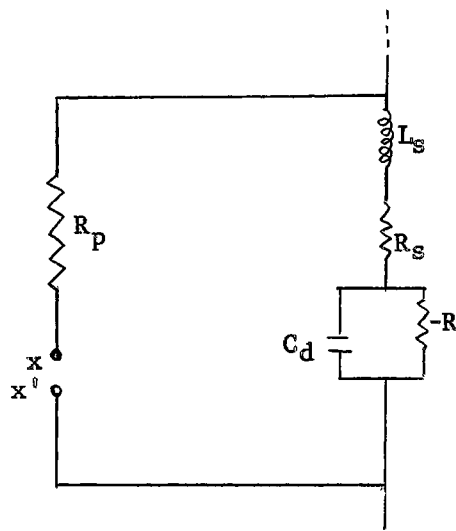
Figure 5-11. Minor Loop Equivalent Circuit

The external circuit at the open loop terminals with the parallel

branch is shown in Figure 5-12.
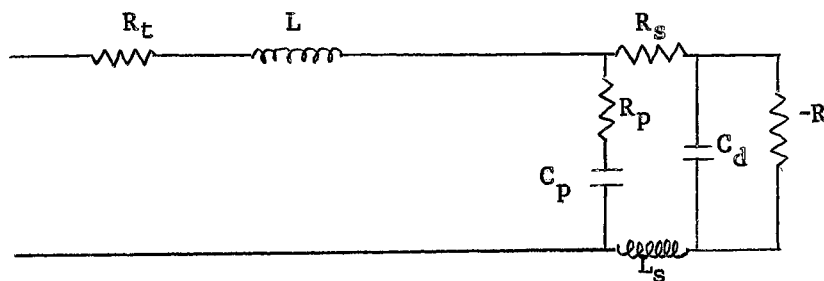
Figure 5-12. Complete Low Frequency Amplifier Circuit

Since W_o is to be low, the lead inductance of the minor loop is in-

significant in the total analysis and is hence omitted. Also the bulk

resistance R_s (usually less than 2Ω) may be neglected to as good an

approximation as R remains linear which gives the final equivalent

circuit shown in Figure 5-13.
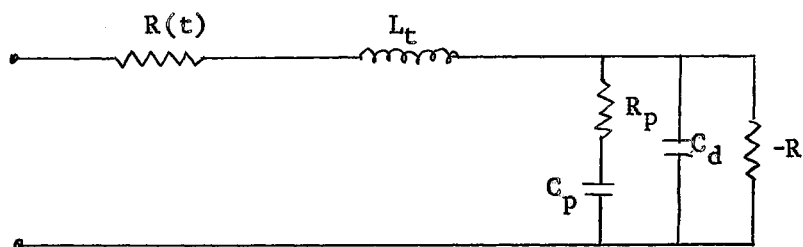
Figure 5-13. Approximate Low Frequency Amplifier Circuit

Since the amplifier will be driven from a sinusoidal source, the series R_p and C_p branch may be converted to equivalent parallel branches as shown in Figure 5-14.[5]
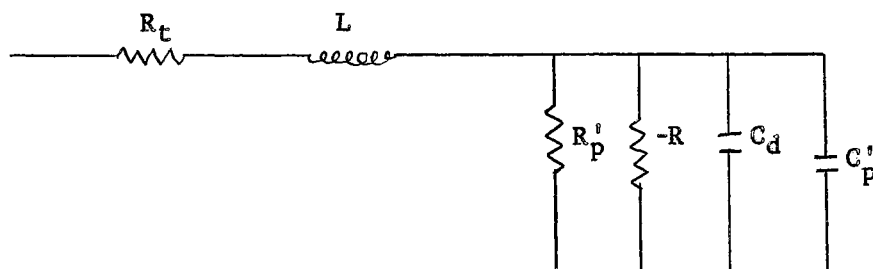
Figure 5-14. Equivalent to Approximate Low Frequency Amplifier Circuit

Where

(a) $$R_p' = R_p \left(1 + \frac{1}{(\omega R_p C_p)^2} \right) = R_p \left(1 + Q_p^2 \right)$$

(b) $$C_p' = C_p \left(\frac{Q_p^2}{1 + Q_p^2} \right) \tag{5-66}$$

These parallel branches may be represented equivalently as shown in Figure 5-15.

