DOCUMENT RESUME

ED 179 394 SE 029 298

AUTHOR	Acker, Frank E.
TITLE	Power Processing, Part 2. Modeling Power Processing Devices and Circuits.
INSTITUTION	Pittsburgh Univ., Pa.
SPONS AGENCY	National Science Foundation, Washington, D.C.
PUB DATE	70
GRANT	NSF-GY-4138
NOTE	212p.; For related documents, see SE C29 295-297; Best copy available
EDRS PRICE	MF01/PC09 Plus Postage.
DESCRIPTORS	*College Science; Curriculum Development; Electricity; Electromechanical Technology; Electronics; *Engineering Education; Higher Education; Instructional Materials; *Science Courses; Science Curriculum; *Science Education; *Science Materials; Scientific Concepts

ABSTRACT

This publication was developed as a portion of a two-semester sequence commencing at either the sixth or the seventh term of the undergraduate program in electrical engineering at the University of Pittsburgh. The materials of the two courses, produced by a National Science Foundation grant, are concerned with power conversion systems comprising power electronic devices, electromechanical energy converters, and associated logic configurations necessary to cause the system to behave in a prescribed fashion. The emphasis in this portion (Part 2) of the two course sequence is on modeling power processing devices and circuits. This publication consists of five chapters which deal with: (1) power diodes; (2) thyristors, or silicon controlled rectifiers; (3) modeling a live-voltage commutated inverter; (4) thermal characteristics of materials; and (5) a free-wheeling diode DC motor drive. An appendix which provides a suggested guide for a time schedule of presentation of material is also included. (HM)

POWER PROCESSING

Part 2

"Modeling Power Processing Devices and Circuits"

Frank E. Acker
University of Pittsburgh

1970

The preparation of this material has been partially supported by Grant GY-4138 entitled "Course Content Improvement Project for Power Processing" made to the University of Pittsburgh by the National Science Foundation.

Preface

This book is intended for use as a text in the senior elective course "Power Processing II", Electrical Engineering Department, University of Pittsburgh. The material presented here has been successfully used as the course content for two trimesters. Although several available books were considered and tried as texts, none were found to be suitable in the light of the objectives of the course, and therefore this book has been written to fulfill the needs of the course.

There are three objectives in the course "Power Processing II" which are: interesting undergraduate students in the power area of electrical engineering, providing the students with factual information and some experience relating to semiconductor power electronics, and to develop the students' ability to model physical problems. Student interest is fostered by the students' growing competence in the semiconductor power area, frequent classroom reference to the current engineering relevance of the types of problems being considered, and by making the problems and laboratory sessions as realistic as possible. Also, the area of semiconductor applications to power processing, the subject material of the course, is an area of great interest and expansion in the present day power industry and is therefore "relevant". Skill in modeling is encouraged in two ways. First, the problems the students are required to work are framed in terms of real circuit elements. The students must decide what idealizations can be made. Secondly, as the course progresses, the sophistication of the modeling required to solve the problems in a reasonable time increases, and hopefully the students' skill will increase as they work the gradually more sophisticated problems.

The author considers modeling the most important aspect of the course as reflected by the subject material and organization of this book. Two of the most valuable attributes of an engineer are his ability to assimilate new technology, and his ability to apply basic science and technology (new or old) to new problems. Without these attributes, the engineer is soon relegated to the position of a competent technician. One of the most, if not the most, powerful tools used to maintain these attributes is the engineer's skill in modeling physical problems; that is, to simplify the problem to the extreme so that the basic parameters and operations become obvious, and then to replace the necessary complexities until the model sufficiently approaches the real physical problem to give valid engineering answers. The author therefore feels that the gain in examining and modeling the problems in some detail far outweigh the disadvantage that less material (fewer circuits, problems, and applications) can be considered in the given time.

There are several reasons for stressing modeling in the particular course on semiconductor power processing. The primary objective in offering the course "Power Processing II" is to interest students in the power area of engineering. By incorporating the learning of a fundamental engineering skill (modeling) into the course, it may be possible to attract more of the "uncertain" students who may not want to commit themselves to a specific area of electrical engineering. The subject material may then provide sufficient challenge to interest these students in power engineering. Also, starting from the students undergraduate background in electronics, logic, and physics, the students actually experience the extension of their knowledge into an unfamiliar technological area (semiconductor power processing) using the tool of modeling as well as using modeling to solve complicated problems. And of course, even the simplest problems in power processing can only be solved by the straightforward application of Kirchoff's laws with utmost difficulty, further impressing upon the student the value of modelling as a problem solving tool.

The laboratory requires some special mention. The laboratory problems do not designate specific experiments to be performed by the students, nor is a "typical" formal laboratory report required of each student. While real-life situations sometimes require a "laboratory report," as in the testing and evaluation of an item or system, the most frequent use of an industrial laboratory is as an aid to finding the answer to a problem. The realistic laboratory problem associated with any problem is "What laboratory experiment should be done?" The student is given the choice of using the laboratory to gather data, confirm his theory, check assumptions, as an aid to understanding device or circuit operation, or any combination of these. The students are not permitted to enter the laboratory without a "plan" in which each student must identify a specific objective for the laboratory experiment, and a detailed plan to carry out the experiment. The students are graded on the basis of how effective their laboratory objective will be in enabling them to solve the problem, and whether their detailed plan has a reasonable assurance of enabling the students to accomplish their immediate objective. After the laboratory session, the students complete their assigned problem, presenting "an answer" which is backed up by laboratory experiment and data. This type of laboratory has proved much more interesting to the students and seems more in keeping with an engineering education than simply "verifying calculations" or "demonstrating effects."

The author gratefully acknowledges the continuing financial support of the National Science Foundation throughout the design and establishment of the course "Power Processing II." The author also thanks Dr. H. B. Hamilton and

Dr. T. W. Sze, University of Pittsburgh, for their aid and comments during the design of the course, and the author is particularly grateful to Mr. Alec H. B. Walker, Westinghouse Research Laobratory, for his continuing assistance in selecting the course material and acting as an expert technical reference to the present state of the art. The author also acknowledges the preliminary course design and notes of Dr. John Choma, Jr., Sacramento State College.

Dr. Frank E. Acker

Table of Contents

Chapter 1. Power Diodes

Chapter Contents

This chapter has three equally important purposes: to acquaint the reader with the circuit properties of power diodes, to review physical models of a semiconductor p-n junction (these models will also be used in discussing the properties of the thyristor in chapter 3), and to provide some elementary examples of modeling both the characteristics of the power diode and the properties of some diode circuits.

The chapter begins with a review of the basic conduction properties of single-crystal semiconductors and the electron-hole model of such a crystal. Next, the electron-hole model of a p-n junction and the voltage-current characteristics of a signal diode are reviewed. The voltage-current characteristics of a power diode will then be compared to that of a signal diode, and the differences in the characteristics will be qualitatively explained on the basis of the electron-hole model as applied to the physical construction of the diodes. Some very simple diode circuits are then analyzed using circuit models or approximations of the voltage-current characteristic of a power diode.

The Electron-Hole Model of a Semiconductor

The semiconductor diode and thyristor are rapidly replacing other nonlinear or time varying control elements in the power electronics area. While the semiconductor devices may not replace all other types of devices such as selenium diodes, high vacuum diode, etc., their widespread and increasing use justifies the review of some of the basic properties of a crystalline, semiconducting material in order to better understand the characteristics of a semiconductor device.

The electrical properties of crystalline germanium and silicon, those semiconductors most useful in electronics at temperatures near room temperature (23°C), can be explained in terms of quantum mechanics, statistical mechanics, and the band theory of solids. However, the mathematics and details of such a treatment are very cumbersome in explaining the gross electrical characteristics of these materials. Accordingly, a "semi-classical" model of semiconductors built on the results of more detailed physical models is frequently used. This model will be referred to as the "electron-hole" model. The bases for this model lie in solid state physics and have sufficient complexity that we choose not to discuss them here. The inquisitive student may begin to consider the underlying physical processes by examining the references listed at the end of this chapter. It must be realized that by skipping the "derivation" of the electron-hole model, we generate a blind spot in our understanding in that we cannot decide from physical principles the limitations of our model. For example, with the electron-hole model above, we would conclude that the tunnel diode cannot work. Therefore, the applicability of the electron-hole model must be decided on the bases of:

a) the successful application the same model in previous similar situations, i.e., history

b) the results predicted by the model agreeing with experiment, i.e., empirical test

Clearly, in any practical situation, empirical test has the last and definitive word.

The Semiconducting Pure Crystal

In a single perfect crystal of germanium or silicon, each atom is located in a regular array (lattice) composed of all of the atoms that make up the crystal. Of course the idea of a perfect crystal is an idealization because there will always be some impurities and lattice flaws in the crystal, especially at the surfaces. The crystals used in semiconductor electronics are usually of such purity and sufficiently free of imperfections that these flaws are responsible for only second order effects in the crystal characteristics.

Germanium and silicon crystalize in the "diamond structure", a three dimensional array in which each atom is bound to its four nearest neighbors. We shall consider a two-dimensional schematic diagram of the three dimensional crystal.

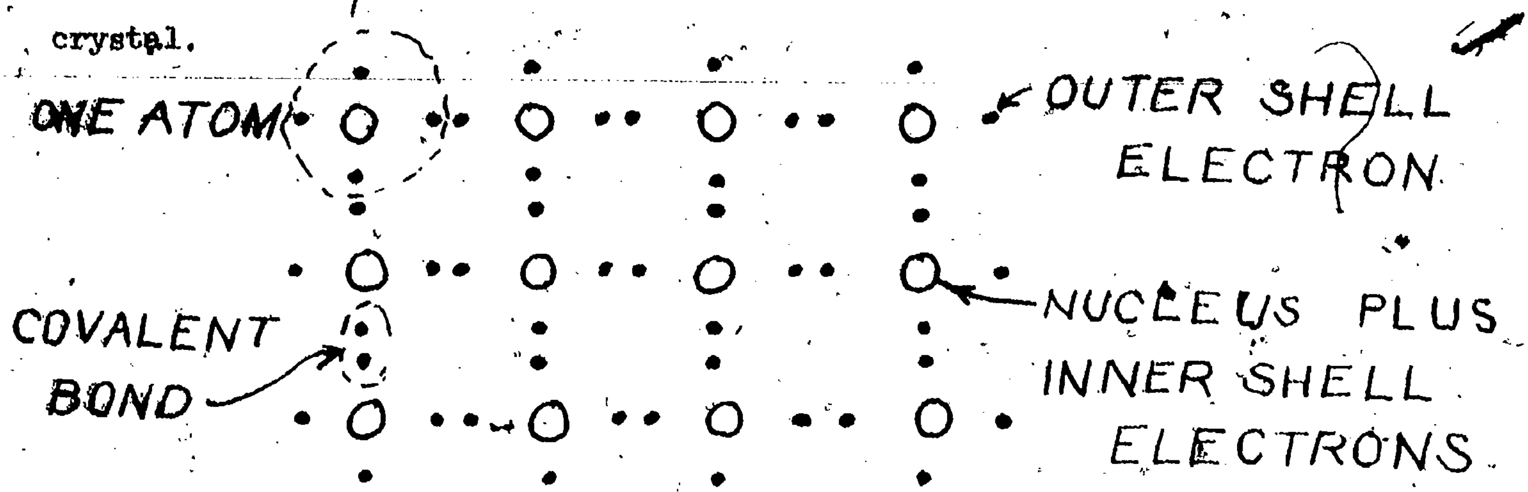


Figure 1 - 1

Note that each atom shares its four outer shell (chemically active) electrons with its four nearest neighbors, thereby effectively completing each atom's outer shell with its preferred number of electrons (eight). This type of shared-electron bonding between atoms is called covalent bonding

If by either some external exciting process or by random chance in the distribution of energy among the electrons, an electron receives sufficient energy to break the covalent bond (about 0.7 electron volts for germanium and about 1.1 eV for silicon at room temperature), the electron is no longer bound but is free to drift within the crystal lattice. When the electron begins to freely drift, it leaves behind a vacancy in its previous covalent bond. Other bound electrons of the adjacent atoms may fill this bond, but will leave behind vacancies in the bonds between other nearby atoms.

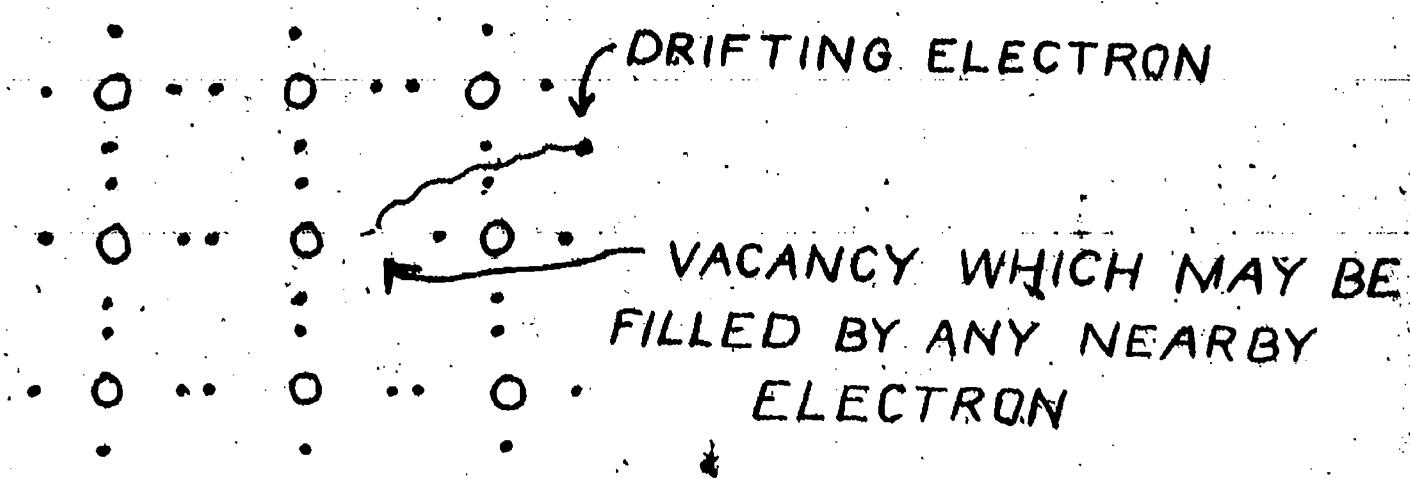


Figure 1 - 2

The vacancy in this model is termed a "hole". As the vacancy is repeatedly filled by bound electrons, the vacancy may propagate from place to place in the crystal. This process is called "hole drift". The hole has associated with it a positive charge due the net positive charge of the atom of which the hole is a vacancy. Despite the fact that the electron has a negative charge and the hole has a positive charge, the two charge "carriers" do not electrostatically attract each other because as the hole and electron drift apart, the many other electrons and holes in the crystal readjust their position slightly so that the value of $\vec{\mathcal{E}}$ (the electrostatic field strength) throughout the crystal is zero on even a microscopic (but not atomic) scale. Thus the hole and electron drift independently of each other. Should an electron and hole drift together by chance, they will annihilate each other (recombine) with the liberation of an amount of energy equal to the energy originally required to break the covalent bond.

Define

n = concentration of free electrons
i.e. $\frac{\text{number of electrons not bound}}{\text{unit volume of crystal}}$

p = concentration of holes

For a crystal of pure germanium or silicon

$$n = p$$

because free electrons and holes are generated in pairs.

Doped Semiconductors

It is possible to make the concentration of free electrons (n) different from the concentration of holes (p) by growing "impure" or doped crystals. If crystals are grown of a mixture of germanium or silicon and a small amount of another element (impurity) having five electrons in its outer shell (such as arsenic or phosphorus), a crystal in which n is greater than p or "n-type" crystal results. In such a material, the impurity atom which occupies a normal lattice site now has one electron which is not bound by a covalent bond.

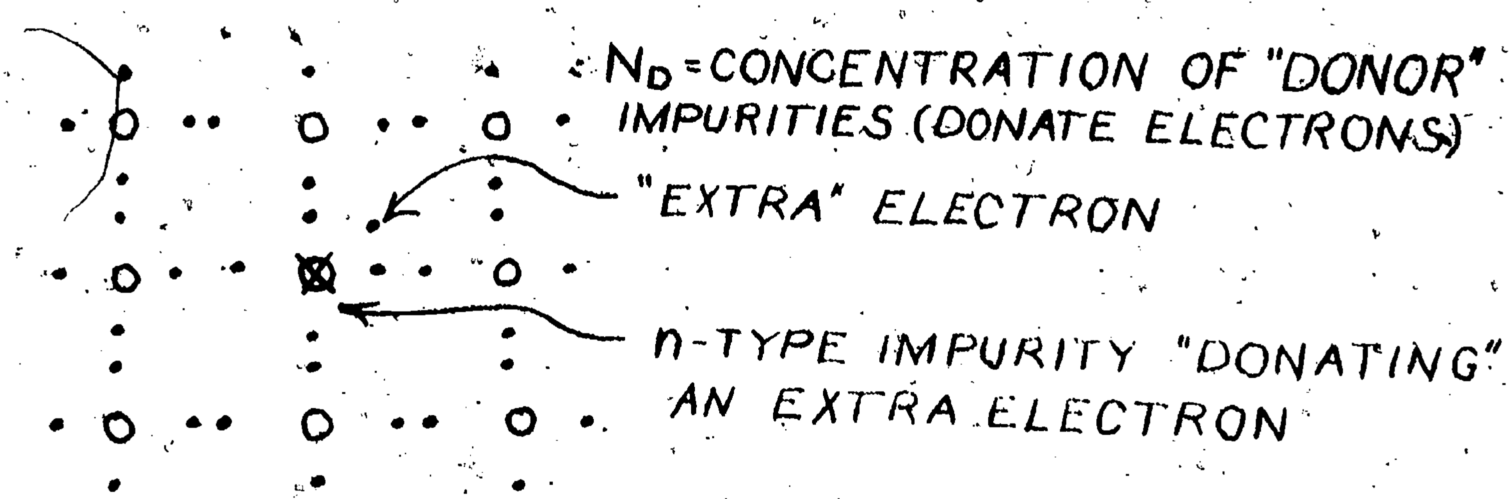


Figure 1 - 3

The energy required to free this "extra" electron so that it may drift about in the crystal is only on the order of one hundredth of the energy required to break a covalent bond. At room temperature, essentially all of these extra electrons are free. Note that when the electron drifts away from the atom, the atom has a positive charge. But this positive charge cannot drift about like a hole because nearby electrons would have to break a covalent bond (requiring lots more energy) to fill this type of a vacancy, and of course, the atom itself is locked in the crystal lattice so that it cannot drift. In a homogeneous material, the free electrons would distribute themselves in such a way as to cancel the electrostatic field produced by the fixed, ionized impurity atoms. However, in a nonhomogeneous material (such as a diode junction) these fixed charges may give rise to macroscopic observable characteristics.

It is also possible to add impurities to germanium and silicon crystals so that the holes will outnumber the free electrons (p-type material). Doping substances such as boron, aluminum, gallium, and indium have only three electrons in their outer shell.

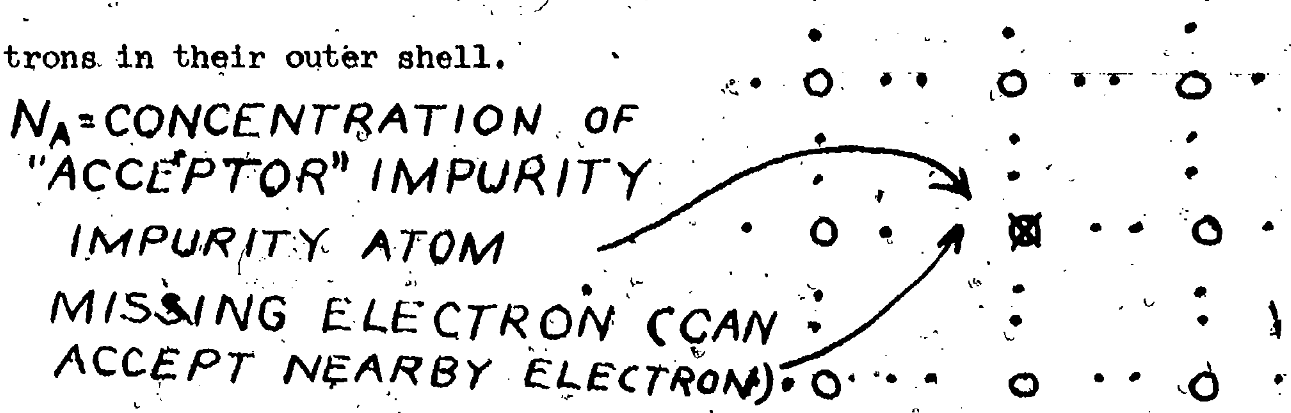


Figure 1 - 4

Only about one-hundredth of the energy required to break a covalent bond and generate a hole-electron pair is required for a nearby electron to "slip over" and complete the outer shell of the impurity atom. Such an action generates a hole which is now free to drift within the crystal. Again, similar to n-type material, almost all of the impurity atoms have their outer shells completed at room temperature. Each impurity atom has a fixed negative charge associated with it (due to its "extra" electron in its outer shell), and in a homogeneous material, the mobile positive charges (holes) move so as to cancel out the electrostatic field due to the fixed charges.

Thus we see that in a doped semiconductor, mobile charges (free-electrons and holes) arise from two sources:

a) broken covalent bonds yielding electron-hole pairs.
b) ionized impurity atoms yielding free electrons or holes depending on the type of impurity.

Suppose we have a material that is n-type due to doping. The concentration of electrons (n) will be greater than in the pure material due to the ionized impurity. The increased population of free electrons increases the probability of a free-electron meeting a hole and recombining. Therefore the concentration of holes will be less in an n-type doped material than in the pure material. In fact, statistical mechanics applied to this problem will give us the result that the concentration of holes times the concentration of electrons is equal to a number which is a function of the absolute temperature

$$n\,p = C(T) = K\,e^{(-Eg/kT)}$$

K = constant with units of concentration^2

Eg = energy to create a hole-electron pair (1.1 eV for Si, 0.7 eV for Ge)

k = Boltzmann's constant

T = absolute temperature

PROVIDED THE SEMICONDUCTOR IS IN A STATE OF THERMODYNAMIC EQUILIBRIUM (the temperature is not changing and no current is flowing).

Summary

In summary, some of the basic properties of the electron-hole model of a semiconductor are:

a) free-electrons and holes are the MOBILE CHARGE CARRIERS in a semiconductor

b) electron-hole pairs may be generated in the material by breaking covalent bonds (GENERATION)

c) free-electrons and holes may come together in the crystal and annihilate each other with the release of energy (RECOMBINATION)

d) semiconductor crystals may be doped with impurities, yielding free electrons or holes and FIXED CHARGES in the crystal lattice

e) at room temperature, essentially all commonly used impurity atoms are ionized

f) in the case of thermodynamic equilibrium, the product of the concentration of holes and electrons is a function of temperature (independent of doping) $np = C(T) = Ke^{(E_g/kT)}$

Conduction in Semiconductors

We next review the basic processes of electrical conduction in a semiconductor crystal. Suppose two copper wires are fastened to an n-type material by some metallurgical process. We then graph the electrostatic potential as a function of distance for this device, assuming the wire-semiconductor junctions are non-rectifying, ohmic contacts.

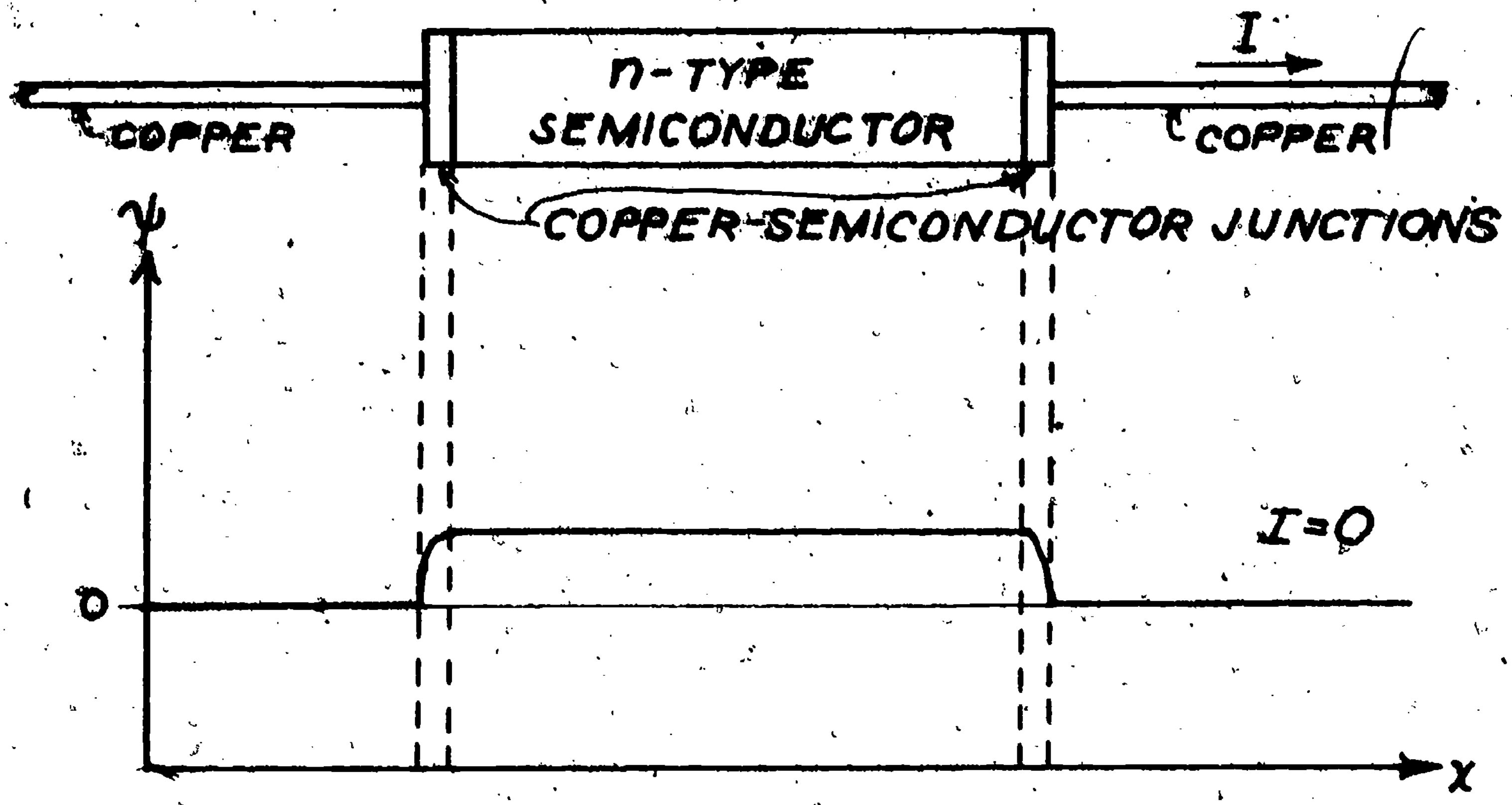


Figure 1 - 5

The difference in electrostatic potential between the copper and the semiconductor is a "contact potential difference" resulting from a difference in the work functions of the two materials. Remember that contact potentials cannot be measured with a voltmeter having leads, since the sum of the contact potentials about any loop is zero (otherwise a current would flow, energy would be dissipated, and the second law of thermodynamics would be violated). We are reassured that the potential difference between the two copper leads (which can be connected to an ordinary voltmeter) is zero.

Consider the case where the semiconductor with attached wires is connected to a battery.

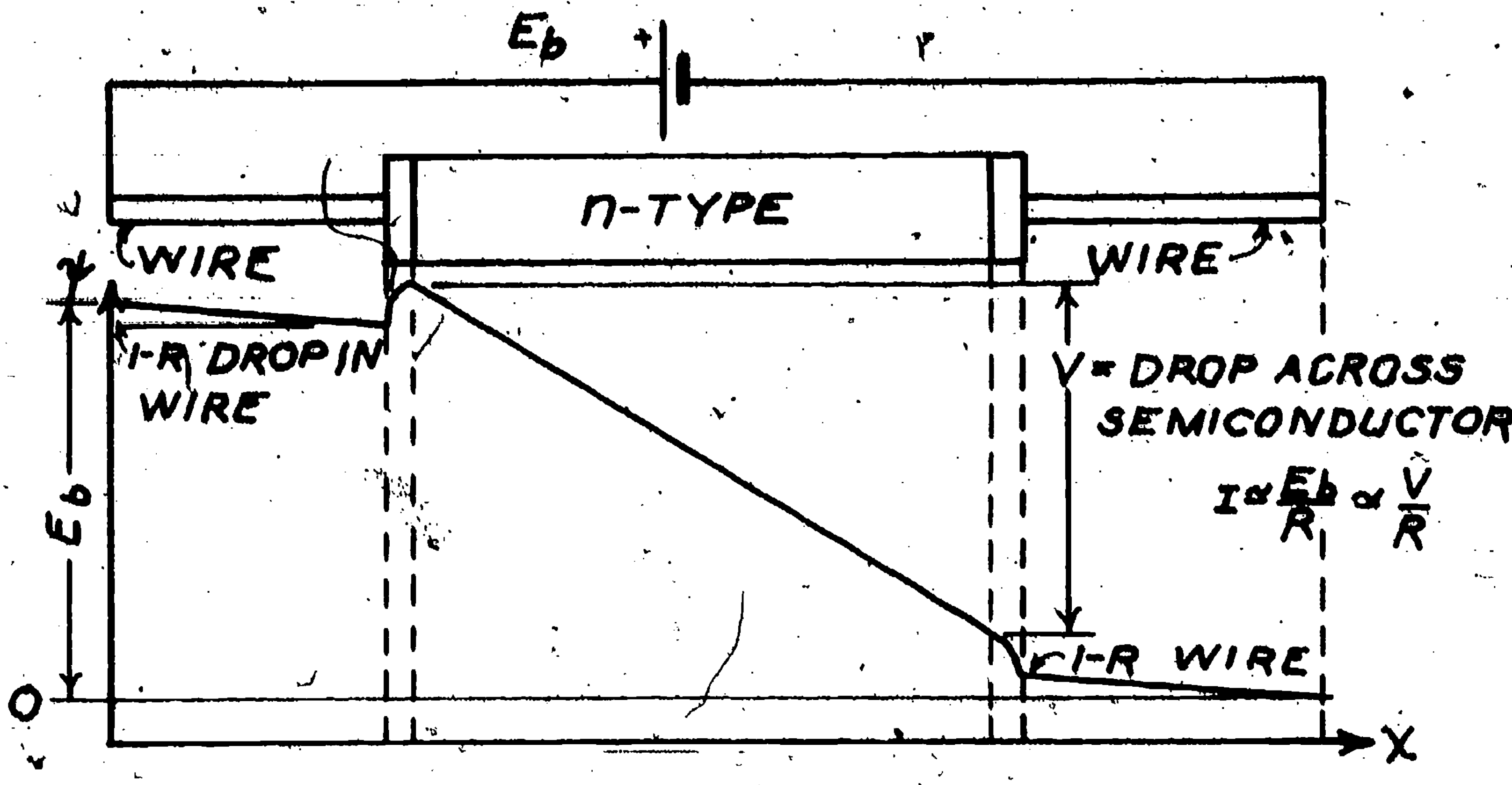


Figure 1 - 6

If we can neglect the effects of the ohmic voltage drops at the junctions, and since the contact potentials cancel, $V \propto E_b - I R_w$. Experiment shows that I is proportional to V, or that the semiconducting crystal behaves like a resistor. The characteristic of resistance coupled to the fact that conduction is by means of electrons prompts us to extend our model in terms of the well known model of electrical conduction by electrons in metals.

To construct a model for a resistor using electron conduction, we reason thus:

a) The current must be proportional to the number of charge carriers, and their velocity.

b) Therefore the application of a voltage to the crystal must change the velocity of the charge carriers, change the number of charge carriers, or result in some combination of the two possible effects.

c) We turn to the theoretical analyses and experimental work of others and accept as fact in our model that:

As the voltage applied to the crystal increases, current at first increases proportional to voltage. This phenomenon is due to the speed of the carriers being proportional to the applied voltage (actually, proportion to the electric field strength). If the magnitude of the voltage is increased sufficiently, the current will increase abruptly with small increases in voltage in a very nonlinear way. This effect (avalanch) is ascribed to the increase in the number of charge carriers.

d) Consider the case where I is proportional to V. We must have the charge velocity increase proportional to the applied value of $\vec{\mathcal{E}}$ (or V/length in a homogeneous crystal). We rationalize by saying each carrier is acted upon by a force $\vec{F} = q\vec{\mathcal{E}}$ where q is the value of the electronic charge. If the charge carrier is an electron, $q = -1.6 \times 10^{-19}$ coulomb. This force accelerates the carrier which soon bumps into the atoms composing the crystal. These collisions change the direction of motion, so that on the average, the carrier has a "drift velocity" or component of velocity in the direction of the applied field that averages to be proportional to the magnitude of the applied field

$$\vec{v} = -\mu_n \vec{\mathcal{E}} \quad \text{for electrons}$$

$\mu =$ mobility, a constant

and $\vec{v} = \mu_p \vec{\mathcal{E}}$ for holes. (by analogy, holes are as good carriers as electrons). The two equations result from the convention that μ is always positive. Also, $\mu_n \neq \mu_p$ because the mechanism of conduction of holes is different than the mechanism of conduction of electrons.

e) Consider a "filament" of unit cross sectional area of the crystal

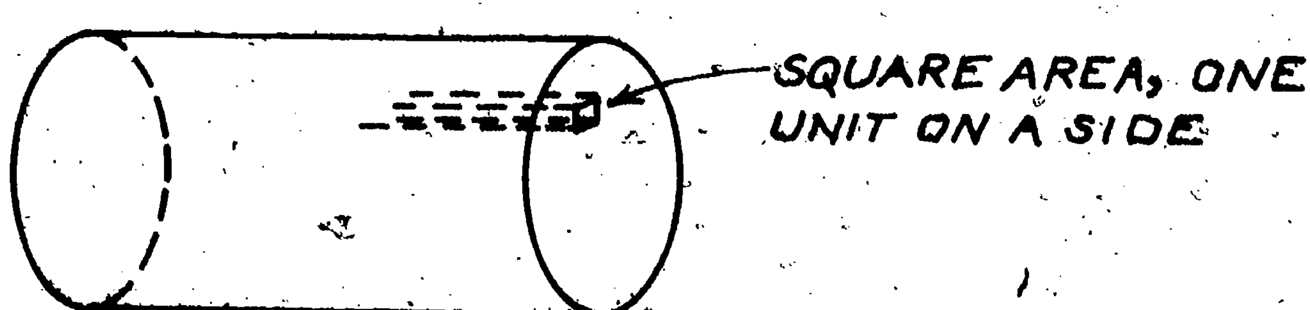


Figure 1 - 7

The number of electrons crossing this area per unit time (flux of electrons due to the battery being connected) is equal to the average electron drift velocity multiplied by the number of electrons per unit volume of the crystal. Of course, this gives the flux density. If the total flux of electrons were desired, one would simply integrate the flux density over the total crystal area.

$$\vec{\mathcal{F}} = \text{electron flux density} = \vec{v}\ n$$

The current is the transport of charge per unit time, and is simply the charge of an electron multiplied by the number of charges crossing the area per unit time. Again, since we are working in a small filament of unit area, we are calculating the current density due to electron drift

$$\vec{J}_n = -q\, n\, \vec{v}. \qquad q = 1.6 \times 10^{-19}\ \text{coul.}$$

Recall that $\vec{v} = -\mu_n \vec{\mathcal{E}}$ and eliminating v

$$\vec{J}_n = +q\, \mu_n\, n\, \vec{\mathcal{E}},$$

which is Ohm's law at a point

$$\vec{J}_n = \sigma \vec{\mathcal{E}} \qquad \sigma = \text{conductivity}$$

or

$$\rho \vec{J}_n = \vec{\mathcal{E}} \qquad \text{where} \quad \rho = \text{resistivity}$$

and

$$\rho = \frac{1}{q\, \mu_n n}.$$

To find the resistance of the crystal, we need only multiply by the length and divide by the total area, or

$$\text{Resistance} = \frac{\rho \ell}{A}.$$

f) Similarly, in a p type material where p is orders of magnitude larger than n, and conduction is by means of holes.

$$\vec{J}_p = q\, \mu_p\, p\, \vec{\mathcal{E}} \quad \text{and} \quad \rho = \frac{1}{q\, \mu_p\, p}$$

In the case that n is not orders of magnitude different from p, conduction by both types of carriers is significant and

$$\vec{J}_{total} = \vec{J}_n + \vec{J}_p$$

$$\vec{J}_{total} = q\,(\mu_p p + \mu_n n)\vec{\mathcal{E}}$$

$$\text{or } \rho = \frac{1}{q\,(\mu_p p + \mu_n n)}$$

In brief summary:

a) the voltage-current characteristic of a homogeneous semiconducting crystal are the same as that of a resistor (it's junctions that cause the non-linearities),

b) there are both positive (holes) and negative (free electrons) charge carriers in a semiconductor. These carriers drift under the influence of an $\vec{\mathcal{E}}$- field, producing a current,

c) the resistivity of a semiconductor is a function of the concentration of the charge carriers.

It is possible for currents in semiconductors to arise from diffusion processes in addition to the drift current previously considered. Diffusion arises in nonhomogeneous cases. Suppose that a number of electrons were "injected" by some unknown process into a semiconducting crystal, analogous to a drop of ink being injected into a bowl of water. In time, the high concentration of electrons (or ink) would decrease as the particles diffused away from the original location due to the fact that the particles are freely drifting, possess random motions, and have a high probability of moving in directions other than together. The eventual dispersion of the electrons or ink droplet represents a current or flux of the particles away from their original location. The diffusion of particles is described by Fick's Law which states that:

the net flux of particles is related to the gradient of the concentration N by a constant

$$\vec{\mathcal{F}} = -D \text{ grad } N \text{ or } = -D\,\vec{\nabla} N$$

where D is the diffusion constant.

In one dimension

$$\mathcal{F} = -D \frac{dN}{dx}$$

For electrons $\quad q = 1.6 \times 10^{-19}$ coulombs

$$\vec{J}_n = q\, D_n\, \vec{\nabla} n$$

For holes

$$\vec{J}_p = -q\, D_p\, \vec{\nabla} p$$

Again, $D_n \neq D_p$ due to the different processes in the motion of holes or electron. As a small point aside, it can be shown that

$$\frac{D}{\mu} = \frac{kT}{q} \quad \text{Einstein relation}$$

Diffusion currents are important near the junction of a diode.

Combining the drift and diffusion currents, the total current consists of four parts

$$\vec{J}_{total} = \vec{J}_{\substack{electron\\drift}} + \vec{J}_{\substack{hole\\drift}} + \vec{J}_{\substack{electron\\diffusion}} + \vec{J}_{\substack{hole\\diffusion}}$$

and I, the total current (steady state neglecting displacement currents)

$$I = \int_A \vec{J}\, dA = I_{\substack{electron\\drift}} + I_{\substack{hole\\drift}} + I_{\substack{electron\\diff.}} + I_{\substack{hole\\diff.}}$$

I must be the same in any plane at any place in the semiconductor or its leads by Kirchoff's current law, $\Sigma I = 0$. However, the relative importance or magnitudes of the four components may vary from place to place in the semiconductor due to either nonhomegenuities in the crystal or in the concentrations of holes and electrons in the crystal.

The Ideal Diode

Consider a single crystal of germanium or silicon in which the doping varies with distance in such a way as to produce both an n and a p type material in the same crystal.

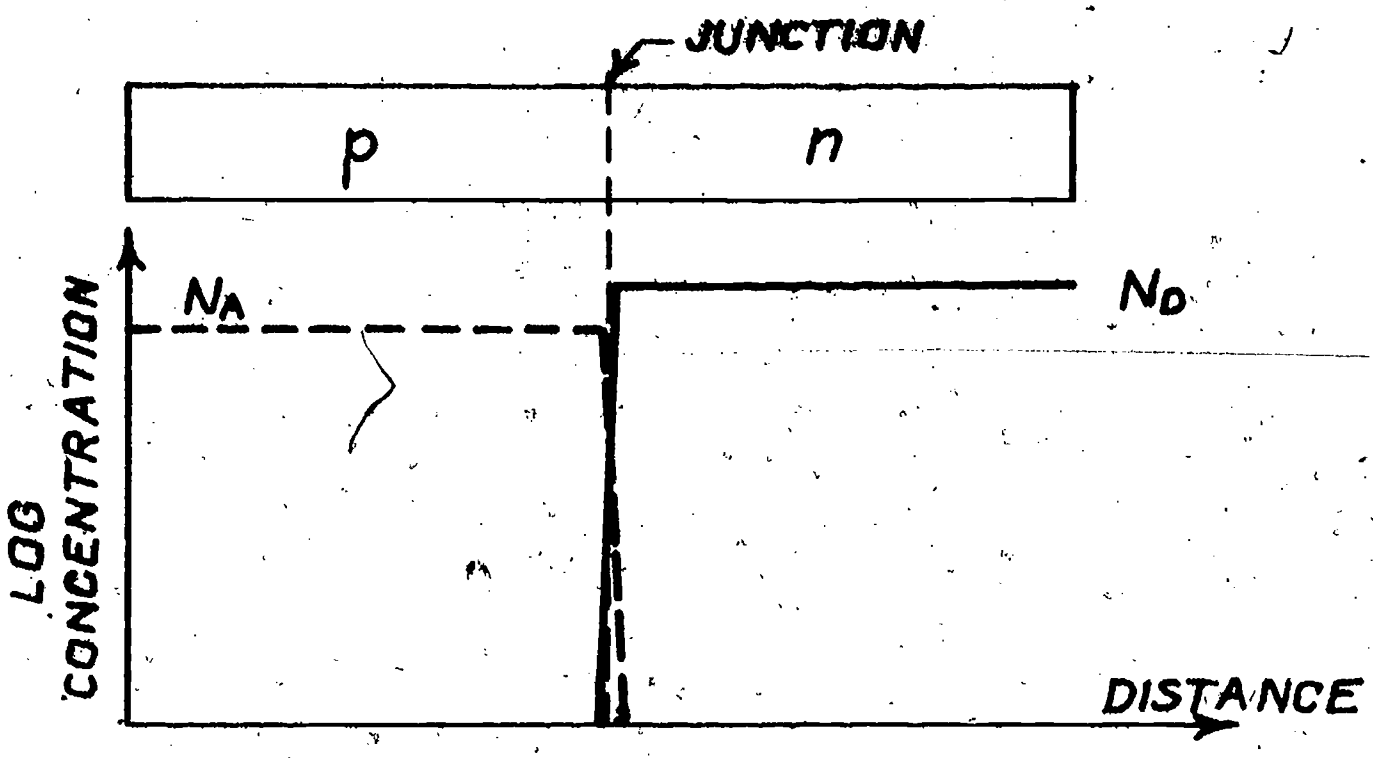


Figure 1 - 8

The region where the material changes from p to n-type is called the "junction" or transition region. Clearly, the crystal is not homogeneous in the junction region. Next, consider the crystal as lying isolated, not connected to any energy source, and in the thermodynamic equilibrium with its surroundings. Far from the junction region (on the order of a few thousandths of an inch), $p = N_A$ in the p-type material and $n = N_D$ in the n-type material. Because $np = C$, a constant, n is determined in the p-type material (and will be several orders of magnitude less than p in the p-type material) and p will be determined in the n-type material (and will be several orders of magnitude less than n in the n-type material).

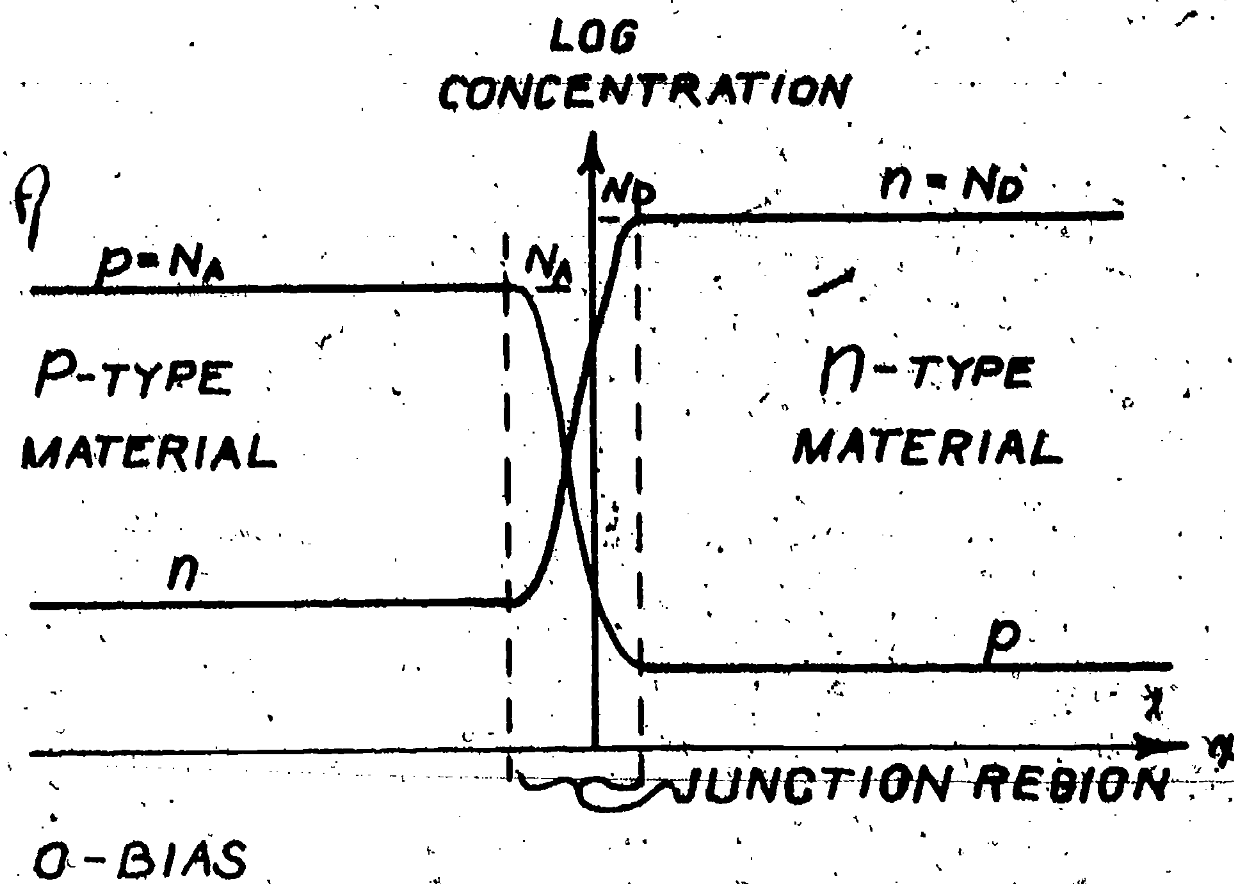


Figure 1 - 9

There is an orders of magnitude change in the concentrations of electrons and holes in the junction region. The large variation in the concentrations of the charge carriers gives rise to diffusion currents. Holes diffuse from the p-type material to the n-type material, and electrons diffuse from the n-type material to the p-type material. The transport of holes and electrons is a diffusion current (I_{diff}) from the p-type material to the n-type material. The carrier concentrations never equalize across the junction, that is, the diffusion current is not a transient phenomenon. A hole, crossing the junction from the p to the n-type material finds itself in a region where there are many electrons. The probability of the hole recombining with an electron drastically increases. The result is that practically all the carriers that diffuse across the junction recombine. Holes are resupplied to the p-type material at its ohmic contact by a fairly complicated process we choose not to describe at present, and by a drift process about to be described. A similar argument holds for electrons.

Thus far, a diffusion current from the p to the n-type material has been qualitatively described. Yet, the total current must be zero, since the crystal is electrically isolated. Therefore a drift current must exist that exactly cancels the diffusion current.

$$I = 0 = I_{diff} + I_{drift}$$

We next consider the source of the $\vec{\mathcal{E}}$ field that "drives" the drift current.

At the junction region, the concentration of holes in the p-type material is decreased (refer to Fig. 1-9) below N_A due to the high rate of diffusion of holes away from that location. Since the ionized acceptor atoms have a negative charge, and the atoms are fixed in the crystal lattice, there is a negative charge density located in the p-type material near the junction. Similarly on the n-type side of the junction, the electron density is lowered and the fixed charge due to N_D causes a positive charge density to result in the n-type region. The charge densities and concentrations are shown in Figure 1-10 on a linear scale (recall that Fig. 1-9 has a logarithmic carrier density scale).

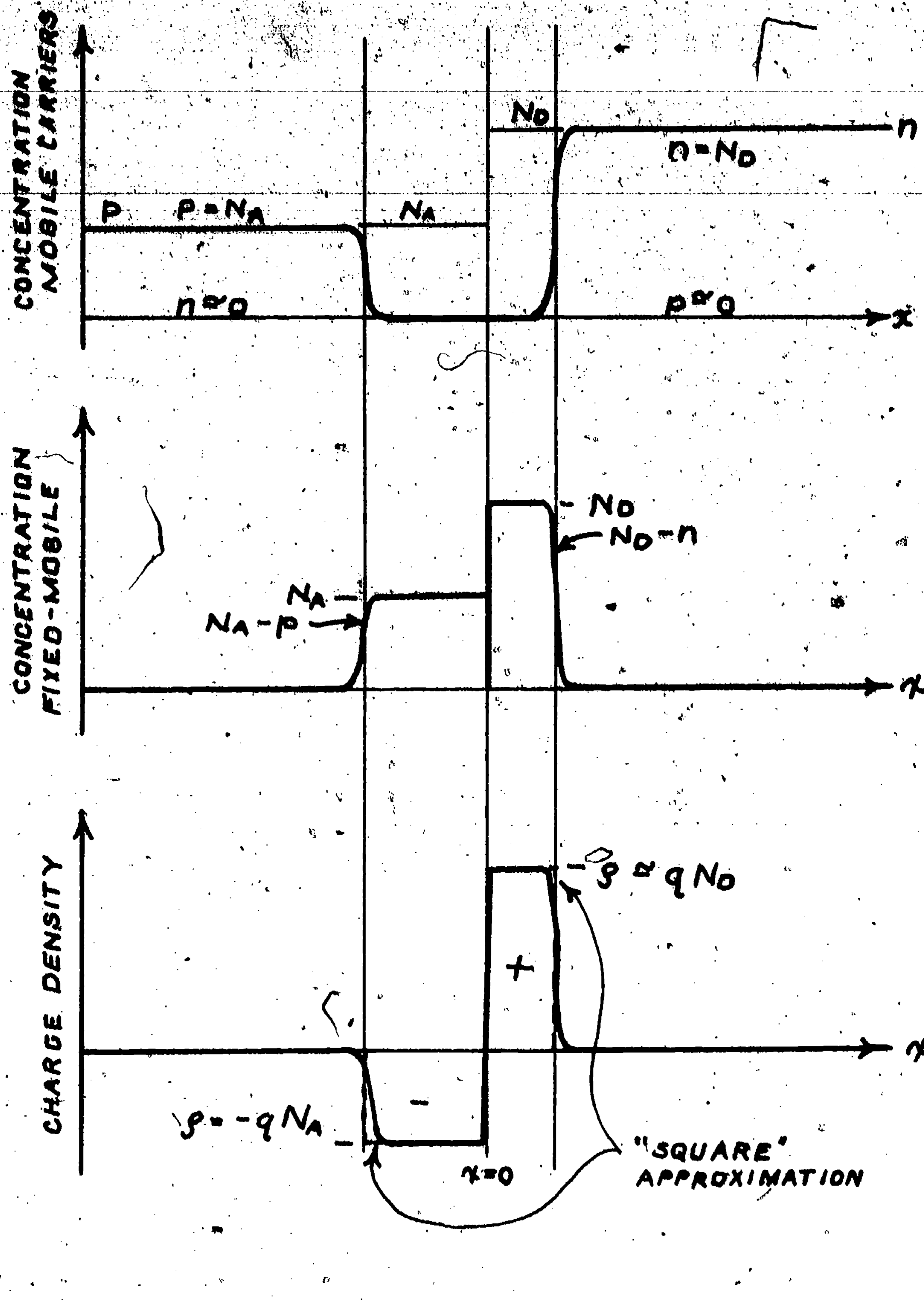

CONCENTRATION MOBILE CARRIERS
P
P = NA
NA
ND
n
n = ND
n ≈ 0
p ≈ 0
x
CONCENTRATION FIXED-MOBILE
ND
ND - n
NA
NA - P
x
CHARGE DENSITY
ρ ≈ qND
+
-
ρ = -qNA
x=0
"SQUARE" APPROXIMATION
x

Figure 1 - 10

The charge densities due to the fixed charges give rise to an electrostatic field which causes electrons to drift toward the n-type material and holes to drift toward the p-type material. Because the electrostatic field sweeps mobile charges out of the junction region, the concentration of mobile carriers is small compared to carrier concentrations in the bulk of the p and n-type materials. The region of low concentration of mobile carriers is known as the "depletion region" or "depletion layer."

Next we choose to calculate the electrostatic potential that exists across the depletion region. First we apply Gauss' Law $\frac{1}{\epsilon}\int_V \rho \, dV = \int_A \vec{\mathcal{E}} \cdot d\vec{A}$ [1]. Then since $-\int \vec{\mathcal{E}} \cdot d\vec{r} = \psi_{AB}$, we can find the potential difference that exists across the junction. To simplify the calculation, we make the justifiable assumption (for the desired accuracy of our qualitative analysis) that the charge density is a constant in the p-type region and a different constant in the n-type region (dashed lines, Fig. 1-10 c), thus neglecting the details of the edges of the depletion region.