———————
[5] H. T. Fristoe, The Use of Q Equations to Solve Complex Electrical Networks, Oklahoma State University Engineering Extension Service, No. 105, 1959.
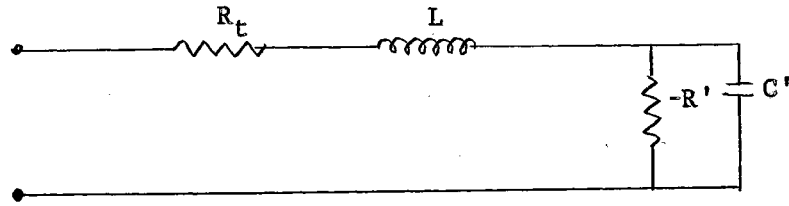
Figure 5-15. Final Equivalent Low Frequency Amplifier

where

(a) $\quad R' = \dfrac{-RR'_{\rho}}{R'_{p} - R}$

(b) $\quad C' = C_p + C'_p$ $\hspace{4cm}$ (5-67)

Since this configuration is exactly similar to that previously

considered, the same stability criteria again apply, that is

$$\frac{L}{R'C'} < R_t < R' \hspace{4cm} (5\text{-}58)$$

One further relationship must be found because the circuit illustrated

in Figure 5-5 is only valid at one frequency if the parallel R_p-C_p branch

is inserted. It is convenient to choose that frequency as W_o and under

this condition the simultaneous solution of Equations (5-66) and (5-54)

yields

$$C_p = \frac{2}{R\omega_o\left[\sqrt{2\zeta^2 + \frac{A}{1-A} - 2\zeta}\sqrt{\zeta^2 + \frac{A}{1-A}} + 4\frac{R_p}{R}\left(1 - \frac{R_p}{R}\right) - 1\right]} \hspace{2cm} (5\text{-}68)$$

For ζ small Equation (5-68) reduces to

$$C_p \approx \frac{2}{R\omega_o\left[\sqrt{\frac{A}{1-A} + 4\frac{R_p}{R}\left(1 - \frac{R_p}{R}\right)} - 1\right]} \hspace{2cm} (5\text{-}69)$$

where R_p is determined from minor loop stability conditions.

The effect upon the amplifier response of these equivalent frequency

varying components is shown in the next chapter.

CHAPTER VI

AMPLIFIER DESIGN AND EXPERIMENTAL RESULTS

Amplifier Design

The design equations derived in Chapter V were used to construct a
tuned series amplifier.

A resonant frequency of 3.0 mc was used since the resultant simpli-
fication in circuit layout avoided the difficulties involved in high
frequency circuitry. These difficulties are a problem in themselves and
are of little consequence in checking the equations which were developed.

The selectivity factor K was chosen to be 1/10 so that fairly sharp
tuning could be obtained and still obtain enough data to yield a
continuous curve with the available equipment.

The diode selected had the following parameters

$$R = 118\,\Omega \tag{6-1}$$

$$C_b \approx 8\,\mu\mu f \tag{6-2}$$

$$R_s \approx 2\,\Omega \tag{6-3}$$

Here it is appropriate to observe that C_d is a function of the junc-
tion voltage in accordance with conventional semiconductor theory. In
this design the voltage variation in C_d was negligible since a large
external capacitance was added. If no capacitance were added, it would
be necessary to obtain the variation in C_d in order to proceed with the
design. The signal generator used had an output impedance of $25\,\Omega$.

$$R_g = 25\,\Omega \tag{6-4}$$

The design proceeded along the following lines. First from Figure
5-9

65

$$\int = BK = \frac{.495}{10} = .0495 \tag{6-5}$$

The W_o chosen was much to low to use only the diode capacitance so an external R_p-C_p branch was added. R_p was determined from the minor loop stability equations.

$$\frac{Ls}{C_D R} < R_p < R \tag{6-6}$$

A first-order estimate of the minor loop inductance was made to be 50 muh. Since C_d and R were given, Equation (6-6) yields

$$46\Omega < R_p < 118\Omega \tag{6-7}$$

A value of

$$R_p = 100\Omega \tag{6-8}$$

was decided upon to allow as much leeway as possible for minor loop inductance although this large value of R_p had a significant effect upon the actual amplifier response curve.