Consider a filament one unit on a side within the crystal (so we can consider densities irrespective of the crystal dimensions). Because of the axial symmetry, only variations in the x direction (along the length of the filament)

[1] Note that by convention ρ is used both for charge density and resistivity. ρ (charge density) has units of coulomb/meter^3 while ρ (resistivity) has units ohm meter.

need be considered and we have only to analyze a one dimensional problem. Starting at x = A (Fig. 1 - 11) in the p-type material and working toward the n-type material, at first no charge is enclosed in the Gaussian surface, then a constant charge density is encountered and $\int \rho \, dV = \rho A x + C$ (unit Area so $\rho A = \rho$)

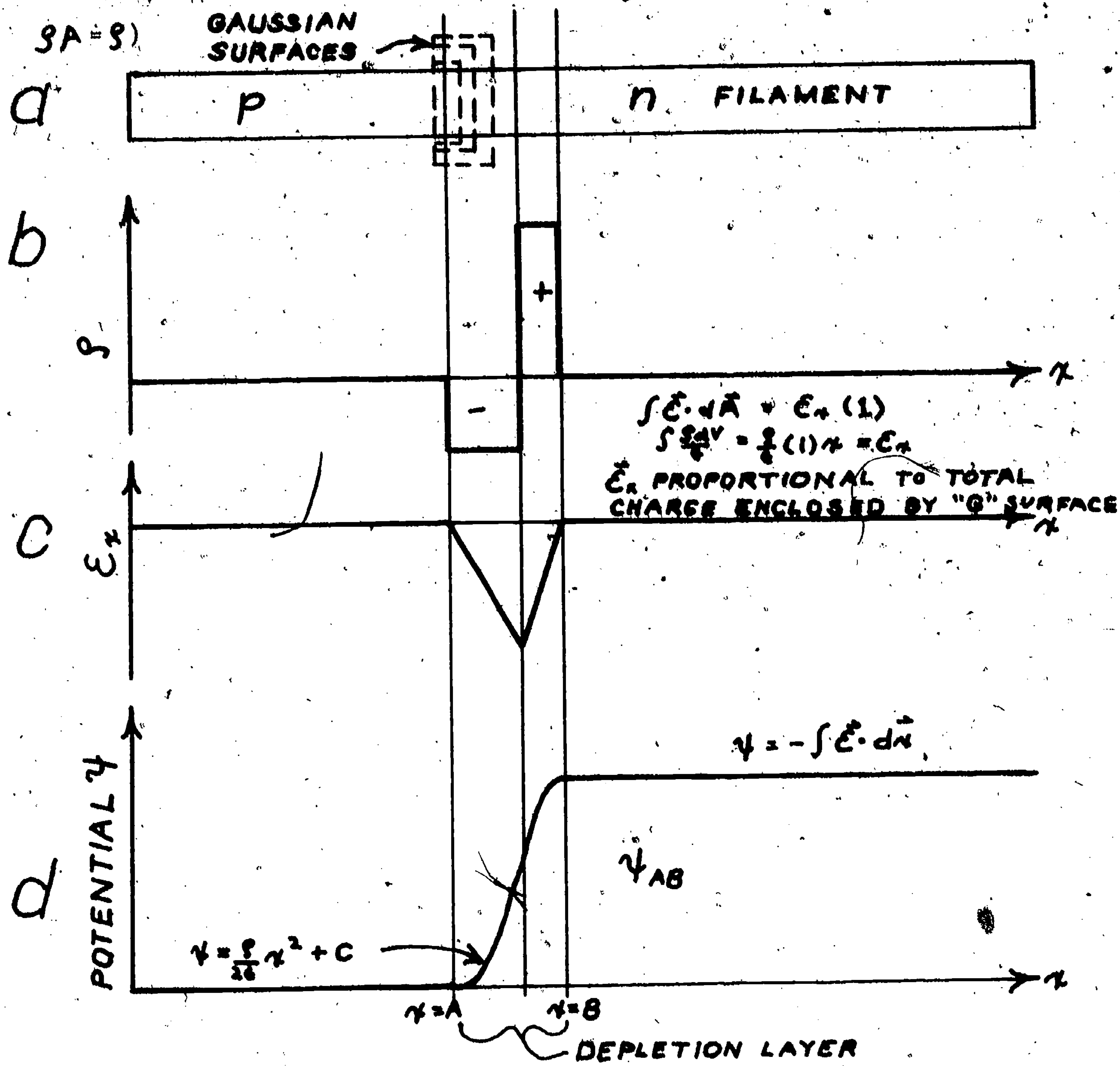
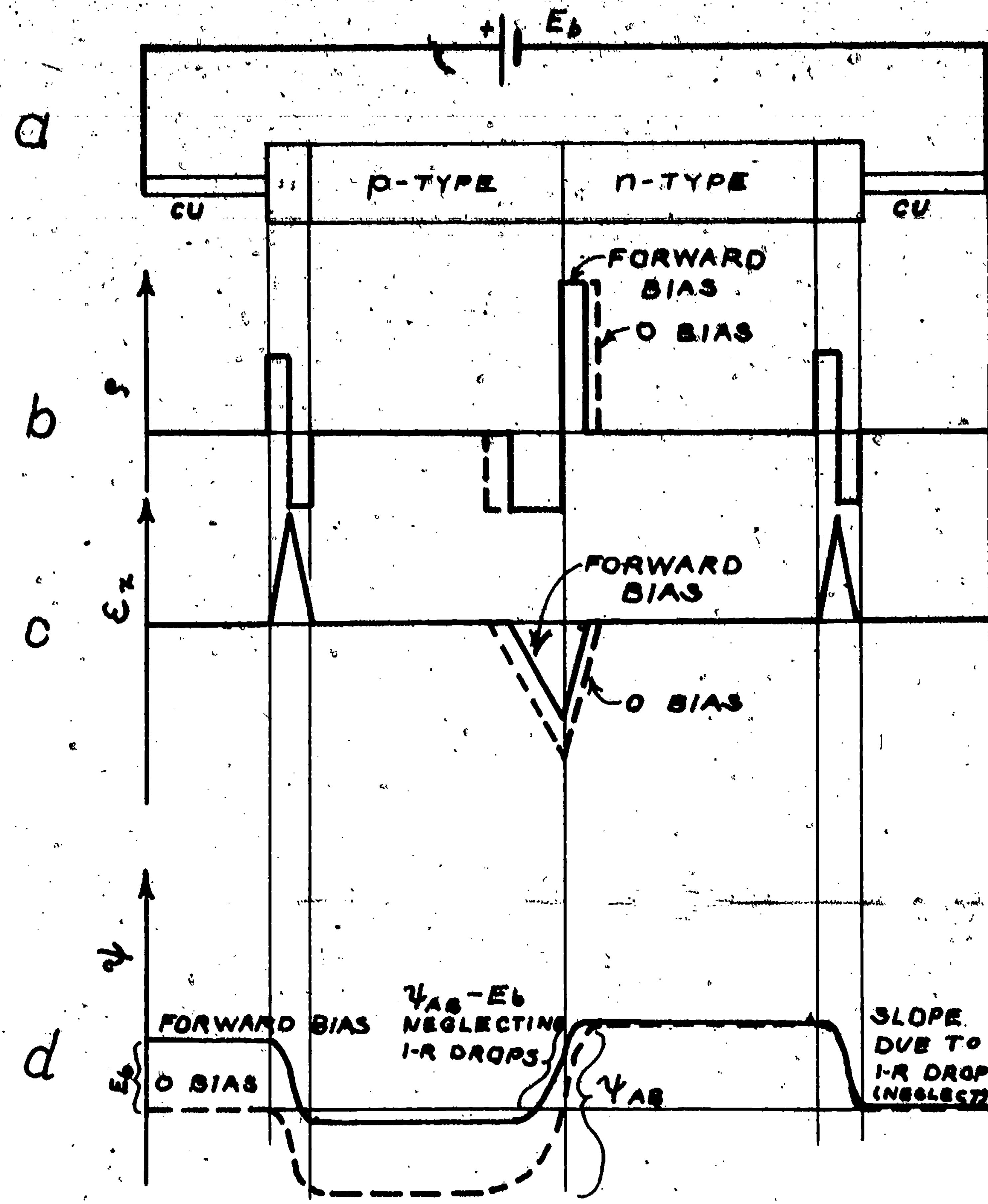


Figure 1 - 11

The field is a linear function of distance (x), and the potential is a parabolic function of distance (x). The potential difference ψ_{AB} between the p and n-type materials is the contact potential between these dissimilar materials. Also, the area of the charge density versus distance for the p-type material (Fig. 1 - 11b) must exactly equal the area of the charge density versus distance for the n-type material. Otherwise an $\mathcal{E}$ -field would exist outside the depletion region causing electron drift far from the junction which does not agree with observed facts or our model if pursued with sufficient detail.

Next consider a battery applied to the device. We have seen that in the case of zero current through the device, the current is made up of a diffusion and a drift component that cancel or balance each other. Applying an external voltage source will upset the balance and a net current will flow. If we can assume that essentially all of any voltage applied to the device appears across the junction, the analysis is greatly simplified. Such an assumption is usually justified in signal type diodes because the distance between the junction and ohmic contacts is small, the current density sufficiently low, and the resistivity of the material is sufficiently small that the bulk of the semiconductor materials far from the transition region has a negligible effect on the electrical characteristics of the device.

Figure 1-12 shows the differences in junction charge, $\mathcal{E}$ -field, and potential between the cases of zero current (zero applied voltage) and forward conduction (forward voltage bias). The depletion layer changes its length dependent on the applied voltage. The forward bias voltage E_b reduces the drift component of current in the junction region because $\mathcal{E}_x$ is reduced. The charge distributions in the copper-semiconductor junctions are simple inventions to make the contact potentials cancel around the loop when E_b is zero. The physical details of the metal-semiconductor junctions are outside the scope of the present course. We have adopted the following commonly used assumptions:



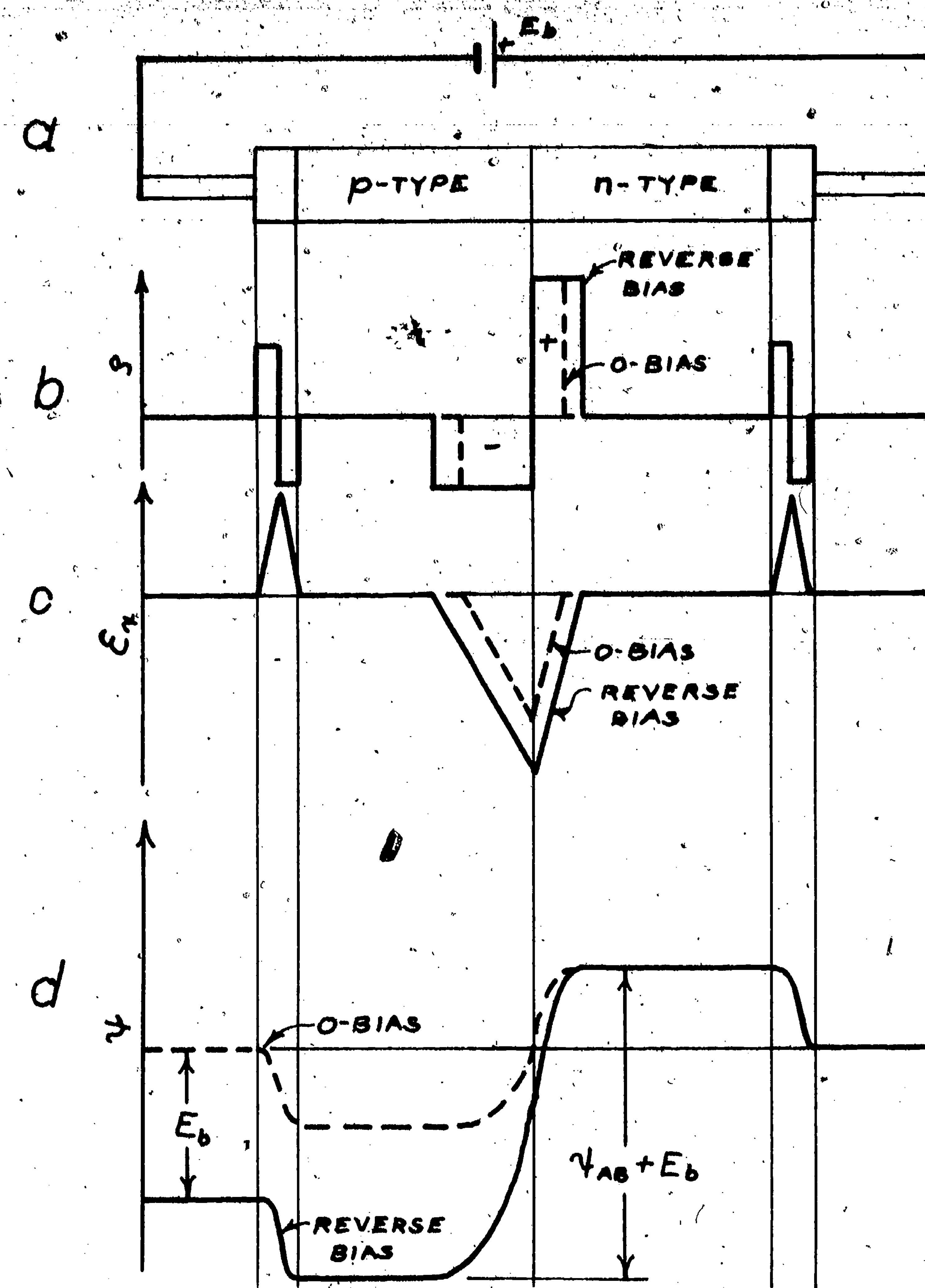
Eb
a
p-TYPE
n-TYPE
CU
CU
FORWARD BIAS
O BIAS
b
c
FORWARD BIAS
O BIAS
d
FORWARD BIAS
O BIAS
ΨAB-Eb NEGLECTING I-R DROPS
ΨAB
SLOPE DUE TO I-R DROP (NEGLECT)

Figure 1 - 12

a) I-R drops in the wires and the n and p-type materials outside the junction area are negligible compared to E_b. These I-R drops are responsible for drift currents outside the junction region (in the wire for example) but do not significantly contribute to the gross I-V characteristics of the device.

b) The metal-semiconductor contact potential differences do not significantly change as a function of current. Note that these potential drops are not insignificant compared to ψ_{AB}, but if they are not a function of current, the change in ψ_{AB} due to the connection of a battery is approximately equal to the battery voltage.

The charge, field, and potential distribution in the diode under reverse bias conditions are shown in Figure 1 - 13.



E_b
a
p-TYPE
n-TYPE
REVERSE BIAS
b
+
O-BIAS
-
c
O-BIAS
REVERSE BIAS
d
O-BIAS
E_b
ψ_AB + E_b
REVERSE BIAS

Figure 1 - 13

As the reverse voltage is increased, the length of the depletion region increases and the magnitude of the field strength (E_x) within the depletion region increases. The drift current (from the n to the p-type material) should increase and the diffusion current (from p to n-type material) should stay about the same value as in the zero-bias case or else decrease.

Consider whether or not is is possible for the drift current to exceed the previously discussed diffusion current. Although the situation is confusing at small reverse bias (because the diffusion and mobility constants for holes and electrons are different), the model is easily understood at large reverse bias. The drift current in the junction region arises from the drifting under the influence of the $\vec{\mathcal{E}}$-field of mobile carriers that have diffused into the junction area. At large reverse bias, all of the holes that diffuse into the junction region from the p-type material and all of the free electrons that diffuse into the junction region from the n-type material are turned back by the high field. Thus it would seem the drift current cannot exceed the diffusion current, and the net current should be zero for reverse bias.

We have neglected an important fact. In considering the cases of forward and zero bias, the concentration of carriers in the junction region exceeded the concentration of minority carriers (p in the n-type material and n in the p-type material) as shown in Figure 1-9. For large reverse bias, the concentration of mobile carriers in the junction is further reduced (Fig. 1-14) due to the large $\vec{\mathcal{E}}$-field in the junction region.

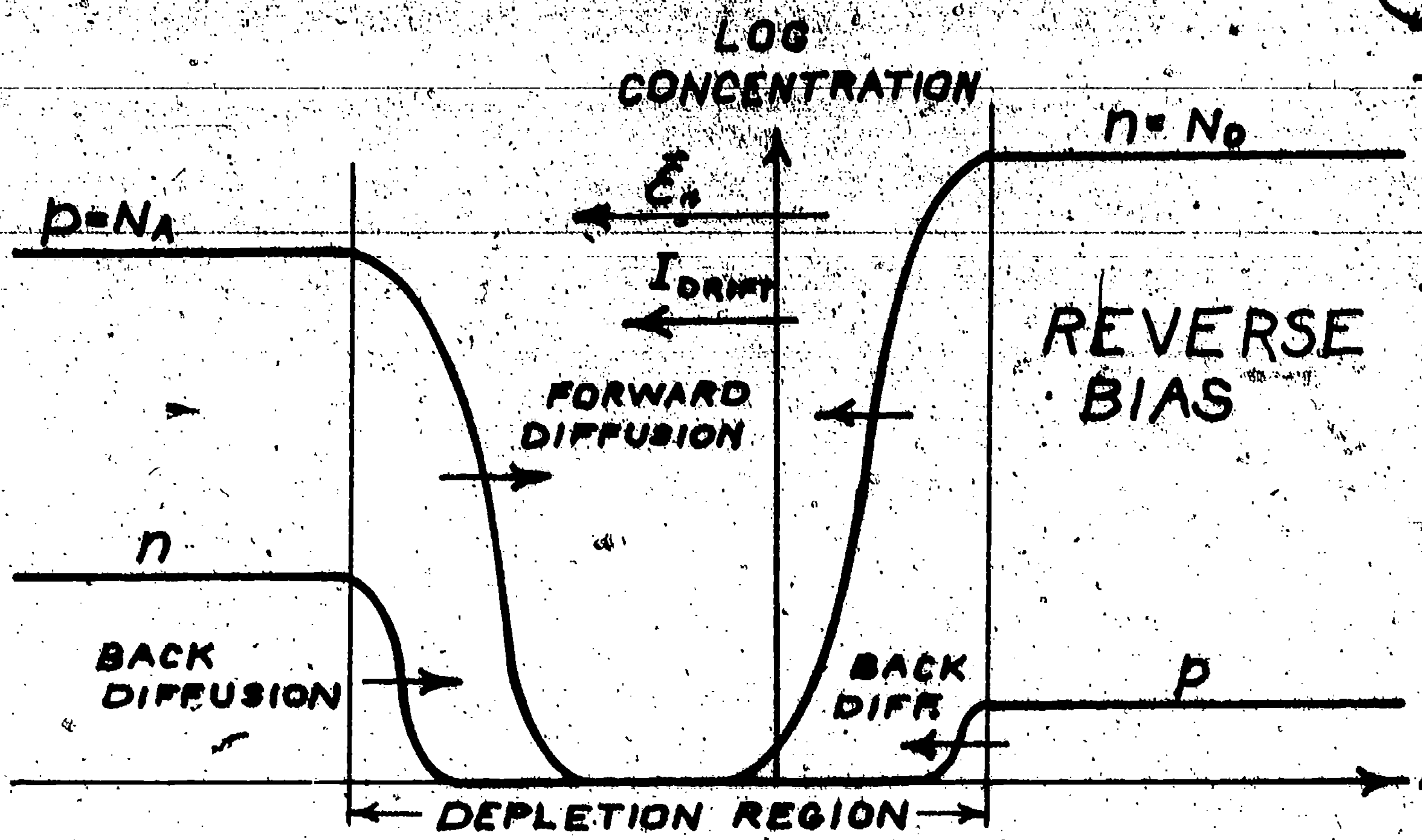


Figure 1 - 14

The concentration of mobile carriers in the junction region becomes so small that the minority carriers also diffuse into the junction region (here labeled "back diffusion"). The minority carriers are of such polarity as to be carried by the drift the remaining way across the junction, giving rise to a reverse current.

$$
\begin{aligned}
I_{reverse} = \; & I_{\substack{diff.\ forward \\ electrons}} + I_{\substack{diff.\ forward \\ holes}} \\
& + I_{\substack{diff.\ back \\ electrons}} + I_{\substack{diff.\ back \\ holes}} \\
& + I_{\substack{drift \\ electrons}} + I_{\substack{drift \\ holes}}
\end{aligned}
$$

Because the concentration of minority carriers is orders of magnitude less than the concentration of majority carriers, we expect that the reverse current should be orders of magnitude smaller than the forward current for the same magnitude of forward and reverse voltage bias applied to the device terminals. Also, changing the magnitude of the reverse bias voltage under conditions of large reverse bias should not change the magnitude of the reverse current, since the current is limited by "back diffusion" and the voltage change does not have much effect on the concentration gradient at the edges of the junction region. These predictions are correct. Careful application of the electron-hole model of the junction plus some application of statistical mechanics would allow us to derive the ideal diode equation which is a good approximation to the V - I characteristics of a signal diode over its normal operating range.

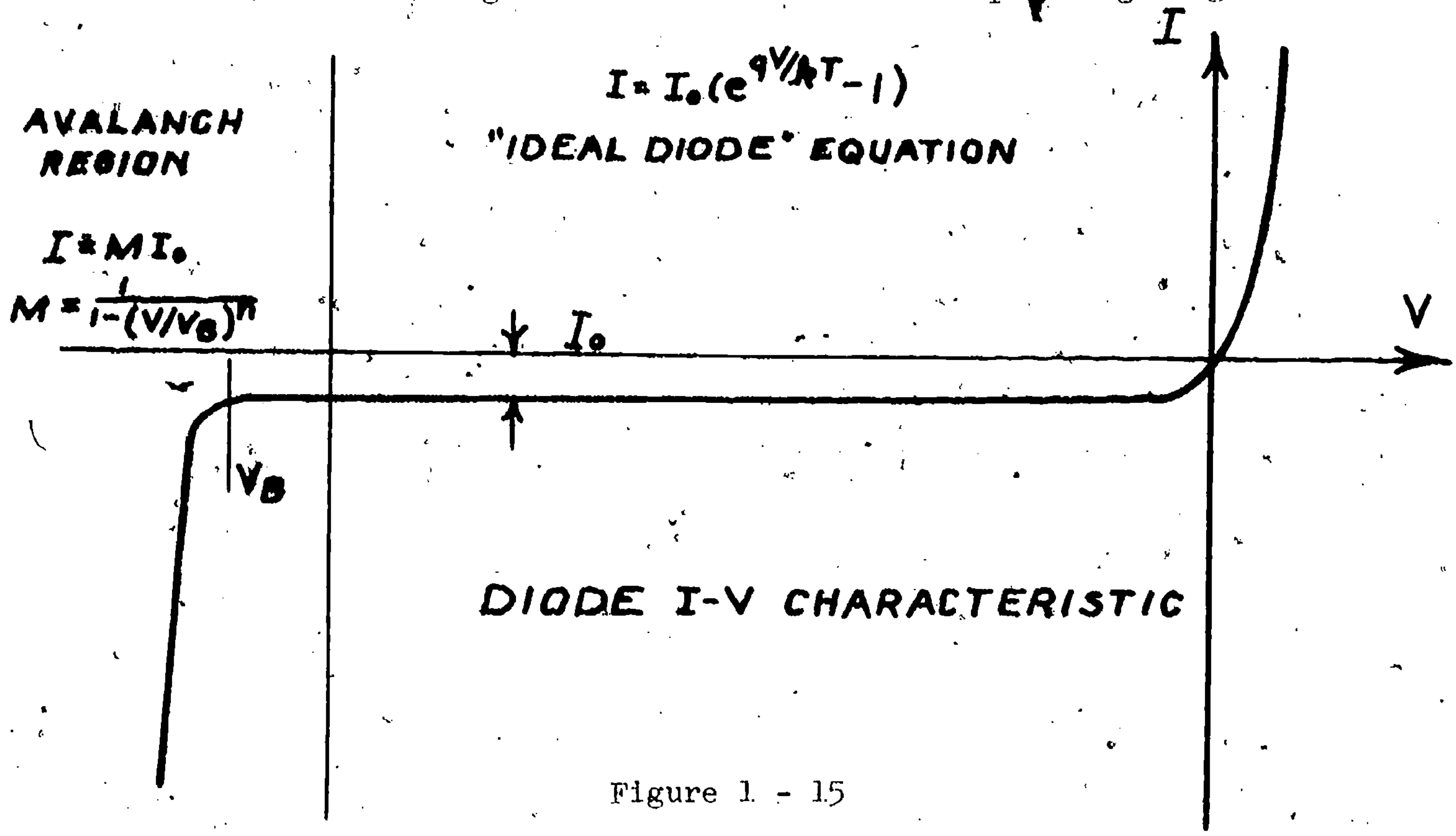


Figure 1 - 15

The ideal diode equation is a valid approximation to the V-I characteristic of any diode provided:

a) ohmic drops are negligible

b) most of the current is conducted through the junction rather than at the surface of the crystal ("Leakage currents" around the junction are caused by contaminants and imperfections in the crystal lattice which often exists at the surface of the crystal).

c) "normal" diffusion-drift processes account for the carrier transport across the junction.

Notice the ideal diode equation is not a good approximation for voltages near the value of the breakdown voltage (V_B) for the diode ("avalanch" region Fig. 1 - 15). The avalanch region occurs (from the point of view of our model) when the reverse bias voltage becomes so large, and the magnitude of the $\mathcal{E}$-field so large, that mobile carriers gain so much kinetic energy between collisions that they can break covalent bonds in the junction region (refer to Fig. 1-13). When a carrier (hole or electron) breaks a covalent bond, two additional carriers result which are also accelerated by the $\mathcal{E}$-field and break more covalent bonds. This process is known as the "avalanch process" and is responsible for drastically increasing the concentration of mobile carriers of the junction region, resulting in a sharp increase in current. Current in the avalanch region is frequently approximated by the relation.

$$I = I_o \frac{1}{1-(V/V_B)^n}$$

n varies between 2 and 4 depending the crystal material.

The avalanch equation is an empirical relation, fitted to the V-I curve in the avalanch region. The avalanch equation has not been derived from fundamental crystal properties.

While avalanch is not necessarily destructive to the diode, unless the diode has been specifically constructed to work in the avalanch region, excessive reverse voltage and subsequent avalanch leads to diode failure. The large simultaneous current and voltage in the avalanch region means that the device is dissipating energy at a rate far exceeding the rate of dissipation in the ideal diode region. Unless some special circuit or construction provision has been made, the temperature at the junction may rise above the melting point of the crystal. When the crystal (or some small area of the junction) melts, it loses its rectifying properties. Even if the current is interrupted before the leads or crystal vaporizes, the device has been ruined because the rectifying properties do not reappear upon cooling. The molten section does not recrystallize with the same structure that existed before melting, and the device exhibits properties similar to that of a resistor after cooling.

Briefly considering the transient behaviour of the diode, there are two predominant phenomena. The fixed charges in the depletion layer act as a stored charge, dependent on terminal voltage, similar to the stored charge on a parallel plate capacitor. Thus the depletion layer acts as a capacitor whose value depends on the applied voltage. Similar devices are used for high frequency tuning circuits (varactor diodes). For signal diodes, the capacitance value is usually on the order of some tens of picofarods. Also, when the voltage across the junction changes, the length of the depletion region changes, and therefore the distribution of mobile charge carriers changes. During the time the distribution of mobile carriers is changing toward a new steady-state, the motion of the carriers gives rise to currents. Because these currents are associated with the presence or absence of mobile carriers near the junction (i.e. "storage" of carriers) the effects of these charges can be associated with another capacitance (in addition to the depletion capacitance) which is both time and voltage varying.

Summarizing:

a) There is a balance of diffusion currents across a p-n junction,

b) The fixed charge densities in the depletion region are approximately equal to the doping densities,

c) The fixed charge densities in the depletion region are directly related (through Gauss' Law) to the value of $\vec{\mathcal{E}}$ in the junction region and the terminal voltage across the device,

d) The ideal diode equation $I = I_o \, (e^{qV/kT} - 1)$ is a good approximation to the characteristics of a signal diode provided:

1) ohmic drops are negligible,
2) there is negligible "leakage current,"
3) the voltage is below the breakdown voltage.

e) Avalanching is related to the value of the field in the junction region and hence to V_B,

f) There are voltage dependent and time varying capacitances associated with the p-n junction.

Power Diodes

The two most prominant ratings for a power diode are the maximum or peak reverse voltage the diode can maintain short of avalanch, and the maximum steady state forward current. Because the cost of a diode increases for increasing peak reverse voltage for constant forward current rating, and the cost increases for increasing forward current rating for constant peak reverse voltage rating, and also because the power dissipated in a diode increases as the current rating increases, diodes are normally chosen to operate near their maximum ratings. Any "safety factor" applied to the ratings will depend on the specific characteristics of the circuit in which the diode is to be used and the quality control of the manufacturer of the diode.

Except for any required safety factor, a diode will normally be operated at the maximum voltages and steady state current the device can tolerate. Therefore, the assumptions made in the case of signal diodes must be re-examined for the case of power diodes.

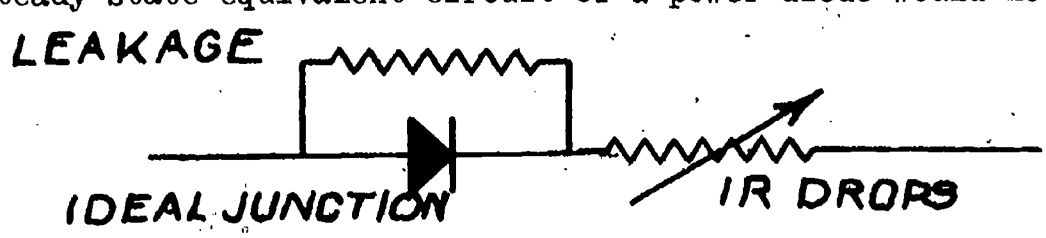
a) I-R drops

If the forward current is the maximum possible steady state value, the current density in the crystal is also as high as possible. The current and current density are limited by the maximum allowable junction temperature specified by the manufacturer such that long term degradation of the diode does not result (about 190° C for S_i) because diode temperature is in part a function of forward current. The large current density gives rise to non-negligible (compared to the total forward voltage across the diode) voltage drops across the p and n-type semiconductor material outside the junction region, and across the metal-semiconductor junctions. Furthermore, because of the high concentration of mobile carriers diffusing across the junction, the concentration of mobile carriers in the crystal far from the junction is increased. The higher than thermodynamic equilibrium concentration of carriers outside the junction area reduces the resistivity of the crystal ("conductivity modulation"). Thus the diode "ohmic drops" which cannot be neglected at high currents vary as a function of current in a nonlinear way.

b) Leakage currents

Surface leakage currents around the junction are not negligible compared to I_o in high power diodes. The leakage currents become significant compared to I_o in power diodes due to the large fields at the junction when the diode is operated at maximum reverse voltage, the increased circumference of the leakage path of larger diodes, and because of differences in the geometry of signal and power diodes arising from the need to dissipate more heat and pass larger currents in power diodes.

The steady state equivalent circuit of a power diode would look like:



In addition to the leakage and "IR" resistances, I_o is larger in power diodes than in signal diodes because the junction area increases as the current rating increases (keeping current density about the same), and I_o is proportional to the junction area. I_o plus the leakage current is frequently specified by the manufacturer for peak reverse current and maximum temperature. A typical specification might be:

P.R.V.	Max. $I_{forward}$	Max. Rev. Current at Max. Temp. and Rated P.R.V.	Max. One Cycle Surge 60 H_z
500	160 A	40 MA @ 190°C	2000A

A maximum current surge for a given time is also frequently specified. The only limit on transient forward current is the melting or fusing of the diode or its leads. Therefore, a current and a time corresponding to a commonly used waveform (usually a 60 H_z sine wave) is specified as a transient surge limit.

The power diode transient behavior depends not only on the conductivity modulation effect, but on the diode capacitance (voltage and time varying) which increases roughly proportional to the junction area and therefore also proportional to the current rating. It is instructive to consider a square wave voltage transient applied to a diode-resistor circuit (Fig. 1-16).

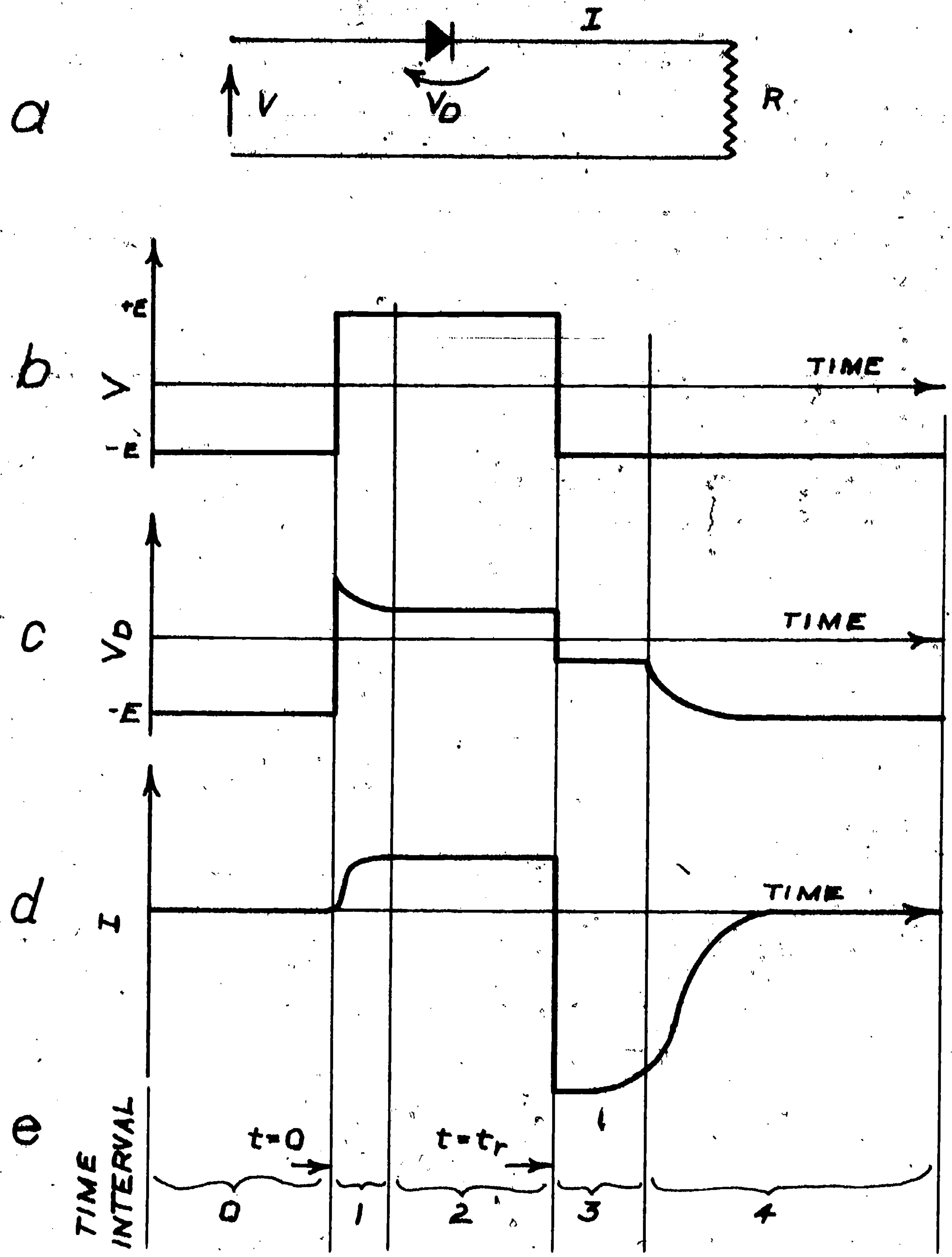

a
I
V
V_D
R
b
V
+E
-E
TIME
c
V_D
-E
TIME
d
I
TIME
e
TIME INTERVAL
t=0
t=t_r
0
1
2
3
4

Figure 1 - 16

time period	
0	Reverse bias voltage E has been applied for a long time. A steady state reverse current I_o + I leakage flows.
t = 0	The voltage applied to the circuit suddenly reverses polarity.
1	Current rises as the new mobile carrier distribution comes to steady-state and the resistivity of the crystal far from the junction decreases due to the high level of minority carrier concentration. Note that it is nonsensical to talk about the "resistance" of the diode.
2	The steady state in the forward bias condition has been achieved.
t = tr	The applied voltage polarity again reverses. Because the voltage across the depletion layer cannot change instantaneously due to the depletion capacitance, a current greater than the steady-state forward current may result.
3 & 4	Carriers are being swept out of the new depletion region and the bulk of the crystal. The motion of the "extra" mobile charges (existing yet from the forward bias condition) causes currents to flow until the excess carriers are recombined and a new steady state is reached. The division between region 3 and 4 is arbitrarily set. Time period 3 is called the "storage time" and persists from the time the voltage V is

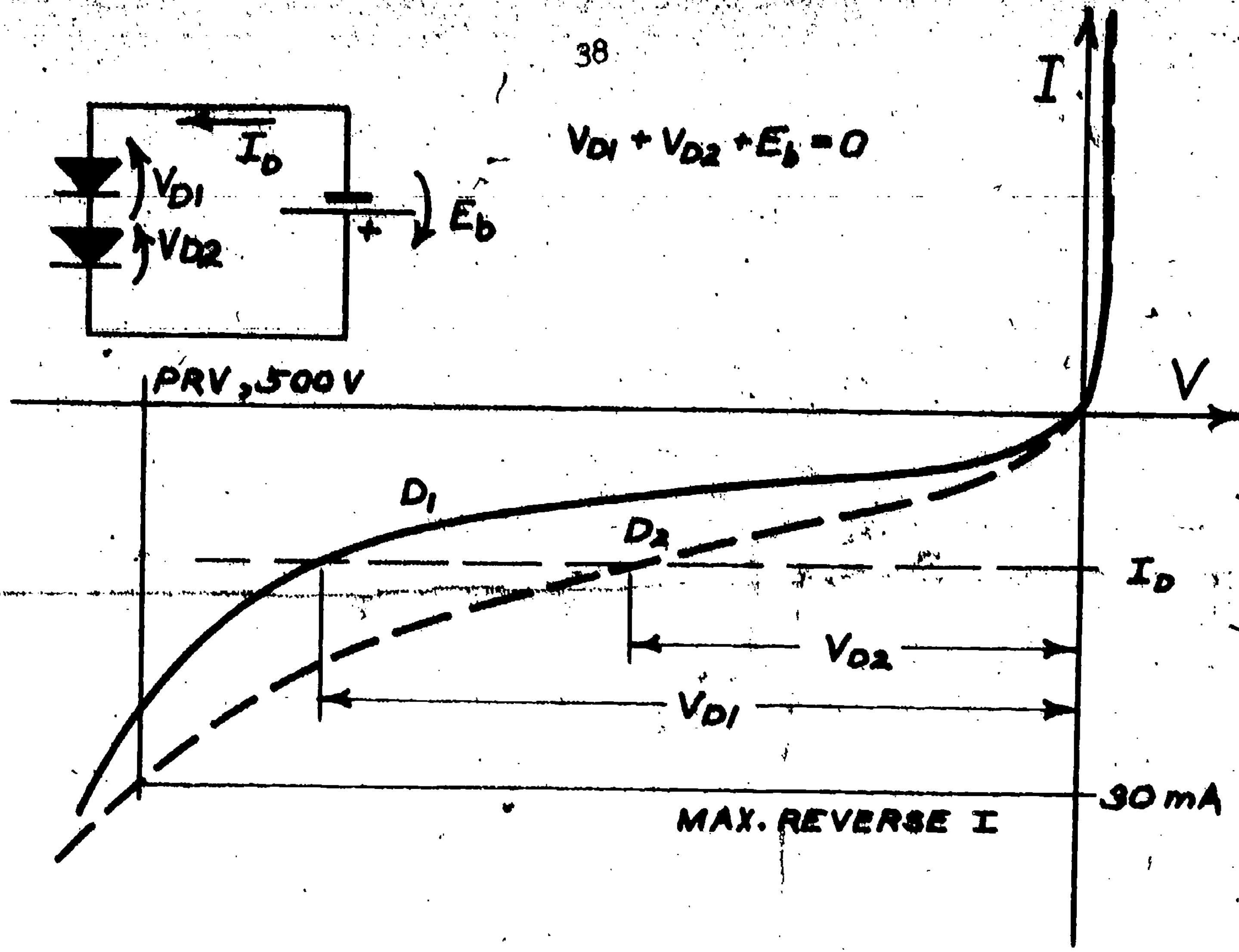
reversed until the current has decreased to 90 per cent of its peak reverse value (in accordance with standard definitions regarding pulse waveforms). The steady state reverse-bias conditions are reached "late" in period 4.

In summary, the differences between power and signal diodes are reasonably predictable in terms of the electron-hole model given the fact that power diodes operate at near maximum possible steady-state current densities and peak reverse voltages. Power diodes have significant "ohmic" IR drops which are conductivity modulated. Leakage currents are significant compared to I_o at maximum reverse voltages. I_o and the diode capacitance increase as the current rating and junction area increases. Finally, it is necessary to bear in mind that the "capacitances" and ohmic "resistances" are current, time, and voltage dependent and are not easily represented in an equivalent circuit.

Diodes in Series and Parallel

Sometimes it is desired to use semiconductor diodes in circuits where the voltages and currents are outside the rating range of commercially available diodes. In such cases, diodes may be connected in series to increase the voltage rating of the rectifier and/or parallel to increase the current handling capability of the rectifier.

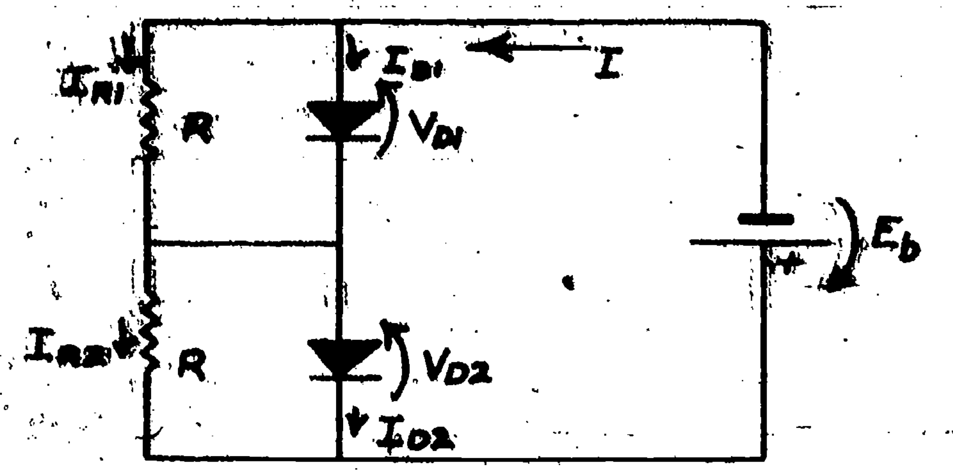
Consider two diodes (same rating, same type number) connected in series. The I-V characteristics of the two diodes will generally not be identical. In the forward direction, both diodes conduct the same amount of current, and some small different voltage appears across each diode. Clearly the current rating of the diode pair is the same as the current rating of one of the diodes. In the reverse direction, the same current flows in each diode, and each diode supports a different reverse voltage. The ratio of the voltages across the diodes will depend on how similar the diode characteristics are, as shown in Figure 1-17. The voltage rating of the diode pair must be larger than the rating of one diode, because part of the total applied voltage will appear across the other diode. However, the voltage rating of the diode pair must be less than twice the voltage rating of a single diode because the voltages do not divide evenly.



Note: The direction of V_{D1}, V_{D2} and I_D are chosen to agree with the conventional way of expressing the I-V characteristics of a diode and the ideal junction diode equation $I = I_o\ (e^{q\ V_D/kT} - 1)$.

Figure 1-17

Thus the series connection of diodes requires either careful matching of the diode I-V characteristics or some additional circuitry to make the voltage divide more evenly. A simple and commonly used solution to this problem is to connect a resistor across each diode. Although a different value of resistor could be placed across each diode to achieve some optimum voltage division, a more practical approach is to place the same value of resistor across each diode, eliminating the problem of matching resistor values to individual diode characteristics, and making replacement of defective units simple. Figure 1-18 shows the effect of placing resistors across the diodes of Figure 1-17.



$$V_{D1} + V_{D2} + E_b = 0$$

$$I = I_{R1} + I_{D1} = I_{R2} + I_{D2}$$

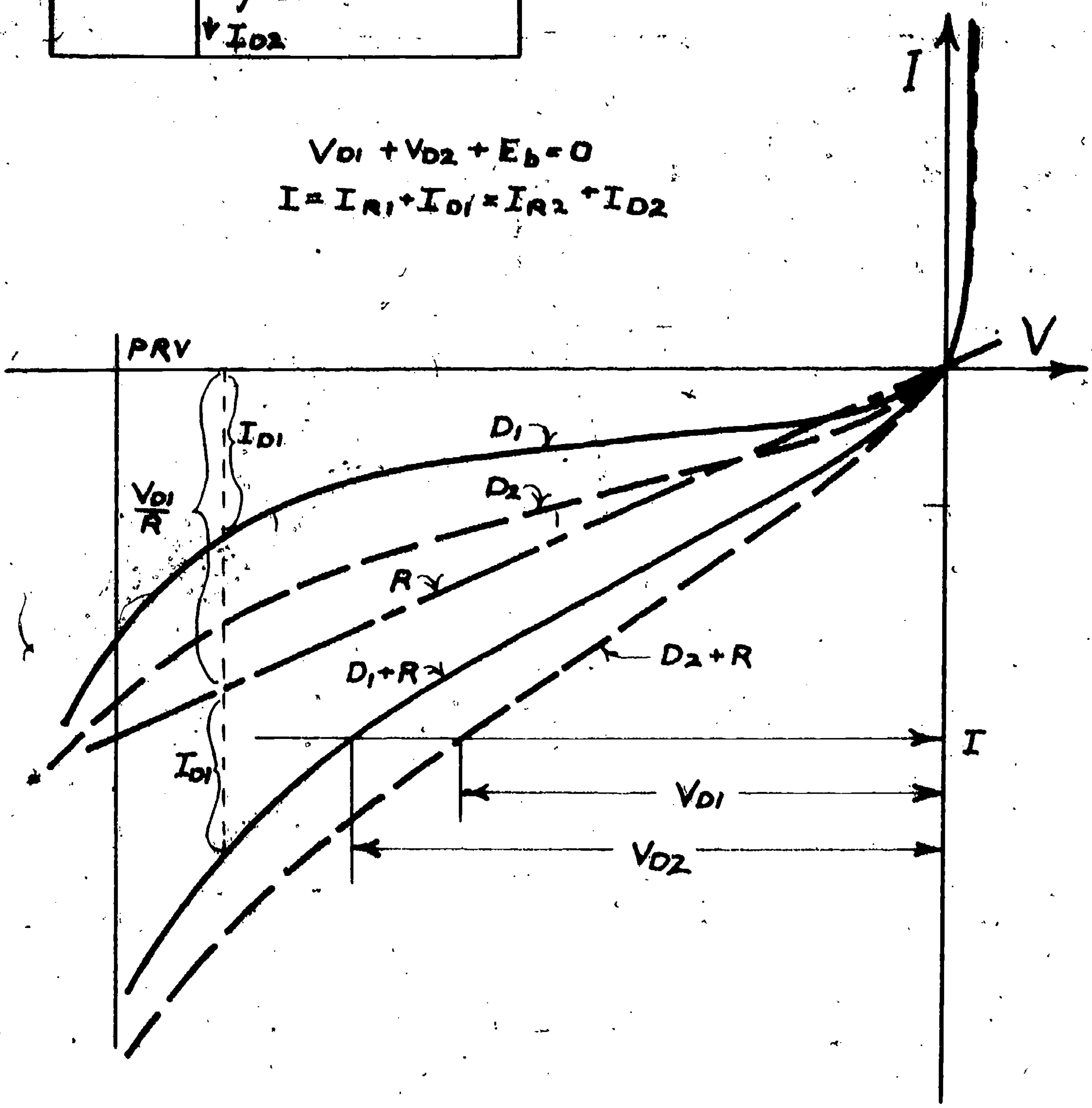


Figure 1 - 18

In addition to the steady state voltage distribution along a string of series-connected diodes, some provision must usually be made for transient voltage changes. Such transients might be caused by switching loads, lightning, or the initial application of voltage to the diode circuit. Assume a voltage transient of such a polarity as to increase the reverse bias is applied to a series string of diodes. At first, all of the diodes pass a changing current which depends upon the transient amplitude, the circuit load, and to some extent, the diodes. The change in current is relatively independent of the diodes due to junction capacitance and carrier storage effects which allow large currents to flow until a new steady-state distribution of carriers is achieved in the diode. Because of minute differences among the diodes, some diodes will approach a new steady-state distribution of carriers before other diodes. These "faster acting" diodes in the string then attempt to control and block the current associated with the voltage transient. Thus the "faster acting" diodes will receive reverse voltages greater than their fair share (the total applied voltage divided by the number of diodes). The voltage across one of the "faster acting" diodes may well exceed the peak reverse voltage rating of the diode and cause the diode to fail. If the diode fails by "shorting out", a common occurrence, the voltage on the other diodes of the string increases causing other diodes to fail until something fuses, disrupting the current.

Diode strings are commonly protected from voltage transients by shunting a capacitor around each diode (Figure 1-19). The capacitors can be thought of as bypassing abrupt voltage transients around the diode string, dividing the voltage transient equally among the diodes by a capacitance divider action, and limiting the rate of change of voltage across the "fact acting" diodes.

The magnitude of expected transients must be estimated in order to choose the capacitor voltage rating as well as the PRV of the diodes, and the entire circuit must be considered in choosing the value of capacitance necessary to limit the rate of change of voltage across the fastest acting diode.

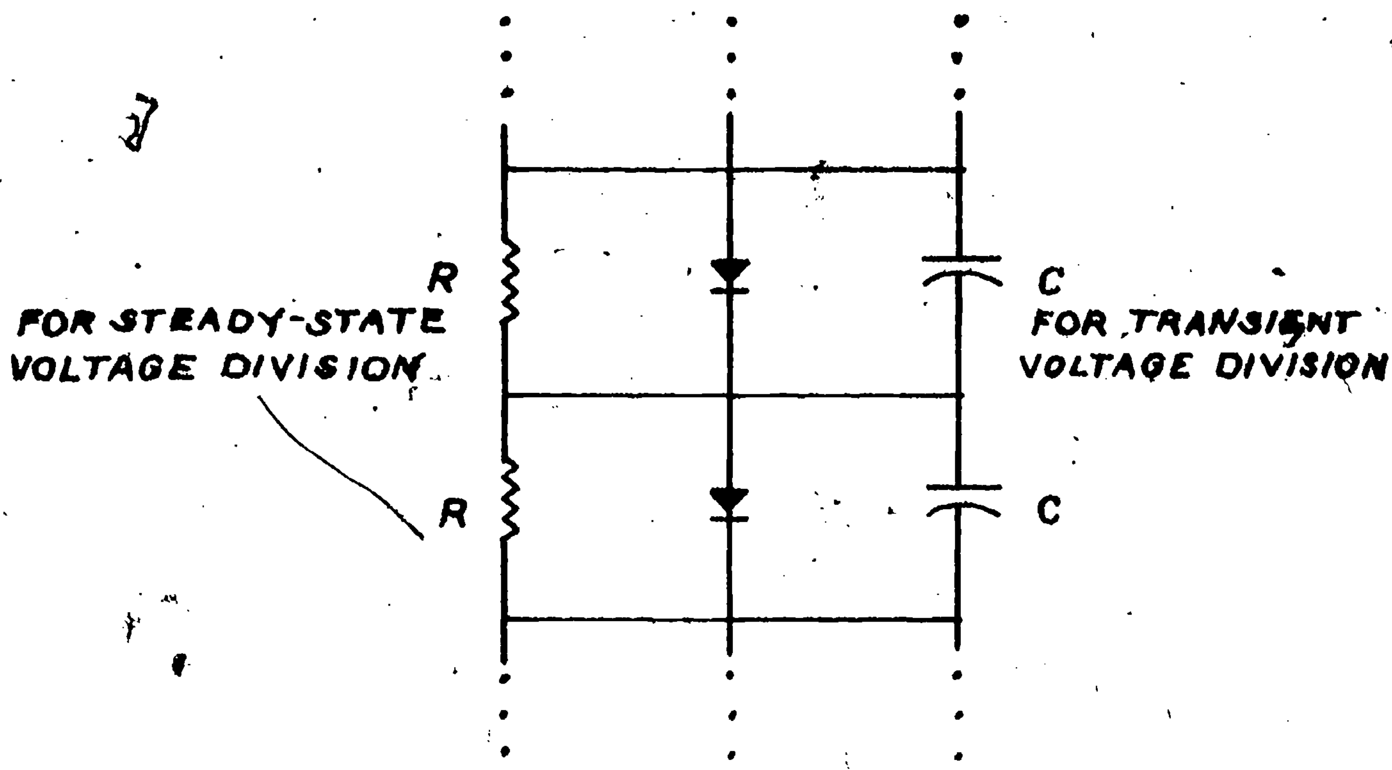


Figure 1 - 19

Diodes are frequently paralleled (Fig. 1-20) to achieve higher current capabilities than present ratings of single commerical diodes would permit. A small resistance is sometimes placed in series with each diode to assure the even division of the steady state current. Transient phenomena are not as important in the case of parallel diodes as in the case of series diodes. Each of the diodes can withstand the peak reverse voltage, and since each diode can withstand large transient surge currents, the transient case where some diodes conduct better than others is usually not of practical importance.