Now the design could have proceeded in either of two ways since an external R_p-C_p branch had been decided upon. Equation (5-59) could have been used to find either A or C_p. If A is specified, then the constant gain term is specified. If C_p is specified, then the corner frequency of the G_d term is specified. For convenience in components, C_p was specified as 400 uuf

$$C_p = 400\mu\mu f \tag{6-9}$$

Now from Equation (5-68)

$$C_p = 400\mu\mu f = \frac{2}{R\omega_o\left[\sqrt{2\int^2 + \frac{A}{1-A} - 2\int\sqrt{\int^2 + \frac{A}{1-A} + 4\frac{R_p}{R}(1-\frac{R_p}{R})}} - 1\right]} \tag{5-68}$$

From Equation (5-68) A was found to

$$A = .36 \tag{6-10}$$

Note that A is this value only at resonance since

$$A' = \frac{R_t}{R'} \tag{5-50}$$

and R' is now a function of frequency.

Now

$$Q_P = \frac{1}{\omega_o R_P C_P} = 1.382 \tag{6-11}$$

Therefore the equivalent parallel resistance R_p' is

$$R_p' = R_P(1 + Q_P^2) = 292\,\Omega \tag{6-12}$$

and

$$C_P' = C_P\left(\frac{Q_P^2}{1+Q_P^2}\right) = 263\,\mu\mu f \tag{6-13}$$

From Equation (5-67)

$$R' = \frac{-R R_p'}{R_p' - R} = -198\,\Omega \tag{6-14}$$

Now R_t is specified since

$$R_t = A'R' = 71 \tag{6-15}$$

Since

$$R_t = R_g + R_L \tag{5-25}$$

and R_g is $25\,\Omega$, then R_L is

$$R_L = 71 - 25\,\Omega = 46\,\Omega \tag{6-16}$$

The last quantity to be obtained was the added series inductance which from Equation (5-55) was

$$L = \frac{(1 - A')}{\omega_A^2 C'} \tag{5-55}$$

When all the quantities on the right side of Equation (5-55) had been substituted in, L becomes

$$L = 6.7\,\mu h \tag{6-17}$$

The d-c biasing arrangement indicated in Figure 6-1 was used for this amplifier. The operating point on the V-I characteristics of the diode was 160 mv which is the inflection point of the characteristics. The R-f coil was to decouple the d-c supply from the signal.

The amplifier circuit shown in Figure 6-1 was constructed using the component values found from the preceeding design. These values were set as close as possible to the theoretical values by using a bridge.
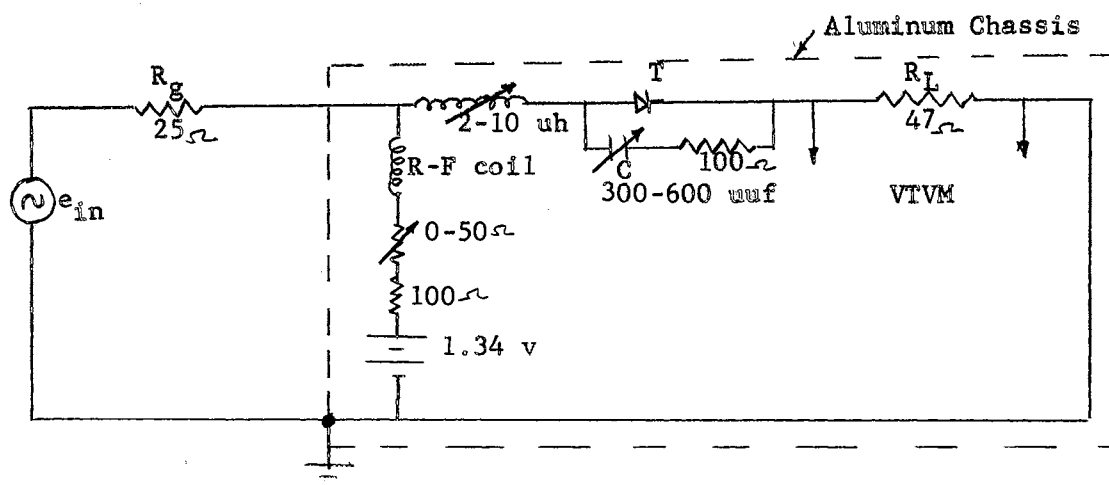
Figure 6-1. Experimental Amplifier