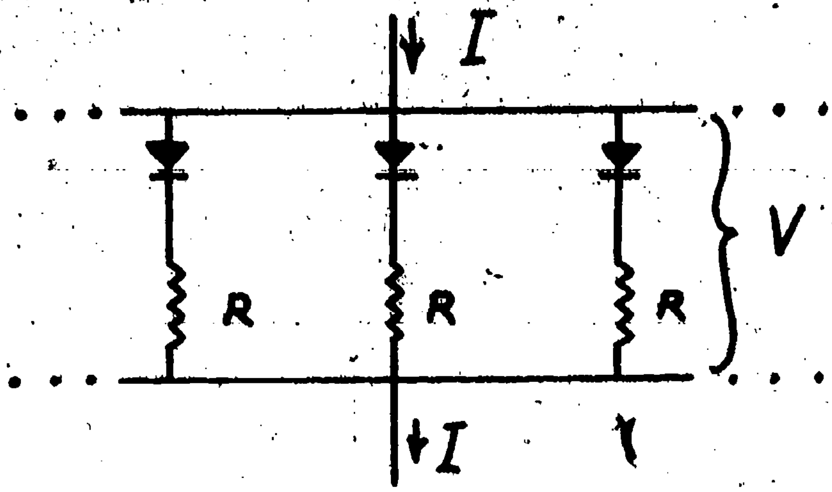


Figure 1 - 20

Modeling a Simple Diode Circuit

We next consider in some detail the analysis of a simple circuit that includes a diode. The emphasis in the analysis will be on the process of modeling a simplified circuit to approximate the real, physical circuit. In engineering, a rigorous solution of a problem in all possible detail and exactitude is never desired. The most elementary real problems would take months or would be unsolvable if the ultimate in accuracy and physical reasoning were required, where even conduction in a copper wire poses formidable problems of quantum mechanics, heat flow, surface phenomena, insulation properties, etc. The degree to which a real interconnection of electrical devices may be simplified (modeled) depends on the question the analysis hopes to answer. Also, since and approximate answer is desired rather than "THE TRUTH," it may be perfectly reasonable as well as desirable to change assumptions, simplifications, and models in the middle of a problem as illustrated by the following problem.

A 1N4590 diode is connected in series with a 2.4 ohm resistor and a 240 volt, 60 Hz source (Figure 1-21). It is given that:

(a) the internal impedance of the source is much less than 2.4 ohms,

(b) the resistor behaves as a pure resistance at 60 Hz (has negligible distributed inductance and capacitance),

(c) and the connecting wires and their physical connections have resistances negligibly small compared with 2.4 Ω.

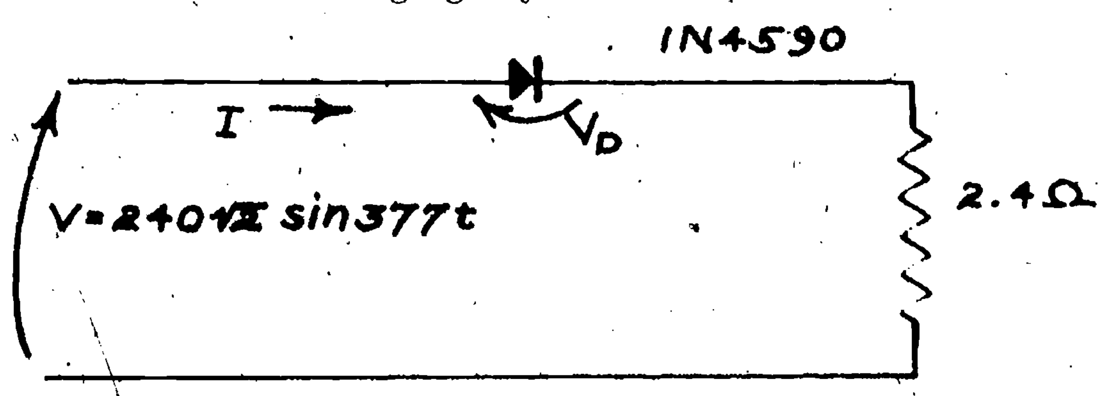


Figure 1 - 21

Question: How much power is dissipated by the diode? We want to choose an appropriate heat-sink on which to mount the diode.

Solution: The fact that the answer is to be used to choose a heat sink gives us some idea of the accuracy of the solution that will be required. Certainly an error of a factor of 2 would be too large, causing a severe "overdesign" in the choice of the heat sink. Similarly, provided some exotic application such as a space-satellite requiring a careful dissipation budget is not contemplated, a calculation to within 1% would be ridiculously accurate for the problem. The desired accuracy will depend on the cost of the heat-sink as a function of its cooling power, but a first-shot aim at an accuracy of 10-20% does not seem unreasonable.

In forming a plan of attack of the problem, we notice that neither the voltage across nor the current through the diode is sinusoidal. Therefore, the power dissipated will not be

$$P = V_{D\,rms} I_{rms} \cos\theta$$

which is valid only for sinusoidal waveforms. We must go back to the fundamental definition of electrical power, which is

$$P = \frac{1}{T}\int_0^T e\,i\,dt$$

where e = voltage between terminals = V_D

i = current = I.

T = time duration of one cycle

P = power dissipated in the device

Recall that by a simple change of variable,

$\theta = \omega t$ where T corresponds

to $\theta = 2\pi$ and ω is the angular frequency, yields the equally valid power formula

$$P = \frac{1}{2\pi}\int_0^{2\pi} e\,i\,d\theta\ .$$

Thus to solve the problem, we must find the voltage across the diode and the current through the diode as functions of time or θ, multiply the two functions together, and integrate over one cycle.

An examination of a manufacturer's data sheet concerning a 1N4590 diode yields the following data and specifications:

Max repetitive peak reverse volts	-400 volts
Max peak reverse volts for one half wave, 60 Hz sinusoidal pulse	-525 volts
Max allowable blocking direct-voltage	400 volts
Max average reverse current at PRV (repetitive), 150A forward current, junction temperature at 110°C	-9.0 m A
Max forward voltage drop, 150 amps average current, 110°C junction temp	+1.35 volts
Max forward full cycle average current	150 A
Max 1/2 cycle, 60 Hz, peak surge current	3000 A

(Westinghouse data sheet 54-166)

First we note that no data are given concerning junction capacitance and carrier storage times. Since 60 Hz is mentioned in the given specifications, and the circuit we are considering does not seem to be an "exotic" application of a diode, a reasonable inference is that the diode capacitive effects are not significant on a time scale of 60 Hz ($\sim$17 m sec). Although we could search further for information from people experienced in working such problems, the manufacturer, or advanced texts and publications, we choose here to assume that such capacitances will be negligible (on the hint that no specifications or data are mentioned on the data sheet) subject to laboratory verification. We shall have to calculate voltage and current waveforms to solve the problem anyway. We shall set up an experiment (after the calculation shows us what should be examined), and if the observed waveforms drastically differ from the calculated waveform in such a way that could be attributed to diode capacitances, we shall have to "sophisticate" our calculations.

We do not have the V-I characteristic of the diode. We could request additional information from a manufacturer, (We would receive families of V-I characteristics as functions of temperature. We might also receive tolerances for the V-I characteristics.) However, we may be able to arrive at some reasonable conclusions using our knowledge of the diode in general and the given diode specifications.

Consider the half cycle in which the diode is forward biased and conducting. A maximum voltage of 1.35 v appears across the diode which is small compared to the source voltage $240\sqrt{2}$. Therefore, to a reasonably good approximation, the diode voltage is negligible and $I = \frac{V}{R}$

$$I = 100\sqrt{2} \sin \omega t \quad \text{from } \omega t = 0° \text{ to } \omega t = 180°$$

During the next half cycle, the maximum reverse current is 9 m A. IR can be at most .009 x 2.4 volts which is small compared to $240\sqrt{2}$ volts. Therefore the diode voltage will approximately equal the source voltage for the half cycle the diode is reverse biased.

$$V_D = 240\sqrt{2} \sin \omega t \qquad \omega t = 180° \text{ to } \omega t = 360°$$

Now if we can make some reasonable approximations about the forward voltage drop (with our known forward current) and the reverse current (with known reverse voltage) waveforms, we shall be able to calculate the power dissipated in the diode. However, we note a glaring logical inconsistency. Our perfectly valid approximations concerning negligible forward voltage drop and negligible reverse current are equivalent to replacing the diode by an ideal switch (Fig. 1 - 22).

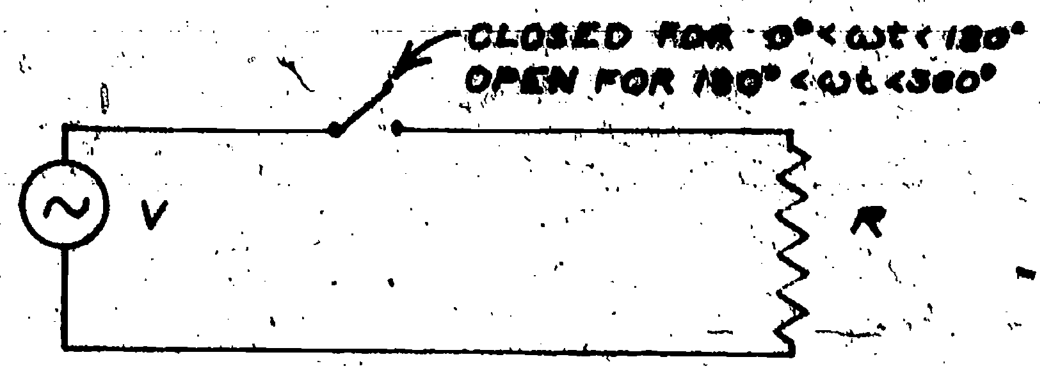


Figure 1 - 22

The switch is a lossless element, and we now propose to calculate the power lost in the switch! Of course, the argument is not inconsistent provided we distinguish between negligible quantities and zero. The principle is emphasized by considering a trivial problem.

Digression

Consider a simple series circuit consisting of an A-C source, a resistor, and an ideal inductor (Fig. 1-23).

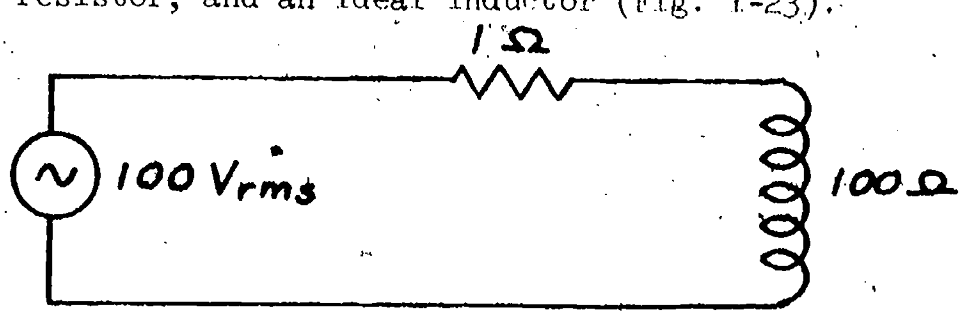


Figure 1 - 23

What is the circuit current, and what power is dissipated in the circuit?

This simple problem can be worked by inspection to better than 1% accuracy if one realizes that the 1 ohm resistor impedance is negligible compared to the 100 ohm inductor impedance

$$I \simeq \frac{100 \text{ volts}}{100 \text{ ohms}} = 1 \text{ amp}$$

$$P = I^2R \simeq 1 \text{ watt}$$

Despite the fact that the resistor has negligible effect on determining the circuit current and can be neglected in determining I, the resistance is not zero and determines the power dissipated in the circuit. Similarly in the diode circuit, the negligible forward voltage drop and reverse current are not zero and power is dissipated.

Next we consider plausible assumptions we might make regarding forward voltage and reverse current. From the review of a diode, we know that as the forward current increases, the forward voltage at first rises rapidly and then more slowly according to $I = I_o(e^{qV/kT}-1)$, and continues to rise faster than the ideal formula would predict due to bulk resistance effects. A sketch of the forward voltage might look like:

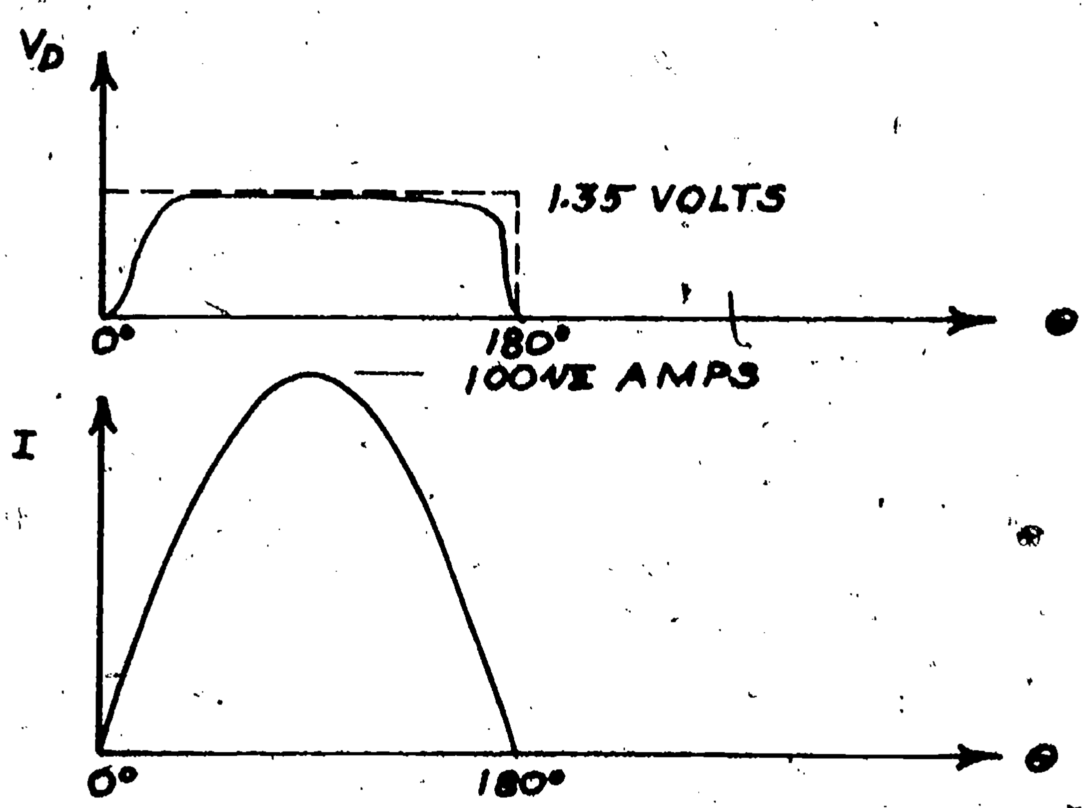


Figure 1 - 24

V_D might not reach the maximum value of 1.35 volts due to variations in diodes and since the maximum current is not being drawn. However, we include a small safety factor in the calculation and assume that 1.35 volts will be the peak forward voltage. Our knowledge of the diode characteristic tells us that the voltage will be near the peak value throughout most of the cycle. Since we are calculating $\int V_D \, I \, d\theta$, we also note that the value of

V_D near 0° and 180° will be less important than near 90° since I will be small at 0°, and 180°. Therefore a reasonable approximation that would greatly simplify the problem and would still seem to provide reasonable accuracy would be to assume V_D = constant = 1.35 volts for $0^\circ < \theta < 180^\circ$. Errors in such an assumption would be more sensitive to a poor choice of V_D max than to the waveshapes near the beginning and end of the half cycle. Similar arguments involving the leakage resistance of the diode apply for the reverse current (Fig. 1 - 25), and we approximate the reverse current I_o = constant = -9 m A.

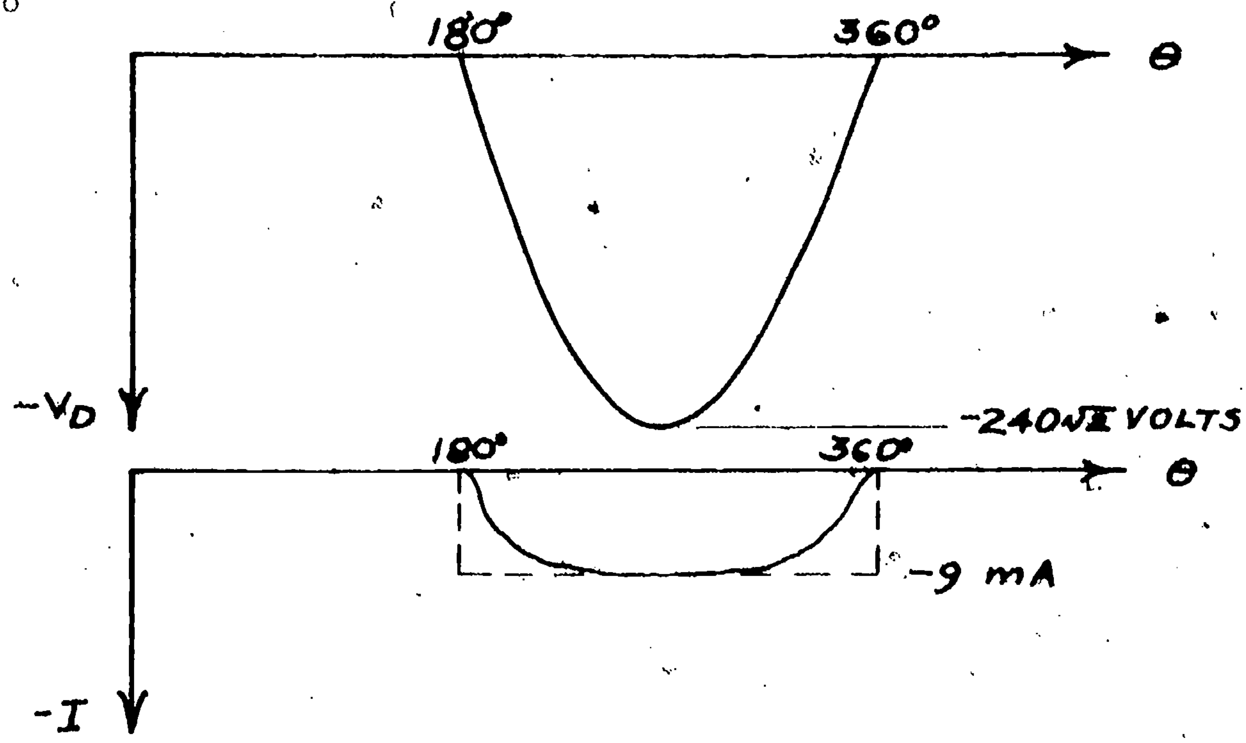


Figure 1 - 25

We now perform the integration, breaking the integral into two parts.

$$P = \frac{1}{2\pi}\int_0^{2\pi} V_D I \, d\theta = \frac{1}{2\pi}\left[\int_0^{\pi} V_D I \, d\theta + \int_{\pi}^{2\pi} V_D I \, d\theta\right]$$

$$P = \frac{1}{2\pi}\left[\int_0^{\pi}(1.35)(100\sqrt{2})\sin\theta \, d\theta + \int_{\pi}^{2\pi}(-0.009)(240\sqrt{2})\sin\theta \, d\theta\right]$$

$$P = \frac{1}{2\pi}\left[135\sqrt{2}\int_0^{\pi} \sin\theta\, d\theta - 2.16\sqrt{2}\int_{\pi}^{2\pi} \sin\theta\, d\theta\right.$$

$$P = \frac{1}{2\pi}\left[135\sqrt{2}\left(-\cos\theta\Big|_0^{\pi}\right. - 2.16\left(-\cos\theta\Big|_{\pi}^{2\pi}\right.\right]$$

Note that we can neglect the power dissipated in the $180^\circ < \theta < 360^\circ$ half cycle compared to the power dissipated in the $0^\circ < \theta < 180^\circ$ half cycle. Thus most of the power is dissipated when the diode is forward biased.

$$P \simeq \frac{135\sqrt{2}\,(2)}{2\pi} = 61 \text{ watts}$$

Although the approximations we have made seem reasonable, we have not used the actual diode characteristic. We have a qualitative feeling that the answer is sufficiently accurate (20%), but there is no way to tell for sure without further data. A laboratory experiment in which the power dissipated, and the critical (in the light of the calculation) forward voltage drop across the diode waveform are measured is in order.

The solving of the simple diode-resistor circuit have brought out some important points regarding modeling.

(a) It is absolutely essential to have some idea of the desired accuracy of a solution. The time and effort required to solve a problem generally increase drastically as the required accuracy increases.

(b) The primary motivation in making assumptions with respect to a given problem is to simplify the path to the solution, saving time and effort. Assumptions are made on the bases of given data (and associated inferences such as the absence of diode capacitance data in the example), similarity of the problem to other successfully solved problems which include assumptions (diode capacitance calculations are <u>almost</u> never included in 60 Hz problems), and observing the relative magnitudes of variables as they are calculated during the solution of the problem.

(c) It is not necessary to stick to the same set of assumptions throughout a problem. It is important to note the reason each assumption is made so that assumptions may be changed at appropriate steps in the solution.

(d) In any real situation, the problem is not solved until the validity of each assumption is demonstrated or experimentally checked.

(e) As the problem is modeled and solved, the critical quanities to be measured in the laboratory become apparent. Without some careful consideration beforehand, the laboratory test may be aimless or lead to the vague conclusion "it doesn't work" instead of illuminating a weak point in the analysis.

As a final comment, sometimes it is necessary to make an assumption solely for the purpose of simplifying the problem and without any real basis. Such assumptions may lead to a better understanding of the problem and assist in finding a path toward the solution. However, each such assumption must be verified and checked and possibly refined or else the answer is not really a solution to the problem. An "it might be" answer is not sufficient in any practical problem.

Exercises

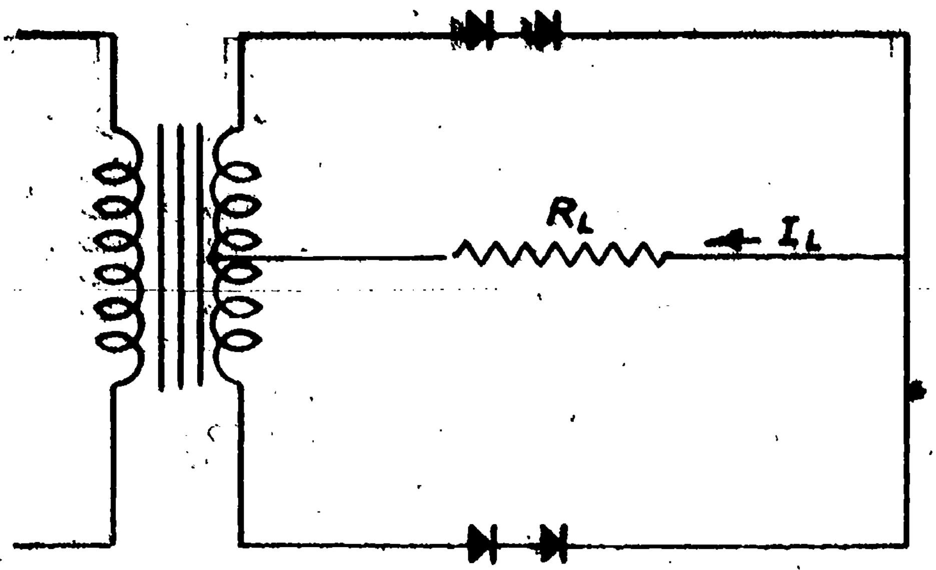
Ex. 1 Plot the resistivity of a silicon crystal as a function of doping density N_D for the range of $N_D = 0$ to $N_D = 10^{20}$ atoms/cm^3. It is given that $\mu_n = 1200$ cm^2/volt sec. and $\mu_p = 250$ cm^2/volt sec.

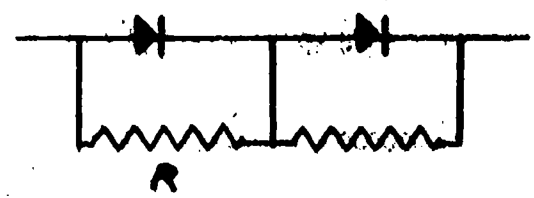
Ex. 2 Assuming a very abrupt junction (the transition from n to p type material occurs in a length of the crystal that is negligible) in a silicon diode. N_A in the p-type material is 2×10^{17} atoms/cm^3 and N_D in the n-type material is 4×10^{16} atoms/cm^3. Plot the value of the depletion layer capacitance Q/ψ as a function of ψ, the electrostatic potential across the depletion region. Also find the incremental capacitance $\partial Q/\partial \psi$ as a function of ψ.

Problems

Problem 1.

Several high power diodes are to be used in a center-tapped transformer rectifier circuit.





Transformer output 710 v_{rms} = phase voltage.

$(\frac{710}{2}$ = center tap voltage)

Diode characteristics

PRV 600 volts

Max forward current 70 $amps_{rms}$

Max junction temp 190°C

Max reverse leakage current at max junction temperature 30 m A

Additional data

Of the available supply of diodes, the reverse current (T_j = 190°C) varies from 25-30 m A.

In the forward direction, the diodes behave as ideal p-n junctions in series with 0.1 ohm resistors.

(a) Since the applied reverse voltage exceeds the PRV for a diode, it is necessary to use two diodes in series. In order that matched diodes do not have to be selected, a resistor is shunted across each diode. Estimate a maximum R such that the peak voltage across any diode is less than 550 volts.

(b) Estimate the power dissipated by each diode.

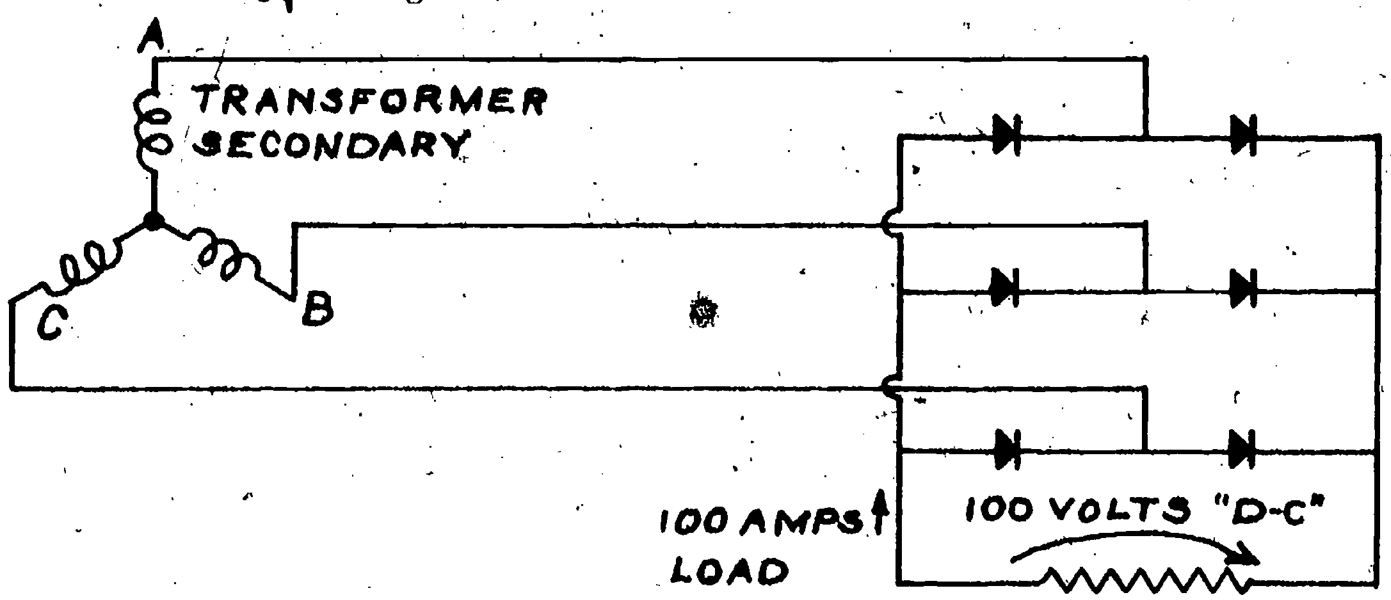
(c) If $I_L = 50$ amps$_{rms}$ what is the value of R_L (engineering accuracy)?

(d) What would be the <u>average</u> value of I_L?

(e) Will the transformer winding resistance and leakage inductance be negligible?

Problem 2

Consider a 3φ bridge rectifier circuit.



The circuit is to deliver 100 amps direct current to a resistive load of 1 ohm. You are asked to find the minimum RMS current rating of the diodes (allow a 20% safety factor) and the minimum voltage rating of the diodes (allow a safety factor of 2.2 x peak voltage in case of transients).

In addition, heat-sink data (power dissipation in the diodes) is frequently given in terms of average current--so calculate the average current through each diode.

Finally, specify the KVA rating of the transformer (a single 3φ transformer).

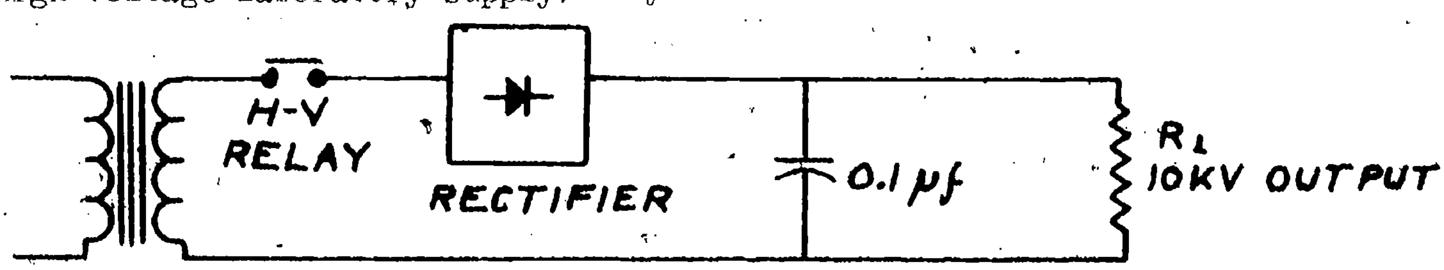
Lab Problem 1

"Semiconductor Diodes" or "The Bigger, the Better"?

You will be assigned two different diodes. One diode will be an instrument diode having a voltage rating of less than 100 volts and a current rating of a few milliamperes. The other diode will be a "power" diode having a voltage rating of over 100 volts and a current rating of several amperes. You are to investigate the electrical differences between the two diodes. You will not have time to investigate all possible differences, so choose a combination of experiments you consider important and interesting.

Lab Problem 2

Design the diode-resistor-capacitor network required for the following high-voltage laboratory supply.



The supply is to be rated at 10,000 volts, 100 m A. It is to be used with various load resistances. The rectifier will consist of a series string of diodes and their associated circuitry (resistors and capacitors). The diodes must be protected from transients caused by opening the high voltage relay. The relay can close during any part of a cycle. You may select the diode type from those available in the laboratory (a list of diode numbers is available), however, only 240 volts rms and a limited number of diodes is available (you may not have enough diodes to construct the entire string).

References

Chapter 1

General References

1.) Millman, M & Halkias, C. C., Electronic Devices and Circuits, McGraw-Hill, New York, 1967.
Excellent review of semiconductor physics and the properties of signal diodes.

2.) Gray, Paul E. and Searle, Campbell L., Electronic Prinicples, Wiley, New York, 1969.
Excellent detailed review of the hole-electron model applied to diodes.

3.) Seymour, J., Semiconductor Devices in Power Engineering, Pitman, London, 1968.
Brief resume of semiconductor physics, excellent description of power diode construction, characteristics, and practical uses.

Solid State Physics

4.) Sproul, R. L., Modern Physics, 2nd Ed., Wiley, New York, 1963.
Background of bond theory of solids, contact potentials, and ideal diodes in an especially readable form for engineers.

5.) Adler, R. B. & Longini, R. L., Introduction to Semiconductor Physics, Wiley, New York, 1963.
The foundation and elementary use of the electron-hole model is explained on an undergraduate level.

6.) Gray, P. E., De Witt, D., Boothroyd, A. R., Gibbons, J. F., Physical Electronics and Circuit Models of Transistors, Wiley, New York, 1963.
A good derivation of signal diode characteristics as the basis of the electron-hole model is presented.

7.) Kittel, C., Introduction to Solid State Physics, 2nd Ed., Wiley, New York, 1956.
This reference is particularly useful in understanding the statistical viewpoint of the band theory of solids and the model of electron conduction in metals.

Chapter 2 Thyristors, or Silicon Controlled Rectifiers

Chapter Contents

The primary purpose of this chapter is to acquaint the reader with the static and dynamic characteristics of the thyristor or SCR. The SCR is modeled as a series of diodes to gain insight into the static V-I characteristic. The triggering characteristics are explained in terms of a two transistor model of the SCR. The transient turn on-turn off characteristics are then rationalized in terms of the combination of the two models. Common methods of triggering are considered and an example of a trigger circuit using a unijunction transistor is presented. Finally, some D.C. turn-off circuits are presented and a particular turn-off circuit is analyzed as an example.

The Thyristor or SCR

The thyristor or SCR is a three terminal semiconductor device. The name "thyristor" comes from the fact that the thyristor is a transistor-like device (having three leads and made of a semiconductor) having properties similar to that of a gas thyratron electron tube (which can be triggered to change from an essentially open circuit device to an essentially short circuit device in a matter of microseconds). SCR is the most commonly used name of the device, standing for silicon (or semiconductor) controlled rectifier. The term SCR is also a descriptive term since SCR's are made of the semiconducting material silicon, and SCR's possesses rectifier-like properties which can be controlled to a certain extent using the third terminal. The electronic schematic symbol for an SCR and the names of the terminals are shown in figure 2-1 along with the V-I characteristic of the anode-cathode terminals. The V-I characteristic is divided into three regions for future reference as the device operation is modeled.

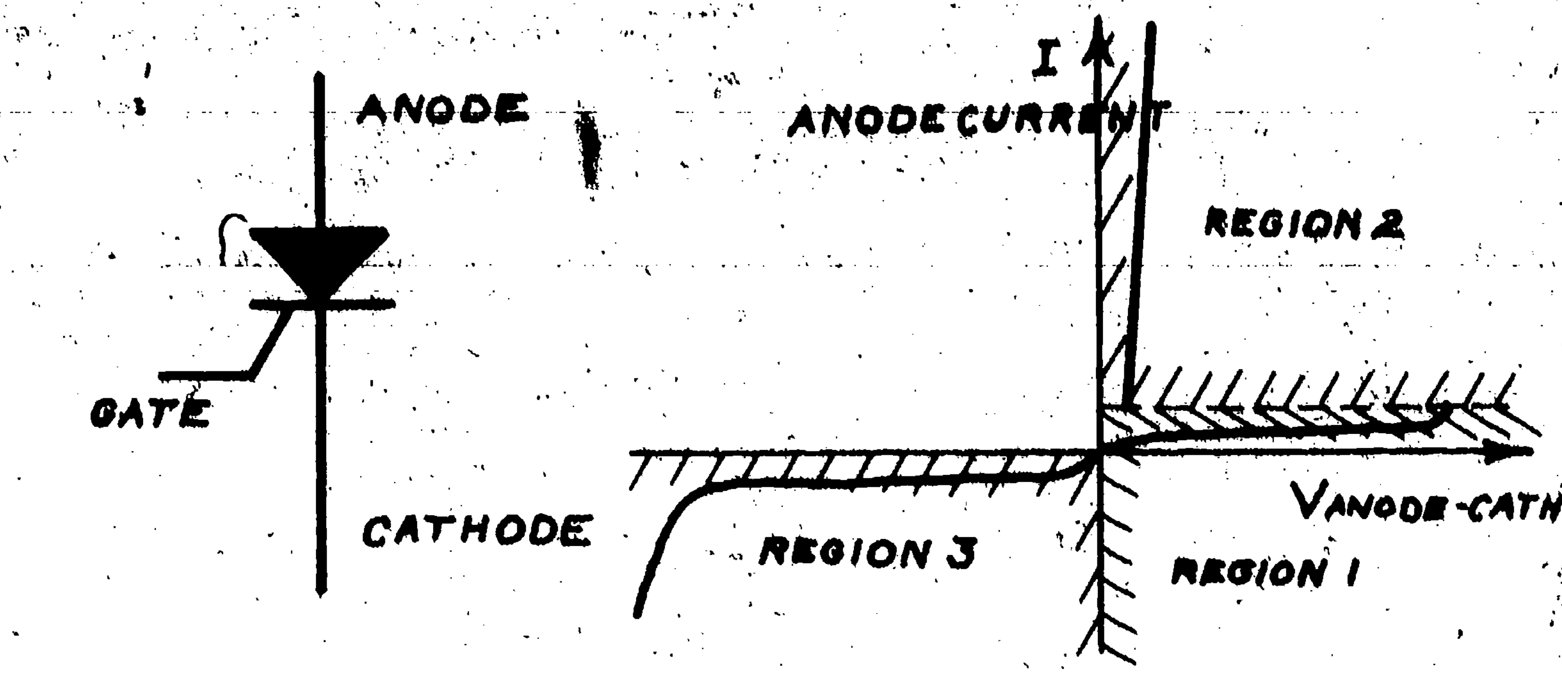


Figure 2 - 1

Next we consider a simple circuit to show the basic usefulness of the SCR (Fig. 2 - 2a).

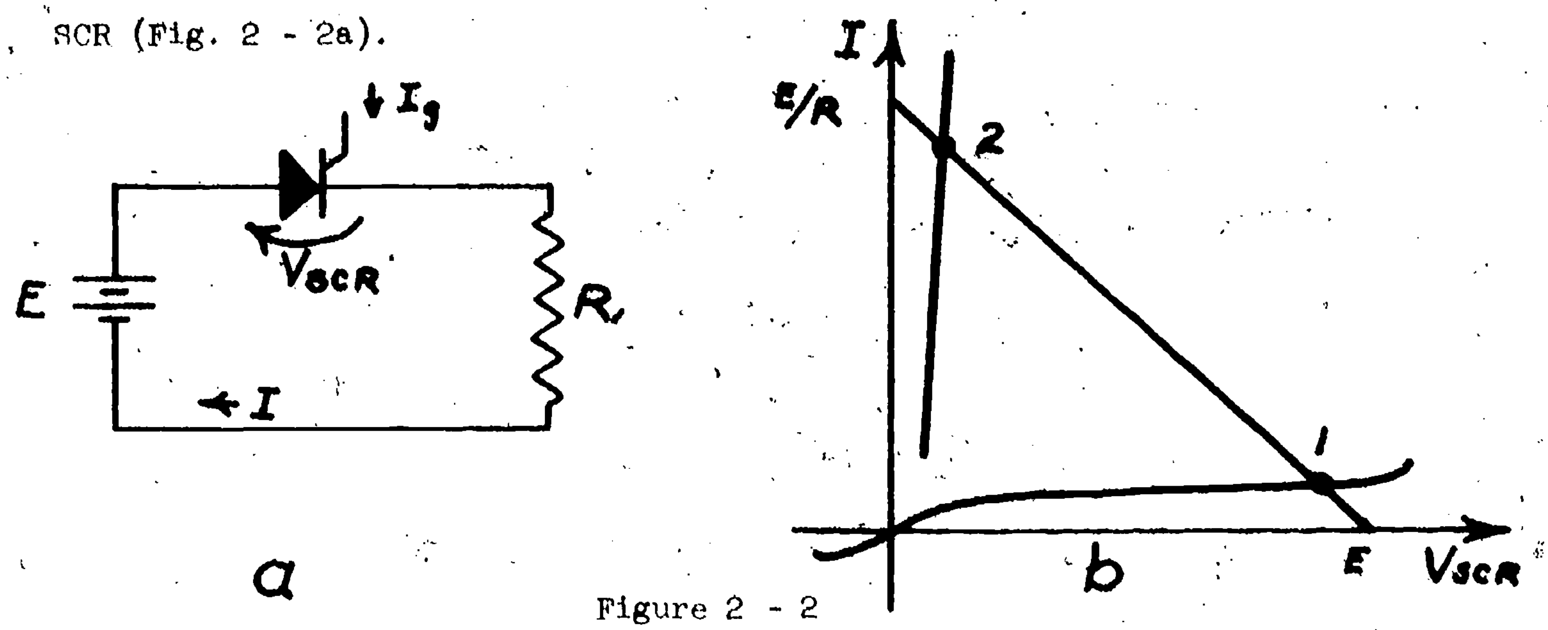


Figure 2 - 2

Figure 2 - 2b shows the load characteristic superimposed on the SCR V-I characteristic. When the circuit is first connected, the SCR is in the "forward blocking" state, the current I has some small (compared to $E/_R$) value, and the circuit state is represented by position 1 in figure 2 - 2b. If a gate current I_g, which is large enough (although possibly hundreds of times smaller than the SCR rated forward current) and of sufficiently long duration (usually in terms of microseconds) is applied to the SCR, the SCR will

be "triggered" into the "forward conducting" state represented by position 2 in figure 2 - 2b. In an SCR, the gate no longer has any control over the circuit operation, and the SCR will remain in the forward conducting state until the battery is disconnected.

We can summarize the gross (somewhat oversimplified) characteristics of the SCR.

(a) The SCR blocks current in the reverse direction (region 3, Fig. 2-1).

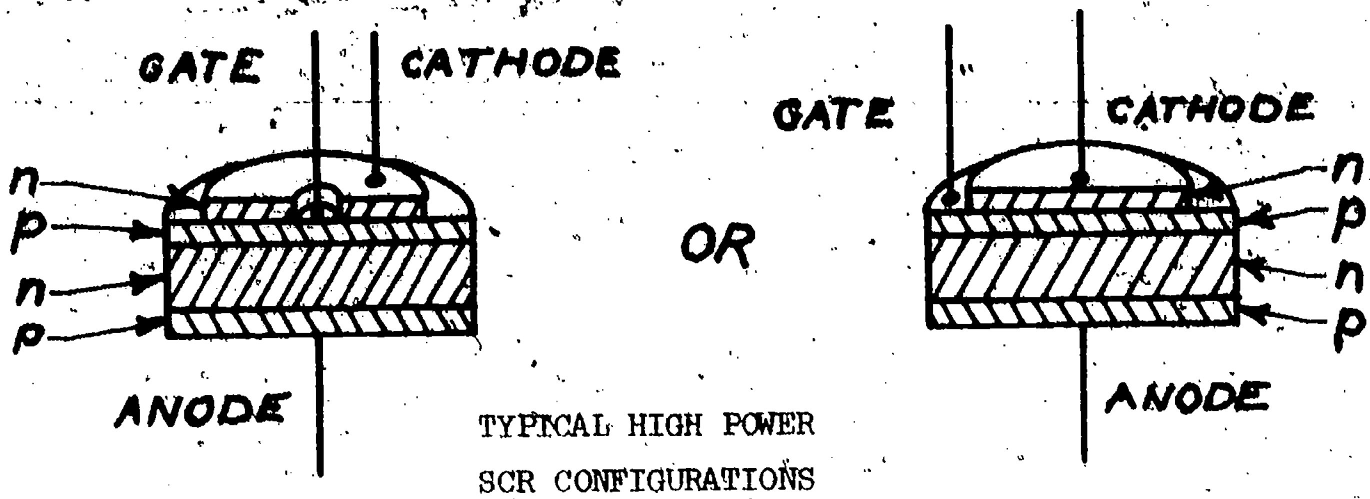
(b) The SCR acts as a switch that can be turned on by the gate in the forward direction (Fig. 2-2).

(c) Once the SCR is turned on (forward conducting), it cannot be turned off unless some external circuit reduces the forward current to zero.

An examination of the ratings of commercially available SCR's shows blocking voltages in excess of a thousand volts and average currents on the order of hundreds of amps. Clearly, SCR's are capable of controlling large amounts of power. The additional facts that SCR's are fast acting (compared to mechanical switches), small and rugged (compared to gas tubes), and do not require power to keep them "turned on" (as transistors require) makes SCR's particularly attractive in the area of power control. There are other related semiconductor devices such as gate controlled switches that can be turned off as well as on by the gate, triacs which can be triggered on in both forward and reverse directions and many more semiconductor switching devices. At present, these devices do not have the power handling capability of SCR's, but are valuable devices at lower power levels. Since these devices can be modeled with relatively simple extensions of the SCR model, we will not consider them in any detail in this book.

SCR Construction

The SCR is made with a single crystal of silicon having four layers of differently doped crystal (Fig. 2-3). The reason for partially cutting away one layer to make the gate connection in the center will become clear when we discuss turn-on transients.



TYPICAL HIGH POWER
SCR CONFIGURATIONS

Figure 2 - 3

Starting from the anode, the anode lead is connected to a fairly heavily doped (on the order of 10^{19} acceptor atoms/cc) p-type material. Next is a comparatively thick, lightly doped (on the order of 10^{14} donor atoms/cc) layer of n-type material. The next layer, (to which the gate is connected) is p-type material, doped with an intermediate density of impurity atoms. Finally, the last layer, to which the cathode lead is connected, is a heavily doped n-type material. The thickness of the crystal is exaggerated in figure 2-3 to show the n and p-type layers. The crystal is normally a thin disk or rectangle which is soldered or tightly pressed (by a powerful spring) against the SCR base in order to encourage the conduction of heat generated in the crystal to the outside heat-sink (Fig. 2-4).

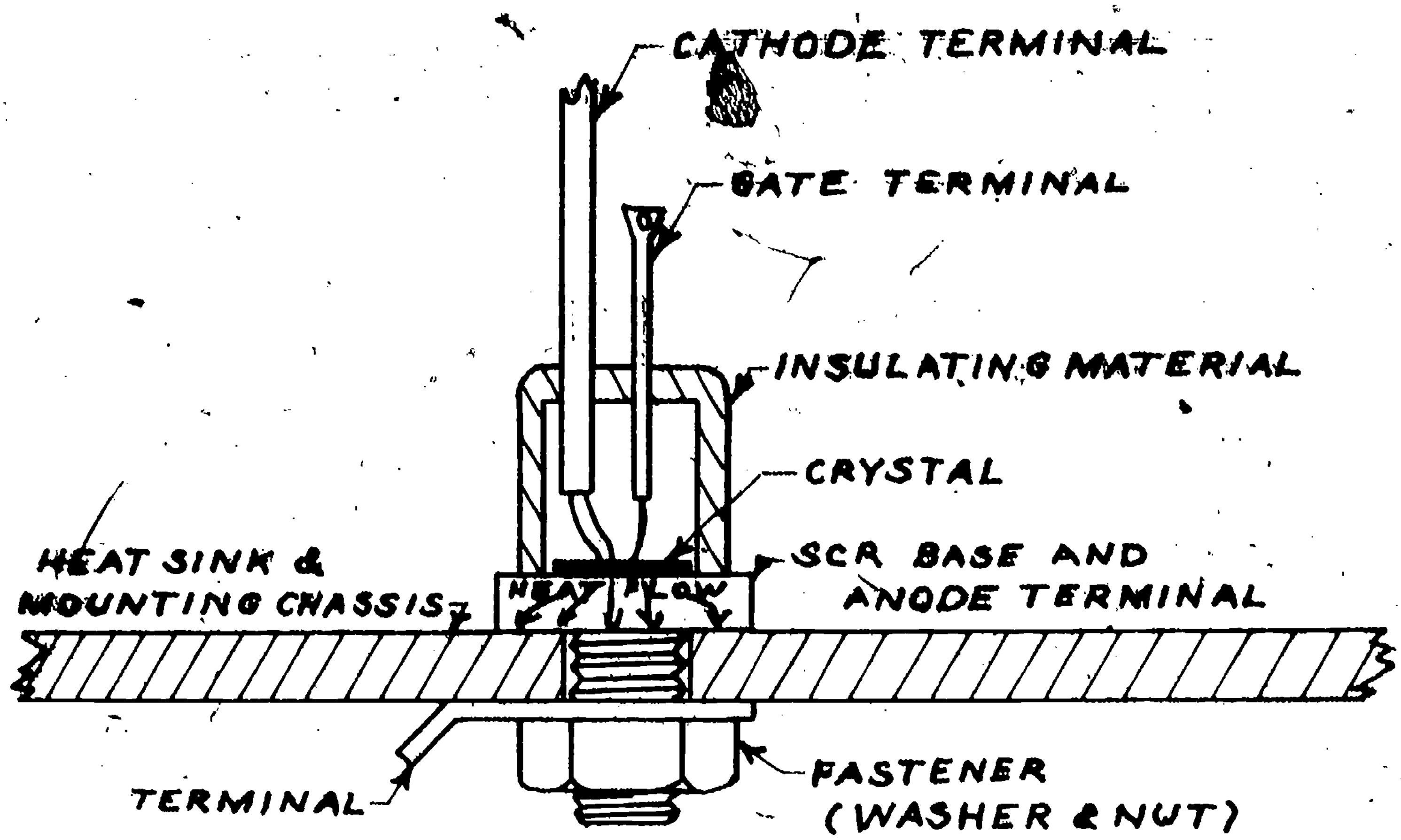


Figure 2 - 4

The SCR as a Series of Diodes

In order to gain some insight into the behavior of the SCR, consider a filament of the crystal as modeled in figure 2-5. Then, as previously in the case of the diode, after understanding the operation of a unit area of the crystal, we need only multiply the appropriate quantities (such as current capacity and capacitive effects) proportional to the area of the crystal.

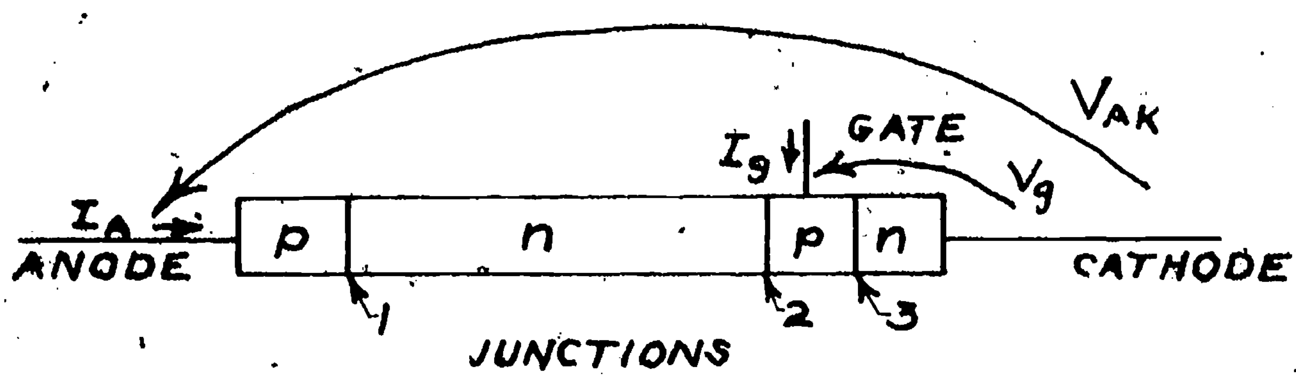


Figure 2 - 5

Figure 2 - 5 shows the commonly used polarity conventions for voltages and currents associated with the SCR and associates, for purposes of discussion, numbers with each of the three p-n junctions in the SCR.

As a first approximation, we may try to think of each p-n junction as a diode (Fig. 2-6). Such a model would rationally explain the SCR forward and reverse blocking characteristics (regions 1 and 3 respectively of Fig. 2 - 1).

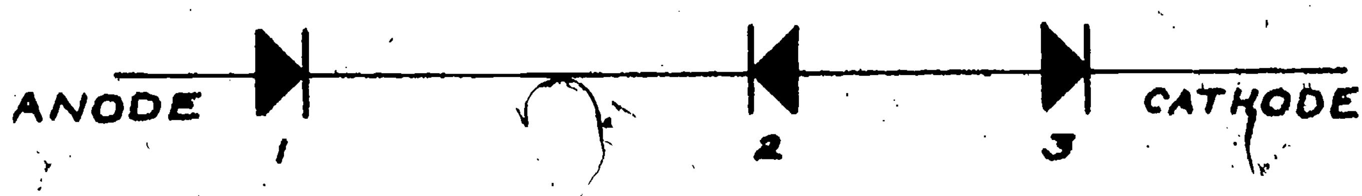


Figure 2 - 6

When the anode potential is positive with respect to the cathode, junction 2 blocks the current (forward blocking, region 1 of Fig. 2 - 1). When the cathode potential is positive with respect to the anode, junctions 1 and 3 block the current (reverse blocking, region 3 of Fig. 2-1).

In order to further improve our understanding of SCR operation, we shall plot the potential as a function of distance in the filament. Such a plot will require the charge density and resulting $\bar{\mathcal{E}}$ -field just as in the case of the diode. To simplify the plots, the metal-semiconductor junctions at the device leads will be ignored. We first attempt to plot the potential distribution for the case of thermodynamic equilibrium.

Figure 2 - 7a forms the first basis for the qualitative plot. We have a "rough idea" of the relative doping densities, and we know the n-type layer between junctions 1 and 2 is thicker than the other layers. The doping density is plotted on a logarithmic scale. Next (2 - 7b) the charge density $\rho(x)$ is estimated on a linear scale. Although the lengths of the depletion regions are unknown, the $\rho(x)$ times Δx areas (on each side of a junction) must be equal

as in the case of the diode. Note that considerable distortion of the charge density amplitudes and depletion layer widths is necessary to display these parameters on a single graph, since the charge densities may vary by more than five orders of magnitude. Using the "square" charge density assumption, we apply Gauss' law to find the $\vec{\mathcal{E}}$-field, $\mathcal{E}_x = \int \rho \, dx$ on a per unit area basis. The $\mathcal{E}_x$ values are triangles (Fig. 2 - 7c) just as in the diode case. Integrating to find the potential, $-\int \mathcal{E}_x dx = \psi$, we get a potential as a function of distance plot consisting of sections of parabolas and having an inflection point at each junction (Fig. 2 - 7d).

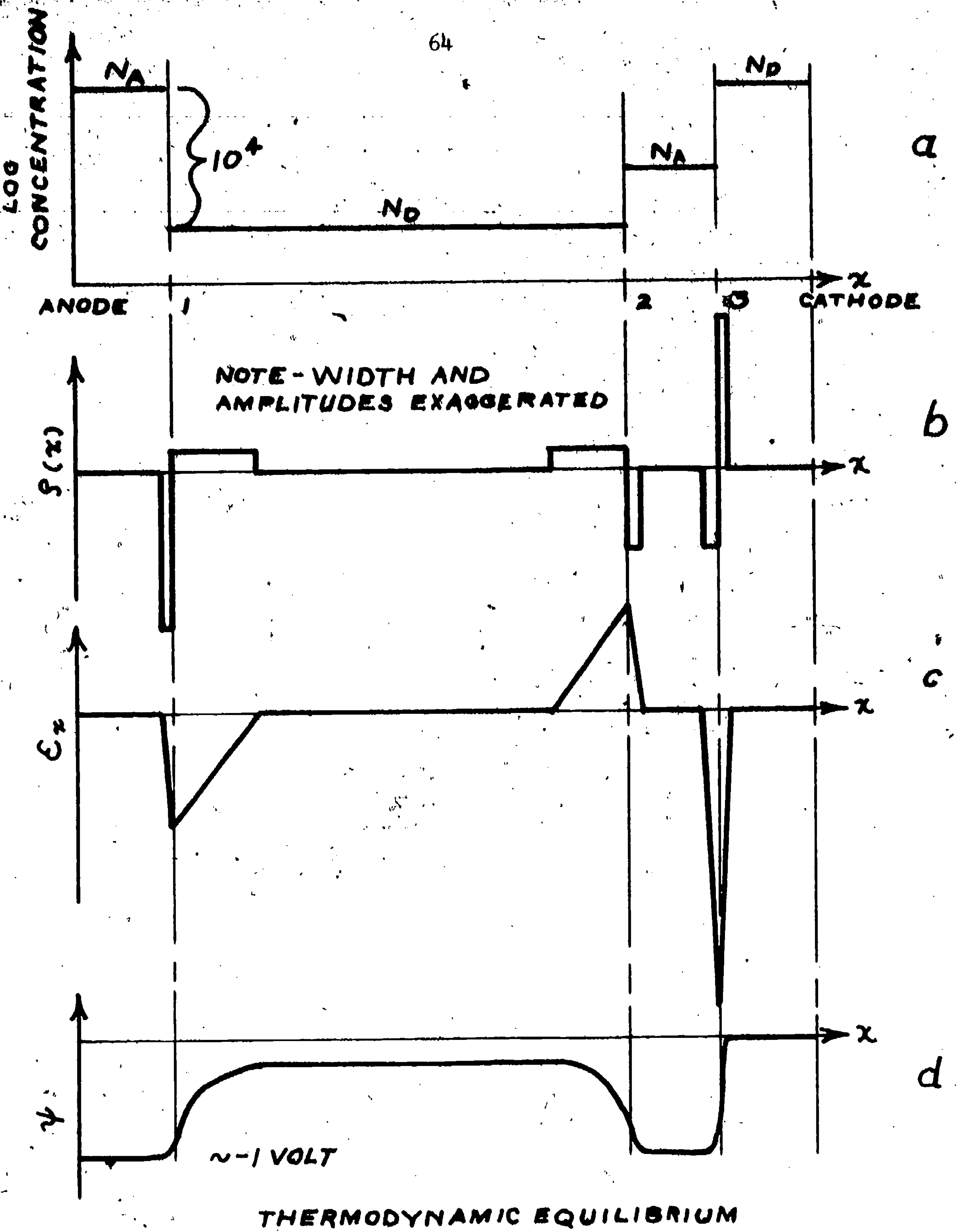

LOG CONCENTRATION
N_A
10^4
N_D
N_A
N_D
x
ANODE
1
2
3
CATHODE
a
NOTE - WIDTH AND AMPLITUDES EXAGGERATED
ρ(x)
x
b
$\mathcal{E}_x$
x
c
ψ
x
~-1 VOLT
d
THERMODYNAMIC EQUILIBRIUM

Figure 2 - 7

In plotting the potential ψ as a function of distance, we can make use of the additional information from solid state physics that the contact potential difference for strongly n and p-type materials is very nearly equal to the ionization energy in electrons volts for the covalent bonds in the crystal (that is, about 1.1 volts for silicon). Thus the electrostatic potential difference from anode to cathode should be about 1.1 volts. Contact potential differences between layers that are not so heavily doped should be less, so in figure 2 - 7d the potential differences across junctions 1 and 2 are drawn smaller than the potential difference across junction 3. While it is not possible to display a reasonably accurate charge density plot (which also distorts the $\mathcal{E}_x$ and ψ plots) due to the large variations in ρ and the necessity of making equal charge-distance areas on each side of a junction to find the depletion region dimensions, such "fudged up" plots as figure 2-7 may help us toward a better qualitative understanding of the depletion region sizes, voltage drops across junctions, and $\vec{\mathcal{E}}$ -field magnitudes in the SCR. Such an understanding will help us in unraveling some of the details of SCR operation, and of course, unless we are aware of such details we can't make reasonable assumptions or models when using the device in a circuit.

Consider the forward-blocking case (Fig. 2 - 8). The anode is made positive with respect to the cathode, and junctions 1 and 3 should be slightly forward biased while junction 2 should be reverse biased. We apply arguments similar to those used in modeling a diode. We assume, since only a small current will flow, that the IR drops in the crystal are negligible. Also, since only a small (compared to rated) current flows, the electrostatic potential differences across the forward biased junctions cannot be very different from the potential differences in the case of thermodynamic equilibrium. Therefore, the applied anode-cathode voltage must add to the potential difference across junction number 2. We now sketch the charge density, $\vec{\mathcal{E}}$ -field, and potential throughout the crystal knowing:

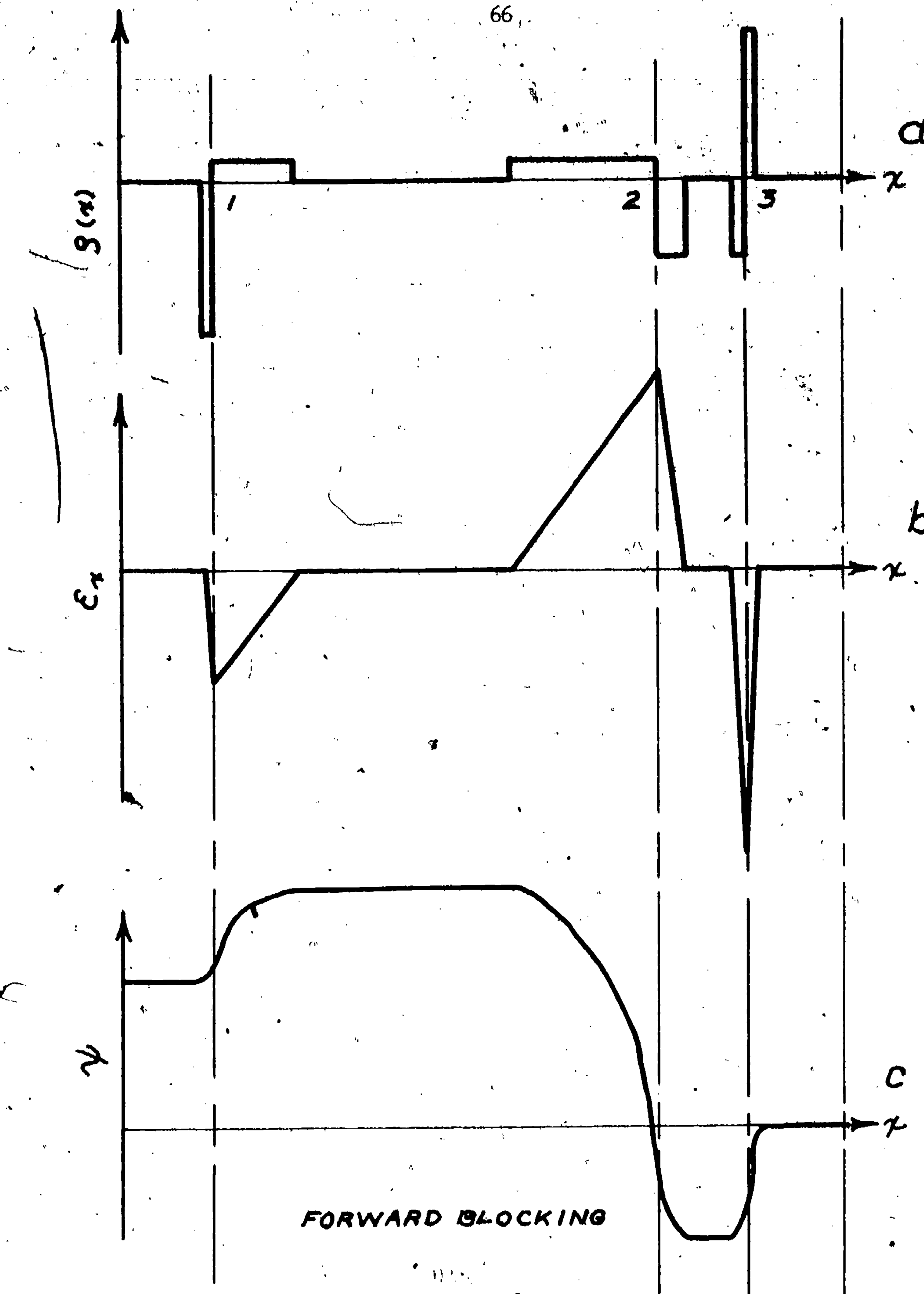


Figure 2 - 8

(a) ρ, $\vec{\mathcal{E}}$, and ψ in all places except junction 2 appears approximately the same as in figure 2-7,

(b) The potential difference across junction 2 is increased by the voltage applied to the device terminals,

(c) We know how ρ, $\vec{\mathcal{E}}$, and ψ must be related from our experience in sketching figure 2-7, that is, charge density will remain the same, the length of the depletion layer must increase, $\mathcal{E}_{\psi}(x)$ will be a triangle, and ψ will be parabolic with an inflection point at junction 2.

Summarizing our understanding of the forward blocking mode; a model consisting of two forward biased diodes and one reverse biased diode yields a V-I characteristic similar to that of a forward blocking SCR. A reverse saturation current will flow in the model as in the real device. At sufficiently high forward voltage (large anode to cathode voltage) the $\vec{\mathcal{E}}$-field at junction 2 will become so large that avalanch occurs, accounting for the abrupt increase in current at high forward voltage (at maximum forward voltage, region 1, Fig. 2-1). We also see, as a consequence of the sketch of figure 2-8, that the lightly doped n-type region between junctions 1 and 2 is responsible for the SCR's ability to block large forward voltages. The light doping yields a wide depletion layer and a smaller peak $\vec{\mathcal{E}}$-field for the same voltage difference than would occur for heavier doping. Thus avalanch is discouraged by light doping. Of course, the ideal diode behaviour is modified by surface leakage effects around the junction as in the case of a power diode.

The reverse blocking mode (anode negative with respect to the cathode) is only slightly complicated by the fact that both junctions 1 and 3 are reverse biased, and we might ask how the applied voltage divides across these two junctions. Junction 2 is slightly forward biased and has a potential drop slightly

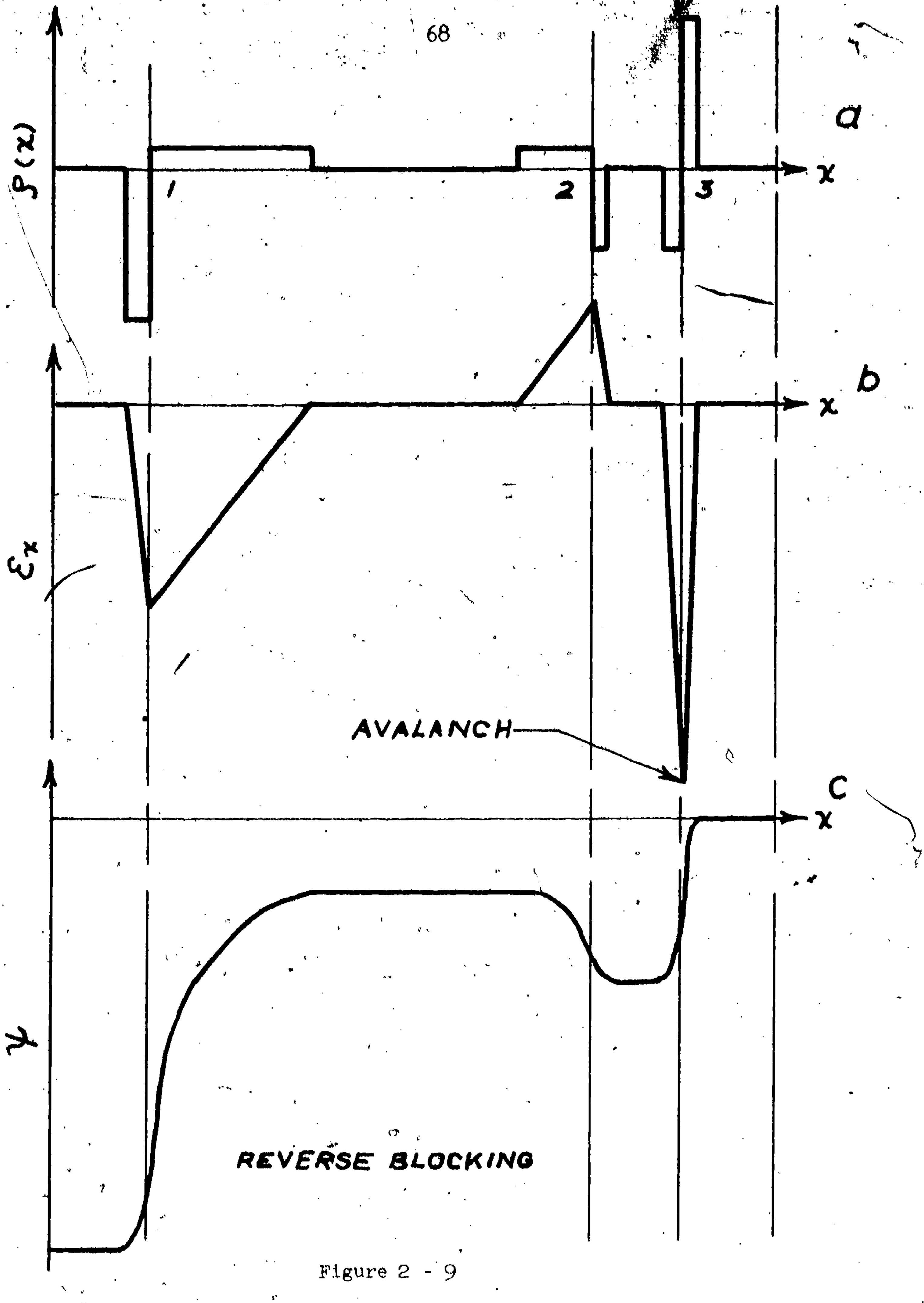


Figure 2 - 9

less than in the thermodynamic equilibrium case. In sketching figure 2 - 9, we expect that as the anode is gradually made more negative with respect to the cathode, the "additional" voltage is first taken up by junction 3 because;

the reverse biased diode having the smaller reverse saturation current will be the diode limiting the current and will have the largest voltage drop. Junction 3 will have the smaller reverse saturation current because;

its p-n layers are more heavily doped (compare N_A between junctions 2 and 3 to N_D between junctions 1 and 2, Fig. 2-7). Recall that the reverse saturation current depends on the concentration of holes in the n-layer and electrons in the p-layer, and the more heavily doped the materials are, the smaller will be the relevant concentrations.

As the reverse voltage increases, the maximum value of the $\mathcal{E}$-field at junction 3 increases to the point where avalanch occurs. The junction does not melt because junction 1 now limits the current. Further voltage increases cause the potential drop across junction 1 to increase as its depletion layer widens while the potential drop across junction 3 remains approximately constant at its avalanch breakdown voltage. This is the situation sketched in figure 2-9. Again we see that it is the lightly doped, thick, n-layer that is responsible for the high voltage rating (compared to transistors) of the SCR, and the model yields the reverse characteristic of a diode which agrees with observed SCR characteristics in the reverse blocking mode.

Finally we consider the forward conducting mode (region 2, Fig. 2-1). The only way this mode can be explained in terms of the series of diodes model is that all three junctions must be forward biased ! Such a phenomenon could not occur in the series connection of three independent diodes. However, we shall briefly consider what might be observed if the junctions were forward biased, keeping in mind that we don't know how to get into or maintain such a state.

The potential vs. distance diagram should be similar to that of the thermodynamic equilibrium case except for slightly smaller junction potential differences and IR drops due to the large (in comparison to blocking currents) forward currents (Fig. 2-10). We would expect the largest IR drop in the crystal to be across the lightly doped, thick, n-type layer between junctions 1 and 2. Not only is this layer the thickest but is has the highest resistivity due to its light doping.

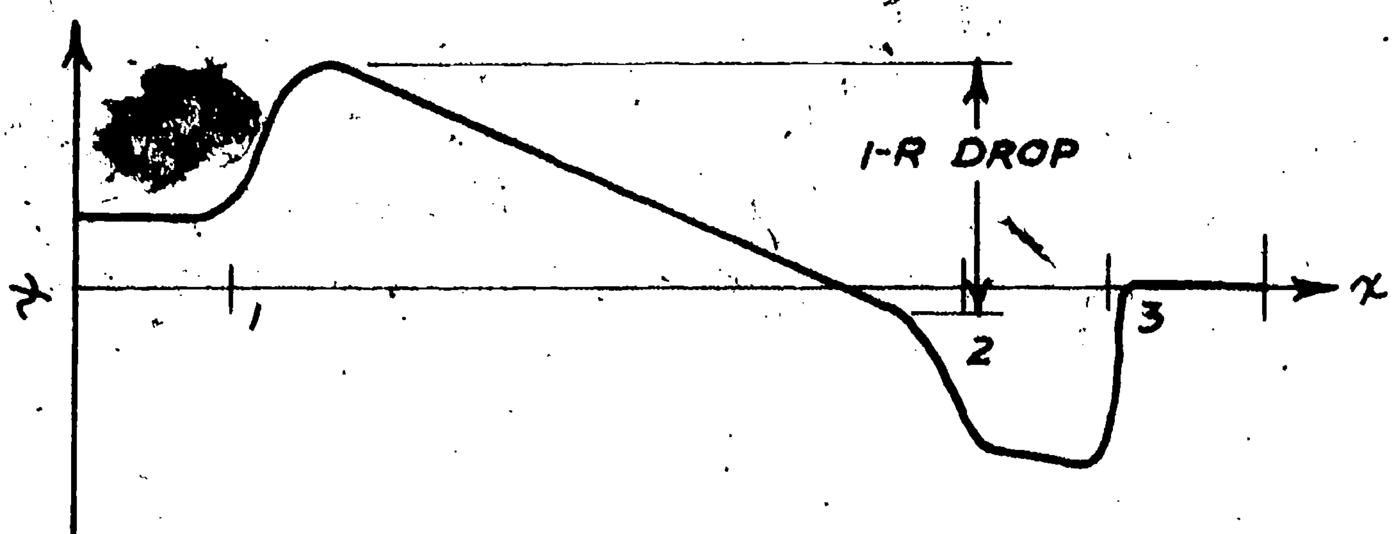
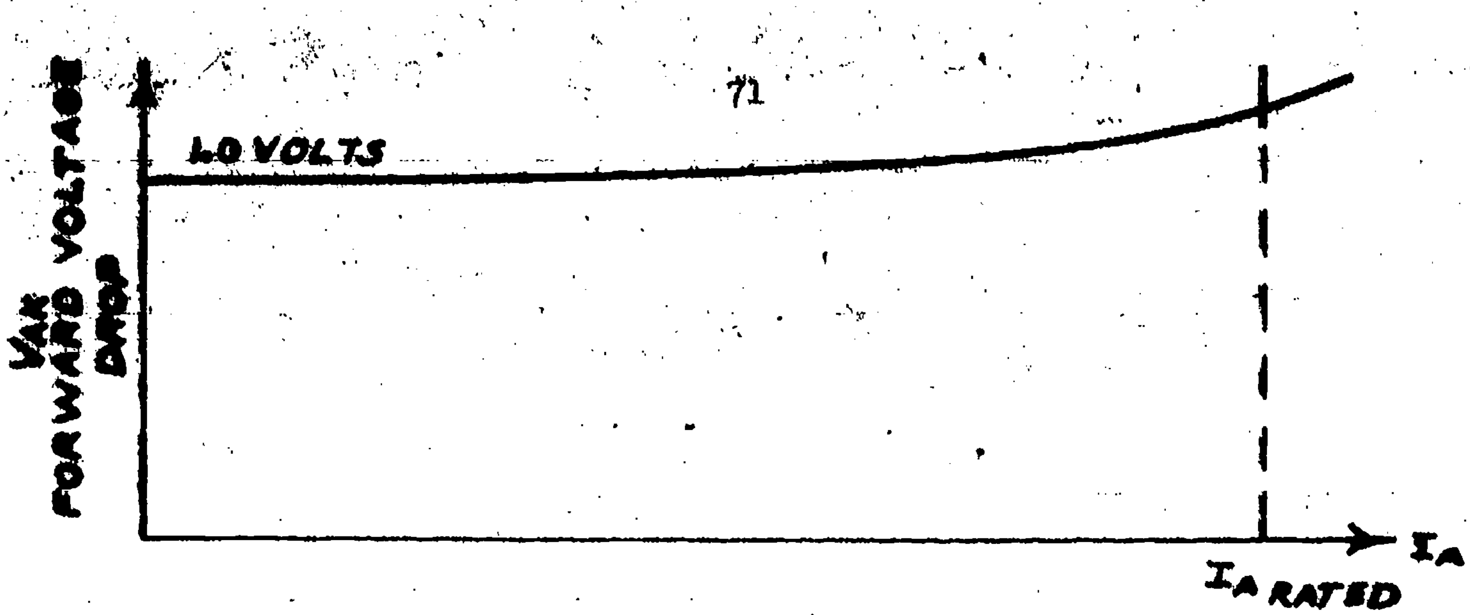


Figure 2 - 10

As the applied voltage increases, the forward drop across the junctions 1 and 3 should decrease, and the current should increase as in the case of a forward biased diode. As junction 1 allows more carriers to diffuse into the n-type layer between junctions 1 and 2, the carrier density in the n-type layer increases. This increase in carrier density should decrease the resistivity of the material, particularly in the lightly doped n-type region. Therefore, the IR drop should not increase proportional to the current. This model agrees with the observed behavior of the forward voltage drop as a function of current for an SCR (Fig. 2-11). Of course there is also a forward



Typical data for 50-100 amp SCR

Figure 2 - 11

drop due to the potential differences across the junctions. The forward voltage drop should be less than for two diodes in series because the potential drop across junction 2 subtracts from the drops across 1 and 3.

Summarizing, the three diode model of an SCR is not satisfactory in explaining the forward conducting mode because junction 2 must be "forward biased" while a very large reverse current passes through the junction. If only the problem of junction 2 could be "gotten over", the shape of the V-I characteristic for forward conduction could be explained in terms of junctions 1 and 3 and the conductivity modulation of the n-type layer between junctions 1 and 2. We have no means of getting into or remaining in the forward conducting mode according to our model. Therefore, we must search for a more sophisticated model in which the junctions are not so independent.

The Two Transistor Model of an SCR

As a next step in understanding SCR operation, we consider modeling the SCR as some interconnection of three layer devices since three is intermediate between two (diodes) and four (the SCR). The simplest three-layed device with which we have a reasonable degree of familiarity is the ordinary bipolar transistor. The transistor normally has one junction reverse biased (collector) so we still cannot model the forward conducting mode. However, we pursue the model of transistors for the forward blocking state since we may gain some insight in getting from the forward blocking to the forward conducting mode, and we can do something with the gate terminal (which was ignored in the diode model). Since junction 2 is the only reverse biased junction in the forward blocking SCR, junction 2 must correspond to the collectors of the transistor model which must also be reverse biased. We then "break" the SCR into a pnp and an npn transistor as shown in figure 2 - 12.

Before mathematically analyzing the two transistor model, we consider the physical properties the circuit should have. This "physical reasoning" step allows some measure of checking the validity of the mathematical results. In the forward blocking mode, the anode is at a large positive potential (tens-hundreds-a thousand volts) with respect to the cathode. Our transistor model will not be good for many hundreds of volts or larger because the high magnitude of the forward blocking voltage depends on the thick n-layer in the SCR's, and transistors do not possess such a layer. In the forward blocking mode, I_g (Fig. 2 - 12) is zero, and I_A and I_K are very small compared to the device current rating. It must be true that transistors 1 and 2 are both in the "cut-off" region of transistor operation because:

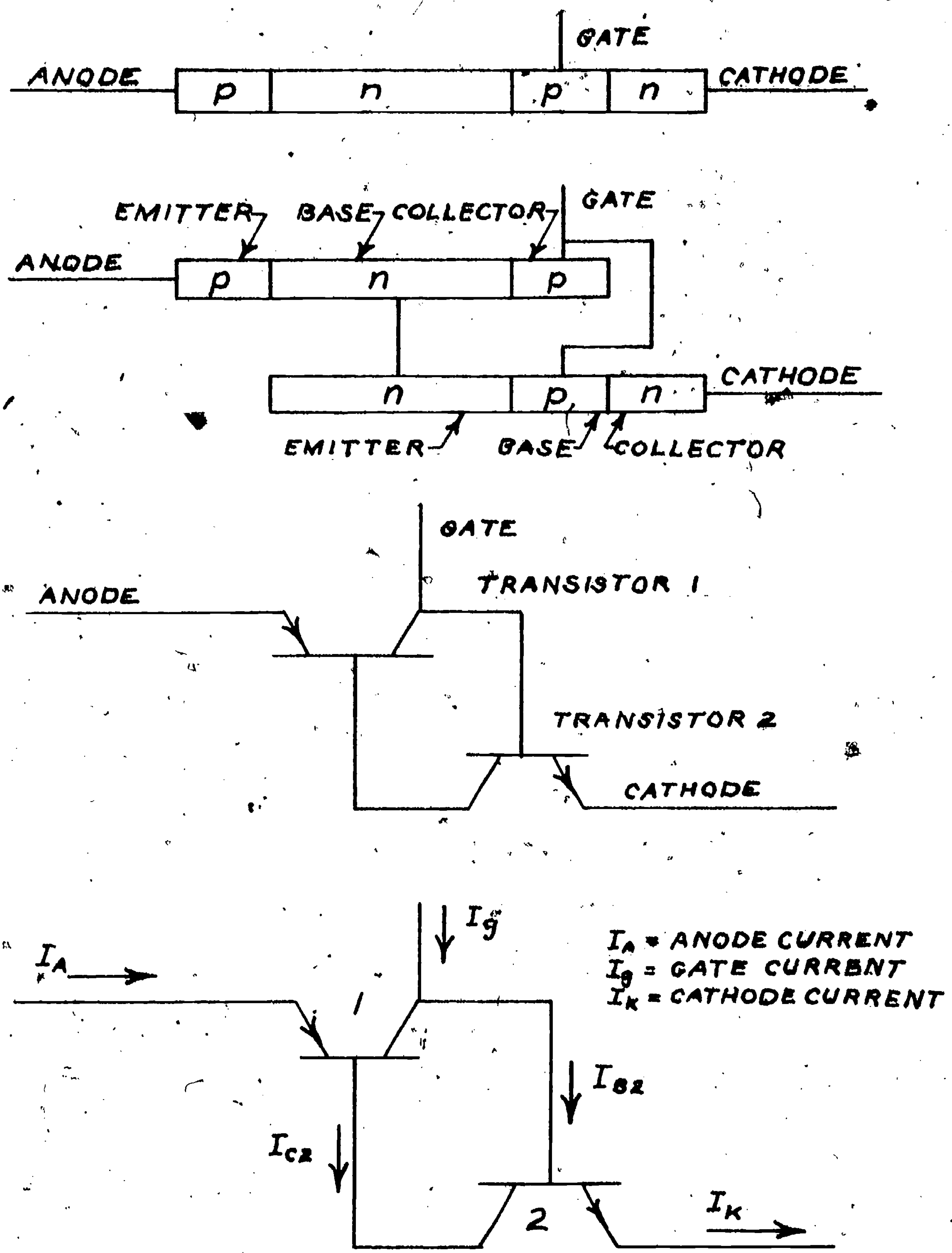

GATE
ANODE
p
n
p
n
CATHODE
EMITTER
BASE
COLLECTOR
GATE
ANODE
p
n
p
n
p
n
CATHODE
EMITTER
BASE
COLLECTOR
GATE
ANODE
TRANSISTOR 1
TRANSISTOR 2
CATHODE
I_g
I_A
1
I_{B2}
I_{C2}
2
I_K
I_A = ANODE CURRENT
I_g = GATE CURRENT
I_K = CATHODE CURRENT

Figure 2 - 12

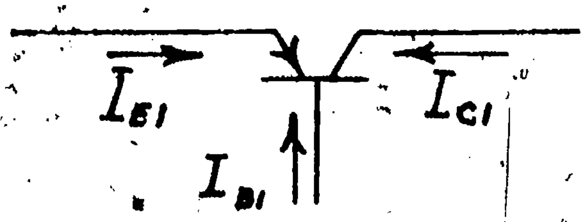
(a) If both transistors were conducting, I_A and I_K would not be small. Therefore at least one of the transistors must be "cut-off,"

(b) If transistor 2 is assumed cut-off, I_{B2} must be small. I_{B2} is limited only by transistor 1, therefore transistor 1 is also cut-off.

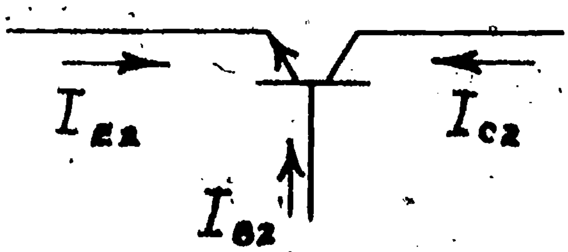
(c) If transistor 1 is assumed cut-off, I_{C2} must be small. I_{C2} is limited only by transistor 2, therefore transistor 2 is also cut-off.

If a current pulse I_g is introduced into the gate lead, transistor 2 begins to conduct and I_{C2} increases. This in turn causes transistor 1 to conduct which further increases I_{B2} which increases I_{C2}, etc. The two transistors are connected in a positive feedback loop, and eventually (in microseconds) arrive at a fully-on (saturated) state so that I_A is no longer blocked. It is this "turn on" or "triggering" mechanism that we wish to further explore.

The basic transistor current relations are:



$$I_{C1} = -\alpha_1 I_{E1} - I_{CO1} \quad (T_1$$
$$I_{C1} + I_{B1} + I_{E1} = 0 \quad (\Sigma i_1$$



$$I_{C2} = -\alpha_2 I_{E2} + I_{CO2} \quad (T_2$$
$$I_{C2} + I_{E2} + I_{B2} = 0 \quad (\Sigma i_2$$

Applying these basic relations to the final circuit of figure 2 - 12 yields the following equations.

Transistor 2

$$(1)\quad I_{C2} + I_{B2} - I_K = 0 \qquad (\Sigma i_2$$

$$(2)\quad I_{C2} = \alpha_2 I_K + I_{CO2} \qquad (T_2$$

Transistor 1

$$(3)\quad I_A + I_g - I_{B2} - I_{C2} = 0 \qquad (\Sigma i_1$$

$$(4)\quad I_g - I_{B2} = -\alpha_1 I_A - I_{CO1} \qquad (T_1$$

We choose to solve for the anode current I_A in terms of the gate current I_g. First we systematically eliminate I_{B2}.

Substitute 1 into 3.

$$(1,3)\quad I_A + I_g - I_K = 0$$

Substitute 1 into 4.

$$(1,4)\quad I_g + I_{C2} - I_K = -\alpha_1 I_A - I_{CO1}$$

Next we eliminate I_{C2} by substituting equation 2 into equation 1,4.

$$(1,2,4)\quad I_g + \alpha_2 I_K + I_{CO2} - I_K = -\alpha_1 I_A - I_{CO1}$$

The terms containing I_K are collected for convenience in eliminating I_K.

Rearranging 1,2,4

$$I_g + (\alpha_2 - 1)I_K + \alpha_1 I_A = -I_{CO1} - I_{CO2}$$

Substitute 1,3 into 1,2,4.

$$(1,2,3,4)\quad I_g + (\alpha_2 - 1)(I_A + I_g) + \alpha_1 I_A = -I_{CO1} - I_{CO2}$$

1,2,3,4 may be solved for I_A in terms of I_g.

$$I_g + \alpha_2 I_g - I_g + \alpha_2 I_A - I_A + \alpha_1 I_A = -I_{CO1} - I_{CO2}$$

$$I_A(\alpha_1 + \alpha_2 - 1) = -\alpha_2 I_g - I_{CO1} - I_{CO2}$$

$$I_A = -\frac{\alpha_2 I_g + I_{CO1} + I_{CO2}}{\alpha_1 + \alpha_2 - 1}$$

This is a very strange result! It states that when the anode is positive with respect to the cathode and the SCR is in the forward blocking mode, the current flows out of the anode lead. Thus the SCR delivers energy to the external circuit as if it were a battery. Of course, the result of the calculation seems absurd only if it is assumed that α_1 and α_2 have values near one. Experience with transistors shows that near the cut-off region, α decreases. To make sense of our result, we must consider the variation of α with collector current in a transistor.

Digression

Let us briefly consider the basic operation of a transistor according to the electron-hole model. The two junctions are spaced closely together (Fig. 2-13) so that the mobile charge carriers that diffuse across the "emitter" junction can further diffuse across the central "base" region and be collected by the high electric field at the "collector" junction.

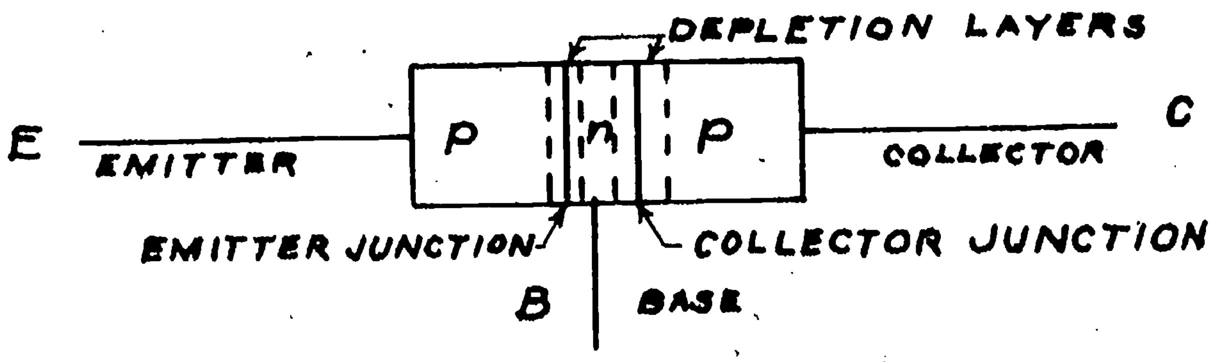


Figure 2 - 13

As the mobile carriers traverse the base region, some carriers will recombine. At low levels of minority mobile carrier concentration, the probability of recombination is proportional to the carrier density. α expresses the fraction of mobile carriers diffusing from the emitter that survive in the base region and are collected at the collector junction. α can be decreased by making the base region longer and by reducing the "lifetime" of a carrier by adding certain impurities which disturb the crystal lattice in a complicated way so as to increase the probability of recombination. The α for transistor 1 (with the long n-type layer) usually ranges from less than 0.1 to about 0.4 depending on the particular SCR being modeled and temperature.

To explain or model the variation of α with collector current, the action of lattice disturbances ("traps" or "recombination centers") is examined. Each lattice flaw which might be due to a physical imperfection or certain impurities in the lattice creates an inhomogenuity in the electric field at that point. For example, in the p-type material of an npn transistor the inhomogenuity may attract and hold or delay at that site ("trap") an electron. Now the electron's charge and the original inhomogenuity attract a hole. The electron and hole recombine, leaving the trap ready to collect another electron. The trapping-recombination effect takes place in addition to the recombination due to the random coming together of holes and electrons and thus increases the rate of recombination. As the density of electrons in the p-type material increases, implying an increase in the current density in a transistor, the traps become saturated. That is, most of the traps have electrons waiting to be recombined. Thus at larger current densities (not due to increased drift but due to increased carrier concentration), trapping becomes less significant as a recombination mechanism. This means that the rate of recombination does not increase in proportion to the current density and that a higher fraction of the carriers will reach the collector, resulting in an increase in α. Trapping is probably important in both the emitter depletion layer and the "base" regions of our model. Figure 2-14 illustrates a variation of the α's with current and temperature. At higher temperatures, there are more carriers because more covalent bands are ionized and therefore a higher percentage of the traps are saturated.

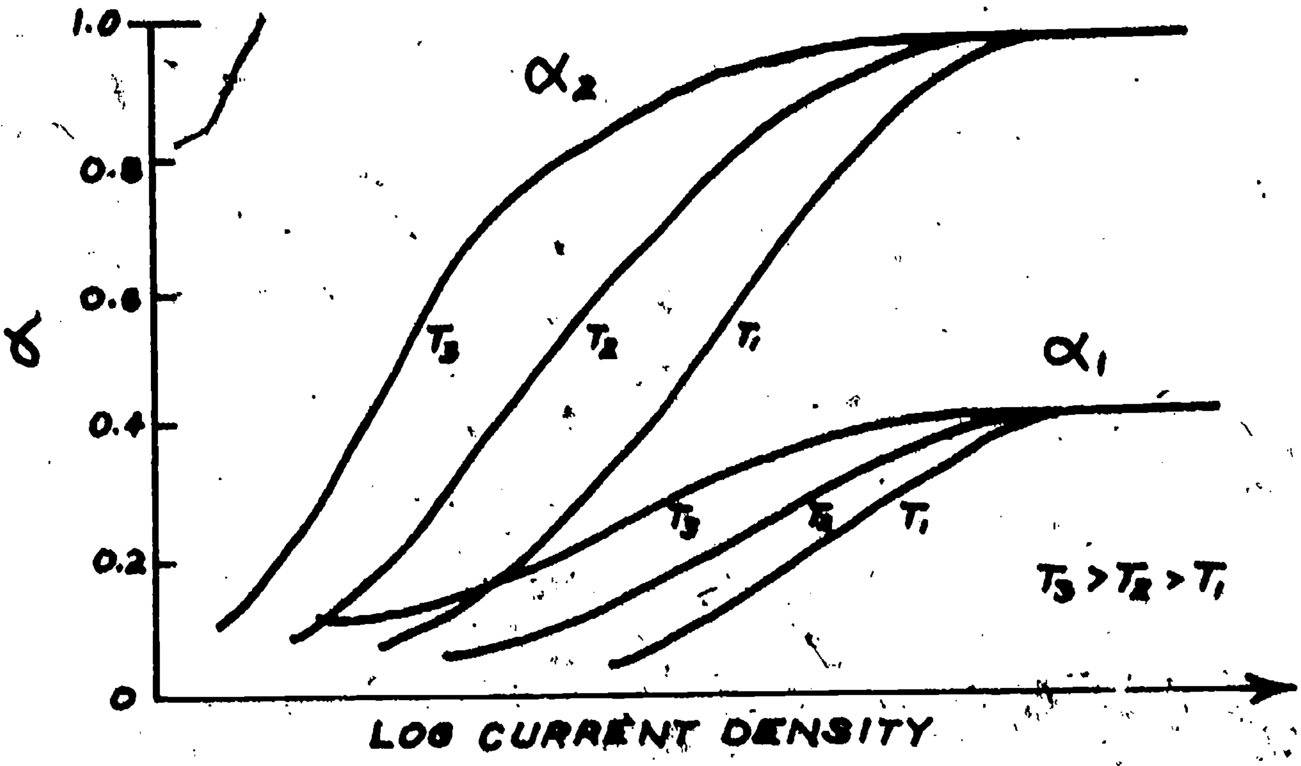


Figure 2 - 14

Thus we see that the α's of the transistors in the model are less than one. If I_g is applied and gradually increased, α_2 and also α_1 (by positive feedback action) increase due to the increasing current density in the transistors. Evenutally as I_g is increased, $\alpha_1 + \alpha_2$ approaches one and I_A approaches $+\infty$. The physical significance of I_A approaching ∞ is that the SCR is no longer blocking so that current can flow freely, and that the gate current I_g can be removed since it is no longer important in determining I_A. Once the current density in the SCR is sufficiently high, the device triggers and stays conducting until the current density is lowered by external circuitry until $\alpha_1 + \alpha_2 < 1$. The value of the anode current that will maintain $\alpha_1 + \alpha_2 \geq 1$ with zero gate current is defined as the "latching current". The "holding current" is defined as the value of the anode current in a particular circuit at a given temperature at which the conducting transistor will switch back to the blocking state. Note that in the "holding current" definition some gate current may be flowing (but not enough to keep the αs high).

It is possible to trigger the SCR by means other than a current pulse into the gate terminal. All that is necessary is to increase the current density at junctions 1 and/or 3 so that the αs increase sufficiently. This can be done by heating the SCR, shining a bright light on a junction and creating electron-hole pairs by photoionization, applying a voltage pulse to the anode-cathode terminal and using the capacitive effect currents of the SCR to trigger the device, or irradiating the SCR with X-rays, gamma rays, neutrons, etc.

We have now gained some insight into the turn-on mechanism using the two transistor model. Obviously such a model is no longer valid in the forward conducting mode because junction 2 is no longer biased like a transistor collector junction. We choose a half-hearted "fudge" by saying that we also have described a mechanism to maintain the SCR in the forward conducting state since the current density is high and if junction 2 began to block, the trigger conditions ($\alpha_1 + \alpha_2 = 1$) would be fulfilled. Hence once triggered (and the anode current is not limited to a value less than the holding current) the SCR cannot spontaneously return to the forward blocking state. We choose to stop the modeling process at this point because we have succeeded to a remarkable degree in explaining the groww V-I characteristic of the SCR and because further modeling is difficult. To model the forward conducting mode of the SCR, we would have to carefully consider and balance diffusion and drift effects not only at the three junctions but also the bulk material between junctions 1 and 3. We also know that conductivity modulation would modify the drift parts of the model. Further, to model the triggering process, we would have to take into consideration at least two models of recombination. While more detailed models would certainly be instructive and even absolutely necessary if we were designing SCR's, the modeling process can go on forever. If in the application of an SCR we find our present models insufficient, we will either extend our models or else have to go deeper into electron-hole modeling.

We next consider the static V-I characteristic of the gate-cathode terminals. In the case that no connection is made at the anode, on the basis of our three diode model of the SCR, we would expect a gate-cathode V-I characteristic similar to the dashed curve in figure 2-15a.

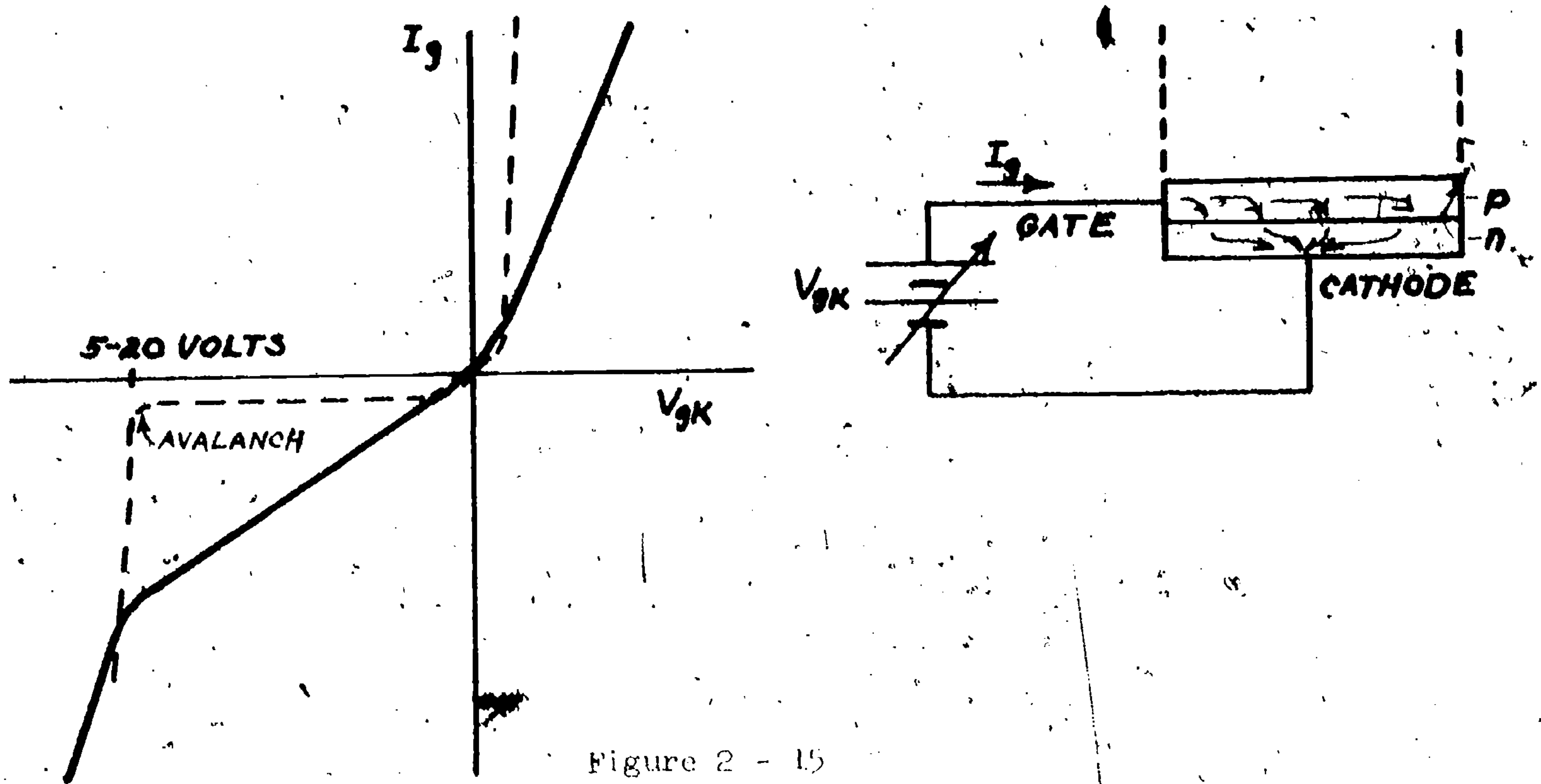


Figure 2 - 15

The actual V-I characteristic is shown by the solid curve in figure 2 - 15a. The differences between the dashed and the solid curve are rather easily explained in terms of our model. Surface leakage current allows more current to flow in the reverse direction than would be expected in the case of an ideal diode. In fact, such leakage paths are specially provided in many SCR's ("shorted emitter " SCR's) for reasons related to the transient characteristics (to be discussed later) of the SCR. In the forward direction, the bulk resistance of the p-type gate layer which is much less heavily doped than the n-type cathode layer must be considered. Therefore at "high" forward voltages or currents, the gate-cathode terminals appear more like a resistor then a diode.

In the case that the diode is forward blocking, again our model predicts that as the gate voltage is slowly varied, the V-I characteristic should be about the same as in the case of no connection to the anode until the gate current begins to be so large that triggering is imminent (Fig. 2 - 16).

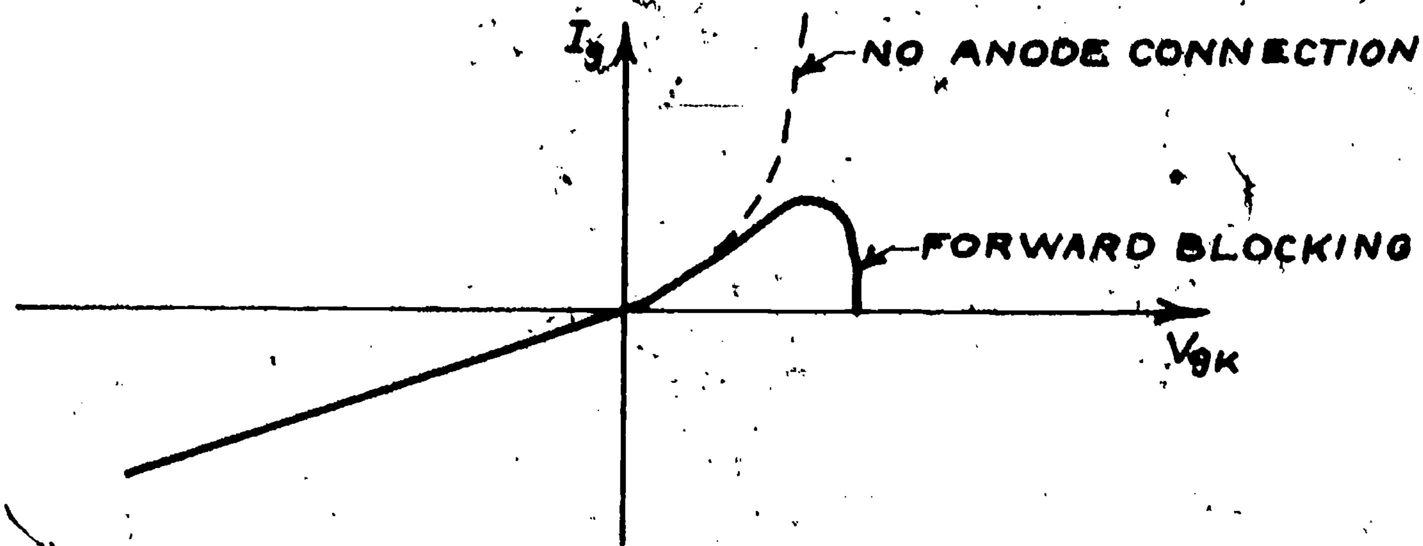


Figure 2 - 16

Once the SCR has been triggered, the gate characteristic changes radically, acting like a source due to the current passing through the junction from the anode circuit. In fact, the gate V-I characteristic will depend to some extent on the anode circuit (Figure 2 - 17).

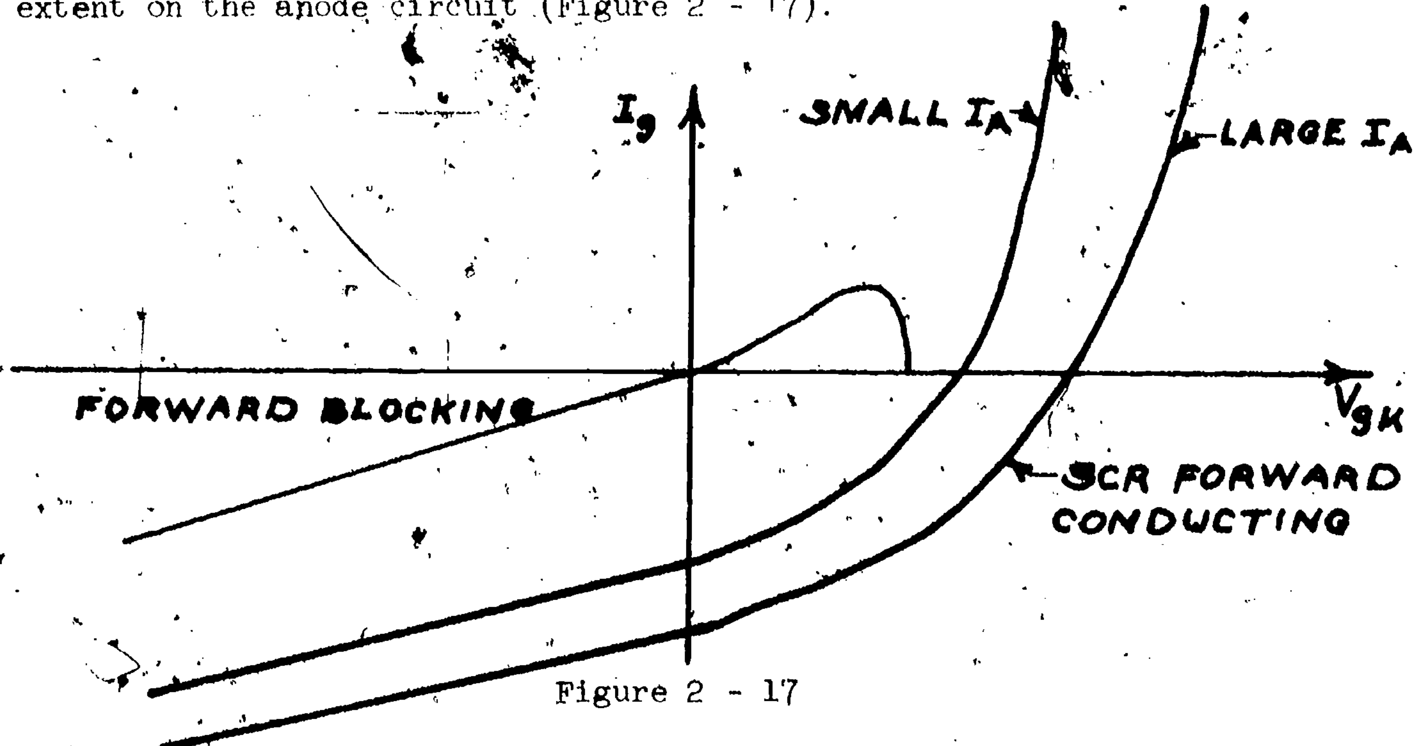


Figure 2 - 17

The gate V-I characteristic under conditions of reverse blocking is left as a problem (Lab problem 2 - 2) for the student.

SCR Transient Characteristics

If an SCR, resistor, and battery are connected in series as shown in figure 2 - 18a, and the SCR is triggered by a step of current into the gate; some time will elapse before the circuit current and the voltage across the SCR reach steady state. While the time required to reach steady state depends upon the partiuclar SCR in the circuit, times on the order of 10 μ sec - 50 μ sec are common. The total time to reach steady state is frequently subdivided into three shorter times (Fig. 2 - 18d), important from design considerations and related to distinct physical processes.

The "delay time", τ_D, can be related analytically and experimentally to the time required for the mobile carriers to cross the SCR regions corresponding to the bases in the two transistor model. Thus τ_D can be considered the time required for the current densities at the transistor emitter junctions to change and begin increasing the transistor alphas. The "risetime," τ_R, is the time required for the positive feedback process of the circuit in the two transistor models to increase the current densities such that $\alpha_1 + \alpha_2 = 1$. The change in current or voltage (0.1 to 0.9 of maximum) used to define rise-time is in accordance with generally accepted conventions regarding pulse waveforms. The rise-time decreases as the gate triggering current increases because large gate currents help to build up the current density at the junctions faster. The sum of the delay and risetimes is frequently denoted as the "turn-on time." Clearly the turn-on time will vary with gate current "drive" and from SCR to SCR. Turn-on times are typically on the order of several micro-seconds.