During the amplifier alignment, very slight changes in L and C were necessary to resonate the amplifier at 3 mc. These changes were less than 2 percent of the values calculated from the design equations. Below is a complete list of the components and instruments used.

LIST OF COMPONENTS AND INSTRUMENTS

Components

G. E. tunnel diode - #1

Ferrite slug tuned inductor -
  L = 1.68 → 12.4 uh
  Q(7.9 mc) = 165 → 22

Resistor - 100 Ω 1/8 W., 1% (2)

Resistor - 47 Ω 1/2 W., 1%

R.F. choke coil - 16.8 mh

Instruments

Hewlett Packard VTVM - Model 400D
  #30964

Booton Radio Corp. Q Meter - Type
  160-A, #3503

General Radio Impedance Bridge
  (Portable) - Type No. 1650-A,
  #1301

Fluke Differential Voltmeter -
  Model 801, #2354

LIST OF COMPONENTS AND INSTRUMENTS (Continued)

<u>Components</u>   <u>Instruments</u>

Wire wound pot - 0-50 Ω   Kepco Power Supply - 0-600 v
                                         Model 815-B, #B-7410

Eveready mercury cell - 1.34 volts

                                     General Radio Standard Sig.
                                         Generator - Type 1001-A, #434

                                     General Radio - Type 874-R20 3 foot
                                         coaxial 50 Ω patch cord

                                     General Radio - Type 874-Q2 adapter

                                     General Radio - Type 1000-P2 40 Ω
                                         Series unit

                                     General Radio - Type 1000-P1 50 Ω
                                         Termination unit

                                     Tektronix Oscilloscope - Type 545
                                         #9546

                                     Tektronix Coaxial line and
                                         Attenuation Proble - X10,
                                         10 meg. and 8 uuf input

                                     Tektronix Plug in unit - Type 53/54L
                                         #666

Experimental Methods

The circuit was mounted in an aluminum chassis with the exception of the diode and external added capacitance which were mounted above the chassis in a lucite base. The diode was mounted into copper strips to reduce lead inductance. The added capacitance and resistance were mounted close to the diode with leads as short as possible, again to reduce lead inductance. The mercury cell was mounted internally.

The signal generator was taken as a voltage input standard and coupled directly to the chassis through its coaxial line. The chassis was plugged directly into the VTVM which was calibrated from the signal generator.

The operating point was set by using a low level d-c voltmeter and varying the 50 Ω pot.

Parasitic oscillation in the amplifier circuit was checked for after each data point of Figure 6-2 by turning the generator output to zero and checking the output voltage, which was read from a high sensitivity VTVM, to insure that it was zero.

Discussion of Experimental Results

The theoretical output of the amplifier was easily found by a graphical addition of the three terms of Figure 5-6 using the method of Bode plots.[1]

The gain magnitudes of the theoretical output were taken directly from Figures 5-7 and 5-8 using the parameter values determined from the design equations and the G_d corner frequency was found directly from Figure 5-10. These were placed in Figure 6-2 along with the actual response curve of the amplifier.