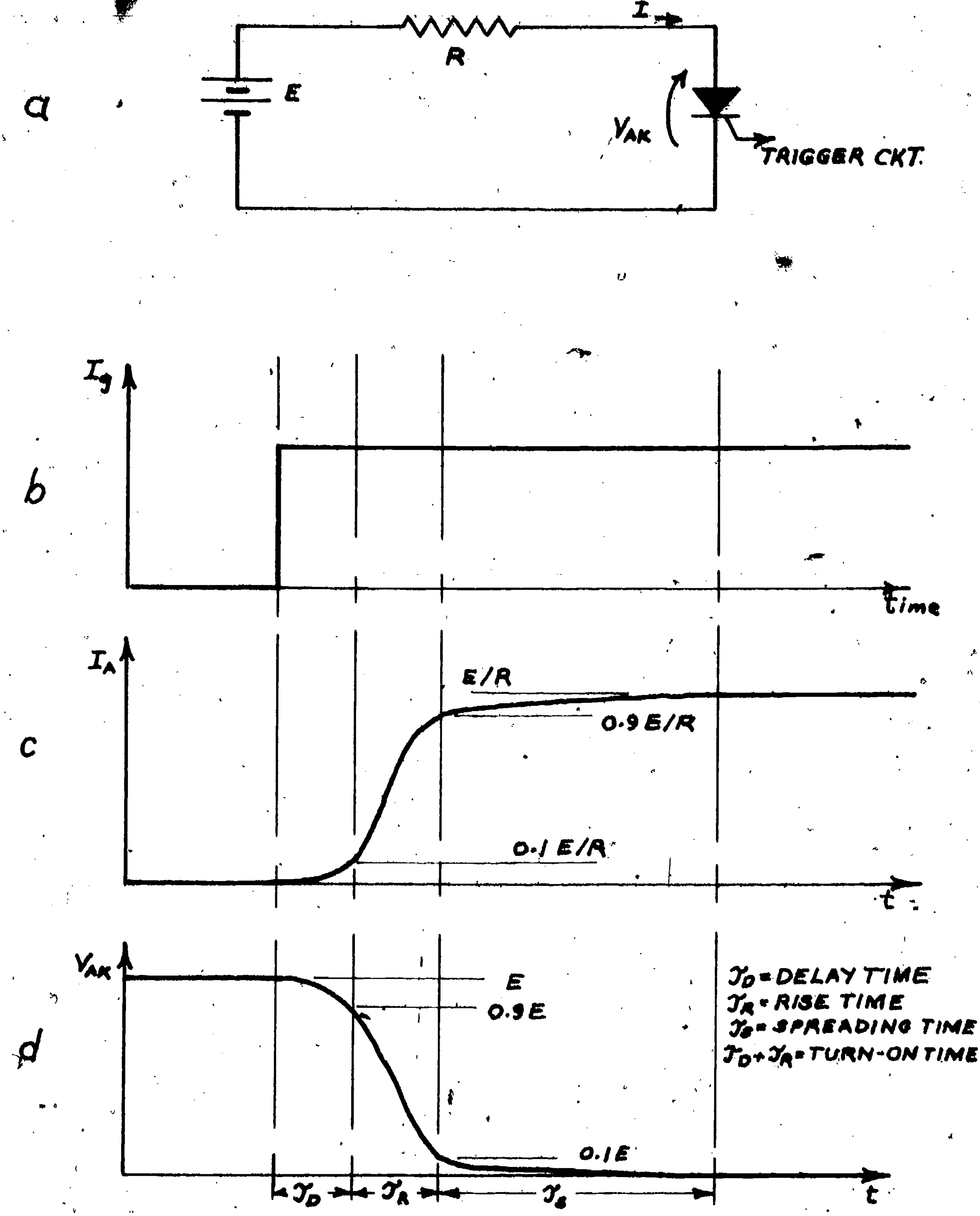

a
I
R
E
V_{AK}
TRIGGER CKT.
b
I_g
time
c
I_A
E/R
0.9 E/R
0.1 E/R
t
d
V_{AK}
E
0.9E
0.1E
τ_D
τ_R
τ_S
t
τ_D = DELAY TIME
τ_R = RISE TIME
τ_S = SPREADING TIME
$\tau_D + \tau_R$ = TURN-ON TIME

Figure 2 - 18

The remaining time until steady state is reached is called the "spreading time", τ_S. Our models for triggering and forward conduction were made on the basis of examining a filament of the crystal. Not all filaments are triggered simultaneously simply because some filaments are nearer the gate than others. Figure 2 - 19 illustrates the "spreading" of the forward conducting state in the crystal.

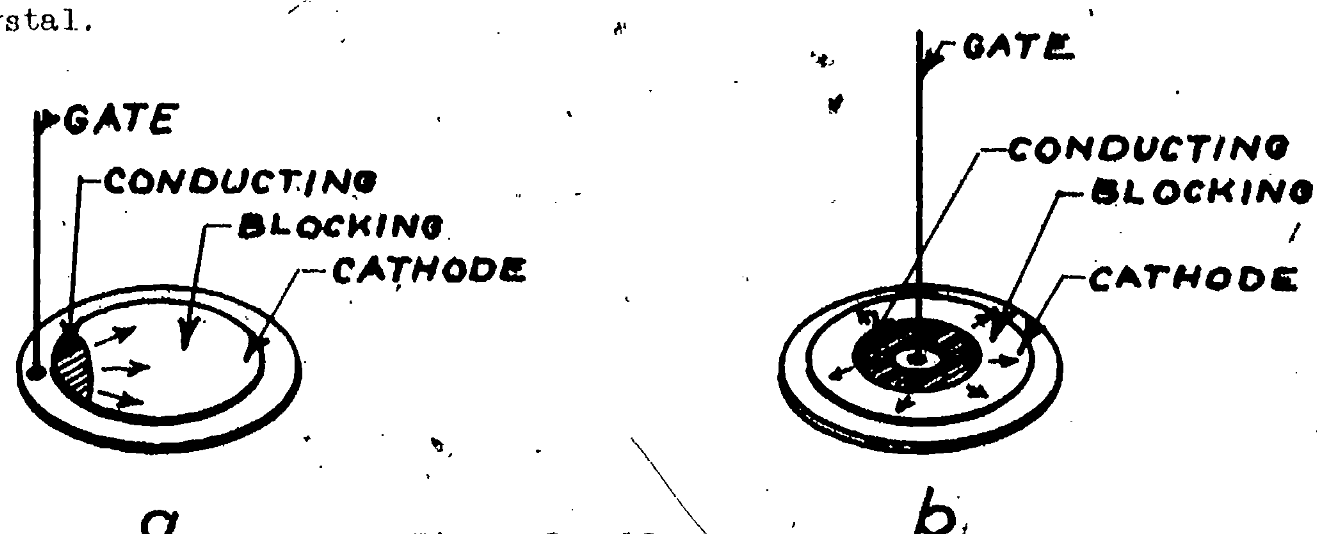


Figure 2 - 19

While at the end of the risetime only a small area of the crystal is conducting, all of the crystal area is conducting at the end of the spreading time.

Note that the anode current is 90% of its maximum value at the end of the risetime, yet only a small area of the crystal is conducting. Under such conditions, the current density in the conducting area may be very large, heating the area to high enough temperatures to damage the junction. In order to prevent such damage, a maximum dI_A/dt is usually specified for an SCR, allowing the conducting area to spread fast enough to accommodate the increasing current. The reason for placing the gate in the center of some SCR crystals (Fig. 2 - 19b) is that the conducting area has only to spread about half as far as in edge triggered SCR's (Fig. 2 - 19a). The dI_A/dt limitation can be reduced as well as the turn on time by "driving the gate hard." If a large current is introduced in the gate lead during the first few microseconds of triggering, the initial triggered area will be larger than if the gate current were gradually increased. The gate current can then be reduced to a lower value until the anode current rises to a value that will maintain the SCR in the conducting state ("latching current"). Figure 2 - 20 shows an "ideal" gate current waveform that allows the largest possible dI_A/dt, has a minimum turn-on time, and has a duration long enough to insure the SCR will stay on for the given anode current waveform.

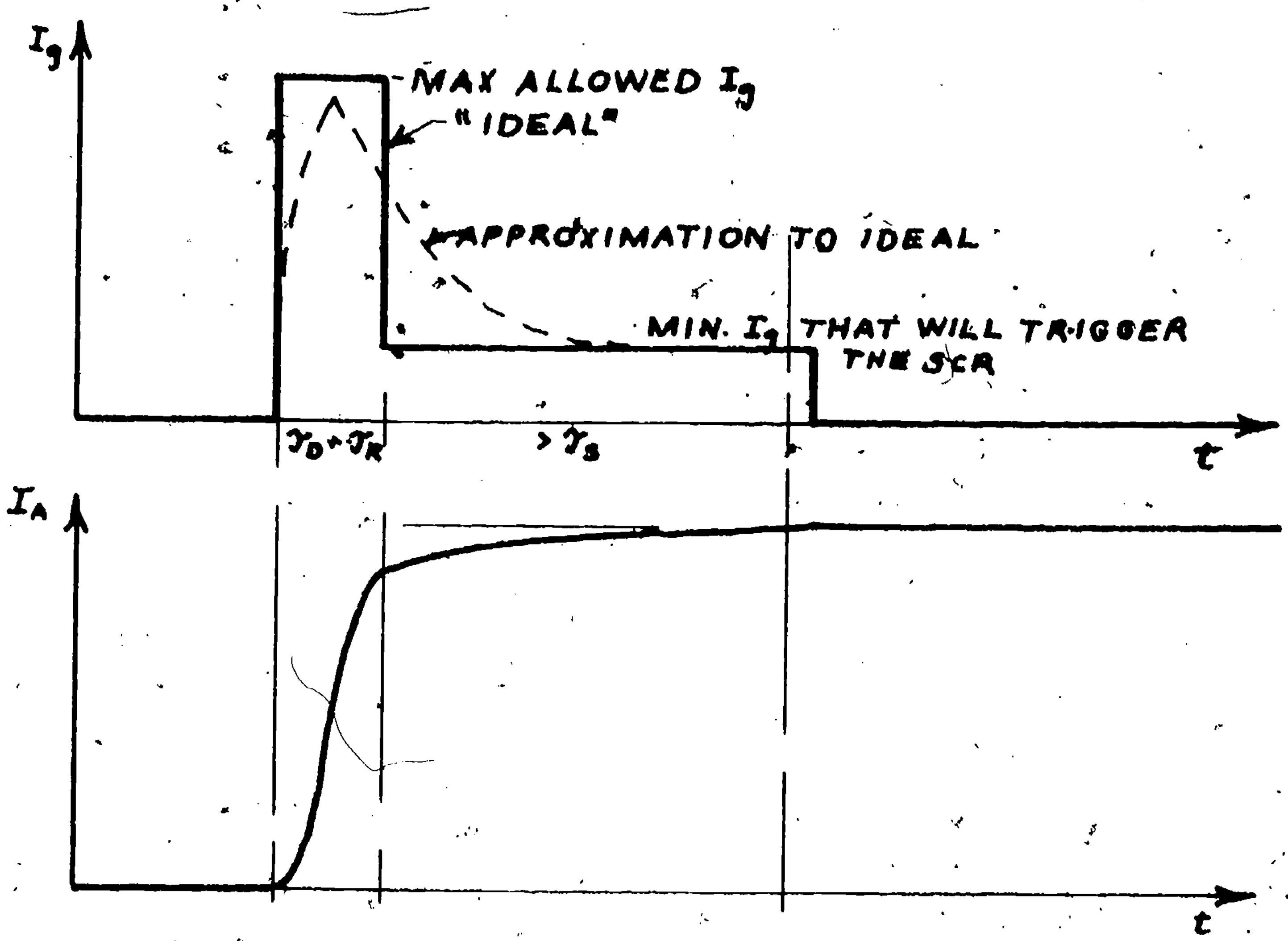


Figure 2 - 20

Such an "ideal" waveform is difficult to achieve, and various approximations to such a waveform are provided by practical trigger circuits. In general, it is very desirable to apply a gate current waveform whose risetime is smaller than the SCR turn-on time.

We now consider a "turn-off" transient. If the voltage waveform shown in figure 2 - 21 is applied to the SCR, the forward blocking mode will be restored. Since a great variety of voltage (and current) waveforms can restore the forward blocking mode, we will examine the essential features of this particular waveform and hope that we can apply our knowledge to other situations.

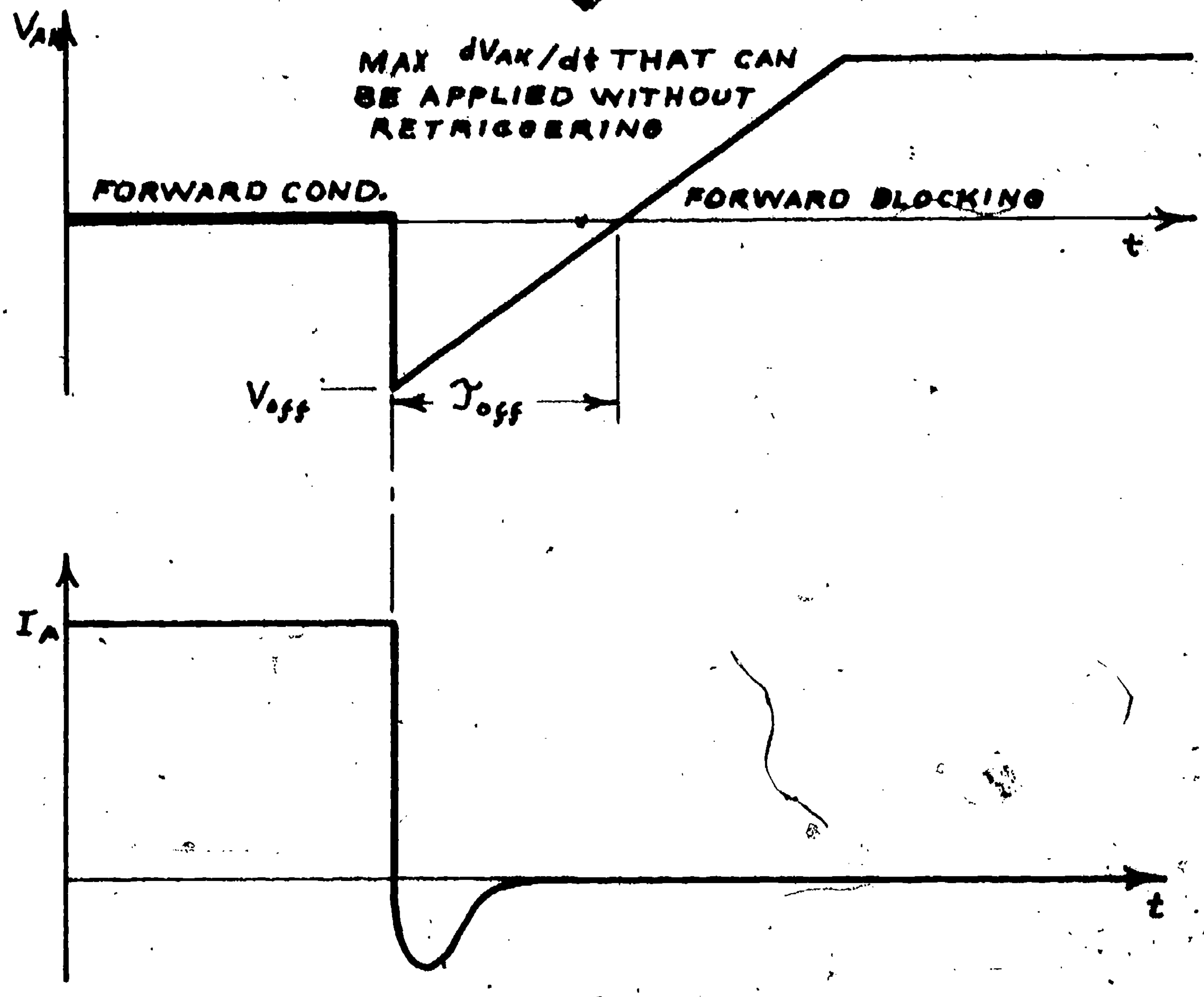


Figure 2 - 21

Figure 2 - 21 shows a large reverse voltage applied to the SCR in order to turn off the SCR. When the reverse voltage is first applied, a reverse current flows, primarily due to the motion of the mobile charge carriers in the base regions of the two-transistor model. Some component of the current is also due to the changing widths of the junction depletion regions. If a forward voltage was applied while the current was still flowing, the current densities and therefore α's would still be large enough that the SCR would re-trigger, and thus the forward blocking mode could not be reached. Also, due to the capacitances of the depletion layers, the rate of increase of

voltage must be limited or else the currents flowing into the junction capacitances will raise the α's, and the SCR will re-trigger. The time required for the stored mobile charges to recombine plus the time for the SCR to recover its forward blocking capability (due to the dV_{AK}/dt limit) is known as the "turn-off time", γ_{off}. The turn-off time depends to a large extent on the circuit in which the SCR is used as well as the SCR, however, "typical" turn-off times are frequently presented in the SCR data sheets. Turn-off time becomes a very important concept in high frequency SCR circuits. Because the turn-off time is usually more than an order of magnitude longer than the turn-on time, the turn-off time usually determines the maximum repetition rate for triggering an SCR.

SCR Trigger Circuits

We have already mentioned an "ideal" triggering current waveform. Before a circuit used to trigger an SCR can be designed in detail, the SCR gate specifications must be considered. While the turn-on and spreading times are usually specified in a relatively straight-forward way, the magnitudes of the voltages and currents that may be applied to the gate are related in a more complicated way. The pertinent gate specifications are:

(a) Specifications to protect the device from damage such as peak forward voltage V_{gK}, peak forward current I_g, peak instantaneous power $V_{gK} I_g$, and maximum average gate power $(V_{gK} I_g)_{avg}$,

(b) Specifications giving the minimum gate currents and voltages required to trigger the SCR. These specifications are temperature dependent.

The gate specifications are further complicated by the fact that there is a significant variation of the gate V-I characteristics and the minimum gate voltages and currents required to trigger the SCR (b) for individual SCR's of the same type (number designation).

The gate specifications may be given as individual numerical ratings, may be summarized in a chart, or may be given as a combination of the two. The particular format used to describe and specify the gate characteristics varies considerably from manufacturer to manufacturer, but the developing of a "typical" gate triggering chart will further clarify the basic ideas.

First we consider the fact that for a given type of SCR, there is a variation in the V-I characteristics of individual units. An individual SCR's V-I characteristic must be between the extreme limits of the V-I characteristics considered acceptable by the manufacturer (Fig. 2 - 22).

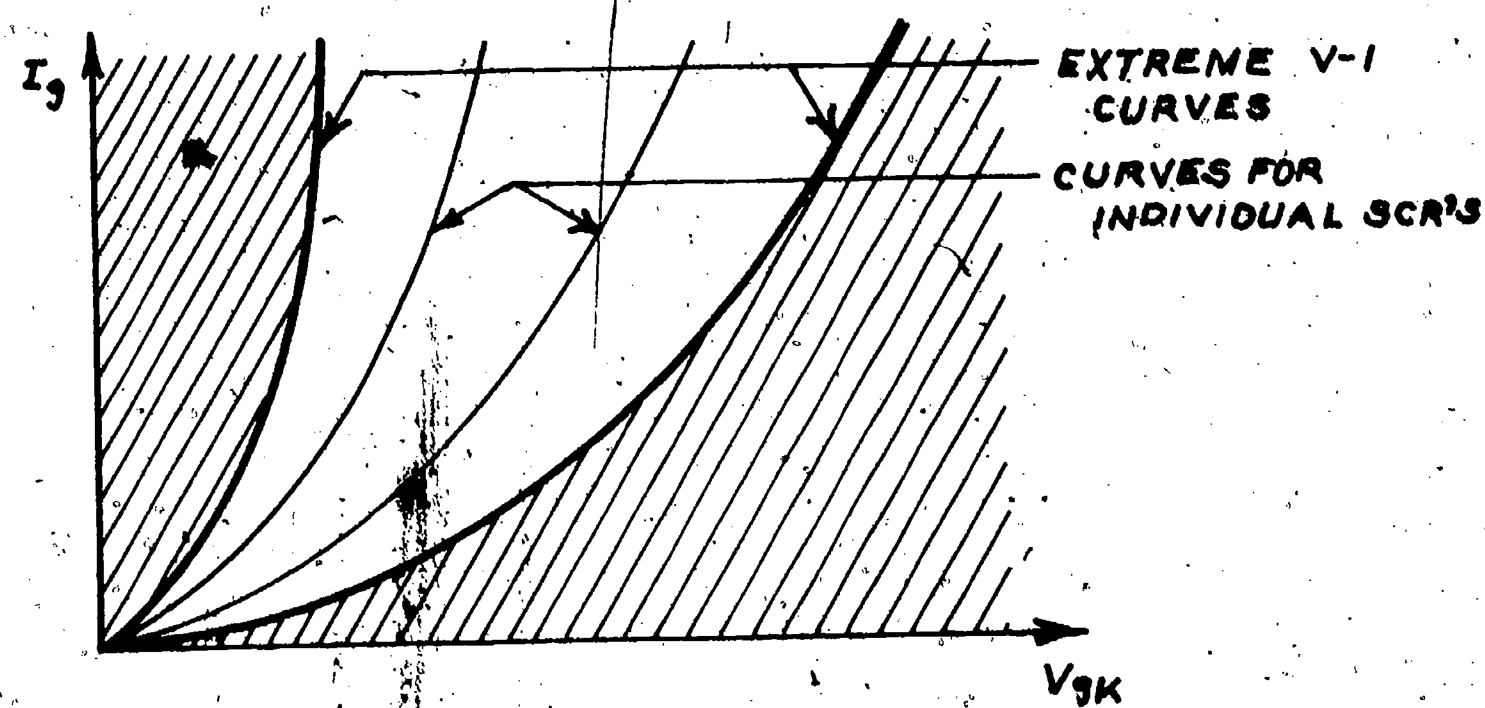


Figure 2 - 22

We further define the region of SCR triggering by denoting the maximum permissible values of forward voltage and current and the maximum instantaneous power. As a helpful note, we also show (dotted line, Fig. 2 - 23) the maximum average power.

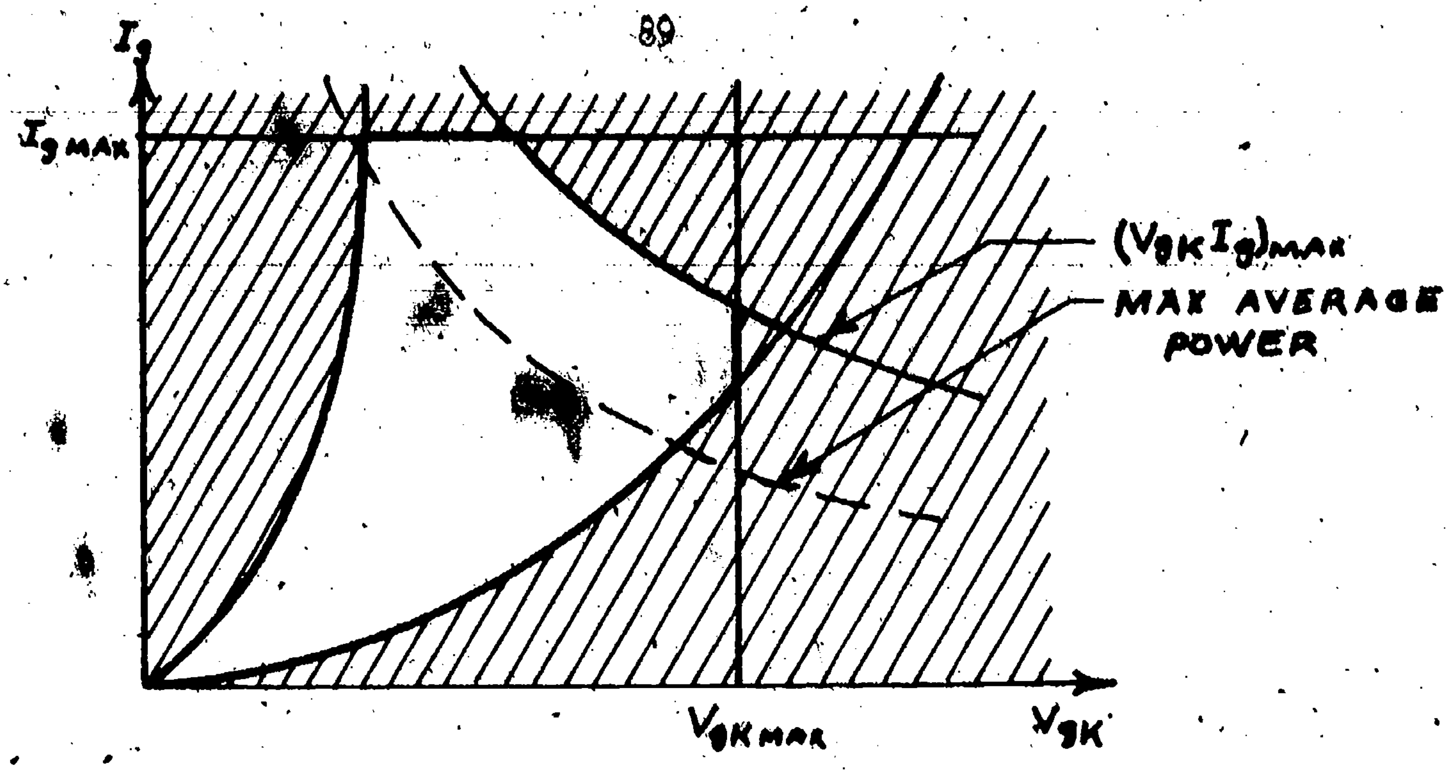


Figure 2 - 23

Finally, we consider the fact that some currents I_g and gate voltages V_{gK} are too small to trigger the SCR. There is considerable variation of the minimum triggering quanitities among the individuals of a given type of SCR (a specific JAN serial number or manufacturer's code catalog number). Because we want the information for design purposes, we must choose the maximum value for the series of SCR's of the minimum current required to trigger or minimum voltage required to trigger the SCR. Thus all of the SCR's of a given type can be triggered with a voltage or current greater than the specified amount. These values are frequently given at several temperatures because the SCR's are "easier" to trigger at higher temperatures. The clear area of figure 2 - 24 is the triggering region for a given type SCR. The actual V-I characteristic lies somewhere in this region, the voltages and currents are large enough to trigger any SCR of the given type, and the voltage, current, and power levels are sufficiently low that the device will not be damaged.

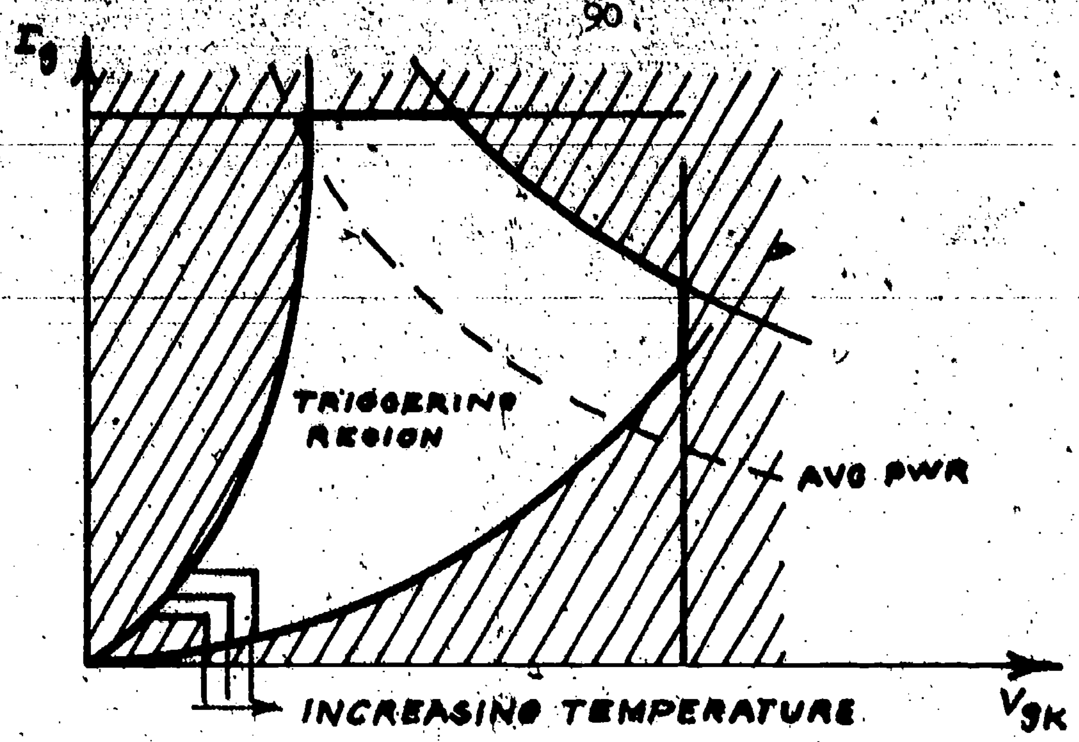


Figure 2 - 24

There is an infinite variety of triggering circuits for SCRs, and there is no single "best one" for all applications. The design of the triggering circuit must include not only the logic for the desired control of the SCR but also the characteristics of the anode circuit. For example, the simple circuit of figure 2-25 is intended to close the relay a specified time after switch "S" is closed. The trigger circuit will usually be chosen as an integral part of the timing circuit thus minimizing the number of components, and the triggering pulse must last long enough for the current in the inductive anode circuit to exceed the latching current.

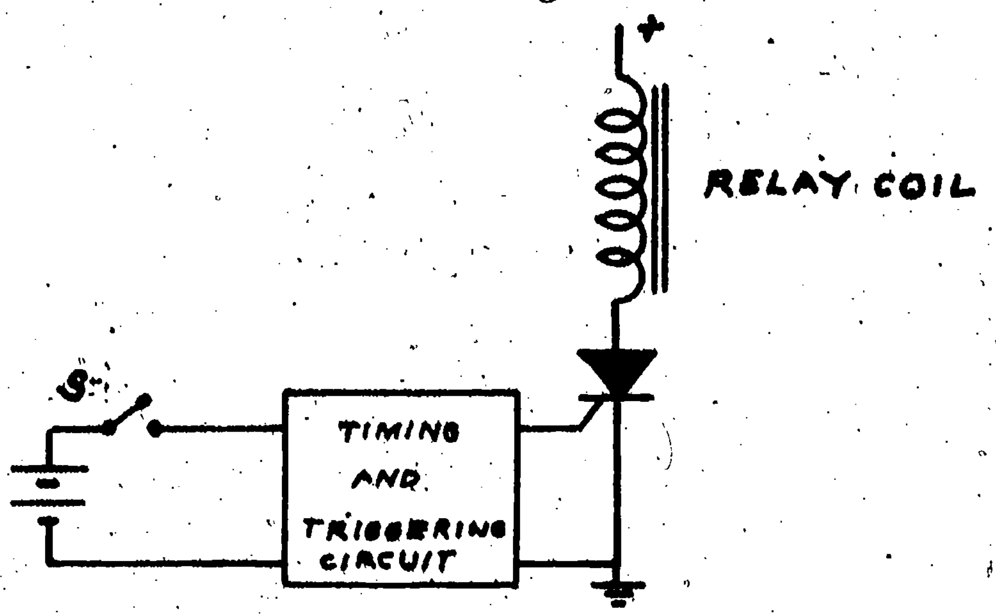


Figure 2 - 25

We shall consider a variety of triggering circuits, not as an attempt to compile a "dictionary" of circuits, but as a spur to the imagination.

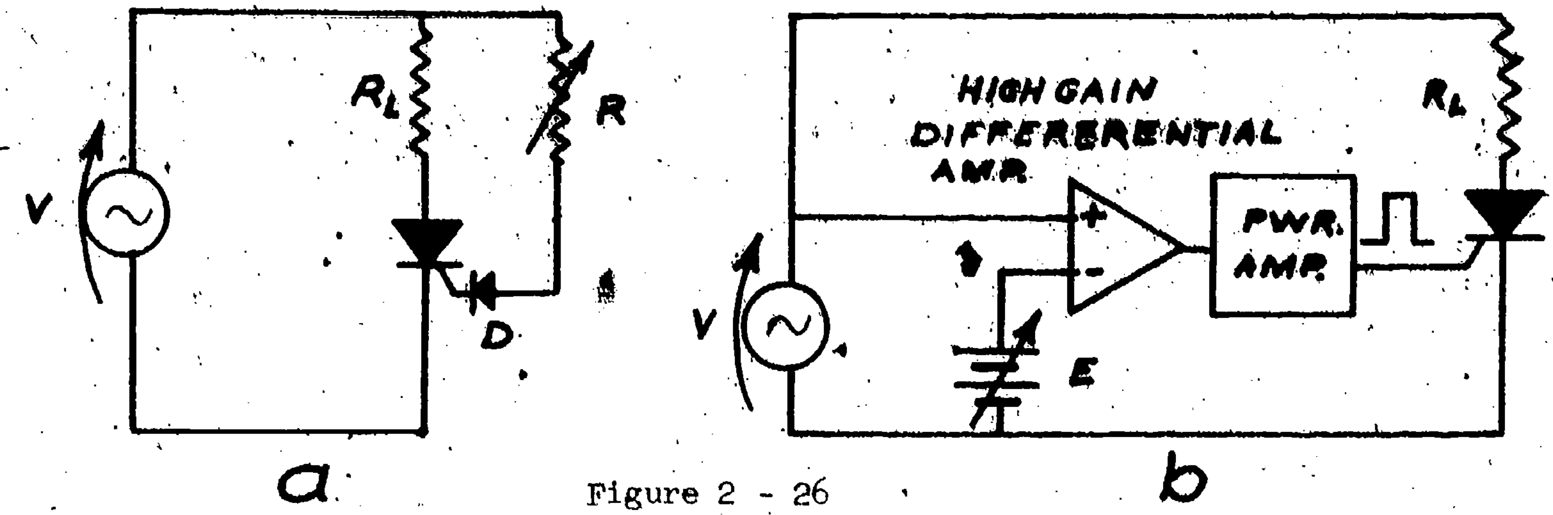
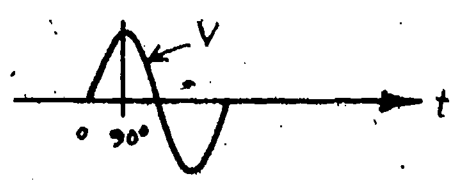


Figure 2 - 26

Figure 2 - 26 a and b illustrates two circuits that control the current through a load resistor R_L by sensing the magnitude of the input voltage. In both cases, the load current flows in the positive half-cycle of the source V only after the source voltage has exceeded some preset magnitude. Obviously, these trigger circuits operate over the electrical angle of $0^{\circ} < \theta < 90^{\circ}$ of the source sinusoidal voltage. If the SCR has not triggered by the time the voltage applied to the gate has reached its maximum value, the SCR cannot be triggered without decreasing the value of R.



The electrical angle at which triggering occurs (assuming a sine wave input) is called the "firing angle". The firing circuit of figure 2 - 26a is an inexpensive circuit having a far from ideal triggering waveform. The trigger voltage rises slowly, unduly limiting the maximum permissible value of dI_A/dt. Also, the voltage at which triggering occurs will depend on the individual SCR and its temperature. The diode D is required to prevent excessive gate currents when the SCR is reverse blocking (remember the gate junction will be in a avalanch state when the reverse voltage exceeds 10-20 volts). Figure 2 - 26b shows a more elaborate voltage sensing circuit using a high gain differential amplifier (frequently an "operational amplifier"). As long as the battery voltage "E" exceeds the supply voltage V by more than a few millivolts, the high gain ($\sim 10^5$ or more) amplifier will be saturated at its maximum negative output voltage. When V exceeds E by a few millivolts, the amplifier switches to its maximum positive output voltage. Thus a "step" of voltage is supplied to the power amplifier that "drives" the SCR gate. This circuit has advantages over the previous circuit of a gate drive amplitude that is independent of source voltage, steep triggering wavefront, and less dependence of the firing angle on temperature. However, the amplifier circuit is clearly more expensive.

Sometimes it is preferable to vary the phase of the controlling waveform rather than vary the level at which the trigger circuit is activated. One particularly convenient phase shifting circuit, having the virtue of an output voltage magnitude that is independent of the phase shift, is the "thyratron phase shifter." The circuit (Fig. 2-27) was originally invented to work with gas thyratron tubes, but the circuit works almost as well with SCR's.

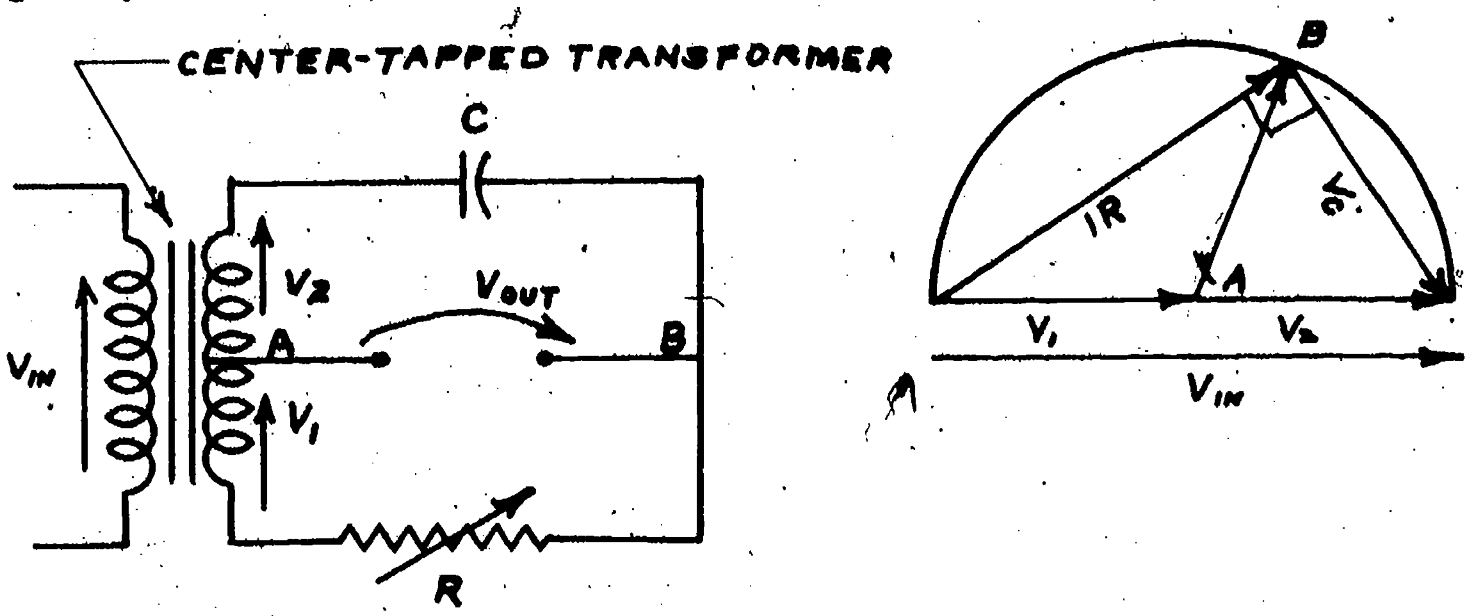


Figure 2 - 27

This circuit is most easily analyzed by constructing its phasor diagram. It is assumed that the transformer is lightly loaded so that its leakage reactance and resistance has no significant effect. Then $\vec{V}_{in}$, $\vec{V}_1$, and $\vec{V}_2$ all have the same phase angle. Knowing that the current through a pure capacitance leads the capacitor voltage by 90^0, and further assuming that the output current (drawn by an external circuit connected to terminals A and B) is negligible compared to the current in resistor R, the phasor right triangle with hypotenuse $\overrightarrow{V_1 + V_2}$ is constructed.

$$\vec{V}_1 + \vec{V}_2 + \vec{IR} + \vec{V}_c = 0 \text{ by Kirchoff}$$

The locus of all points B for different R's must be a semicircle having radius $|V_1| = |V_2|$ by elementary geometry. Thus the phase of $\vec{V}_{out} = \vec{V}_{BA}$ can be varied relative to the phase of $\vec{V}_{in}$ from 0^0 to 180^0, and $\vec{V}_{out}$ will have the same amplitude (the radius of the semicircle) regardless of the phase shift. It is possible to have phase shifts in the range $180^0 < \phi < 360^0$ by interchanging the positions of the capacitor and resistor.

The gradually rising sine wave out of a phase shifter makes a poor SCR gate pulse. While a large variety of ways to "sharpen the wavefront" of the gate pulse exist, such as the circuit of figure 2 - 26b or schemes using monostable multivibrators, there is a class of semiconductor devices specifically designed for this purpose that may be applicable. These devices (called "breakover diodes", "Shockley diodes", DIAC's, etc.) have the property that for low voltage levels (0-25V), they block the current. If higher voltages are impressed across the diode, they "break down" in a few microseconds. The voltage across the diode rapidly decreases, and large currents (on the order of amps) may flow. Figure 2 - 28 shows a circuit using a "DIAC" (G-E) to modify the wave front of a sine wave.

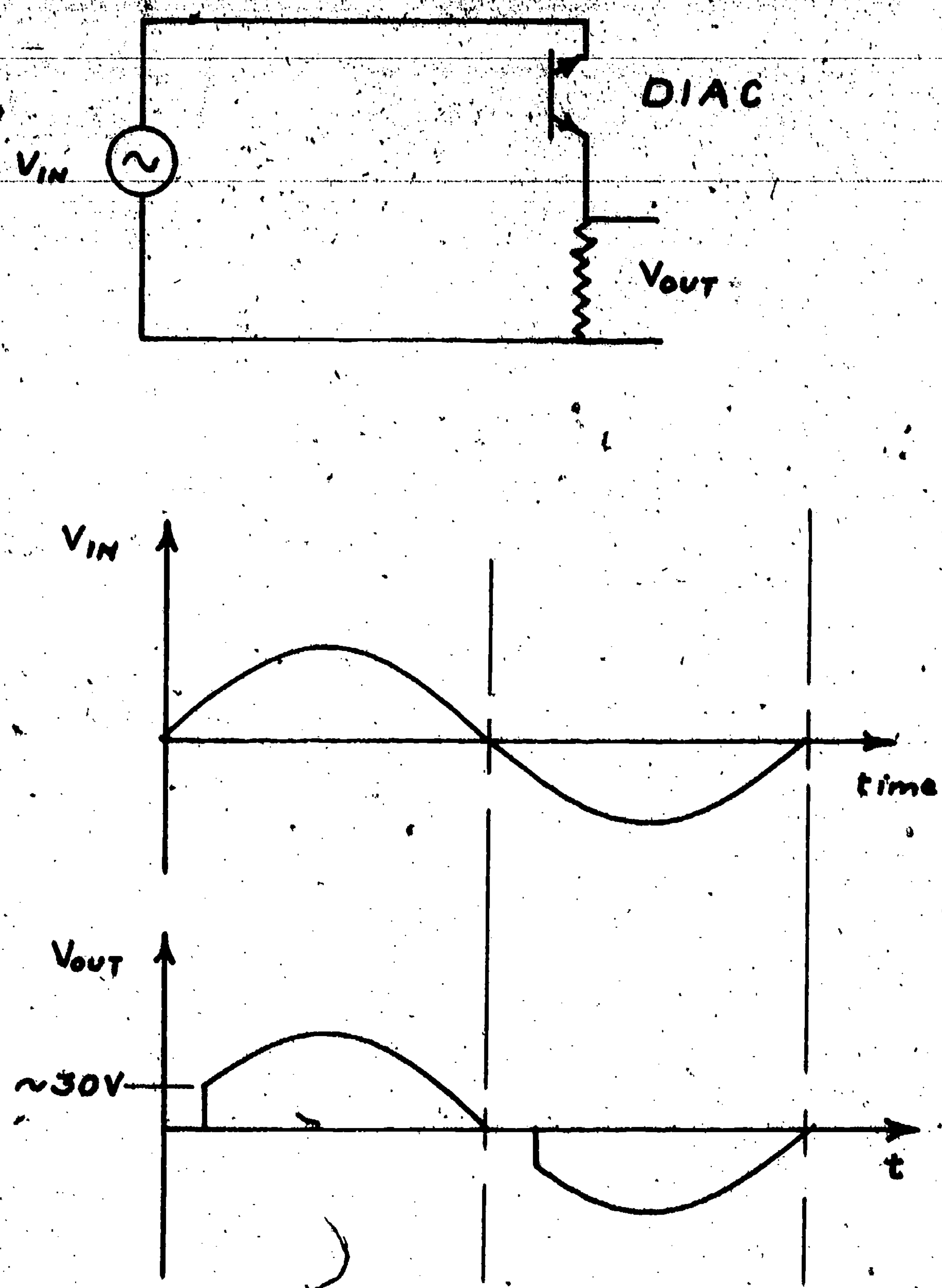


Figure 2 - 28

Another device that deserves special mention in regard to SCR triggering circuits is the "unijunction transistor" (also called a "double-base diode"). This device can be used in circuits to provide time delays, integrators, oscillators, and level detectors. In addition, the output waveforms of relatively simple unijunction circuits are reasonable approximations to the "ideal" triggering waveform for the SCR. The unijunction transistor (UJT) is a three terminal device whose generalized V-I characteristics are shown in figure 2 - 29.

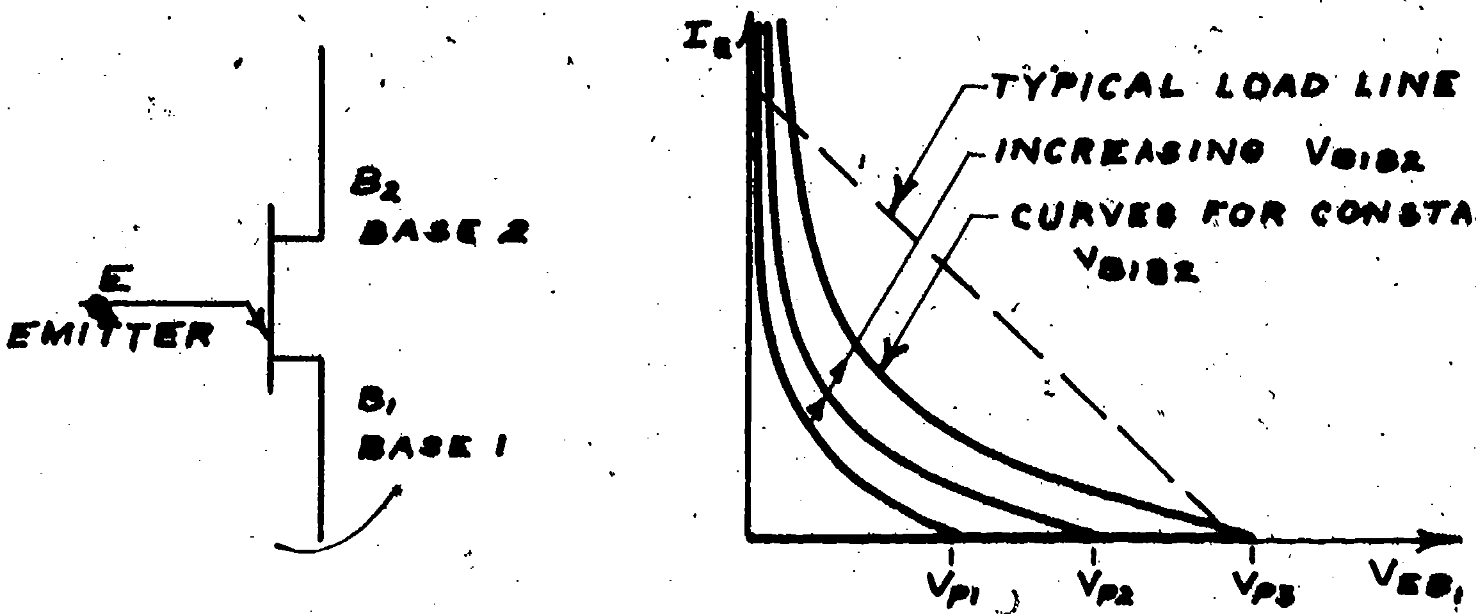


Figure 2 - 29

Note that the characteristics usually presented for a UJT are the emitter-base 1 characteristics. The peak emitter-base voltages are proportional to the base-to-base voltage. The constant of proportionality "η" is called the "intrinsic standoff ratio" and usually has a value near 0.5-0.6. Once the emitter-base voltage has exceeded the peak value, large emitter currents may flow.

In order to gain some insight into the UJT characteristics, we consider the physical electron-hole model (Fig. 2 - 30). The bar of material between base terminals is a lightly doped, n-type semiconductor. With no connection made to the emitter terminal, the bar acts as a resistor, usually having a resistance value on the order of thousands of ohms. The emitter terminal is connected to a p-type "dot" near the middle of the bar.

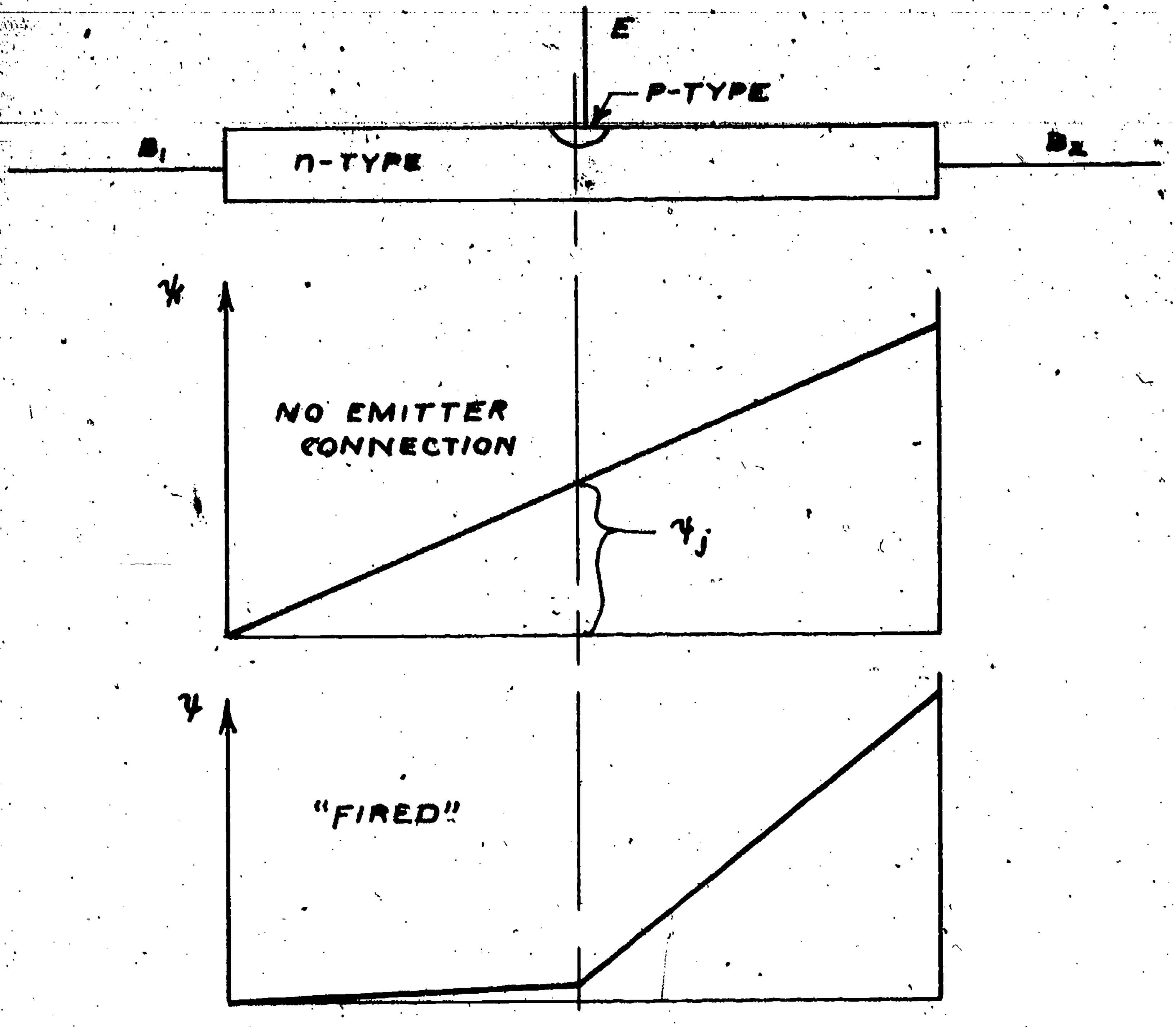


Figure 2 - 30

When a fixed potential difference is applied to terminals B1 and B2 (B2 positive), the potential distribution along the resistive bar appears as in figure 2 - 30b. If (after having taken into account the metal-semiconductor contact potentials) the potential difference V_{EB_1} is smaller than ψ_j, the p-n junction will be reverse biased, and only a small leakage current will flow. As the potential difference V_{EB_1} is increased, eventually the p-n junction will become forward biased. As holes diffuse from the p-type material into the n-type bar, they are swept toward B1 by the drift field due to the potential difference $V_{B2\ B1}$ applied to the bar. The mobile carriers then decrease the resistivity (conductivity modulation due to increased carrier density) of the bar between E and B1. This increases the current flow through E, reducing the resistivity of the bar between E and B1, increasing the current I_E, etc. Eventually (microseconds), the external circuit limits the current flow, and the potential difference between E and B1 is small. The "Resistance" or $\frac{V_{E\ B1}}{I_E}$ of the emitter-base terminals is typically on the order of a few tens of ohms after the UJT has fired. If the emitter-base 1 voltage is decreased until the p-n junction is reverse biased, the initial state (fig. 2-30b) will be restored after the carriers between E and B1 have been "swept out" by the electrostatic field or have recombined.

As an example of a UJT triggering circuit, we consider a circuit that could be (and has been) used as a light dimmer (Fig. 2-31).

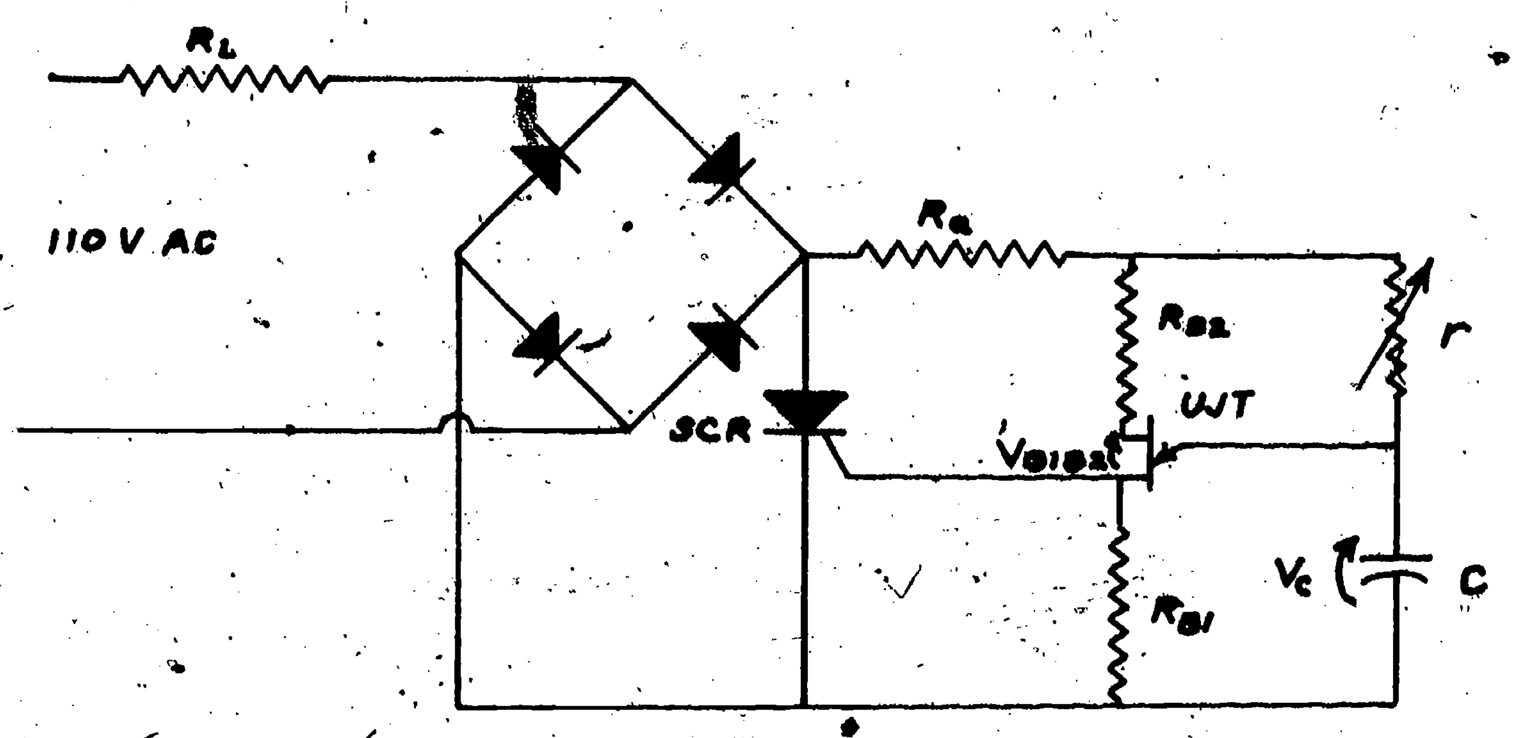


Figure 2 - 31

The current through the load resistor (R_L) can be varied by adjusting the value of the resistor "r". To analyze the circuit, we first consider the voltage waveform that exists across the SCR assuming the SCR is never triggered. This voltage is a full wave rectified sine wave as in figure 2 - 32a. If the SCR is triggered once during each half cycle, the voltage across the SCR should appear as in figure 2 - 32b. We assume perfect on-off action of the diodes and the SCR for simplicity. R_a serves to lower the voltage V_{BB} across the UJT to its normal working voltage (around 40 V). R_{B2} and R_{B1} are chosen on the basis of the UJT specifications so as to limit the current when the UJT is conducting.

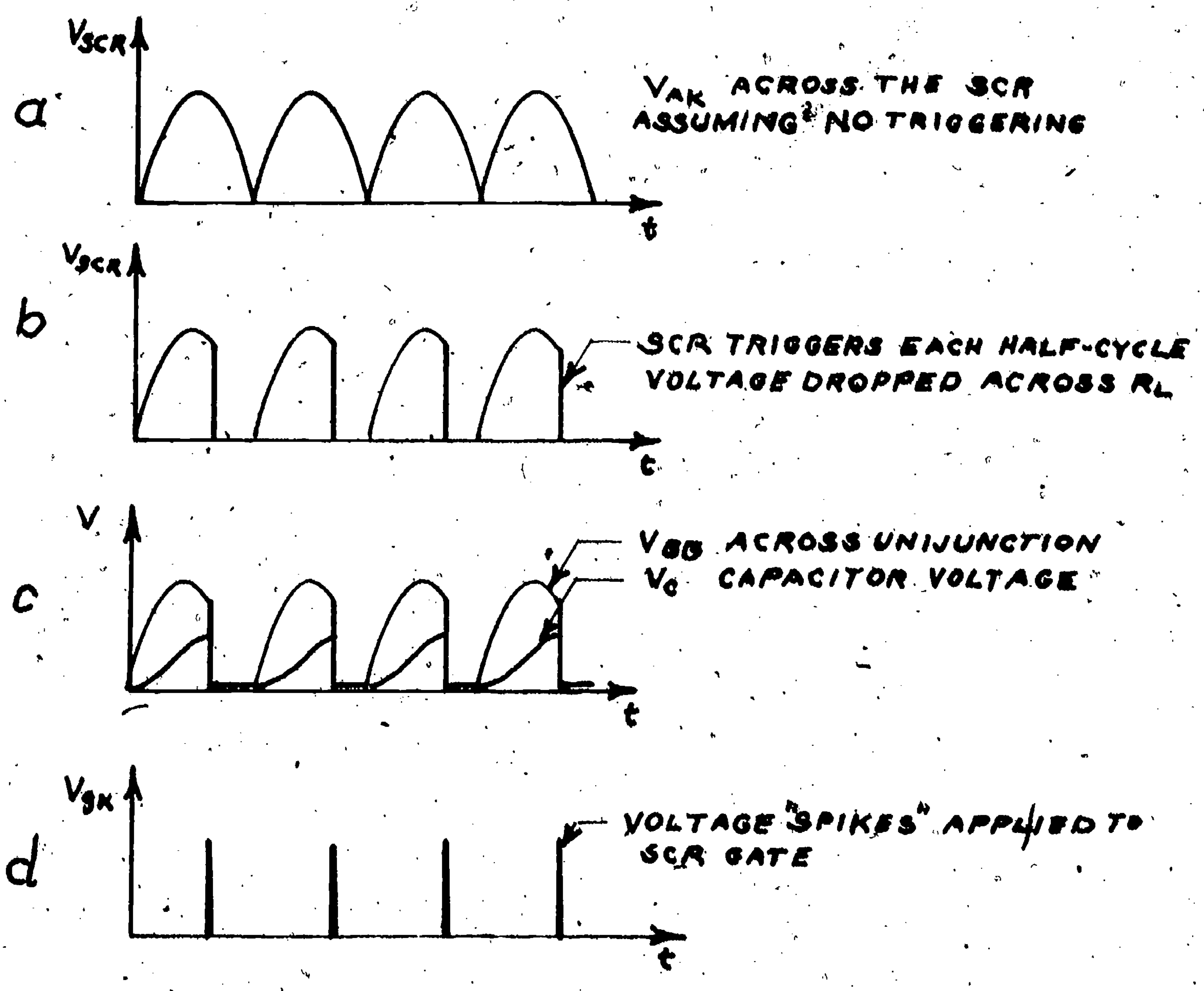


Figure 2 - 32

The resistances R_{B2} and R_{B1} are usually much less than the resistance of the UJT between terminals B2 and B1. Thus the UJT base-to-base voltage V_{BB} is essentially equal to the voltage across r and C combined. The capacitor voltage V_c continues to increase during each half cycle as long as $V_{BB} > V_c$. Eventually, the capacitor voltage reaches the value ηV_{BB}, and the unijunction fires. η is assumed to be 0.6 in figure 2 - 32c. When the UJT fires, the capacitor discharges through the emitter terminal of the UJT, sending a current pulse into the gate of the SCR. If the capacitor and the UJT peak voltage are large enough, the SCR will be triggered. The gate current and pulse duration are determined by the UJT characteristics, the SCR gate characteristics, the value of R_{B1} and of course, C. Once the SCR has been triggered, the voltage across the triggering circuit is just the forward voltage drop across the conducting SCR. Therefore the capacitor does not charge up during the remaining part of the half-cycle. At the end of the half-cycle, the SCR anode current goes to zero since the source voltage goes to zero. The SCR turns off and the circuit is prepared to repeat the cycle. This example, outlined in the crudest of models, shows one of the many ways UJT's are utilized in trigger circuits. The UJT is a particularly versatile device, and the student who is seriously interested in electronics, logic circuits, or analog computing circuits is encouraged to further study this device.

It is frequently desirable to provide direct current isolation between the gate of an SCR and the trigger circuit. A pulse transformer connected to the gate (Fig. 2 - 33) is one commonly used method providing the desired isolation.

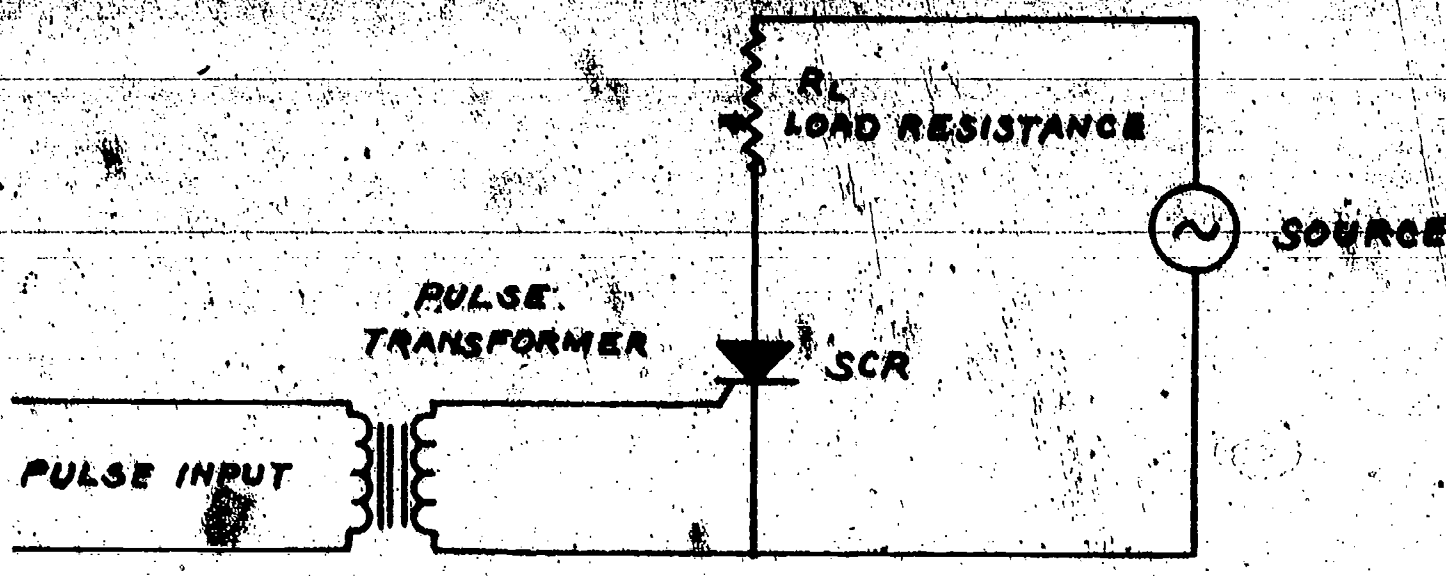


Figure 2 - 33

In the cases where the size and weight of the pulse transformer are critical, a "picket fence" gate voltage waveshape may be chosen. The picket fence (Fig 2-34a) is an approximation to a desirable triggering waveform such as shown in figure 2 - 34b.

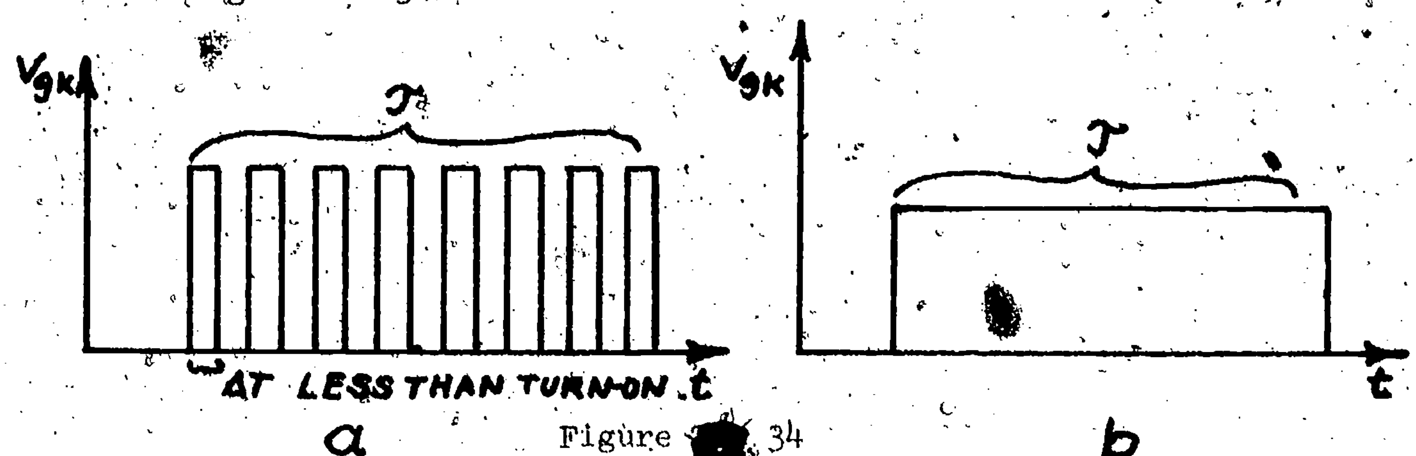


Figure 2 - 34

The size and weight of a pulse transformer are related to the volt-time integral of the pulse waveform the transformer must deliver to a given impedance. This statement is based on the fact that iron-core (even exotic ferrous materials) have a maximum useful flux density "B_{max}" for a given application. The flux "ϕ" is roughly equal to the flux density times the area of the iron core.

$$\phi = BA$$

The voltage "e" appearing across the transformer terminals is related to the flux by Faraday's low

$$e = -n\frac{d\phi}{dt}$$

Integrating,

$$\int_{t_1}^{t_2} e\,dt = -n(\phi_2 - \phi_1) \approx -nA(B_2 - B_1)$$

As the volt-time integral of a pulse increases, the maximum $B_2 - B_1$ being limited by the characteristics of the iron core, n A must increase. If n increases, more turns of thicker wire (keeping winding resistance the same) imply a larger, heavier transformer. Also, A increasing implies a larger transformer. Obviously the volt-time integral of one spike of the picket fence is very much smaller than the volt-time area of the entire gate pulse. The transformer flux "recovers" or is "reset" between the spikes of the picket fence. Hence a smaller, lighter pulse transformer may be utilized with the picket-fence waveform than with a single gate pulse.

The triggering circuits presented in this chapter have not been presented in sufficient detail to represent more than the germ of an idea. Even the UJT circuit, which was primarily intended to introduce the student to the characteristics of a UJT, was modeled in only the crudest terms. A much more careful modeling would be required to design such a circuit. The basic ideas this section on triggering circuits is intended to convey are:

a) The designer of a trigger circuit must interpret the gate data carefully,

b) the characteristics of the anode circuit as well as the gate circuit must be considered,

c) the trigger circuit may contain a significant portion (or all) of the logic circuitry and sensing circuits used to control the SCR.

d) the trigger circuit to be used must be selected from the infinite variety of possible trigger circuits on the basis of the specific application.

A really intelligent choice of a trigger circuit requires experience, creativity, and most important; a careful evaluation of the "desirable features" of the circuit. For example, using a picket-fence triggering waveform involves knowing the value of transformer weight reduction in terms of circuit complexity, cost, and reliability.

SCR Turn-off Circuits

Once the SCR has been triggered, the gate loses control over the device. In order to turn the SCR off the anode current must be reduced below the latching current (reducing current density and the α's so that $\alpha_1 + \alpha_2 < 1$) or by making the anode negative with respect to the cathode (instituting reverse blocking after the current transient due to carrier storage and junction capacitance has passed). When the SCR anode circuit is connected to an AC source, the alternations of the voltage during each half cycle serve to turn off the SCR. If the SCR is connected to a DC source or if the SCR must be turned off during a half cycle of an AC source, some sort of "turn-off circuit" must be employed.

The circuits that reduce the anode current to a value below the latching current are said to "starve" the SCR. Two simple "starvation mode" turnoff circuits are shown in figure 2 - 35.

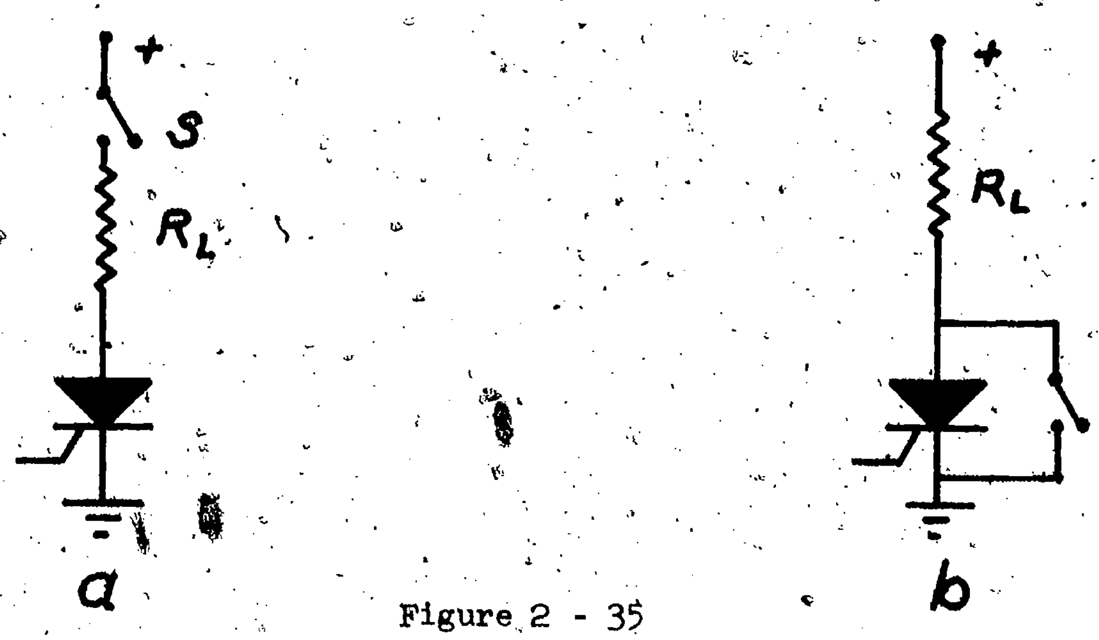


Figure 2 - 35

Opening the switch of figure 2 - 35a obviously reduces the anode current to zero. Closing the switch in figure 2 - 35b bypasses the load current around the SCR. After the SCR has turned off, the switch is opened. These circuits are seldom used except at low power levels (less than some hundreds of watts capability) where switching transistors can take the place of the switches. A germanium transistor would be preferred over a silicon transistor for the circuit of 2 - 35b since germanium transistors have a lower collector-emitter voltage drop when driven "on".

Applying a reverse voltage to the SCR until it has turned off requires circuitry more complicated than the switch used in the starvation made turn-off. Such circuits are turned "commutating circuits" because they are used to "switch" the SCR anode current. The commutating circuit may be an integral part of the SCR application circuit (as in the case of the phase-controlled rectifier-inverter of chapter 3) or it may exist as an SCR appendage, functioning only to turn the SCR off. There is a large variety of commutating circuits including a) underdamped LC circuits (Fig. 2 - 36a) which are said to be self-commutating because no other switch device is necessary, b) capacitor commutated circuits (Fig. 2 - 36b) where a charged capacitor is switched across the conducting SCR (The switch is usually another SCR), c) combinations of the two schemes as in figure 2 - 36c, d) external pulsed circuits such as shown in figure 2 - 36d, and so on in endless variety.

TURN-OFF CIRCUITS

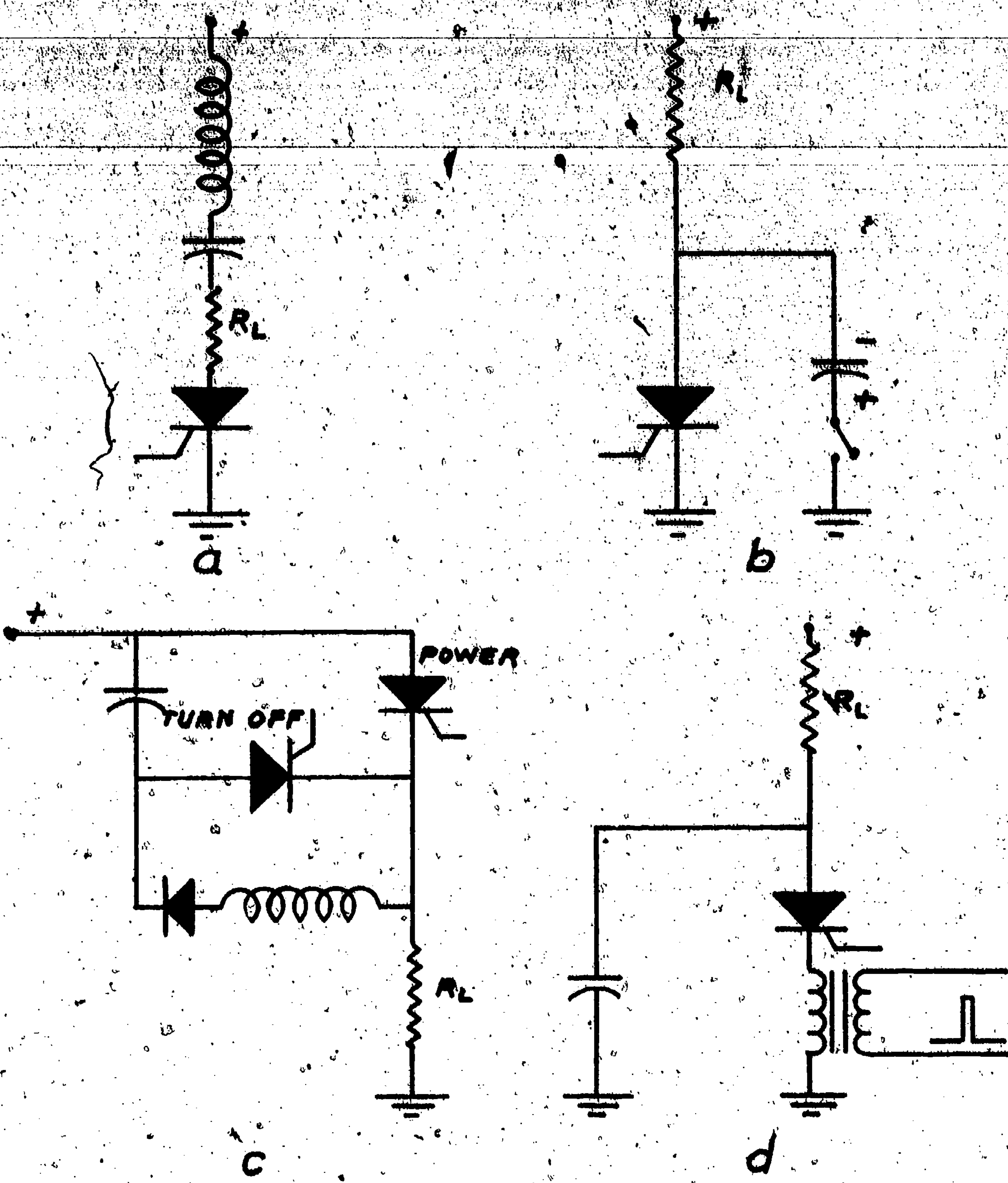


Figure 2 - 36

For the sake of example, we consider a particular commutation circuit (Fig. 2 - 37). This particular circuit is symmetric; however, the circuit can be made assymmetric by choosing SCR_2 of a much smaller current capacity than SCR_1 and choosing the resistor connected to SCR_2 to have a much higher resistance value than the resistor connected to SCR.

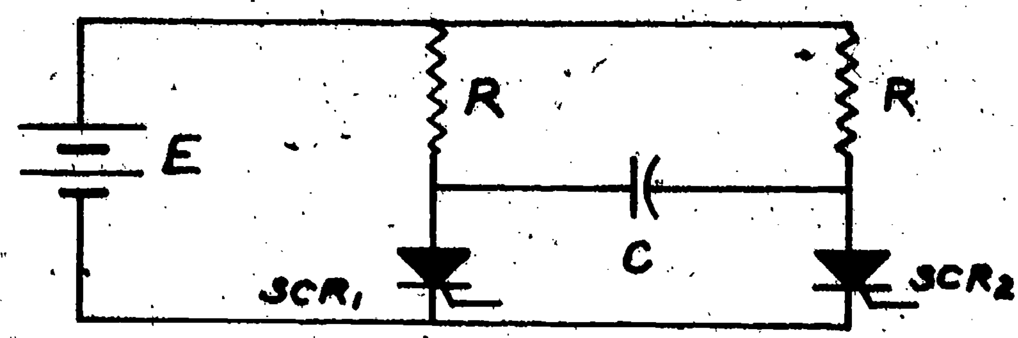


Figure 2 - 37

The purpose of the circuit of figure 2 - 37 is to shift the flow of current from one resistor to the other by alternately triggering the SCR's. The SCR's do not have to be triggered periodically, but may be triggered at any time intervals that are long compared to RC. The desired current waveforms are shown in figure 2 - 38.

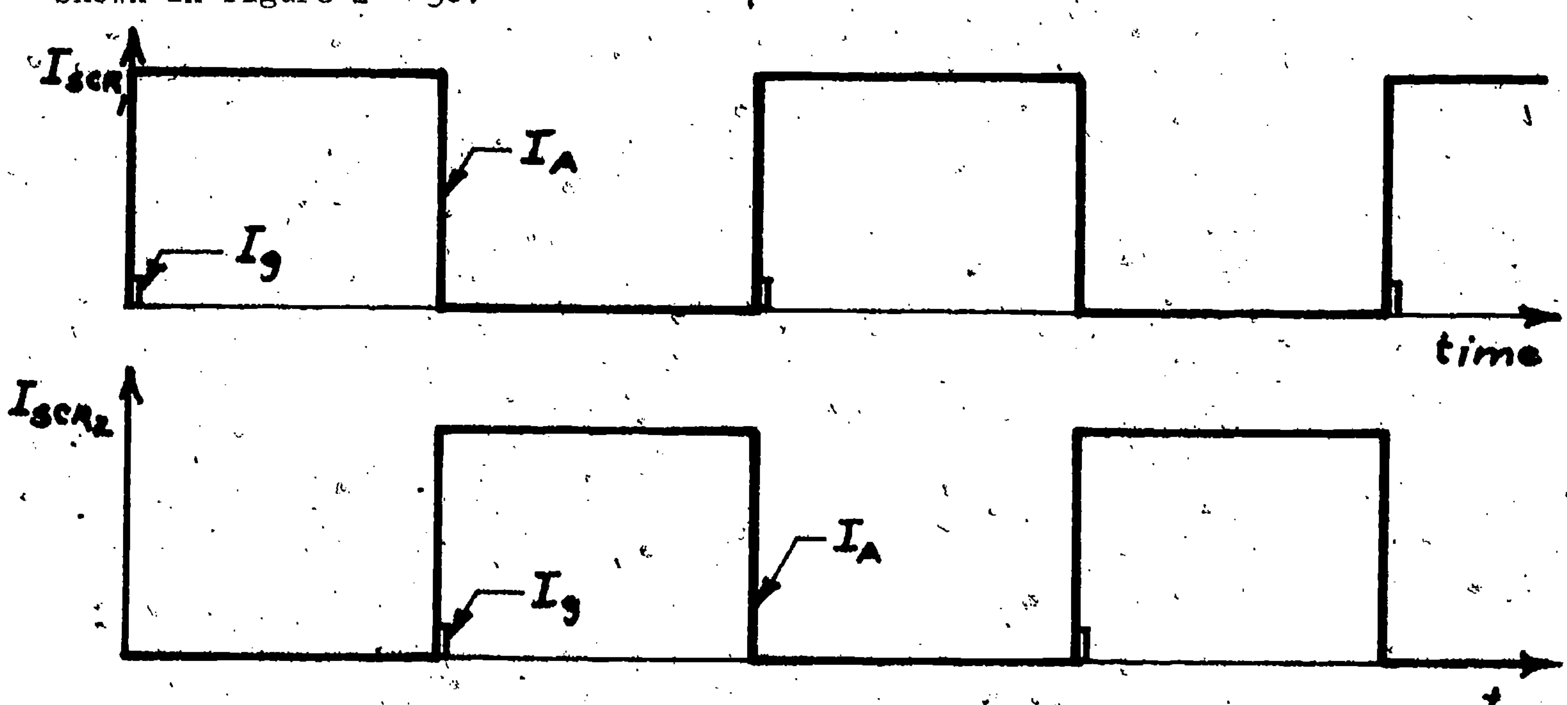


Figure 2 - 38

We shall define the initial conditions of this circuit as SCR_1 conducting, SCR_2 forward blocking, and assume this situation has existed for some long time period. A transient will be initiated by triggering SCR_2. The expected steady state is SCR_2 conducting, SCR_1 forward blocking. The decision to choose these initial conditions and the expected steady state come from examining the circuit and knowing the purpose of the circuit. We had five possible steady states: 1) both SCR's conducting, 2) both SCR's blocking, 3) SCR_1 conducts while SCR_2 blocks, 4) SCR_2 conducts while SCR blocks, 5) the circuit might oscillate. Knowing the function of the circuit, we eliminate possibilities 1, 2, and 5 as circuit malfunctions at best. By examining the circuit and observing its basic symmetry we suppose that states 3 and 4 represent likely steady states. We arbitrarily choose 3 as the condition of the circuit and will examine the transient involved in getting to state 4. The identification of initial and steady states may seem ridiculously trivial in this simple circuit, however, such a process aids in understanding and modeling the circuit. We are "lucky" that the problem is so trivial, for many circuits exist having hundreds of possible steady states and simplifying the problem by choosing the appropriate states is no easy task.

Let us further assume that E is much larger than the forward voltage drop across an SCR when it is conducting, and that the SCR current when the SCR is forward blocking is very small compared with $^E/R$. Such assumptions <u>may</u> permit us to treat the SCR's as perfect switches during some part of the time under consideration. We start with the initial conditions (Fig. 2 - 39).

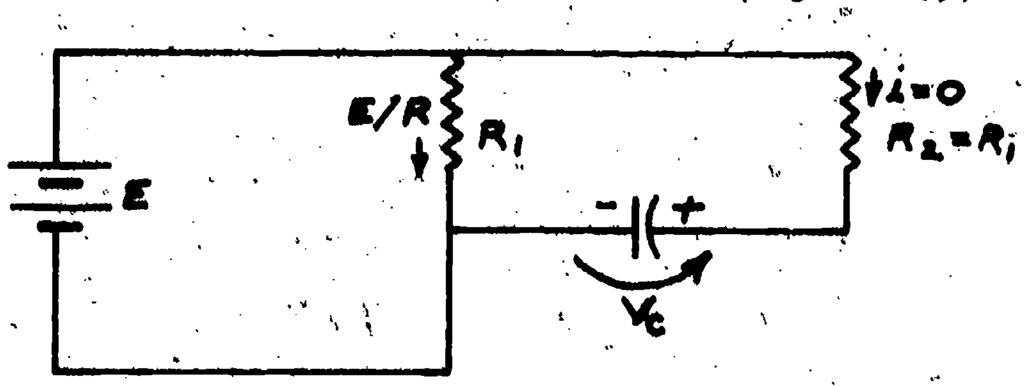


Figure 2 - 39

Initially, a current = E/R is flowing through R_1 since SCR_1 is conducting. The only current through R_2 would be the blocking current of SCR_2 which is negligible (by assumption). The capacitor is charged to the voltage E with the polarity shown in figure 2 - 39. Next, a triggering pulse is sent to the gate of SCR_2 which turns on in a few microseconds. Note that before turn-on, the forward blocking voltage on SCR_2 was $E = V_c$. When SCR_2 turns on, V_c is applied to SCR_1 (Fig. 2 - 40), turning off SCR_2.

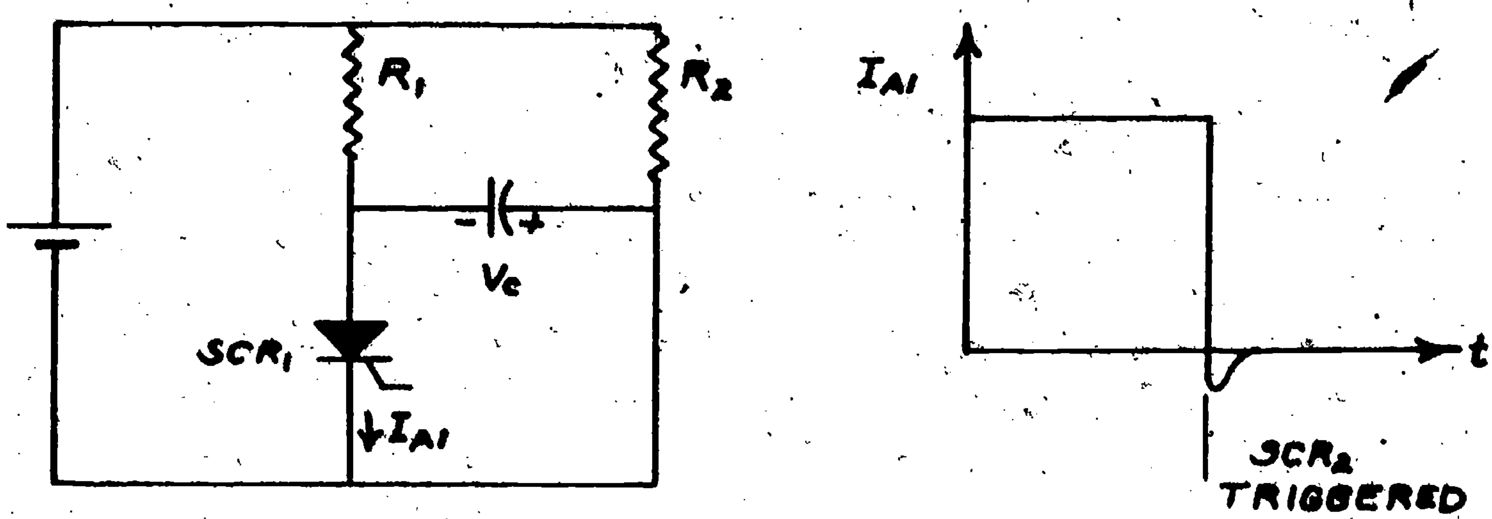


Figure 2 - 40

C partially discharges through SCR_1 and R_1 as the mobile carriers in SCR_1 are being swept out of the semiconductor and recombining. When SCR_1 no longer accepts current, the voltage across capacitor C continues to change with time constant R_1C. If the rate of change of voltage dV/dt is below that value that will retrigger SCR_1 (known form the SCR data sheet = $\frac{d\,V_{SCR}}{dt}_{max}$), and if the capacitor is large enough that the SCR current ceases before the voltage polarity across the SCR is in the forward direction, the SCR_1 will be able to block the forward voltage, and steady state will be reached in about 5 time constants ($R_1C = \tau$) is shown in figure 2 - 41.

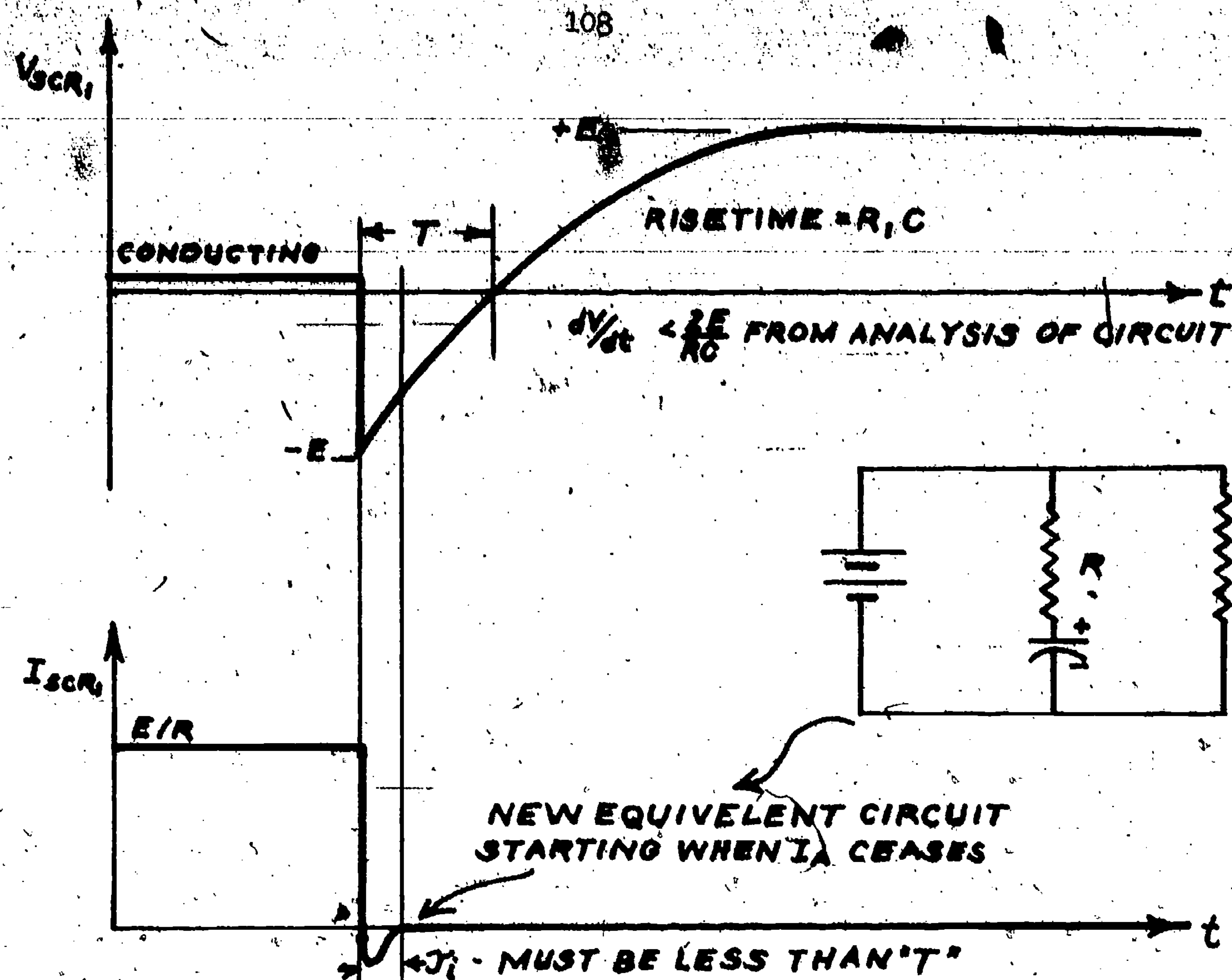
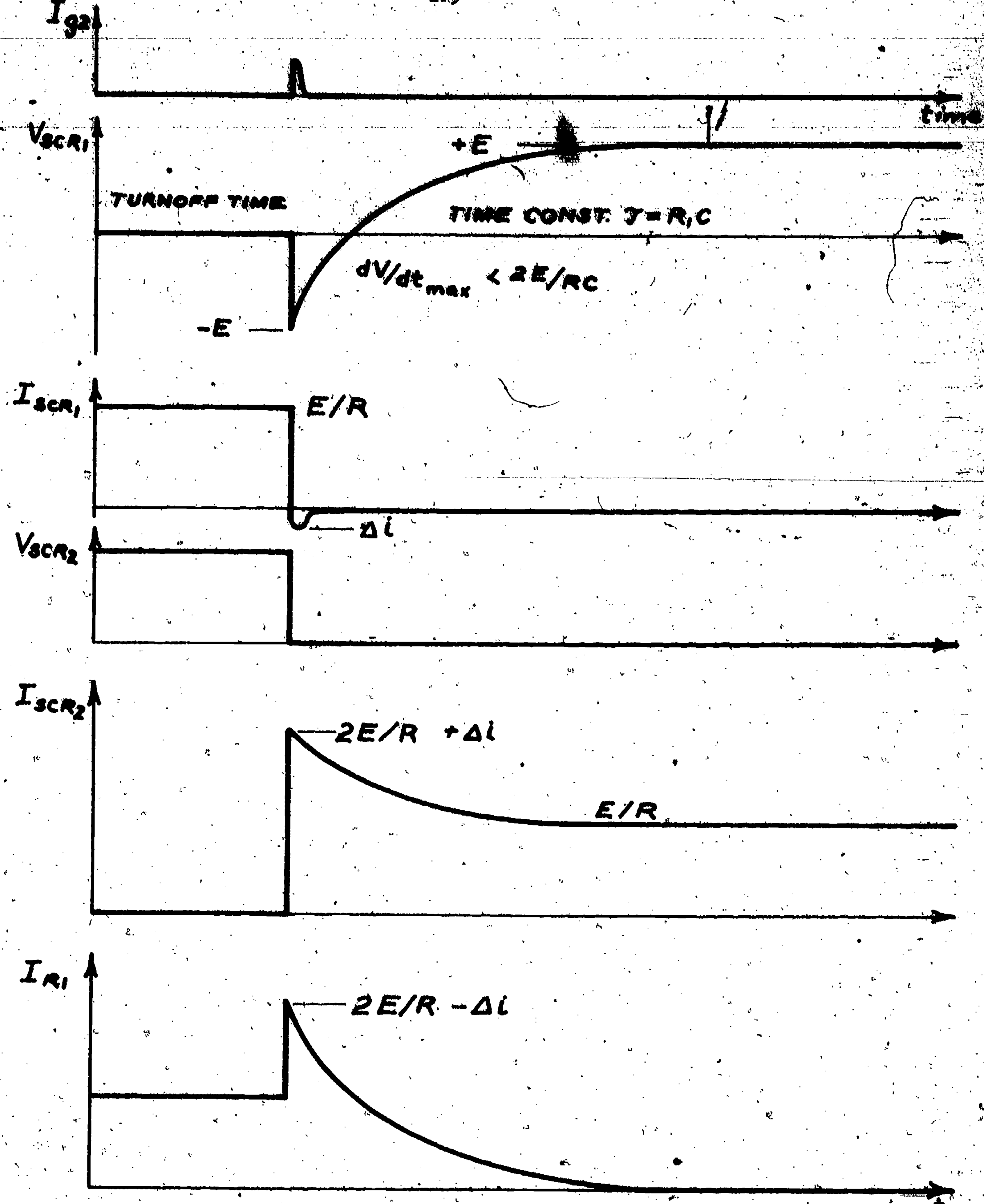


Figure 2 - 41

Now, using Kirchoff's voltage and current laws, we can plot all the significant voltages and currents in the circuit (Fig. 2 - 42).



I_{g2}
time
V_{SCR_1}
+E
TURNOFF TIME
TIME CONST. τ = R1C
dV/dt max < 2E/RC
-E
I_{SCR_1}
E/R
Δi
V_{SCR_2}
I_{SCR_2}
2E/R +Δi
E/R
I_{R_1}
2E/R -Δi
t

Figure 2 - 42

In designing the circuit, the maximum $\frac{dV}{dt}$ across the SCR_1 gives one design equation, ie:

$$\frac{2E}{RC} < \frac{dV}{dt}_{max}$$

That the capacitor is sufficiently large to turn off the SCR (absorb the reverse current during turnoff) is usually determined by trial and error. Finally, the turn on $\frac{dI_A}{dt}$ may have to be limited by placing some inductance in series with each SCR. It is assumed that R is a known load. We have considered in this example one of the simplest turn-off circuits. Reference 3 at the end of this chapter is especially recommended for suggestions of a variety of turn-off circuits.

Summary

In this chapter, several models of the SCR have been presented as an aid in remembering and understanding the circuit characteristics of the SCR. Once the SCR has been modeled, not only do the published data and specifications of the SCR become more intelligible, but device behavior not normally specified on the available data sheets can be anticipated. Triggering circuits have been presented and discussed, not from a detailed design or analysis point of view, but to impart some idea of the kinds of things that are typically considered in choosing or designing such circuits. The unijunction transistor has been presented in some detail because it is a commonly used device in logic and trigger circuits for SCR's. The UJT's characteristics are drastically different from those of the transistor, and few students are familiar with the device. Finally, in the analysis of an SCR turn-off circuit, an important concept in modeling non-linear circuits is presented, which is, a single circuit may have a variety of possible steady-states, and the particular steady-state reached at the end of a transient depends on the initial circuit conditions and the disturbing (or triggering) signal. In this context, switches that are time dependent, voltage dependent, or current dependent are considered nonlinear elements as are diodes, SCR's, transistors, UJT's, etc.

Exercises

1. Given that the doping densities of an SCR starting at the anode are:

anode	$N_A = 10^{19}$ atoms/cc,	
	$N_D = 10^{15}$,	
gate	$N_A = 10^{17}$,	
cathode	$N_D = 10^{19}$,	

and that the SCR is forward blocking with $V_{AK} = 500$ volts, find the width of the depletion region for the blocking junction (2), and the gate-cathode junction (3).

2. Estimate the dielectric strength of silicon knowing that in the SCR of exercise 1, when the SCR is reverse blocking 500 volts, the gate-to-cathode voltage is 14 volts as measured by an FET VOM (input impedance $= 10^9$ ohms).

3. Show that the rate of change of voltage across the SCR of figure 2 - 42 during "recovery" is 2E/(RC).

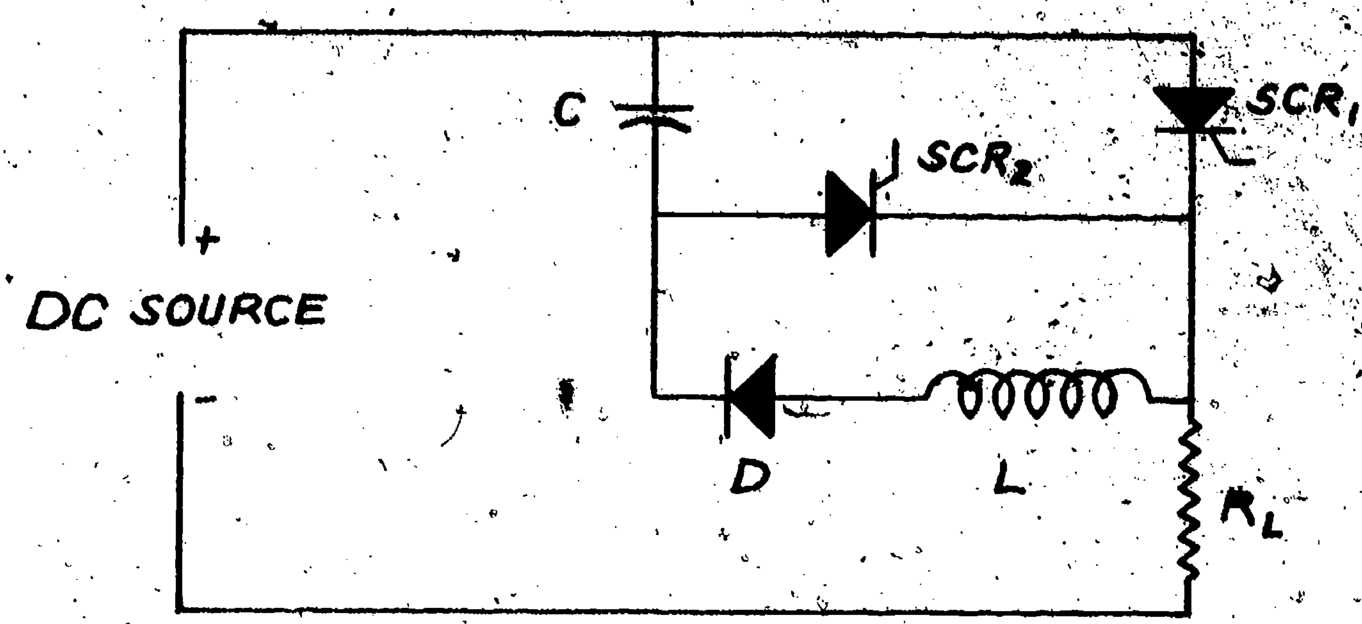
Problem 1

The following circuit is a commonly used SCR "turn-off" circuit in which SCR_2 is triggered in order to turn off SCR_1. You are asked to plot the appropriate voltage and current waveforms that describe the intended operation of this circuit and answer the following questions.

a) Does the circuit design depend on the value of R_L in ways other than selecting the current capabilities of the two SCR's?

b) Sometimes the diode is replaced by a third SCR that is triggered simultaneously with SCR_1. What could be the advantage of such a scheme?

c) Will the circuit operate properly the first time the source is applied or must some additional circuitry be added to establish the "proper steady state"?



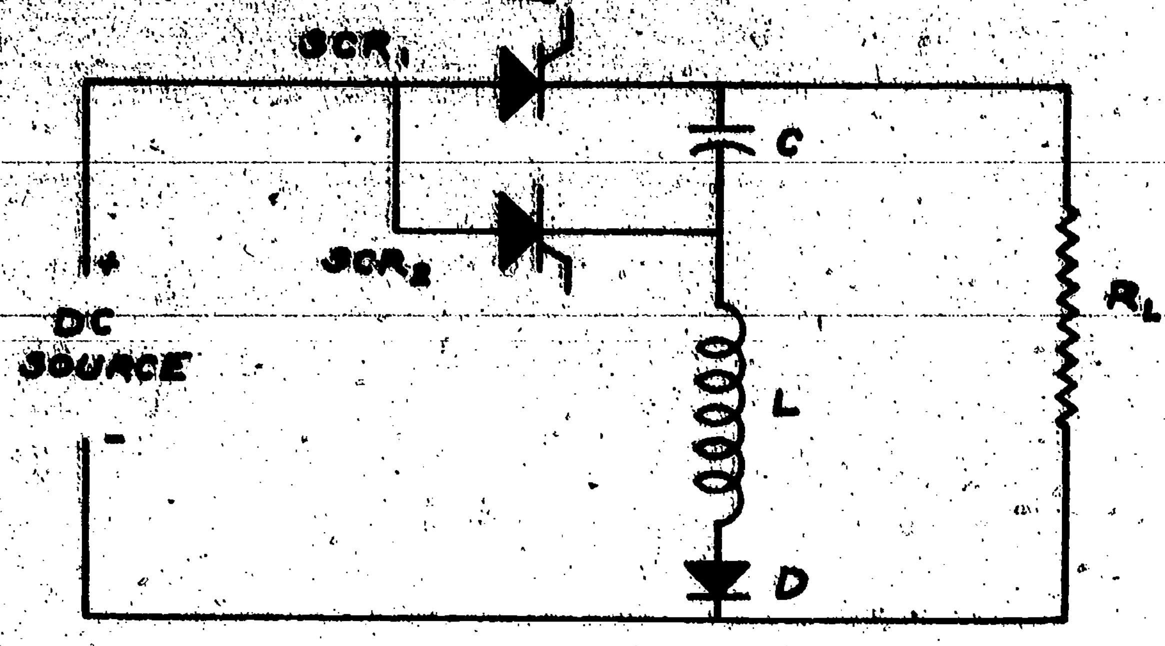
Problem 2

The following circuit is proposed as an SCR turn-off circuit. SCR_2 is triggered in order to turn off SCR_1. Later SCR_1 is triggered to turn off SCR_2. You are asked to plot the appropriate voltage and current waveforms that describe the circuit operation and answer the following questions.

a) How will the diode voltage rating be related to the DC supply voltage?

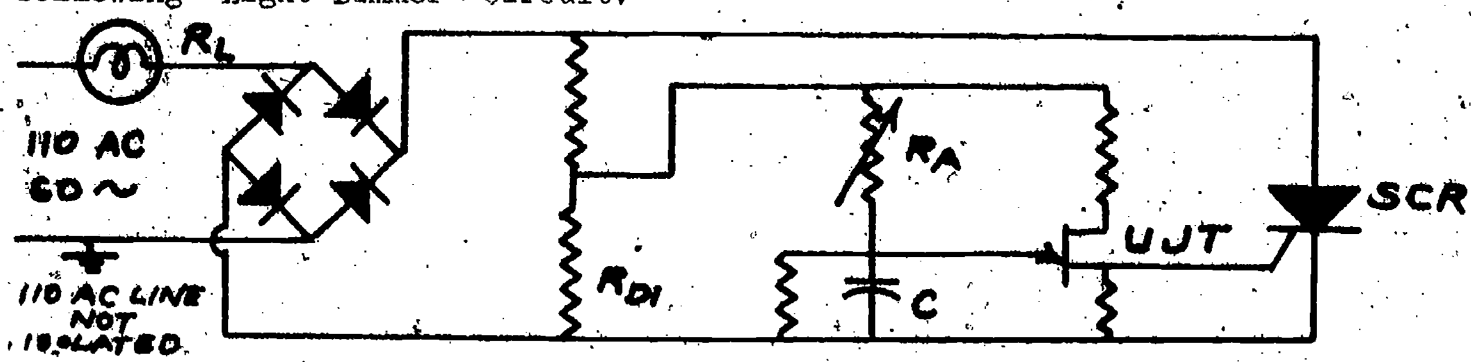
b) How should the circuit be modeled if the time period between SCR firings is very long?

c) Will the circuit operate properly the first time the source is connected or is some additional circuitry required?