There are three distinct regions of analysis on the response curve. The first is from .01 mc to 1.15 mc where the response is flat. The governing gain term in this region is G_c which is a function of the $\frac{R_t}{R'}$ ratio. The value at resonance of G_c is -5db and assuming R' is constant gives the predicted curve as shown. However, R' is a function of R_p'

$$R' = \frac{-R R_p'}{R_p' - R} \tag{6-14}$$

_____

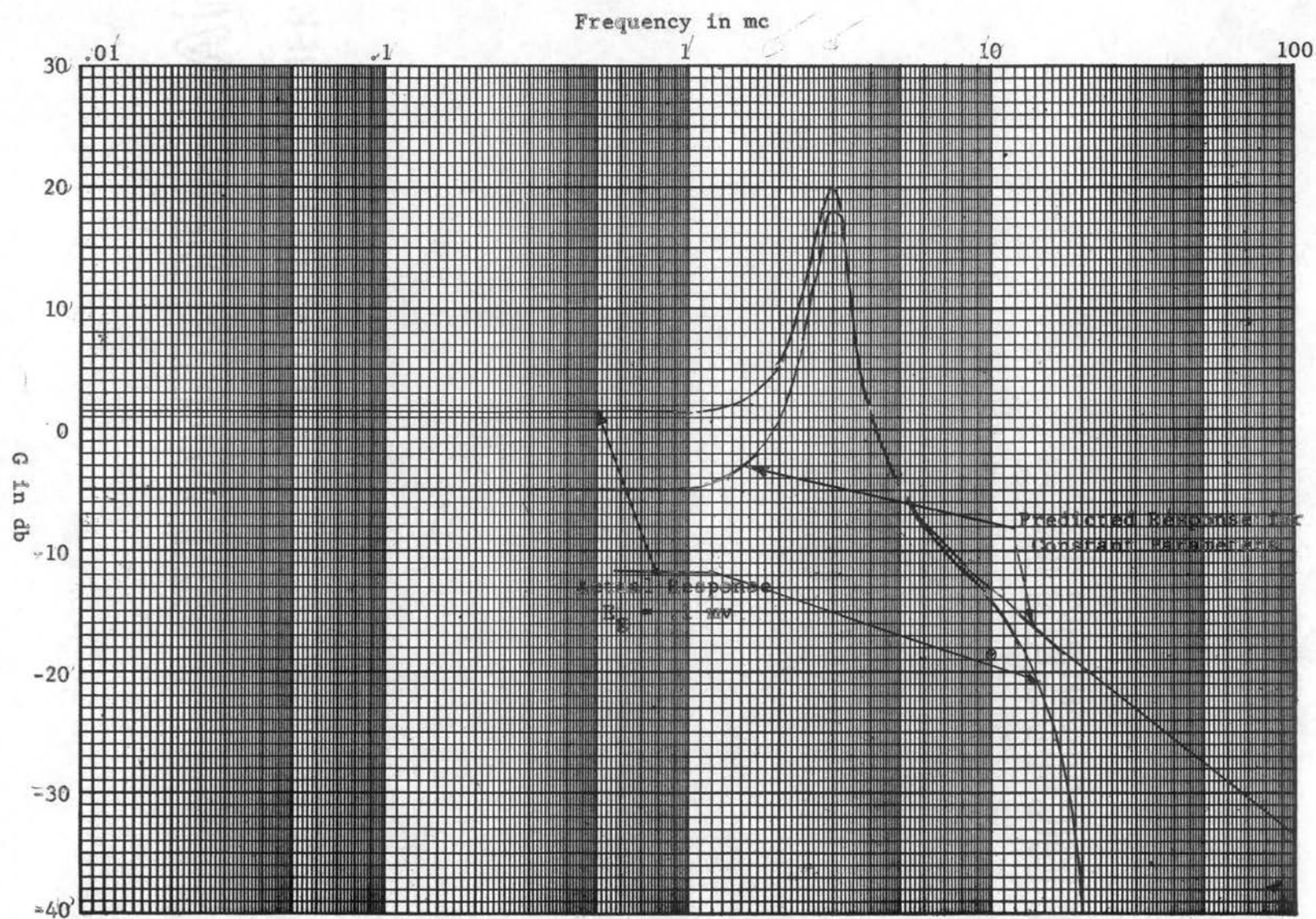[1] Brown and Schultheiss, _op. cit._

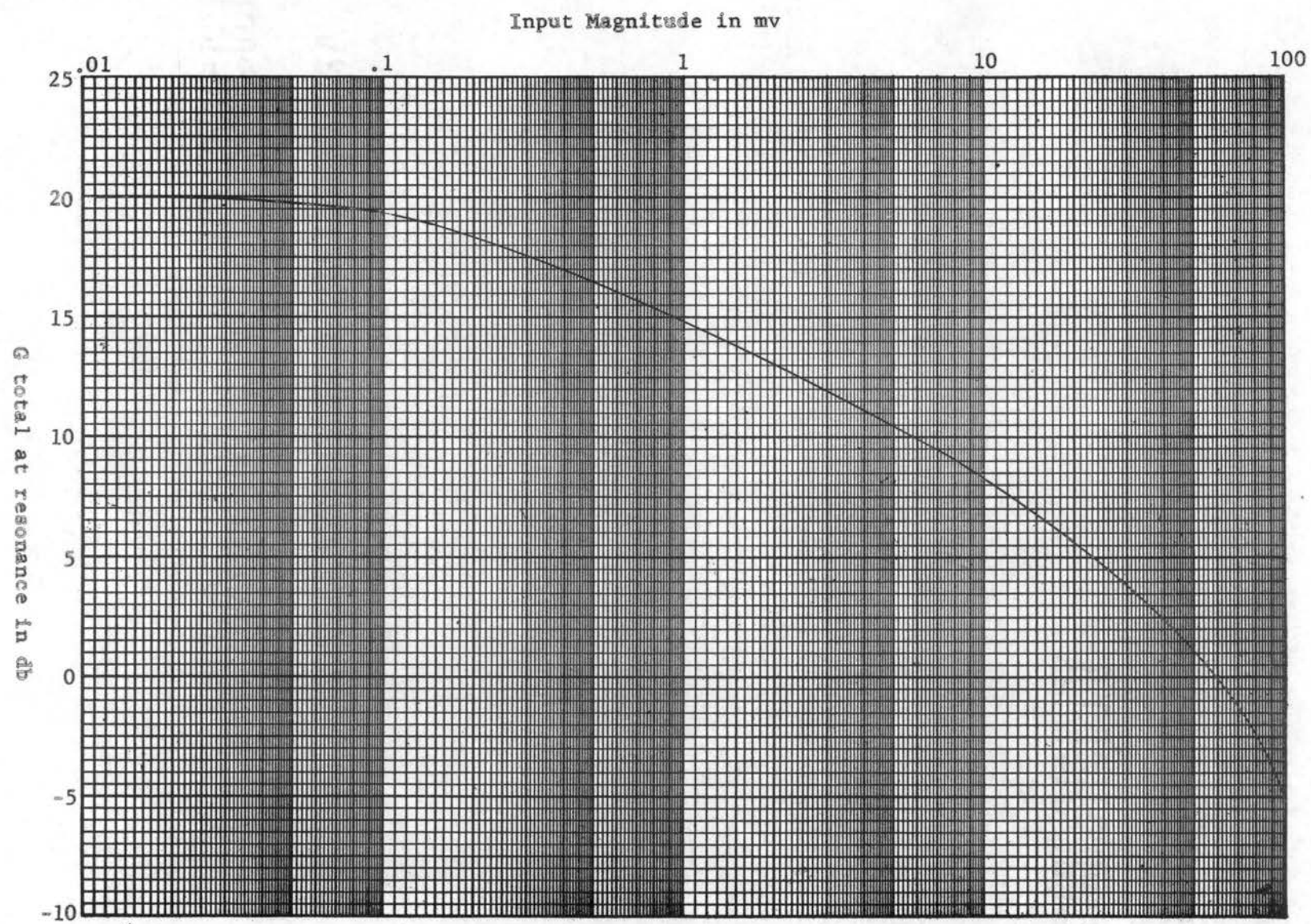Figure 6-2. Predicted and Experimental Results of 3 mc Tuned Amplifier

Figure 6-3.   Variation in Gain With Driving Signal Magnitude for 3 mc Amplifier

and $R_p^!$ is an inverse function of frequency which tends to infinity as the frequency tends to zero. This means $R^!$ tends to drop in magnitude from its resonance value thus making G_c approach the value

$$G_c \longrightarrow \left| \frac{\frac{R_t}{R}}{1 - \frac{R_t}{R}} \right|^2 = 2.03 \tag{6-18}$$

and

$$G_c \, db \longrightarrow 3.4db \tag{6-19}$$

The actual magnitude of G_c is approximately 2 db and the apparent inconsistency in the first region is resolved.

The second region is in the near vicinity of the resonant peak. The predicted and actual responses are in good correlation in this area except for an actual wider bandwidth and smaller resonant peak than predicted. The reason for this is again traceable to the frequency vary-ing components which were assumed constant in constructing the predicted response. The frequency varying equivalent $C_p^!$ starts contributing gain in the G_d term sooner than theory would predict since

$$G_d = 1 + (\omega R' C')^2 \tag{6-20}$$

$C^!$ is principally $C_p^!$ which is a direct function of frequency and thus the corner frequency at which G_d contributes 3db of gain is essentially lower than predicted.

The third region to consider is that for frequencies appreciably higher than resonance. The actual response is approaching -18db per octave fall-off while the predicted curve is asymtotic to a fall-off of -6db per octave.

The reason for this is that the equivalent $R_p^!$ has become lower than the negative resistance of the diode and consequently from Equation (6-11) the entire circuit has become passive.

Another curve of some interest is shown in Figure 6-4 which illustrates the non-linear relation between the power gain and the magnitude of the driving voltage. The data was taken at the resonant frequency and clearly shows that if the amplifier is overdriven, the gain falls rapidly. The reason for this is that the negative resistance of the diodes is not constant for large signal application. This may be seen from Figure 4-4 by noting that the average or effective negative resistance depends upon the driving magnitude. The end points of the negative resistance at I_p and I_v are soon reached and eventually the effective resistance may even be positive. Figure 6-3 suggests some application in automatic gain control since the gain is a function of input amplitude.

Some final points should be observed. While the device is capable of large power gain, nevertheless, the output power is limited. Power output at W_0 in Figure 6-3 was approximately 22 mm watts while the maximum output power before the power gain dropped below unity in Figure 6-3 was 0.158 m watts. This means that the device should only be used in low level applications such as the first amplification stage in radar or television.

Also the device is essentially at its best at higher frequencies than were used in this thesis since no parallel R_p-C_p branch must be added.

If this is the case, then the design equations developed should follow almost exactly the theoretical predicted response obtained from the graphical method. In any case the equations should be accurate enough for most engineering applications.

CHAPTER VII

SUMMARY AND CONCLUSIONS

The phenomenon of electron tunneling was qualitatively shown through quantum mechanical methods. This effect was then utilized, after some preparatory semiconductor theory, in a qualitative and quantitative description of the Esaki tunnel diode. This analysis explained the origin of the negative resistance region in the narrow p-n junction.

By assuming the negative resistance region was linear, a small-signal analysis of a series tuned amplifier circuit was outlined, based upon stability criteria and a graphical analysis.

Design equations were developed which gave all circuit parameters, in terms of a selectivity factor, 1/K where K was the ratio of the desired bandwidth to the desired resonant frequency.

An amplifier was then constructed based upon the derived design considerations and its response was compared to its theoretical predicted response.

The developed design relationships appeared valid based upon the experimental results. The deviations observed seem to be due to the additional circuitry required for low frequency operation. Even with these deviations the design equations give very good correlation between the predicted and actual response especially in the significant region of the resonant peak.

However, the non-linear relation between gain and input magnitude would seem to limit the application of the amplifier to low power level applications, especially if large amounts of gain are to be obtained.

The amplifier circuit seems best suited for much higher frequency operation than was used since the circuit is simpler and no approximation is involved in the design equations.

A study of the non-selective amplifier would be useful since this amplification mode has a number of useful applicatons. For example, it might be considered in such uses as very low voltage operational amplifier or as a video amplifier.