Laboratory Problem 1

We are specifically interested in SCR triggering circuits. It is desired to control the current through a load resistanced R_L, representing a 150 watt light bulb (110V). The SCR controller is to be used to vary the "dimness" of the light from no light at all to full "lighting power". An examination of several encyclopedias of electronic circuits produces the following "Light Dimmer" circuit.



You are to design a working circuit. It is not necessary to minimize cost. The data sheets for the available diodes, SCR's, and UJT's can be obtained at the laboratory. You are specifically asked to present a "typical" firing angle vs resistance R_a curve, and a light bulb current vs R_a curve. There are two small difficulties you must consider. If you construct the above circuit, it will be difficult to make oscilloscope measurements in the triggering circuit because of grounding problems (all available "scopes" have one input terminal grounded), and the value of R_L will change as a function of current.

Laboratory Problem 2

Some types of triggering circuits continue to supply pulses to the SCR even when the SCR is reverse blocking. Such pulses add to the average power dissipated in the SCR gate circuit. Furthermore, the gate characteristics may seriously affect the triggering circuit. Therefore it is desired to determine the gate V-I characteristics while the SCR is reverse blocking. Will the gate characteristics depend on the anode-cathode voltage? Will the connection of a resistor between the gate and cathode terminals significantly affect the blocking capability of the SCR? Try to relate your answers to SCR "models". Data sheets for a 100 amp, 500 volt SCR are available at the laboratory.

References

1). J. Seymour, Semiconductor Devices in Power Engineering, Sir Isaac Pitman and Sons, London, 1968.
This book contains a brief but clear summary of the characteristics of the SCR.

2). F. E. Gentry, F. W. Gutzwiller, Nick Holonyak, Jr., and E. E. Von Zastrow, Semiconductor Controlled Rectifiers, Prentice-Hall, Englewood Cliffs, N.J., 1964.
This book is an excellant and detailed review of SCR theory and characteristics.

3). F. W. Gutzwiller, Ed., G-E SCR Manual, General Electric, Electronics Park, Syracuse 1, N.Y., 1967.
This is a "designer's" manual containing a good summary of device operation, typical applications, terminology, and summarized spec. sheets of G-E SCR's.

4). Robert Murray, Jr., Ed., Westinghouse Silicon Controlled Rectifier Designer's Handbook, Westinghouse Electric Corp., Youngwood, Pa., 1964.
Similar to the GE Handbook, this handbook devotes considerable attention to heat sink specifications.

Chapter 3 Modeling a Line-Voltage Commutated Inverter

Introduction

Using the medium of analyzing a circuit that inverts DC power into AC power, this chapter introduces two major ideas. First is the usefulness of average values of voltage and current in problems where the same cycle is repeated again and again in a steady state, and second is the idea of "iterative modeling" where the model is successively refined until the answers yielded by the model have the required accuracy. Secondary objectives of the chapter are a discussion of Kirchoff's Laws as applied to a switching circuit, a review of the modeling ideas presented in Chapter 1, and modeling a circuit containing an SCR.

The Line Voltage Commutated Inverter

While "inverter" is frequently used as a generic name for power conversion circuits, we adopt the following commonly used definitions:

Rectifier - changes AC to DC

Inverter - changes DC to AC

Converter - changes DC to DC at a different voltage

A converter could be made by connecting a rectifier to an inverter. These are not universally accepted definitions but are sufficiently common and logical that their meaning will usually be clear in context.

A line-voltage commutated inverter circuit is shown in figure 3 - 1. The phrase "line-voltage commutated" implies that the AC line voltage in some way turns off or switches the current from one SCR to the other. The circuit is intended to transfer energy from the battery to the AC line. The circuit may also be used to transfer energy from the AC line to the battery (an efficient battery charger), and when so used is known as a "phase-controlled rectifier" because the charging current depends on the phase of the SCR triggering signals with respect to the phase of the line voltage. In either case, the SCR's are triggered alternately and 180° (of line voltage ωt) apart. The phase of the triggering signal may be varied from 0° - 180° with respect to the line voltage phase. Knowing the purpose of the circuit, we now begin an analysis of the circuit operation.

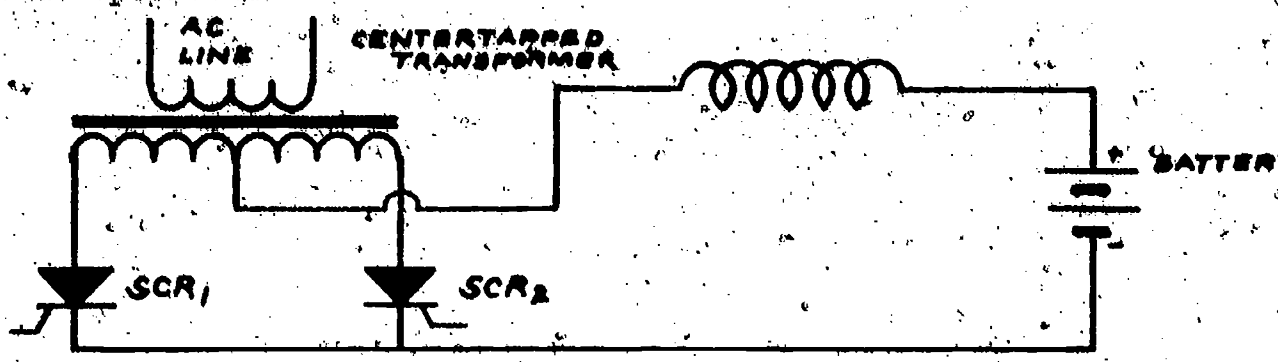
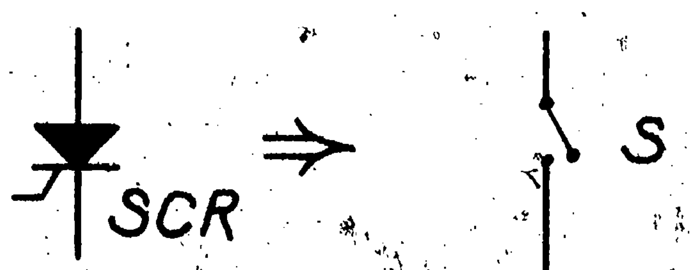


Figure 3 - 1

Modeling the Circuit - #1

Despite the fact that the circuit contains only a few elements, the circuit operation is not obvious. We begin analysis by modeling or idealizing the circuit. The assumptions or models are spelled out so that results inconsistent with assumptions can be easily identified.

1. The particular voltage and current values are not specified, so we choose to work at voltages large compared to the forward voltage drop across the conducting SCR and currents large compared to the blocking (forward or reverse) current of the SCR. RMS voltages between the transformer tap and one side of 100 volts or larger and currents of 10 amps RMS or more will easily meet our requirements and give some "feeling" of the magnitudes and sizes of components. Much smaller voltages or currents would suggest transistor circuitry instead of SCR circuitry. Our first assumption is that each SCR can be modeled as an ideal switch.



2. The inductor has a resistance associated with its physical winding. Although we are not yet certain of the purpose of the inductor, as a first try we assume the resistance is negligible (in fact, zero). We also assume the inductor is "ideal", that is, its iron core does not "saturate" and the flux in the inductor core is always proportional to the inductor current.

n = number of turns of wire

φ = inductor magnetic flux

I = current in inductor winding

$$L = n\frac{\phi}{I}$$

Thus e_L, the voltage across the inductor terminals is equal to $-L\frac{dI}{dt}$

since:

by Faraday

$$e = -n\frac{d\phi}{dt}$$

but $\phi = L\,I/n$ by assumption

and $\frac{d\phi}{dt} = \frac{L}{n}\ \frac{dI}{dt}$, L and n constant

therefore,

$$e = -L\frac{dI}{dt}$$ by substitution.

3. We assume the internal resistance of the battery is negligible (zero). Normally, the iR drop due to a battery internal resistance is small compared to the battery emf. The internal resistance of a lead-acid automobile battery for example is only a few thousandths of an ohm for a voltage of 6 volts and an ampere hour capacity of 100 A. hr. At the recommended charge-discharge rate of 10 amps, the internal iR drop changes the terminal voltage by less than ½%.

4. We assume an "ideal" transformer with negligible leakage reactance and winding resistance. Examination of transformer design methods suggests that the impedance due to leakage reactance and winding resistance combined is usually

less than 5% PU (Per Unit - ref 3 p. 404). We neglect the voltage drops due to this impedance as a first try. If we further assume the core losses are negligible and the transformer iron doesn't saturate, we can use the ideal transformer relation:

$$N_1 i_1 + N_2 i_2 + N_3 i_3 = \oint \vec{H} \cdot d\vec{l} = 0$$ by assumption directly from Amperes Law

$$\oint \vec{H} \cdot d\vec{l} = \sum_j n_j i_j$$

where the current directions are defined as in figure 3 - 2.

Figure 3 - 2

5. Finally, we assume that the AC line is "stiff", that is, has negligible source impedance.

Now that the elements have been idealized, we begin to analyze the modeled circuit. We first label the voltages and currents of interest for ease in discussion (Fig. 3 - 3).

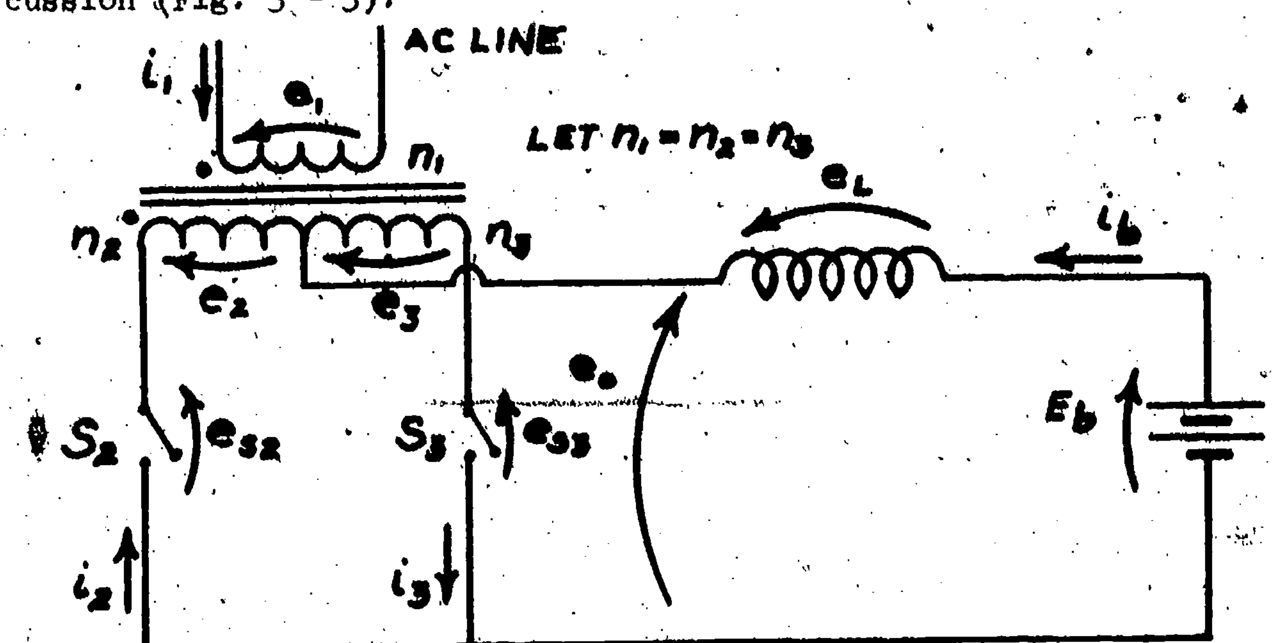


Figure 3 - 3

We define e_o as shown in figure 3 - 3 as a convenience in working with the switching details.

Transformer relations

a) $e_1 = e_2 = e_3$

b) $i_1 + i_2 + i_3 = 0$ (since n's are the same)

Kirchoff's voltage equations

c) $e_o = e_L + E_b$

d) $e_2 = e_{s2} - e_o$

e) $e_3 = e_o - e_{s3}$

Kirchoff's current equation

f) $i_b + i_2 = i_3$

These equations are not sufficient to solve the circuit. We must have additional information as to the switch behavior. Not only must we find a way of expressing the electrical characteristics of the switch which are:

S open $\rightarrow$ e_s = anything, $i_s = o$,

S closed $\rightarrow$ $e_s = o$, i_s = anything,

but we must know <u>when</u> each switch is open or closed. Examining the circuit of figure 3 - 1, some facts become clear.

<u>$i_2 \leq o$</u> or SCR_2 is reverse blocking

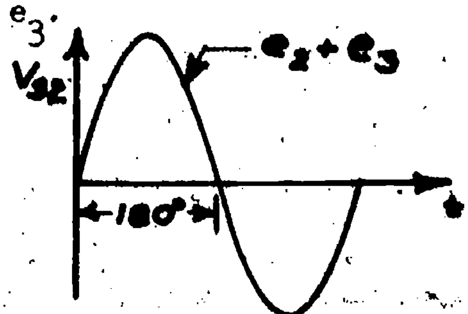
<u>$i_3 \geq o$</u> or SCR_3 is reverse blocking

Also, considering the loop composed of e_2, e_3, s_2, and s_3, it is clear that <u>S_2 and S_3 cannot simultaneously be closed,</u> because one of the SCR's (depending on the polarity of $e_2 + e_3$) will be reverse biased and will therefore turn off. Furthermore, if the inductance is assumed large enough to maintain a nearly constant current through the battery (recall there is no R in the circuit, therefore $^{di}/_{dt} = - \frac{E}{L}$ and a large L implies a small $^{di}/_{dt}$), <u>S_2 and S_3 cannot</u>

simultaneously be open. In fact, S_2 and S_3 cannot both be open provided only that the inductance is large enough to prevent the current i_b from decaying to zero. We have come across another possibly simplifying assumption.

6. Assume i_b is approximately constant because L is large.

The remaining two switch states, S_2 closed with S_3 open and S_2 open with S_3 closed, seem reasonable. Once a switch has closed, the inductor will assume whatever voltage is necessary to keep the current flowing (with value I_b) through the switch. Therefore the switch will remain closed regardless of voltage e_2 and e_3 until the other switch is closed. Then it must open. When can a switch be closed? SCR_2 (assumed open) is forward blocking throughout the first 180° of $e_2 + e_3$.



Voltage across S_2 if S_2 is open and S_3 is closed.

Therefore S_2 can be closed only within $0^\circ < \theta < 180^\circ$. From $180^\circ < \theta < 360^\circ$, the SCR_2 is reverse blocking and cannot be triggered. Similarly, S_3 can be closed only within $180^\circ < \theta < 360^\circ$.

We are now in a position to begin plotting voltage and current waveforms. The waveforms will be considered for three different "firing angles" (α) at which SCR_2 is triggered, namely 45°, 90°, and 135°. Firing angles of 0° and 180° will not be considered because $e_2 + e_3$ will not be large compared to the forward voltage drop across the SCR and SCR turn-off will be more complicated. Thus at 0° and 180° firing angles, modeling the SCR as a switch would be unrealistic. We begin (fig. 3 - 4) by plotting the currents using equations (b), (f), the switching sequence, and the constant current $i_b = I_b$ whose value is not yet known. Switch 2 is closed firing angle α. I_b flows through s_2 until s_3

is closed at $\alpha = \theta + 180^\circ$. Note that the line current of the transformer (i_1) is a square wave while the voltage is a sine wave. This means the centertap side of the transformer does not look anything like a linear "impedance."

Next we consider the voltage waveforms. Using equations (d) and (e) and knowing that e_{s_2} or e_{s_3} is zero when the appropriate switch is closed, we plot e_o. We can plot e_{s_2} by solving equations (d) and (e) to eliminate e_o:

$$e_2 + e_3 = e_{s2} - e_{s_3} \quad ,$$

and knowing when e_{s_2} and e_{s_3} are zero. We note that the above equation is just the Kirchoff voltage sum around the loop containing the two switches and the centertap transformer winding.

We have yet to determine the values of I_b, e_L, E_b. Consider first the determination of E_b. From equation (c), we can relate the voltages E_b and e_L to e_o. Further, we can use the fact that $e_L = -L \frac{di}{dt}$ for e_L. In terms of the polarities assigned in figure 3 - 3:

$$e_L = -L \frac{di_b}{dt} \quad ,$$

and

$$e_o = e_L + E_b = -L \frac{di_b}{dt} + E_b .$$

Obviously i_b cannot be absolutely constant regardless of how large L might be. However, it will not disturb any of our analysis if we allow i_b to fluctuate by some small amount (such as a fraction of a percent of the value I_b). We can rearrange the above equation and integrate.

$$\int_{t_1}^{t_2} (e_o - E_b)\,dt = -L \int_{i_1}^{i_2} d i_b = -L \Delta i_b$$

Changing variables

$$\frac{1}{\omega} \int_{\theta_1}^{\theta_2} (e_o - E_b)\,d\theta = -L \Delta i_b .$$

Consider the form of e_o in figure 3 - 4 for the firing angle $\alpha = 45^\circ$. Divide the cycle into two parts: $45^\circ < \theta < 135^\circ$, and $135^\circ < \theta < 360^\circ$ plus $0^\circ < \theta < 45^\circ$. In the region $45^\circ < \theta < 135^\circ$, Δi is changing from its original value, $i_{b_{45}}$ to a new

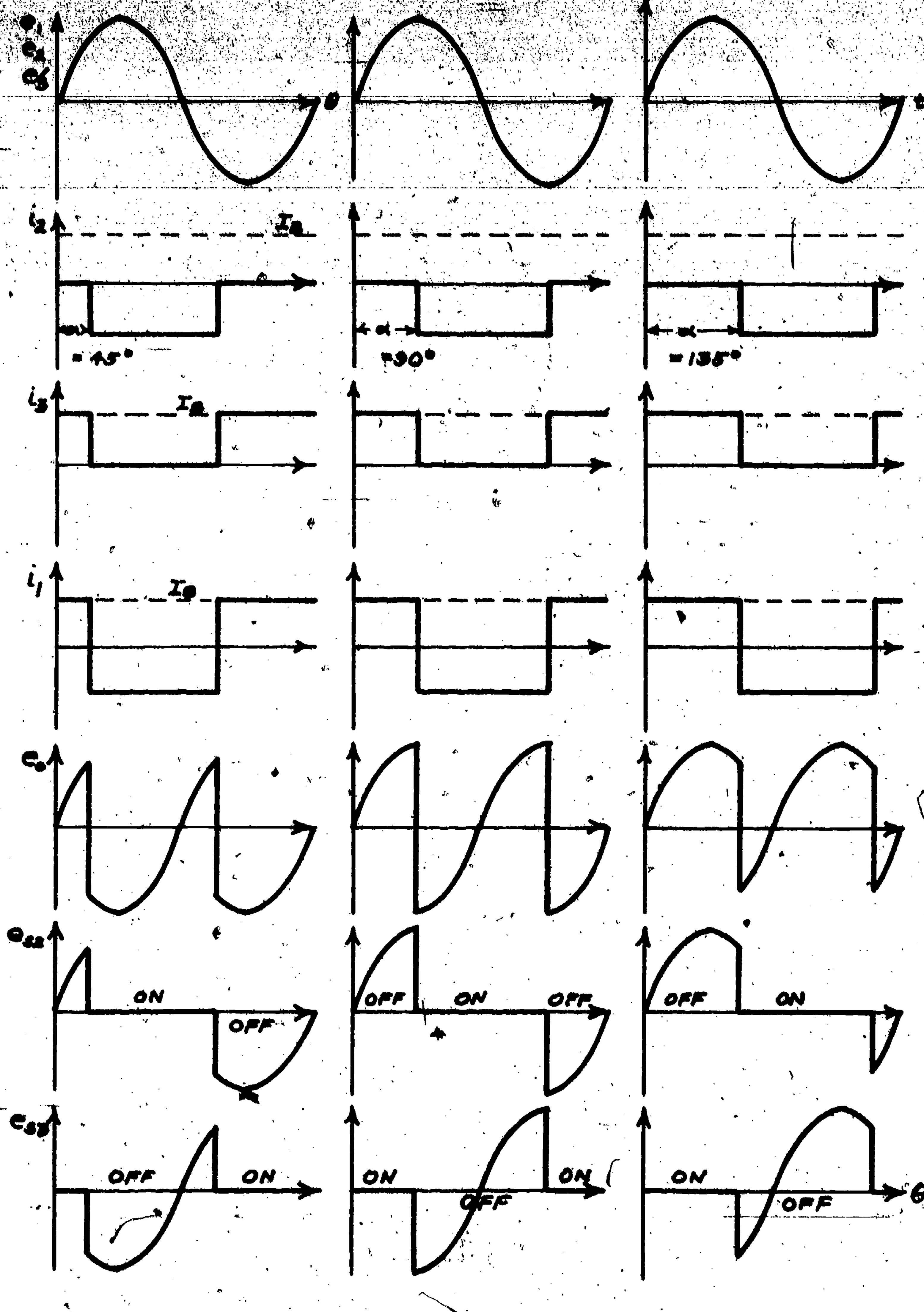

θ
i_2
I_a
= 45°
= 90°
= 135°
i_3
I_a
i_1
I_a
e_o
ON
OFF
OFF
ON
OFF
OFF
ON
OFF
ON
ON
OFF
ON
ON
OFF
OFF
θ

Figure 3 - 4

value i_{b135}. The polarity and amount of this change depends on the value of E_b.

$$\Delta i_1 = i_{b135} - i_{b45}$$

During the next part of a cycle, i_b again changes

$$\Delta i_2 = i_{b135+180} - i_{b135}$$

If we are in a steady state, cyclic process, Δi_1 must equal $-\Delta i_2$, bringing i_b back to its original value at the beginning of the cycle. This is obvious from the definition of a cyclic process, that is:

Definition: In a cyclic process, $f(\theta) = f(\theta + 2\pi)$ for any and all variables. In our set of equations, there is only one value of E_b that will satisfy the cyclic requirement. We could find such a value for each α and determine the voltage waveform of e_L. I_b is not determinable by any obvious means. Should E_b be different from the value such that $\Delta i_2 = -\Delta i_1$, the current would "ratchet" toward plus or minus infinity. We would no longer have a cyclic process or a steady state. In actuality, a steady state and a cyclic process would be attained if we considered the iR and other voltage drops in the circuit that increase with current. The circuit current value is determined by these drops and the value of E_b. We suspect that if E_b has in fact the value calculated above ignoring iR drops, I_b will be zero when iR drops are considered.

Average Voltages and Currents

Before considering resistances and stray inductances all over the circuit, we pause to consider a shortcut, both conceptual and effortwise. Reconsider equation (c), that is:

$$e_o = e_L + E_b$$

where e_o is known to a first approximation. Knowing that the average voltage across an inductor in a cyclic process must be zero (see the following digression), we average the equation over a cycle.

$$\frac{1}{2\pi}\int_0^{2\pi} e_o\,d\theta = \frac{1}{2\pi}\int_0^{2\pi}(e_L+E_b)\,d\theta$$

$$\frac{1}{2\pi}\int_0^{2\pi} e_o\,d\theta = \cancelto{0}{\frac{1}{2\pi}\int_0^{2\pi} e_L\,d\theta} + \frac{1}{2\pi}\int_0^{2\pi} \overset{\text{const}}{E_b}\,d\theta$$

$$\frac{1}{2\pi}\int_0^{2\pi} e_o\,d\theta = E_b$$

Furthermore, because of the repetitive nature of e_o during a cycle (refer to figure 3 - 4),

$$\frac{1}{2\pi}\int_0^{2\pi} e_o\,d\theta = \frac{1}{\pi}\int_\alpha^{\alpha+180^\circ} -e_{1\,max}\sin\theta\,d\theta = \frac{1}{\pi}\int_\alpha^{\alpha+180^\circ} e_o\,d\theta$$

$$= \frac{e_{1\,max}}{\pi}\cos\theta\Big|_\alpha^{\alpha+180^\circ} = E_b$$

This simple expression gives the same value of E_b that would have been found by requiring Δi_1 to equal $-\Delta i_2$.

Digression

In a cyclic process, all ideal inductors, linear or nonlinear, must have a zero cyclic average voltage across their terminals. Consider for example an iron-core inductor (a more complicated case than an air-core inductor). First we topologically "pull out" the winding resistance (Fig. 3 - 5).

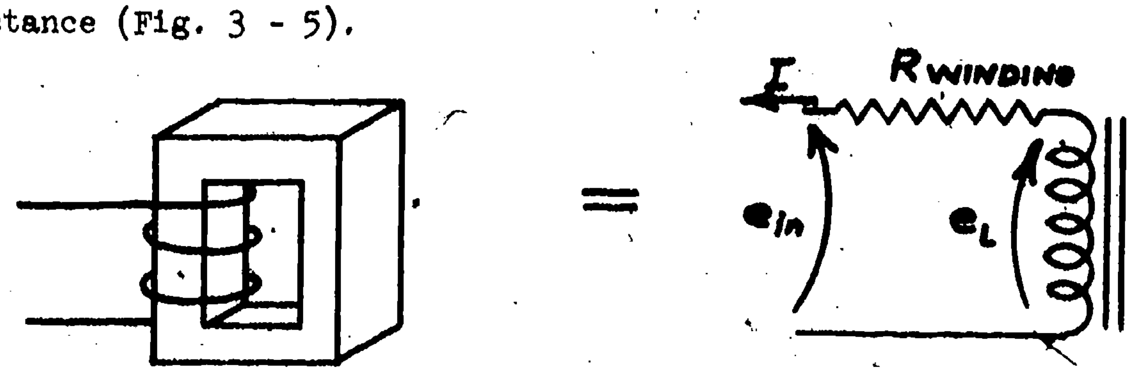


Figure 3 - 5

By Faraday's law, $e_L = -n\frac{d\phi}{dt}$. The iron core has a flux ϕ associated with the value of the current and the history of the core. The ϕ-I curve of the iron is shown in figure 3 - 6.

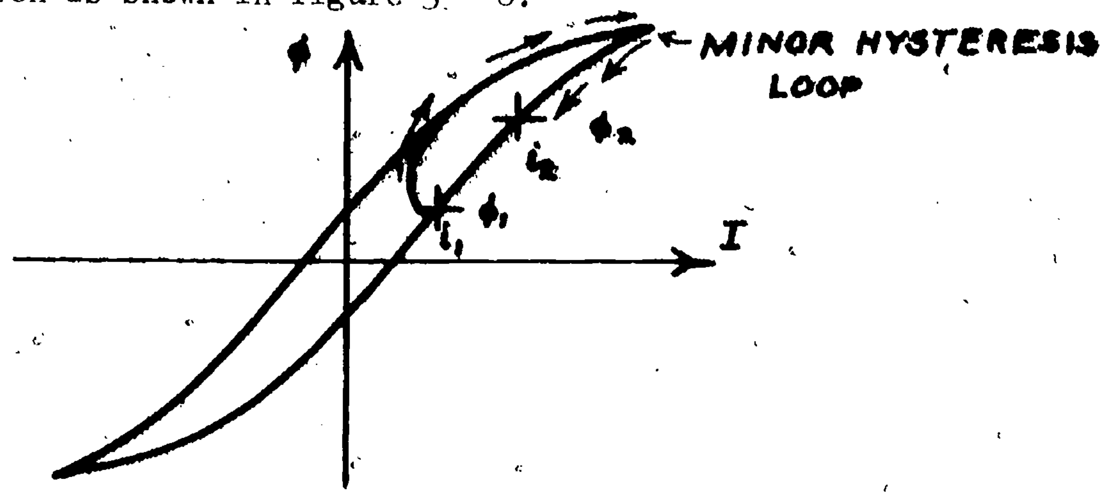


Figure 3 - 6

Suppose at the beginning of a cycle, $\phi = \phi_1$ and current i_1 is flowing. During the cycle, the flux undergoes some excursion and stops at flux value ϕ_2. In a cyclic process, all variables must return to the same value at the end of the cycle as at the beginning. In the example of figures 3 - 5, 6; if $\phi_2 \neq \phi_1$, $i_2 \neq i_1$. If the process is cyclic, $e_{L_1} = e_{L_2}$, $\phi_1 = \phi_2$, and $i_1 = i_2$.

Thus

$$\int_{cycle} e_L\, dt = -n\int_{\phi_1}^{\phi_2} d\phi = -n(\phi_2 - \phi_1) = 0$$

and, dividing by the period of a cycle;

$$\frac{1}{T}\int_T e_L\, dt = \frac{0}{T} = 0, \quad = \text{average value of } e_L.$$

or changing variables;

$$\frac{1}{2\pi}\int_0^{2\pi} e_L\, d\theta = 0 \quad = \text{average value of } e_L.$$

Thus the cyclic average of e_L must be zero. Writing Kirchoff's voltage law for the given loop;

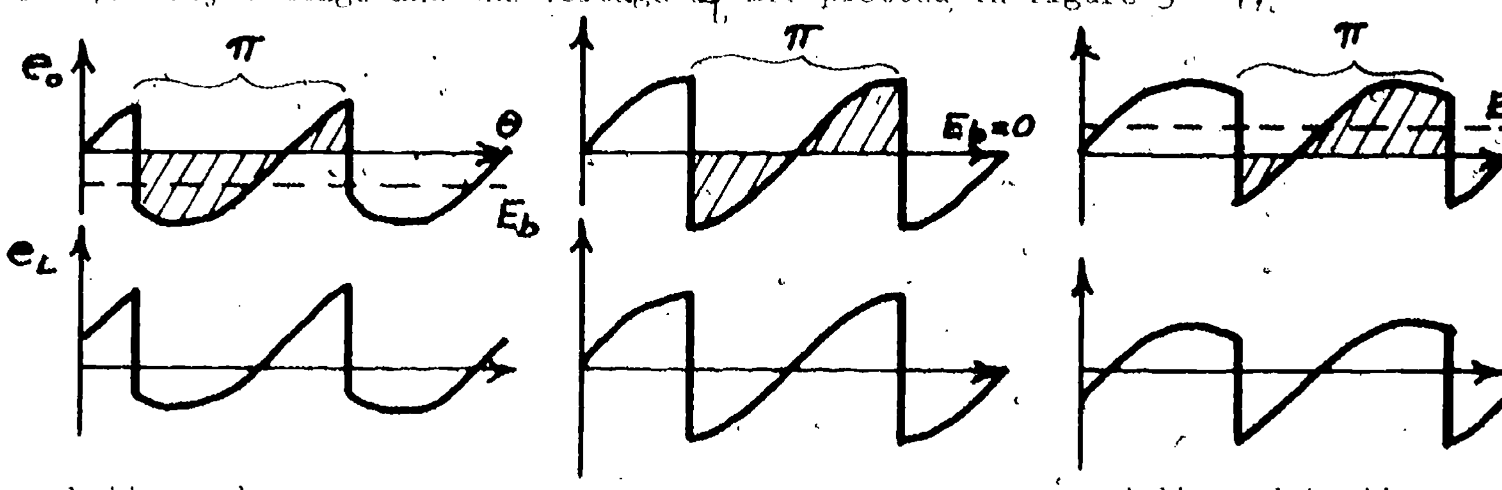
$$e_{in} = iR + e_L.$$

Averaging over a cycle:

$$\frac{1}{2\pi}\int_0^{2\pi} e_{in}\, d\theta = \frac{1}{2\pi}\int_0^{2\pi} iR\, d\theta + \frac{1}{2\pi}\int_0^{2\pi} e_L\, d\theta \; (= 0)$$

Note that it is not necessary for the average value of e_{in} or i to be equal to zero. Thus the winding resistance must be negligible or brought outside the terminals of the "ideal" inductor to use the "zero cyclic average of the voltage" shortcut, and the argument is only valid when the circuit variables describe a repetitive cycle (steady state).

The battery voltage and the voltage e_L are plotted in figure 3 - 7.

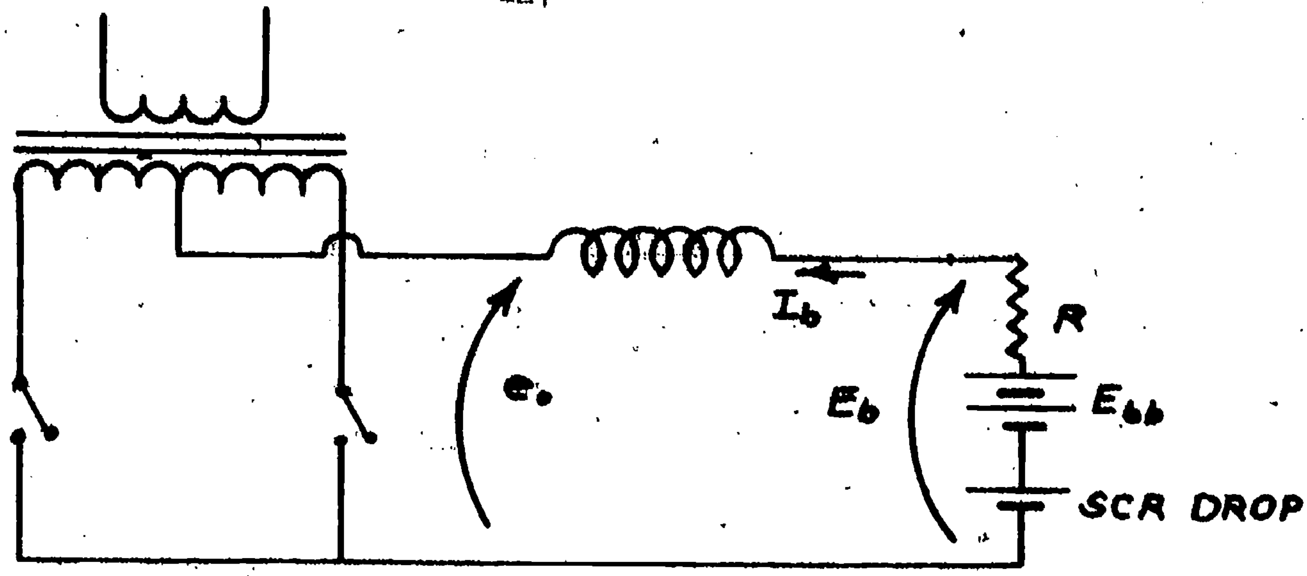


battery charge | no energy transfer | power delivered to line

Figure 3 - 7

Modeling the circuit -#2

I_B can be calculated using the same artifice used in separating the winding resistance from the ideal inductor. We move the forward voltage drop across the SCR (about 1 volt) around the circuit loop to the battery (fig. 3 - 8). The battery internal resistance, the inductor winding resistance, the resistance of one side of the centertapped transformer winding resistance, and the transformer line-side winding resistance are all lumped in a single resistor next to the battery. Note that the transformer leakage reactance is ignored because i_b is constant throughout most of the cycle so that $L_{leakage} \frac{di_b}{dt}$ will be negligible except during the SCR switching.



$R = R_{transformer} + R_{inductance} + R_{battery}$

Figure 3 - 8

Because the current I_b is constant, the iR drop is also constant. Nothing changes the analysis we have made so far. We merely apply Kirchoff's and Ohm's law to the loop containing the battery E_{bb}.

$$E_b - \text{SCR drop} - E_{bb} = I_b R$$

$$I_b = \frac{E_b - \text{SCR drop} - E_{bb}}{R}$$

Obviously e_o and E_b are no longer physically measurable quantities because of the fictional R.

Modeling the Circuit - #3

It is a common occurrence in inverter and phase controlled rectifier circuits that the transformer leakage reactance is not negligible. This leakage reactance limits the rate of change of current in the SCR's (a very favorable effect) and delays the actual turn-on time of the SCR's an amount "γ" (commutation angle) in addition to the firing angle α. The leakage reactance of concern is associated with the center-tapped side of the transformer. The leakage reactance is "pulled out" of the ideal transformer as a pair of lumped inductors (Fig. 3 - 9).

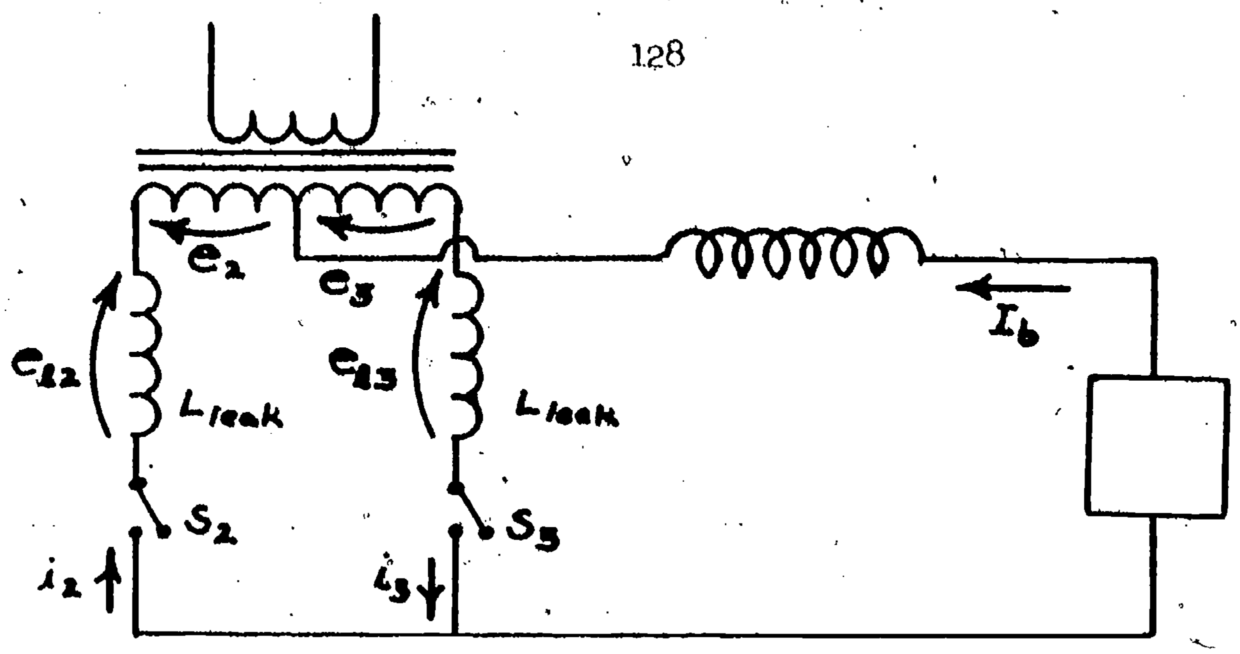


Figure 3 - 9

Because I_b is still assumed nearly constant, the leakage inductances have no effect whenever one switch is open and the other switch is closed ($L^{di}/dt \approx 0$). We must revise some of our original thinking regarding the possible states of the switches. When SCR_2 is closed, the current $-i_2$ increases gradually, its rate of change limited by the leakage inductance. While $-i_2$ increases, i_3 must gradually decrease, its rate of decrease limited by the leakage inductance. Thus, <u>during switching, both SCR's are conducting</u>. We express our concept more precisely:

$$I_b = \text{const} = i_3 - i_2,$$

differentiating with respect to time

$$\frac{di_3}{dt} = \frac{di_2}{dt} \quad ,$$

assuming the two leakage inductances have the same value, the voltage magnitudes across each inductance will be the same since

$$\left.\begin{aligned} e_{l2} &= -L_{leak}\frac{di_2}{dt} \\ e_{l3} &= +L_{leak}\frac{di_3}{dt} \end{aligned}\right\} \qquad \text{Lenz's Law + Faraday}$$

$$\therefore e_{l2} = -e_{l3},$$

during the time both s_2 and s_3 are closed

$$e_2 + e_3 = e_{\ell 2} - e_{\ell 3} \quad \text{by Kirchoff,}$$

since $e_2 = e_3$

$$e_2 = e_{\ell 2},$$

and therefore

$$e_2 = -L_{leak}\, di_2/dt$$

$$\int_{t_1}^{t_2} e_2\, dt = -L_{leak}\left(i_{2(t_2)} - i_{2(t_1)}\right).$$

We consider the meaning of these equations by graphing e_2, $e_{\ell 2}$, i_2, and i_3 during a switching operation. We plot $-i_2$ instead of i_2 to show the forward current through SCR_2 (Fig. 3 - 10). Assume initially SCR_3 is conducting (S_3 closed) and at firing angle α, SCR_2 is triggered (S_2 closed). When S_2 closes, the current $-i_2$ increases while i_3 decreases as the current is transferred from S_3 to S_2. Initially, $i_{2(t1)} = 0$. After switching $i_{2(t2)} = -I_b$. The voltage across the leakage inductor is e_2. The switching time $(t_2 - t_1)$ must last until $\frac{1}{L}\int_{t1}^{t2} e_2\, dt = I_b$. When the switching time (t2 - t1) is expressed as an angle γ, it is known as the commutation angle. The value of the angle γ depends on the firing angle α because e_2 is small for α's near 0° or 180° (giving larger γ's) and increases to a maximum at $\alpha = 90^\circ$ (giving the smallest γ).

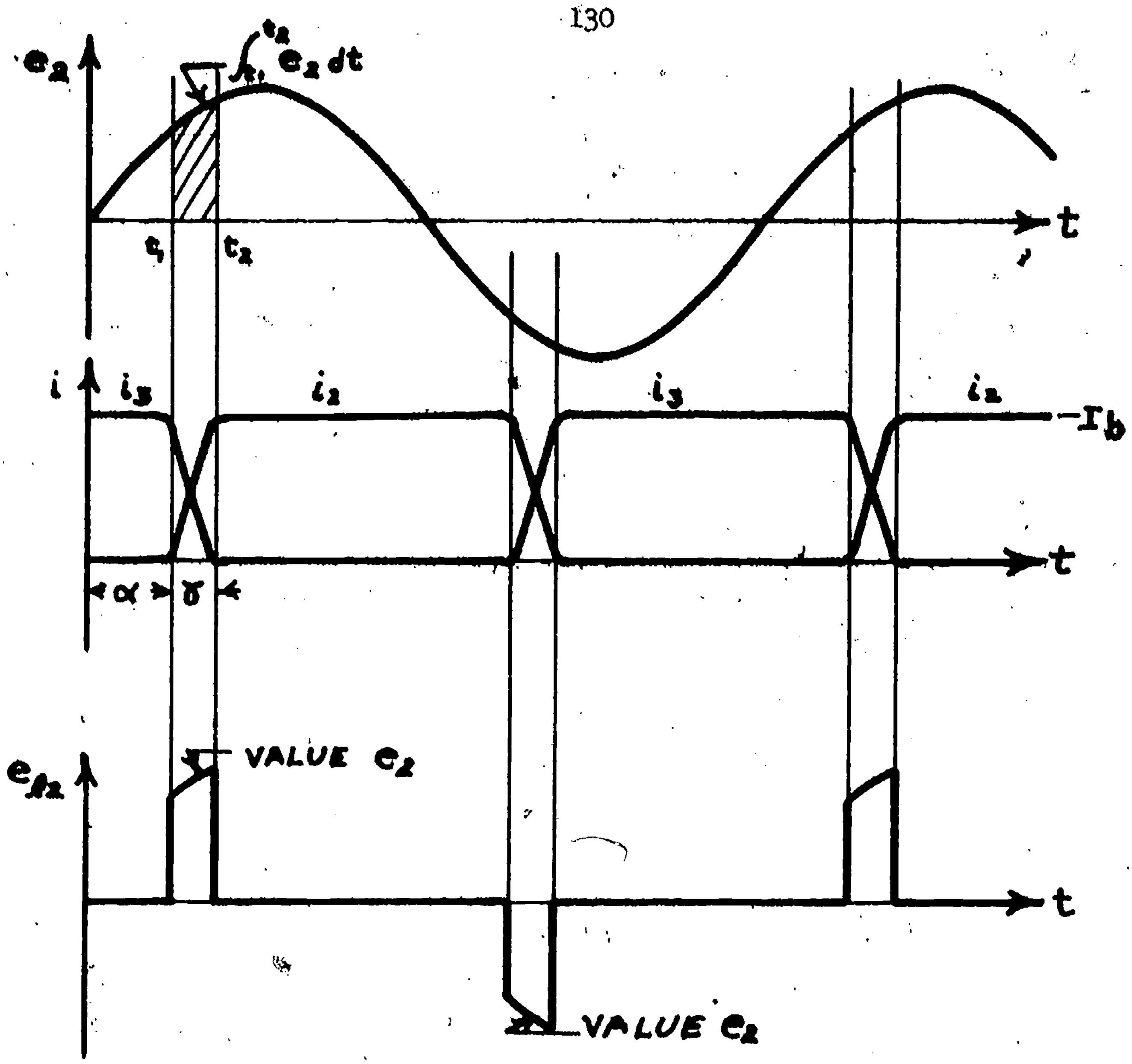
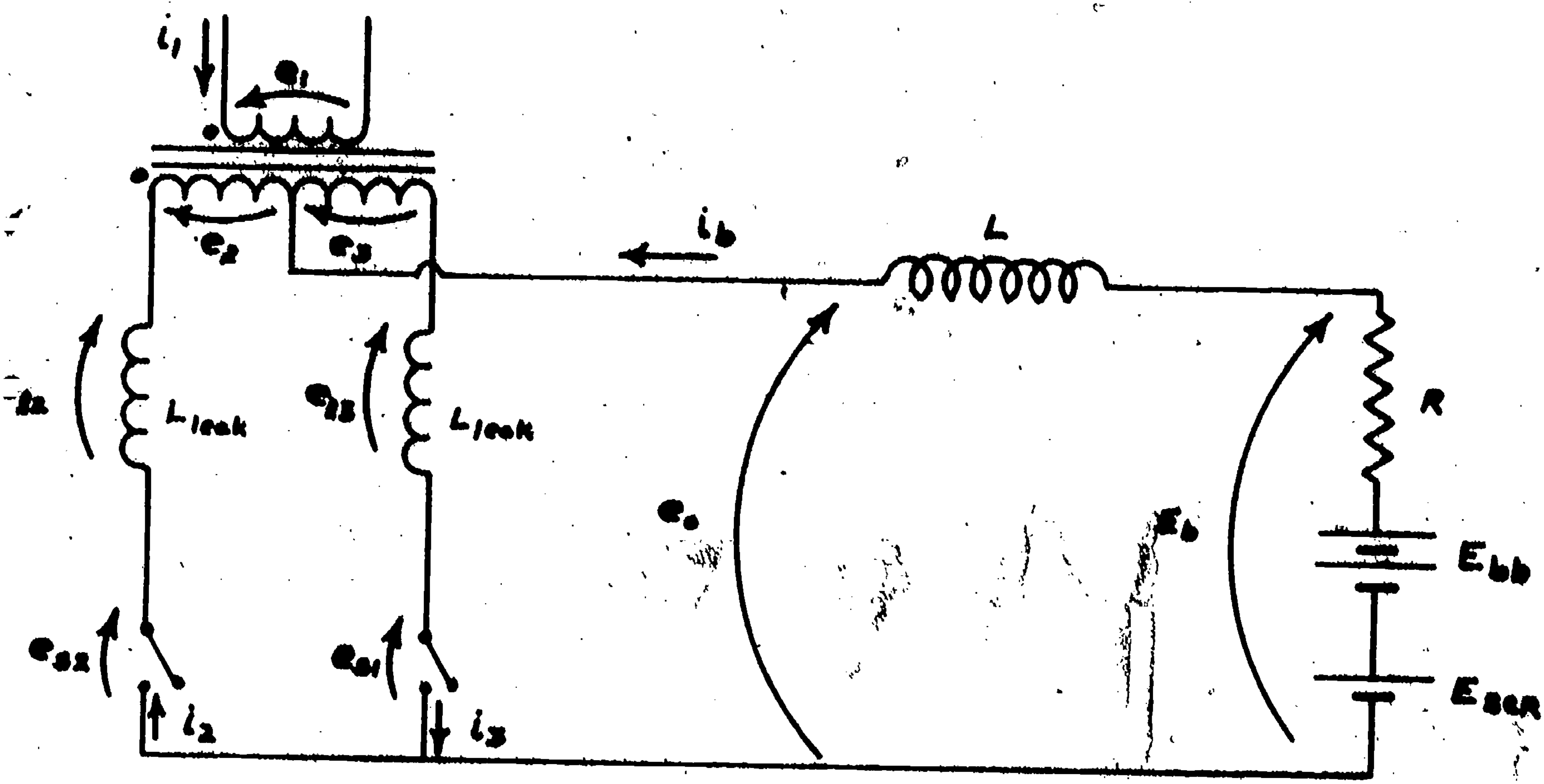


Figure 3 - 10

The voltages dropped across the leakage inductances subtract from e_o and thus modify the value of the battery voltage. We correct figures 3 - 4 and 3 - 7 taking into account the leakage inductances in figure 3 - 12. The analysis between switching times remains precisely the same. The model we have now analyzed is shown in figure 3 - 11.



E_{SCR} ≅ 1 volt forward drop across a conducting SCR.

R = lumped equivalent resistance composed of transformer and inductor winding resistances and the battery internal resistance

L_{leak} = ½ leakage inductance of the centertaped side of the transformer.

Figure 3 - 11

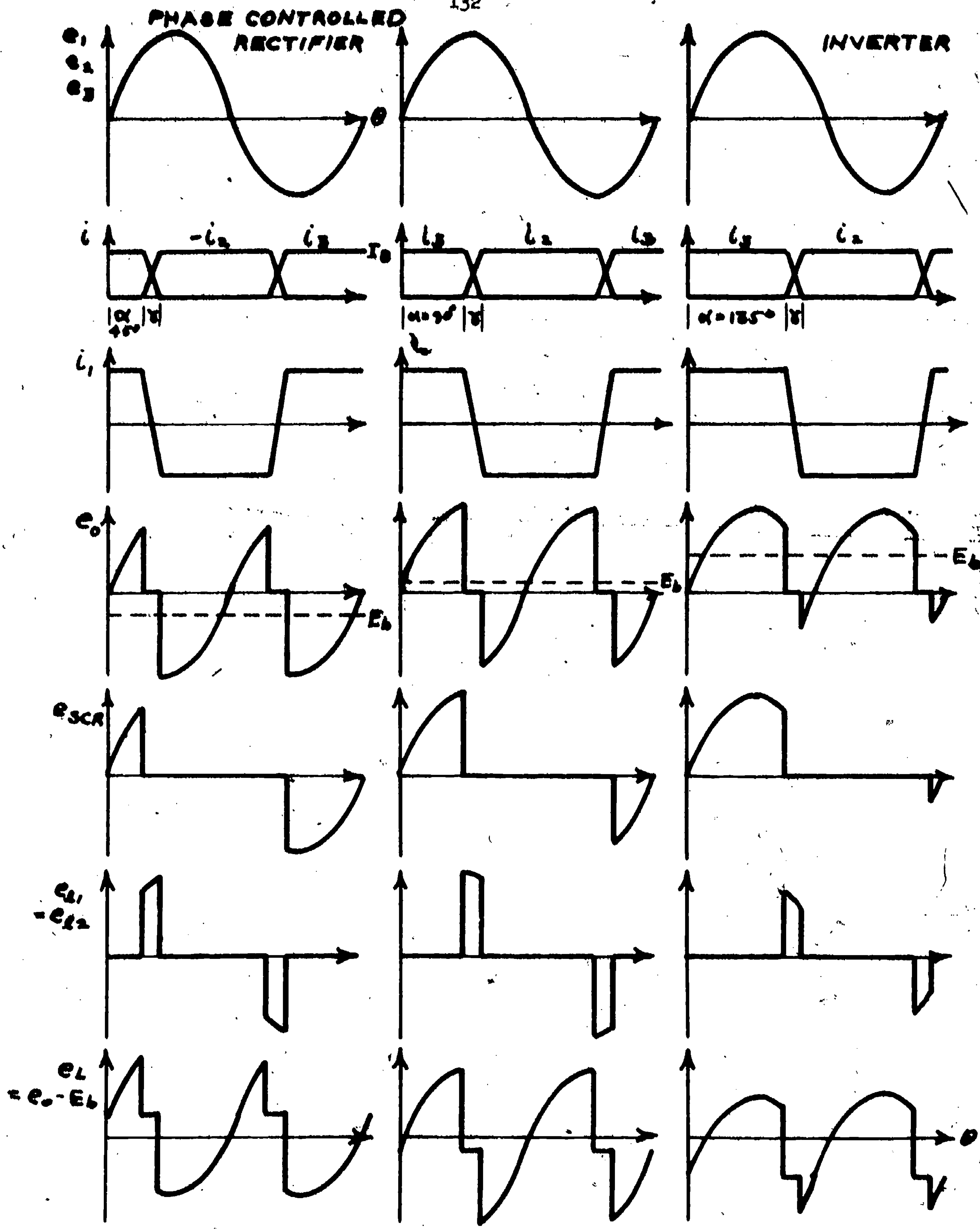


Figure 3 - 12

Let us review the assumptions made in devising model #1 of the circuit. 1) All through the analysis we have assumed large currents compared to the SCR blocking current and large voltages compared to the SCR forward voltage drop during conduction. We have relaxed the requirement that the SCR forward voltage drop be negligible in model #2. 2) We have taken into account all winding resistances. It is also not necessary that L be a linear inductor (avoided in the "averaging trick"). 3) We have accounted for the internal resistance of the battery. 4) We have also considered the transformer winding resistance and the leakage inductance on the center-tapped side in model #3. We have not considered the leakage reactance on the line side. 5) We still require the line to be "stiff" so that voltage harmonics ("messing up the sine waveshape") do not occur because of the non-sinusoidal current. 6) We still require L to be large enough that i_b can be considered constant. We could further relax the assumptions using more detailed circuit models, however, we stop at this point having shown the methods of refining the model and because further refinements are more complicated (but certainly possible).