The non-linear relationship between input magnitude and resultant gain suggests some applications in automatic gain control.

A number of devices in addition to the tunnel diode such as multivibrators, gas tubes (thyratrons, neon bulbs, etc.), PNPN transistors, superregenerative amplifiers, tetrodes, Dynatron and Transistron oscilla-tors, and parametric amplifiers exhibit a negative resistance between some two terminals. It is the author's suspicion that feedback in general (especially positive feedback) may be analyzed as a negative resistance effect. Thus, it might prove advantageous to make a study of the general properties of negative resistance so that circuits incorporating these elements might be better understood from a stability viewpoint.

SELECTED BIBLIOGRAPHY

Brown, J. L. and Peter M. Schultheiss, *Introduction to the Design of Servomechanisms*, John Wiley & Sons, Inc., New York, 1958.

Chang, K. K. N., "Low-Noise Tunnel Diode Amplifier", *Proceedings, I.R.E.*, Vol. 47, p. 1268, 1959.

Davidsohn, U. S., Y. C. Hwang and G. B. Ober, "Characterization of Tunnel Diodes and Circuit Stability Considerations", *Electronic Design*, March 17, 1960.

Esaki, L., "New Phenomenon in Narrow Ge P-N Junctions", *Physical Review*, Vol. 109, p. 603, 1958.

Fristoe, H. T., *The Use of Q Equations to Solve Complex Electrical Networks*, Oklahoma State University Engineering Extension Service, 1959.

General Electric Research Laboratory, "Tunnel Diodes", *Technical Information Sheet*, New York, 1959.

Gurney, R. W., *Elementary Quantum Mechanics*, Cambridge, 1934.

Hoffman, Banesh, *The Strange Story of the Quantum*, Dover Publication, New York, 1959.

Lesk, I. A., N. Holonyak, Jr., U. S. Davissohn and M. W. Aarons, "Germanium and Silicon Tunnel Diodes--Design, Operation and Application", *1959 Wescon Convention Record*.

Shockley, W., *Electrons and Holes in Semiconductors*, D. Van Norstad Publishing Company, New York, 1956.

Sklar, Bernard, "The Tunnel Diode--Its Action and Properties", *Electronics*, November, 1959.

Sommers, H. S., Jr., "Tunnel Diodes as High Frequency Devices", *Proceedings, I.R.E.*, Vol. 47, p. 1201, 1959.

Wolfendale, E., (Editor), *The Junction Transistor and Its Applications*, New York, 1958.

Ziel, Aldert Von Der, *Solid State Physical Electronics*, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1957.

VITA

JOHN C. PAUL

Candidate for the Degree of
Master of Science

Thesis: A STUDY OF THE ESAKI TUNNEL DIODE AND DEVELOPMENT OF SOME
ASSOCIATED AMPLIFIER DESIGN RELATIONSHIPS

Major Field: Electrical Engineering

Biographical:

Personal Data: Born in Fairview, Oklahoma, March 25, 1935 the son
of John C. and Mable A. Paul.

Education: Attended city schools in Fairview, Oklahoma. Graduated
from Fairview High School, Fairview, Oklahoma in May, 1952.
Received the Degree of Bachelor of Science in Electrical
Engineering from Oklahoma State University in May, 1959.
Completed the requirements for the Master of Science Degree
in May, 1960.

Experience: Employed one year as worker on oil well drilling rig,
1953-54, Warren-Bradshaw Drilling Company, Tulsa, Oklahoma.
Served two years in U. S. Army from August, 1954 to August,
1956. Duty included one year as electronic maintenance
technician on Nike surface-to-air guided missle launcher
section, Fort Bliss, Texas. Employed one semester (1959) by
Oklahoma State University as laboratory instructor in
Industrial Electronics. Employed one summer (1959) as Engineer
with Halliburton Oil Well Cementing Company, Duncan, Oklahoma.

Professional Organizations: Student member of the American Insti-
tute of Electrical Engineers and Institute of Radio Engineers,
Engineer-in-Training No. 729, Mid-North Chapter, Oklahoma
Society of Professional Engineers.