Summary

In analyzing the line voltage commutated inverter or phase controlled rectifier we have used the method of "iterative modeling". The circuit was simplified to a point where it could be understood and analyzed with reasonable facility. This was the most difficult step. If the circuit had been over simplified it wouldn't work, and another model would have had to be invented. The model was then refined by steps or iterated until the model and its solutions gave sufficiently accurate and valid results.

The iterated modeling technique is an important and powerful analysis technique. Each model brings an increased understanding of the circuit operation and the importance and roles of the circuit components as the solution is approached in a step-by-step logical manner. Such understanding is extremely valuable in design and synthesis of circuits as well as analysis. In addition, the required sophistication of the model or the next step in the analysis can be indicated by numerical calculation of given or estimated circuit parameters or by laboratory experiment. Thus the analysis is only as complicated as is necessary for the given purpose. In each step, the preceding analysis gives the engineer a basis from which to work and extend his knowledge in that each model can be checked and compared to physical reasoning. If an error is made, one need only go back to the previous models and try again, therefore, past efforts are not lost in the case of an error.

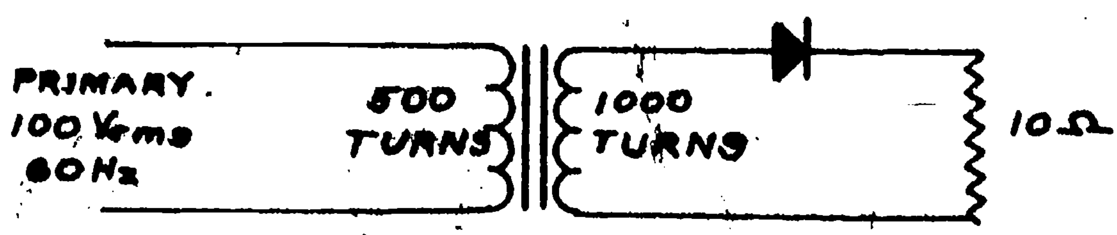
In contrast, the approach of including all possible circuit elements in the circuit model and writing Kirchoff's loop and node equations all over the place has serious disadvantages. Such a method is an all-or-nothing method. If the equations become so complicated the engineer cannot solve them, there is no obvious next step. If an error is made, one must begin again and examine each step because there are few or no physical reasoning checks along the way of manipulating the many equations. After (and if) an answer is found, some sort of simplified modeling is still necessary to assure the correctness of the answer, and finally, more elements than necessary may have been included in the circuit model, adding unnecessary complexity to the solution.

The "shortcut" of averaging voltages over a full cycle, presented in the circuit analysis, allows one to eliminate the inductors and transformers (exercise 2) from the voltage loop equations. A similar "shortcut" for current equations and capacitors exists (exercise 1). In the particular problem solved in this chapter, the averaging scheme enabled us to avoid solving differential equations consisting of Kirchoff loop equations where the coefficients varied as a function of time (the switches).

Exercises

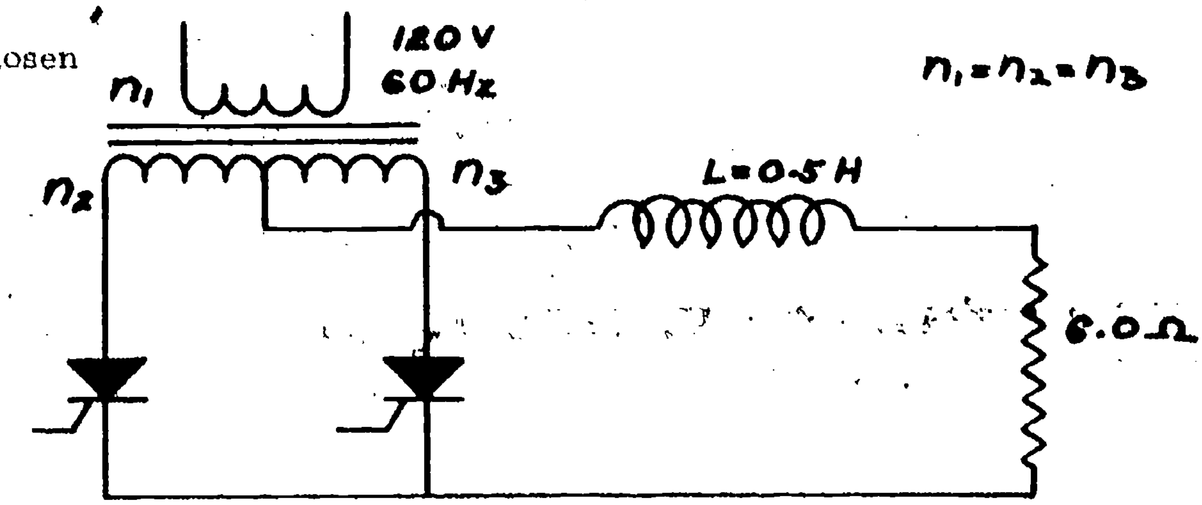
1. Prove that the average cyclic current through a capacitor must be zero. What about the case where surface leakage currents exist around the capacitor plates (capacitor current "leakage")?

2. Show that the Faraday voltage at a transformer terminals must have a zero cyclic average regardless of core losses and nonlinear transformer loads (like diodes, etc.). Calculate the average primary current in the following circuit. Assume the iron core does not saturate.



Problem 1

Determine the maximum power dissipated in each SCR of the phase-controlled rectifier diagrammed below so that appropriate heat sinks for the SCR's may be chosen



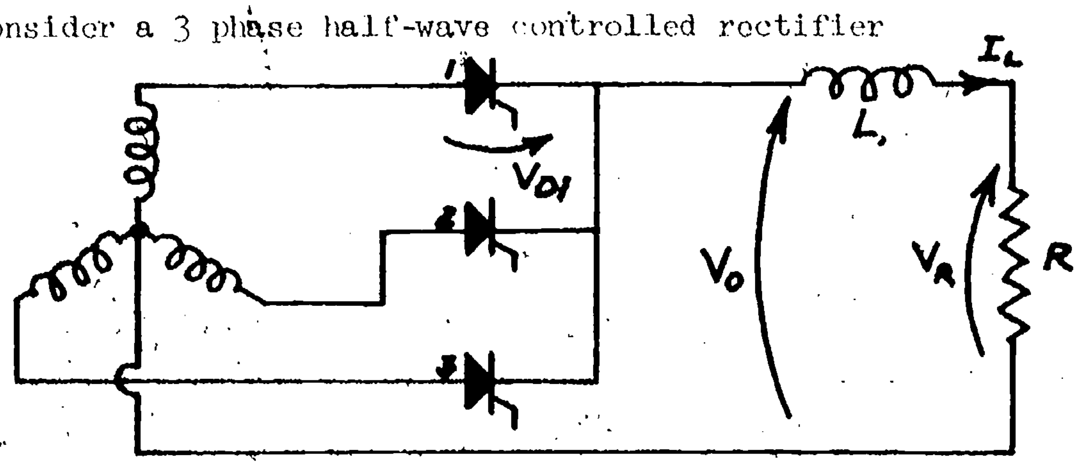
Winding resistance of inductance L = 2.0Ω

Total winding resistance of the center-tapped winding = 3Ω.

Total leakage inductance of the center-tapped winding = 10 mh.

Problem 2

Consider a 3 phase half-wave controlled rectifier



Assume L so large that I_L is constant throughout a cycle.

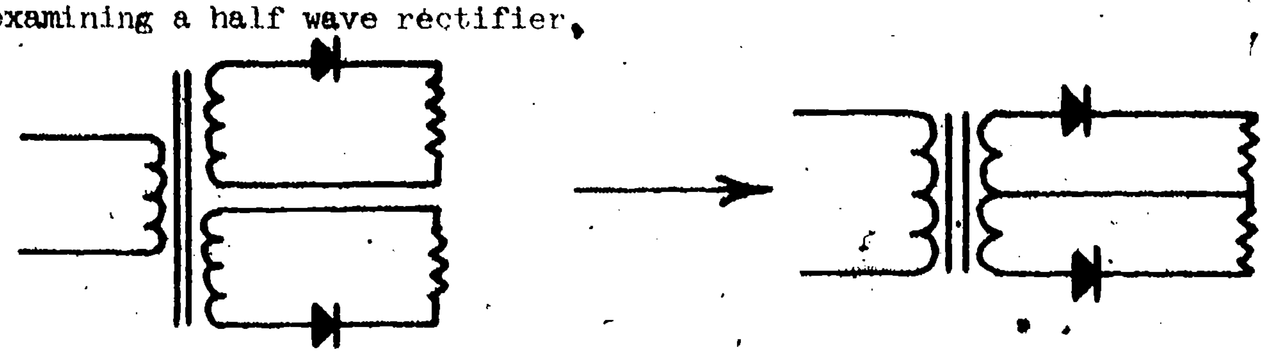
The transformer leakage reactance is not negligible resulting in a commutation angle γ (about 5^o).

a) Plot I_{D1}, I_{D2}, I_{D3} on the same graph as a function of time.
Plot V_o, V_{D1}, and V_R as functions of time.
Explain your reasoning, show α and γ.

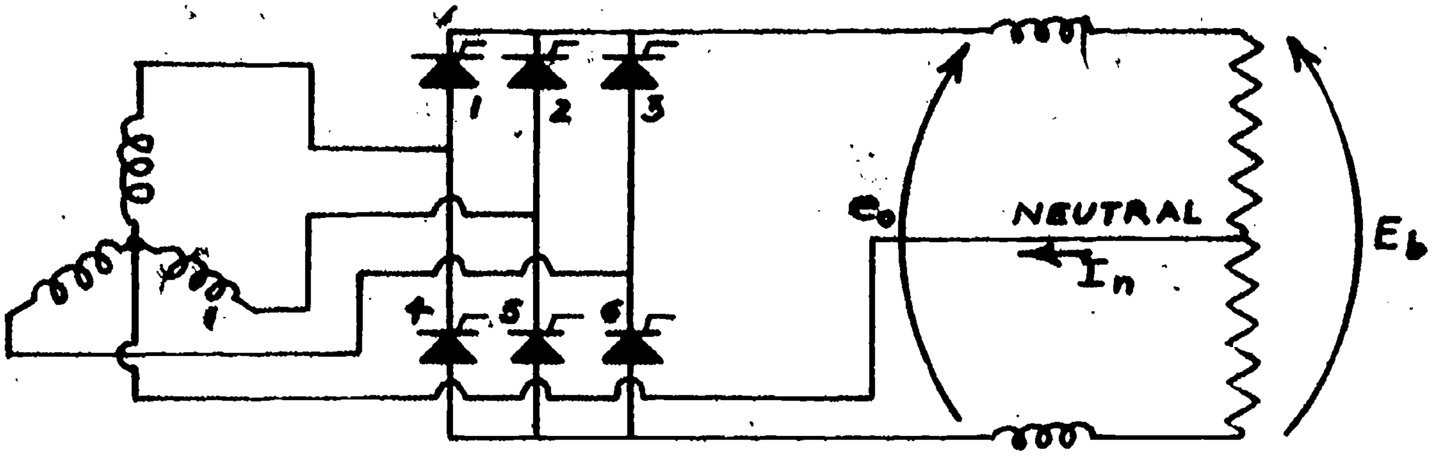
} for firing angle $\alpha < \pi/2$

b) Plot I_R (current through the load resistor R') as a function of firing angle α from $0<\alpha<\pi$.

As we can see how to "invent" a bridge rectifier (single phase) from examining a half wave rectifier,



so we can see how the "Graetz Circuit" was invented for three phase (maybe).

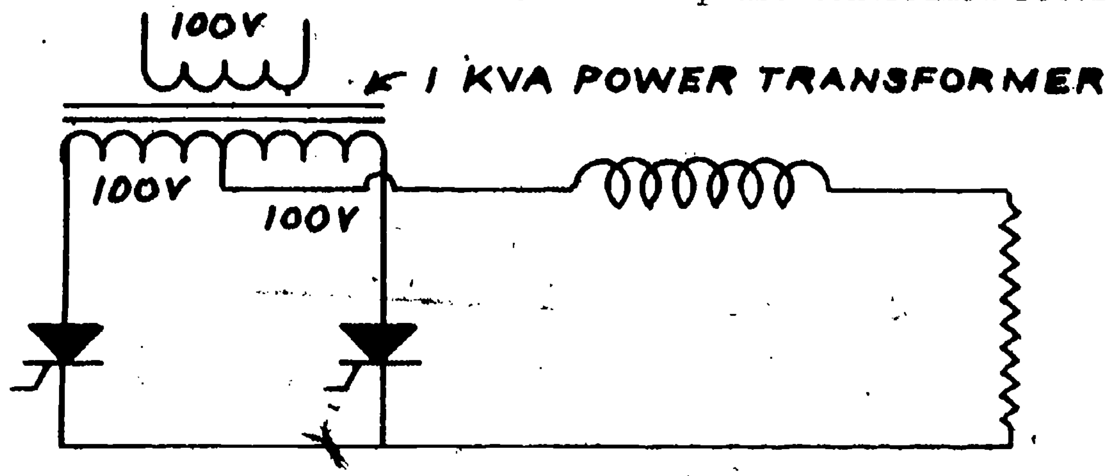


a) What is the firing order (sequence in which the thyristors are triggered) of the thyristors?

b) It is claimed that $I_n = 0$ and that the neutral wire can be removed. Is this true? If so, what does e_o look like (neglect leakage reactance for simplicity)?

Lab Problem 1

A phase controlled rectifier is being designed (circuit below). A working model could be thrown together in the laboratory to check the design using a 110/220 V center-tapped 1 KVA power transformer, an assortment of available SCR's, a large variable inductor and assorted resistors. Your specific problem concerns the SCR triggering circuit. How precisely must the SCR triggering pulses be spaced, that is, if SCR_2 is triggered 175° after SCR_1 instead of 180°, what happens? How about 170° or 150°? Because precise trigger circuits tend to cost more than sloppy trigger circuits, you must determine the specific effects such a "disymmetry" in the firing angles of the SCR's will cause so that you could make an intelligent selection of trigger circuits for a specific application. Show the critical waveform changes, efficiency changes, and component rating changes that would be caused by an unsymmetric triggering of the SCR's. As an additional question, how should the transformer design be improved or changed from that of an ordinary power transformer if the transformer is to serve in a phase-controlled rectifier?



Lab Problem 2

The following inverter circuit is useful at high frequencies where the physical size of L is not so formidable. The circuit operates best for small values of R. The basic idea is that SCR_1 is triggered. When the current pulse through R has ended, C is charged. Then SCR_2 is triggered. C discharges through R giving a current pulse in the other direction. The alternating current pulses occur at the "frequency" or more precisely, the repetition rate of the SCR. gate pulses.

Plot the important current and voltage waveforms that describe this circuit assuming a "negligibly small" R. Answer the following questions.

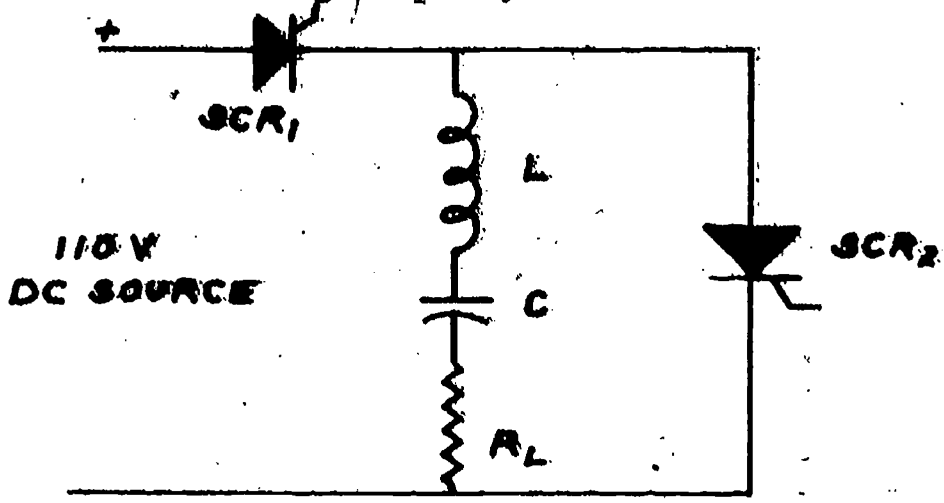
a) How small is "small R"?

b) What does "operates best" mean?

c) What happens if SCR_2 fires while SCR_1 is still conducting?

d) Does the magnitude of the output current depend on the triggering repetition rate?

e) What is the highest frequency at which circuit can operate for a given set of values of L, C, and R? What is ultimately the limiting factor on increasing frequency?



Experimentally verify the correctness of your analysis.

References

1.) B. D. Bedford and R. G. Hoft, Principles of Inverter Circuits, John Wiley & Sons, Inc., New York, 1964.
This reference presents an excellent analysis of a large variety of inverter and phase controlled rectifier circuits.

2.) J. Seymour, Ed, Semiconductor Devices in Power Engineering, Sir Isaac Pitman and Sons, Ltd., London, 1968.
This reference contains a brief outline of the operation of a variety of commonly used phase controlled rectifiers ("thyristor converters").

3.) John H. Kuhlman, Design of Electrical Apparatus, 2nd Ed, John Wiley & Sons, Inc., New York, 1940
This reference contains summary tables and methods of calculation of the characteristics of power transformers.

Chapter 4 Thermal Characteristics of Materials

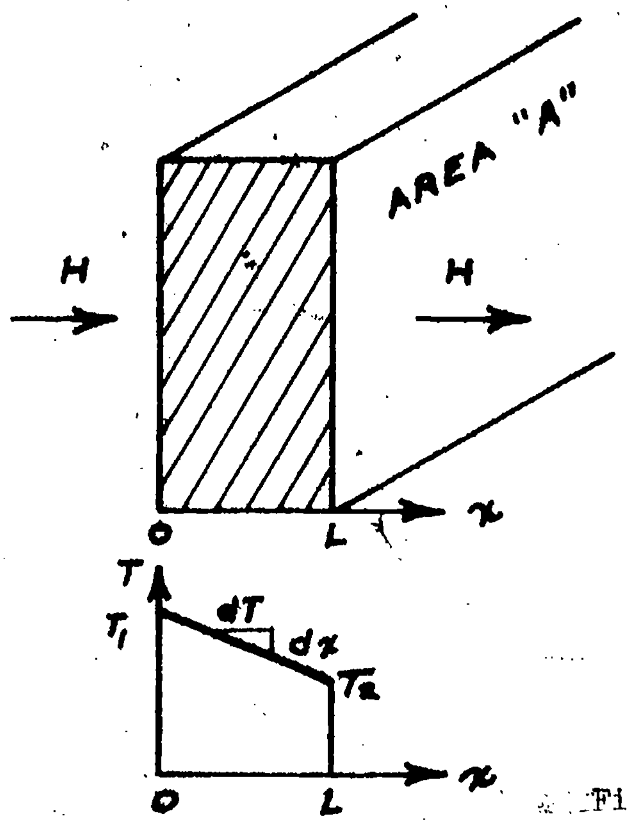
Introduction

The basic principles of heat transfer are reviewed in this chapter with special emphasis on heat-sink considerations. Four different kinds of model-examples are presented. In deriving an electrical model of a heat flow problem, we consider "modeling by analogy". Considering the steady state and transient heat-flow and temperature distribution in a material, we use a "lumped parameter model" of a distributed or continuous system in which the material is artificially broken up in to many pieces which are considered as individual units. We use the "superposition of models" in a linear system when considering the transient thermal behavior of a material where a model for heat conduction is superposed on a model of heat capacity. Finally, a very "specialized model" of the thermal characteristics of an SCR under conditions of a particular (but common) current waveform is considered.

Steady-state Constant Heat Flow

The heat generated in a semiconducting device such as a diode or SCR must travel from the crystal through some bonding material, through the case of the device, possibly through a mechanical fastener system, through a chassis and/or heat-sink, and is finally dispersed in some environment. The temperature of the crystal depends on the thermal properties of these "components" of the system and on the ambient temperature of the environment. We frequently have a choice of the size and type of heat-sink or chassis and the type of fastener system that is to be used in a practical situation. In order to be able to make an intelligent choice of these "thermal system" components, we consider in some detail the thermal properties of these components and the way they are commonly specified.

We begin by reviewing the simplest case, that is, steady state, constant, unidirectional heat flow by conduction through a homogeneous material. Recall from the elementary physics of heat flow that the rate of flow of heat energy passing through a homogeneous material (steady state or not) is proportional to the rate of change of temperature with distance (Fig. 4 - 1).



$H = -kA\frac{dT}{dx}$

H = power or heat flow (watts)

A = area (m^2)

T = temperature (^oC)

x = distance in the direction of the heat flow (m)

k = constant = "thermal conductivity" (watts/m^oC)

H is assumed uniform over A.

Figure 4 - 1

"H" is constant and has the same value for any x such that $0< x <L$ because:

- the flow of heat is in the x-direction (no heat is flowing up, down, or to the sides because of the uniform distribution of H and T over the area A), and
- in the steady state, the temperature distribution and values do not change (for constant H) thus no heat is "used" to change the temperature of the material as a function of time.

It is a trivial matter to integrate the heat conduction equation under the given conditions.

$$H = -kA\,dT/dx .$$

$$H\,dx = -kA\,dT .$$

Integrating both sides of the equation:

$$H\int_0^L dx = -kA\int_{T_1}^{T_2} dT,$$

$$HL = -kA(T_2 - T_1).$$

Rearranging the terms by algebra,

$$\left(\frac{L}{kA}\right)H + T_2 = T_1 .$$

In a physical situation where the dimensions and thermal characteristics do not change as a function of time (we already assumed they do not change as functions of temperature - an excellent first approximation), the constant $(\frac{L}{kA})$ is redefined as "θ", the "thermal impedance" (for reasons more apparent later), thus:

$$H\theta + T_2 = T_1 . \qquad \theta \text{ in } {}^{\circ}C/\text{watt}$$

Consider several slabs of material having different k's and L's in intimate (perfect) contact with each other (Fig. 4 - 2). Each slab has a θ associated with its k and L which may be different than the θ's of the other slabs.

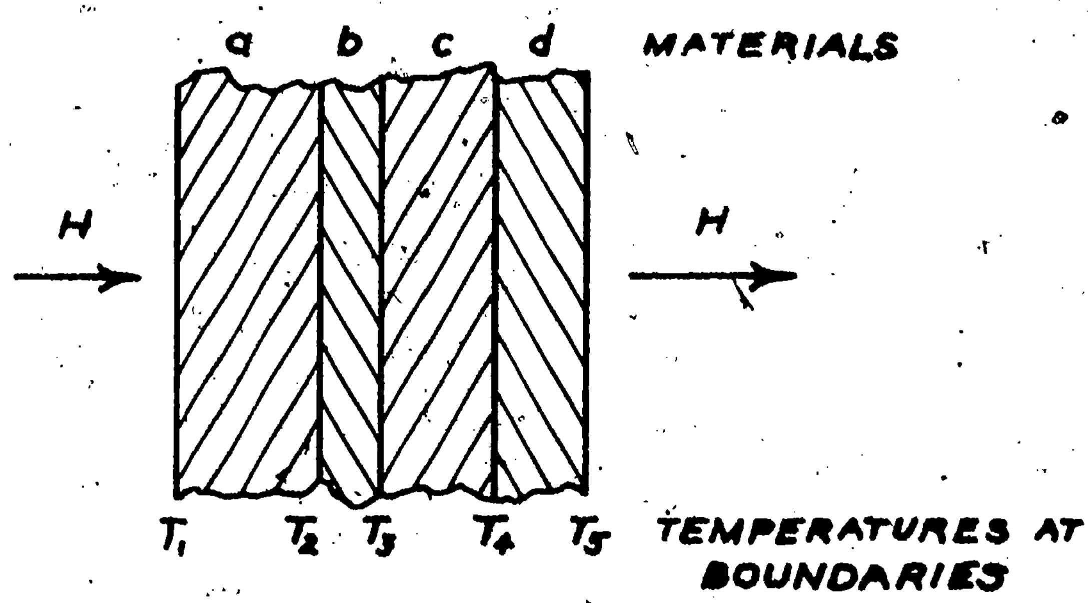


Figure 4 - 2

Consider the first slab "a",

$$H\Theta_a + T_2 = T_1 .$$

The next slab "b", yields a similar equation,

$$H\Theta_b + T_3 = T_2 .$$

which may be simultaneously solved with the previous equation to eliminate T_2, yielding

$$H\Theta_a + H\Theta_b + T_3 = T_1 .$$

For the slab "c",

$$H\Theta_c + T_4 = T_3 ,$$

yielding

$$H\Theta_a + H\Theta_b + H\Theta_c + T_4 = T_1 .$$

For the slab "d",

$$H\Theta_d + T_5 = T_4 ,$$

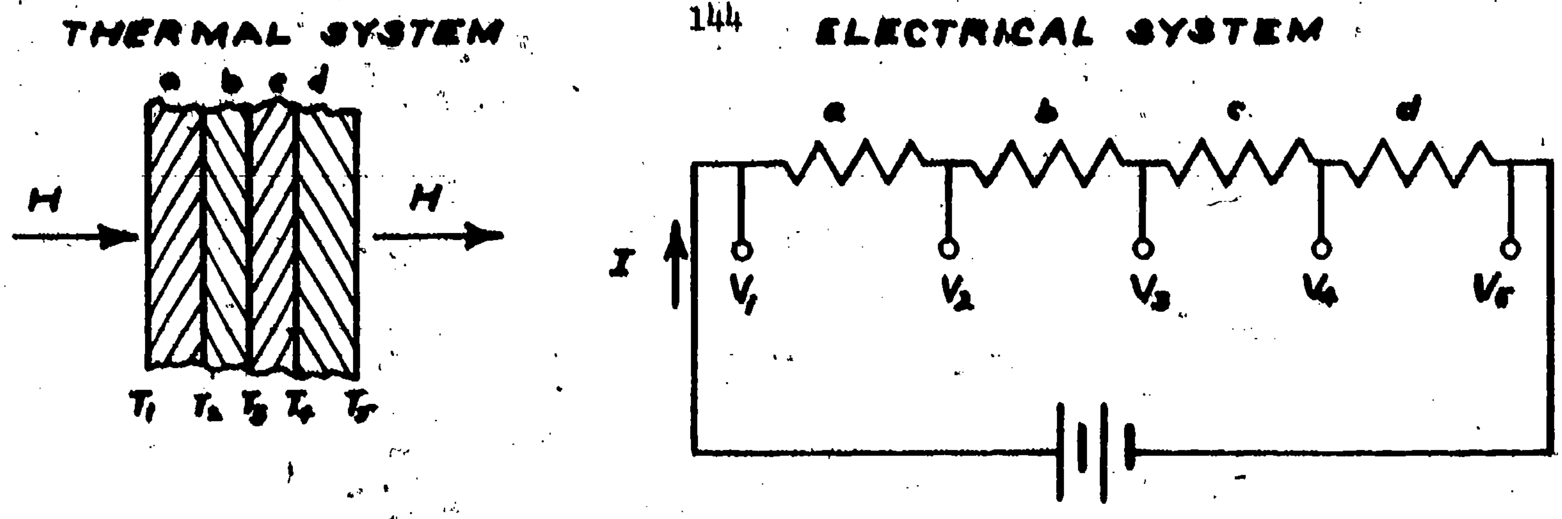
yielding

$$H\Theta_a + H\Theta_b + H\Theta_c + H\Theta_d + T_5 = T_1 \quad \text{etc.,}$$

or

$$H(\Theta_a + \Theta_b + \Theta_c + \Theta_d) = T_1 - T_5$$

We note that the equation relating power and temperature difference for the series of four slabs of heat conducting material is similar to the equation relating current to potential difference for a series of resistors (Fig. 4 - 3). We are encouraged (being electrical engineers) to make an electrical analogue to the heat flow system.



H	=	heat flow	→	I	=	current flow
T	=	temperature	→	V	=	electric potential
θ	=	thermal resistance	→	R	=	electric resistance

Figure 4 - 3

The reason for calling θ the "thermal resistance" is now apparent. The case of three dimensional heat flow is slightly more complicated than the one dimensional problem we have considered, however, if we choose areas in the three dimensional problem which are normal to the heat flow direction, we can intuitively see that the same general form of solution, i.e.

$$P\,\theta + T_2 = T_1,$$

will always result over any temperature range for any material that can be said to have a constant thermal conductivity "k".

The notion of thermal resistance can be used in almost all practical semiconductor power dissipation-heat flow problems for the elements of the thermal system from the crystal where the heat is generated to the heat-sink. The heat-sink, however, is a very complicated element of the system which transfers heat to the environment by means of convection, radiation, and conduction. Commonly used, commercially available, convection cooling (and forced air flow) heat-sinks are specified in two ways. A graphical curve of temperature difference between the mounting surface of the semiconductor's case and the ambient air temperature versus "cooling power" may be presented as shown in figure 4 - 4. These curves for the various heat-sink sizes and configurations are the result of empirical test. The "natural convection" characteristics of figure 4 - 4 could be roughly approximated by a straight line (dashed line, Fig. 4 - 4). The slope of the straight line $(\frac{T_1 - T_2}{H})$ is known as the "thermal resistance of the heat-sink". Thus, despite the fact that the heat-sink cools by means other than conduction, it can be said to have an emperically measured thermal resistance.

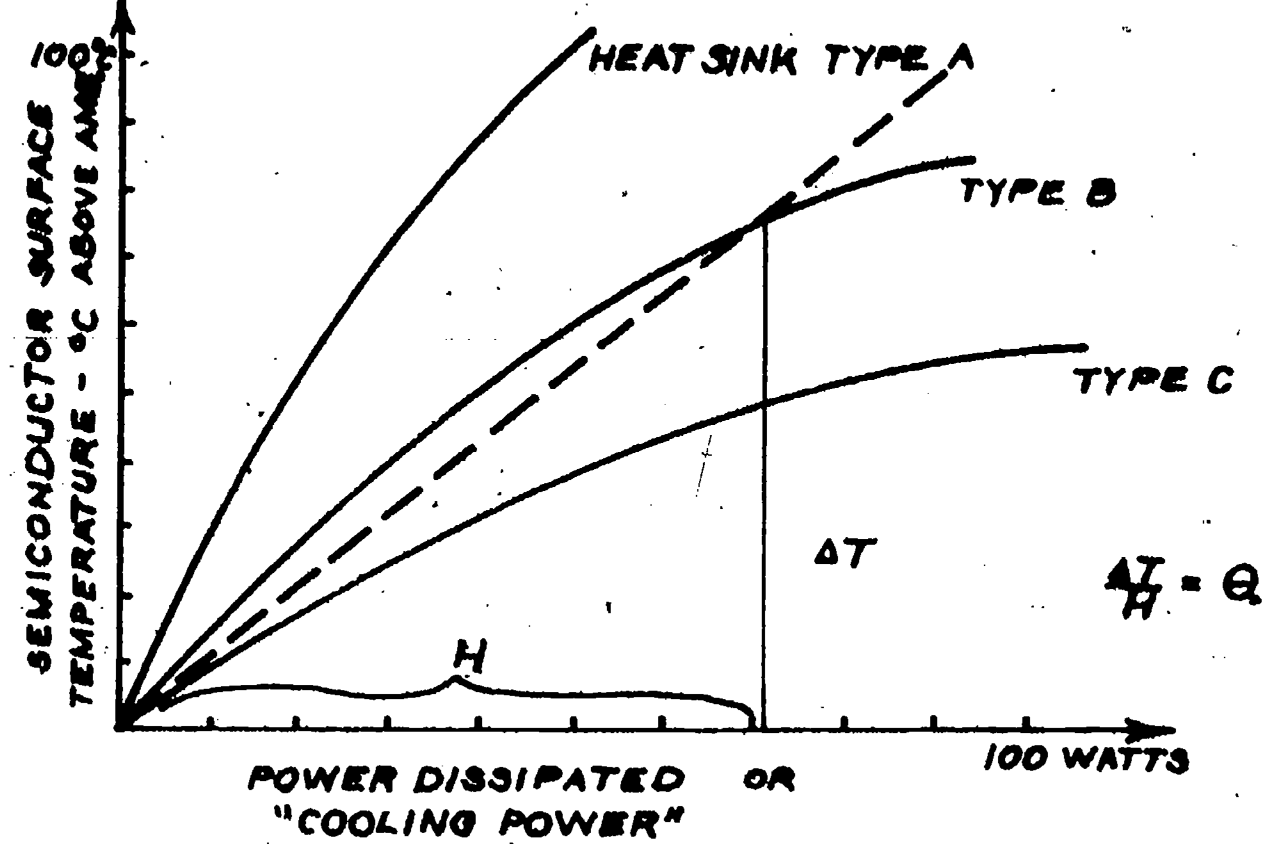


Figure 4 - 4

The fact that in general, an object placed in air transfers heat by conduction, convection and radiation approximately proportional to the temperature difference of the object and the ambient air temperatures is called "Newton's law of cooling".

As an example of using thermal impedances, consider a stud mounted SCR on an anodized aluminum heat-sink. The thermal impedances for the components of the thermal system are obtainable from manufacturers' data sheets or could be measured in the laboratory. The question is, "What is the maximum power the SCR can be permitted to dissipate?"

Given:

Maximum allowable junction temperature (T_j) = $100^{\circ}C$

Maximum ambient temperature (T_a) = $25\ ^{\circ}C$

θ_j - junction to mounting stud = $1.5\ ^{\circ}C/watt$

θ_s - stud to heat-sink when the stud is fastened with the maximum permissable stud torque (dry threads)

= $0.35\ ^{\circ}C/watt$ dry

= $0.25\ ^{\circ}C/watt$ using "joint compound"

= $3.0\ ^{\circ}C/watt$ using insulating mica washer with joint compound.

θ_h - heat sink $1.7^{\circ}C/watt$

We shall consider the case where the SCR is mounted "dry".

$$H(\theta_j + \theta_s + \theta_h) = T_j - T_a$$
$$H(1.5 + 0.35 + 1.7) = 100 - 25$$
$$H = \frac{75^{\circ}C}{3.6^{\circ}C/WATT} \cong \underline{21\ WATTS}$$

Unless the SCR and heat-sink mounting instructions are followed with care, the rated thermal impedances will not be attained. Mounting a stud-mounted SCR will usually involve a torque wrench. Press-fitted SCR's require a close adherance to dimensional tolerances ($\pm$ .005 is typical). Other types of SCR's require special mounting hardware such as the GE Press-Pak or IR "Hockey Puks". Any surface finish of the heat-sink such as anodizing or paint should be removed directly under the SCR, and no burrs or knicks may be permitted under the mounting surfaces. Heat-sinks cooled by natural convection are designed to have their fins placed vertically. If the heat-sink is mounted horizontally or near other objects that might restrict the air flow or heat the air, the heat-sink must be de-rated. If several semiconducting devices are to be mounted on the same heat-sink so as to maintain uniform semiconductor temperatures, the devices cannot be equally spaced since as the air warms and rises along the heat-sink fins the top of the heat-sink will not be cooled as effectively as the bottom.

Heat-sinks are made of a variety of alloys, the particular material depending on the application. Copper is frequently chosen where the size (volume) of the heat-sink is important. Magnesium may be chosen where weight is a critical factor. Aluminum is usually chosen when cost is important. Aluminum heat-sinks are available with clean surfaces or with an anodized surface. Anodizing increases the thermal emissivity of the heat-sink with a thin coating on the aluminum. Painting an aluminum heat-sink with an oil base paint (any color) has about the same effect as anodizing on emissivity, however, the paint layer acts as an insulator in terms of thermal conduction. If the heat-sink gets rid of heat by a convection process, paint is inferior to a black anodized coating. If the heat-sink primarily radiates (as in a vacuum), paint may be used effectively. Commercially available heat-sinks frequently have clean, bright, surfaces. It is frequently cheaper in terms of dollars cost/watt dissipation to use a larger unfinished heat-sink than a smaller, more expensive, anodized heat-sink.

Transient Heat Flow

The fact that the temperature of a device that is dissipating power takes some time to reach steady state becomes important in fast pulse circuits (where "fast" means pulse times shorter than the time required for the device to achieve thermal equilibrium) and in cyclic processes where the power dissipated varies as a function of time within the cycle. Consider figure 4 - 5 in which the power dissipated as a function of time is typical for the SCR's in the inverter circuits of the previous chapter. Such a curve would result in the case of constant forward current through the SCR for half a cycle followed by a half cycle of negligible power dissipation while the SCR is blocking.

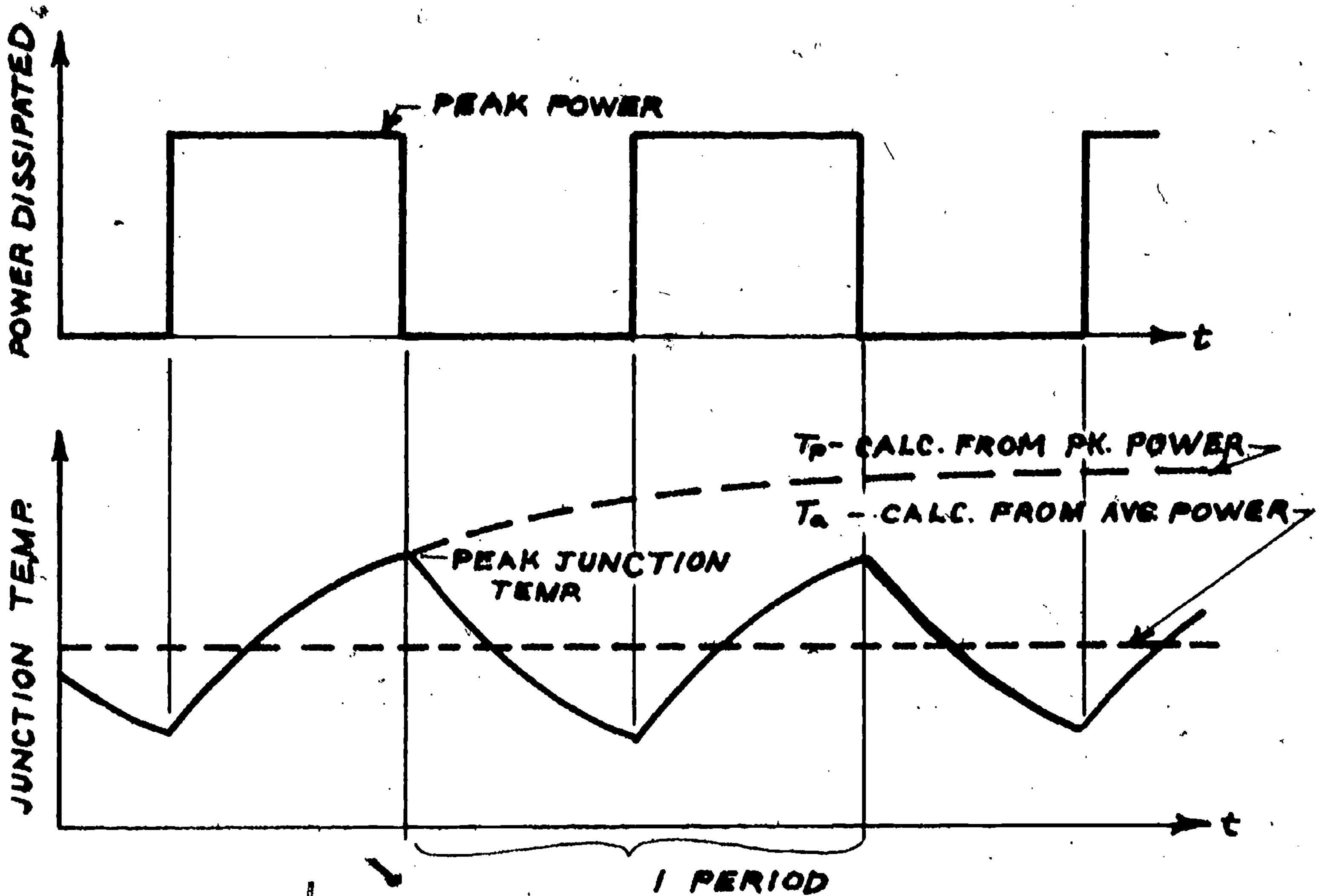
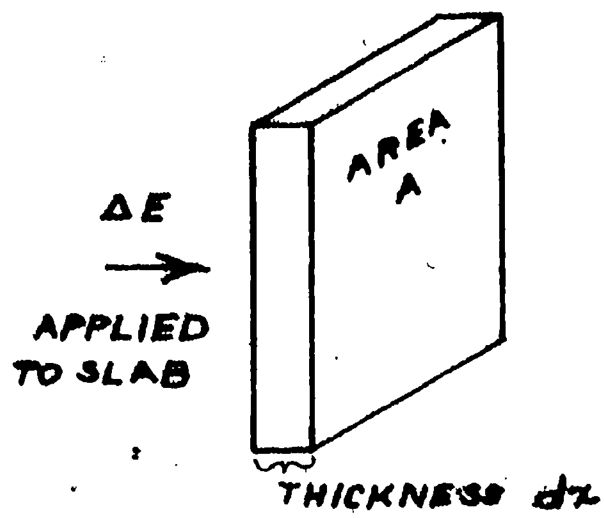


Figure 4 - 5

If the device were operated for a long time at the peak power level, a device junction temperature T_p could be reached for a given thermal system. If the repetition rate of the circuit is increased so that the period of a cycle was very short compared to the time required for the system to reach thermal steady-state, the junction temperature would be T_a as calculated from the average power dissipated during a period. For time periods in between these long and short extremes, the peak junction temperature lies in between T_p and T_a. In many practical situations, knowing just where between T_p and T_a the actual peak junction temperature will lie makes a significant difference in the choice of an SCR or the components of its heat-dissipation system.

The reasón that some time is required for the thermal system to reach equilibrium or steady-state is related to the "heat capacitance" of the system. We know from basic physics that if we apply thermal energy uniformly to a piece of material (a "slab") that is thermally isolated from the surrounding world, the temperature of the material will increase in proportion to the energy supplied (Fig. 4 - 6).



and $m = \rho A \cdot dx$

$\Delta E = \int H\,dt = cm\,\Delta T$

ΔE = energy supplied
H = power supplied
t = time
c = specific heat - depends on material
m = mass of slab
ΔT = resulting temperature change.

ρ = density of material
A = area of slab
dx = slab thickness

Figure 4 - 6

The quantity $cm = c\rho A\,dx = C$ where "C" is known as the "heat capacity" of the slab. Note that the temperature is uniform throughout the slab (steady-state or else k, the thermal conductivity is infinite). The heat capacitance has an analogy in the electrical system. Comparing the analogous quantities:

Thermal		Electrical
H = heat flow	→	I = current flow
T = temperature	→	V = electrical potential.
$\frac{\Delta x}{kA} = \theta$ = thermal resistance	→	R = electrical resistance

C = thermal capacitance $\longrightarrow$ C = electrical capacitance

since,

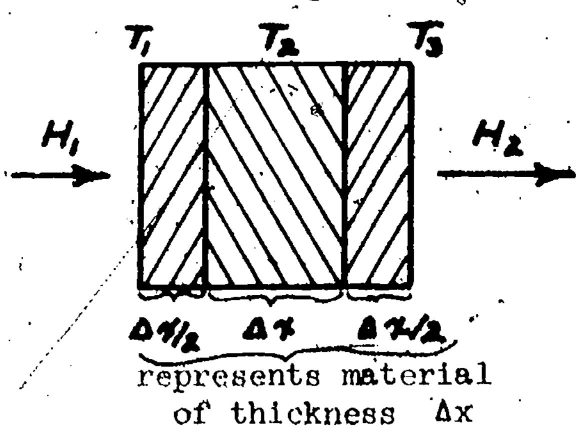
$$\Delta E = \int H\,dt = C\Delta T \longrightarrow \Delta Q = \int I\,dt = C\,\Delta V$$

or

$$\Delta T = \frac{1}{C}\int H\,dt \longrightarrow \Delta V = \frac{1}{C}\int I\,dt$$

Note that the equations are analogous term by term. We have not only shown that electrical capacitance is analogous to heat capacity, but that electric charge Q is analogous to thermal energy E. Clearly, <u>analogous does not mean "equal to"</u>!

To <u>model</u> a real material, both the specific heat and the thermal conductivity must be considered. We must find a way to add together or superpose these two models. In the case that the slab is very thin so that the thermal capacitance and resistance are "small" such that the temperature variation across the slab is small compared to the required temperature <u>accuracy</u>, we make a first approximation as follows. Represent the thin slab as a three layer sandwich as shown in figure 4 - 7.




The end materials have zero heat capacity and have thermal conductivity "k".

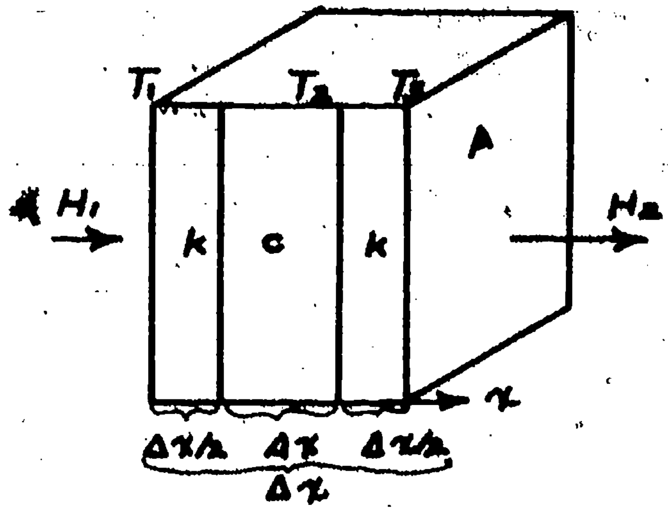

The middle material has infinite thermal conductivity and has specific heat c.

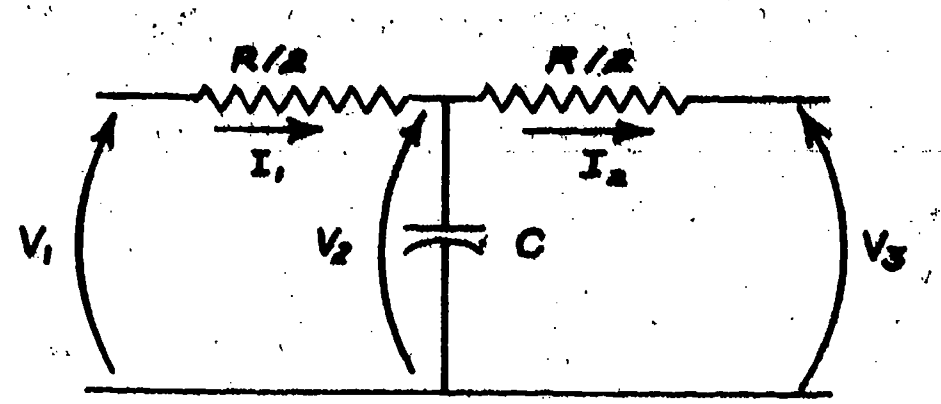
Figure 4 - 7

The rationale of such a model depends upon T_1, T_2, and T_3 not being very different. Regarding heat capacity, as heat energy is applied to the slab, the temperatures change, and T_1 will change in a different manner than T_3. If T_1 and T_3 are nearly the same, T_2 will be some sort of "average temperature" of the whole slab and will to a first approximation describe the variation of all three temperatures as functions of time as energy is added to or removed from the slab. Also, regarding the thermal resistance model, if the temperatures are nearly the same, the temperature distribution in the slab can be modeled by two straight lines (from $T_1 - T_2$ and $T_2 - T_3$) to a reasonable accuracy. Furthermore, if T_2 is nearly the same as T_1 and T_3, the slab must be close to its steady-state condition for the amount of heat flow through the slab, and and thus the thermal resistance model which has been considered for steady-state is valid.

We must define what we mean by T_1, T_2, and T_3 being "nearly the same". Because of the variety of temperature scales available (Celsius, Farhenheit, Kelvin, Rankine) and the particular definition relating to freezing and boiling water of the commonly used Celsius (centigrade) scale, a per-cent type definition such as "$\frac{T_3 - T_1}{T_3} < .05$" is not very meaningful. We could compare the temperature difference across a slab to the accuracy we require, for example, we could require "$T_3 - T_1 < \beta\ 3^{\circ}C$" where β is some factor relating the error to the temperature distribution in the slab, but it is not obvious how β is to be determined. Thus it appears that before the accuracy question can be resolved, the temperature distribution must be found.

In figure 4 - 8, the equations describing the behavior of the thermal model and the analogue of the thermal model are developed side-by-side.





WHERE $R = \frac{kA}{\Delta x}$

$C = c\rho A \Delta x$

END MATERIALS

$$H_1 = -\frac{kA}{\Delta x/2}(T_2 - T_1) \qquad I_1 = -\frac{2}{R}(V_2 - V_1)$$

$$H_2 = -\frac{kA}{\Delta x/2}(T_3 - T_2) \qquad I_2 = -\frac{2}{R}(V_3 - V_2)$$

CENTER MATERIAL (USING CONSERVATION OF ENERGY / CHARGE)

$$\Delta E = (H_1 - H_2)\Delta t = c\rho A \Delta x \Delta T_2 \qquad \Delta Q = (I_1 - I_2)\Delta t = C \Delta V_2$$

Figure 4 - 8

If we are given sufficient boundary conditions, such as T_1 and T_3 as functions of time, the other parameters (such as H_1, H_2, T_2) may be found as functions of time using the equations shown in figure 4 - 8.

We again raise the question of accuracy of the model. Suppose that during some transient process the temperature distribution in a homogeneous material is given by the solid curve in figure 4 - 9.

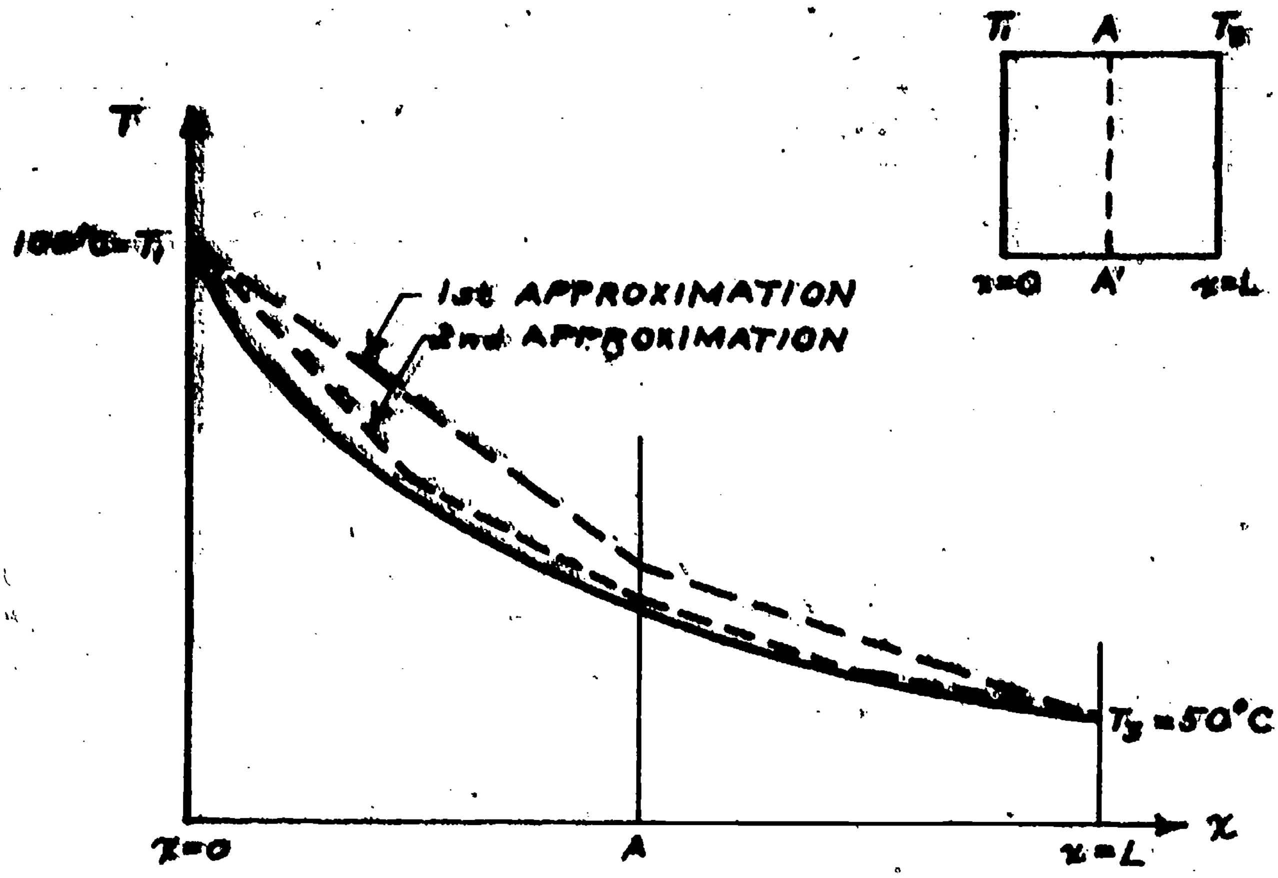


Figure 4 - 9

The temperature distribution throughout the material might be estimated using the model of figure 4 - 8 as the upper of the dashed curves in figure 4 - 9. Such a curve might be sufficiently accurate for the purpose at hand. Suppose further that such a temperature distribution is either not sufficiently accurate or that we can't yet tell if the estimated distribution is sufficiently accurate. We can divide the material into two thinner slabs along the plane AA'. Because the model is more accurate for each of the new slabs (the temperature difference across the slabs is reduced), the total temperature distribution of the two slab system is more accurate (the lower dashed curve of figure 4 - 9). Considering the electric analogy (Figure 4 - 10), we recognize that this subdividing technique is exactly the same as that used in making lumped parameter models of transmission lines and in designing delay lines.

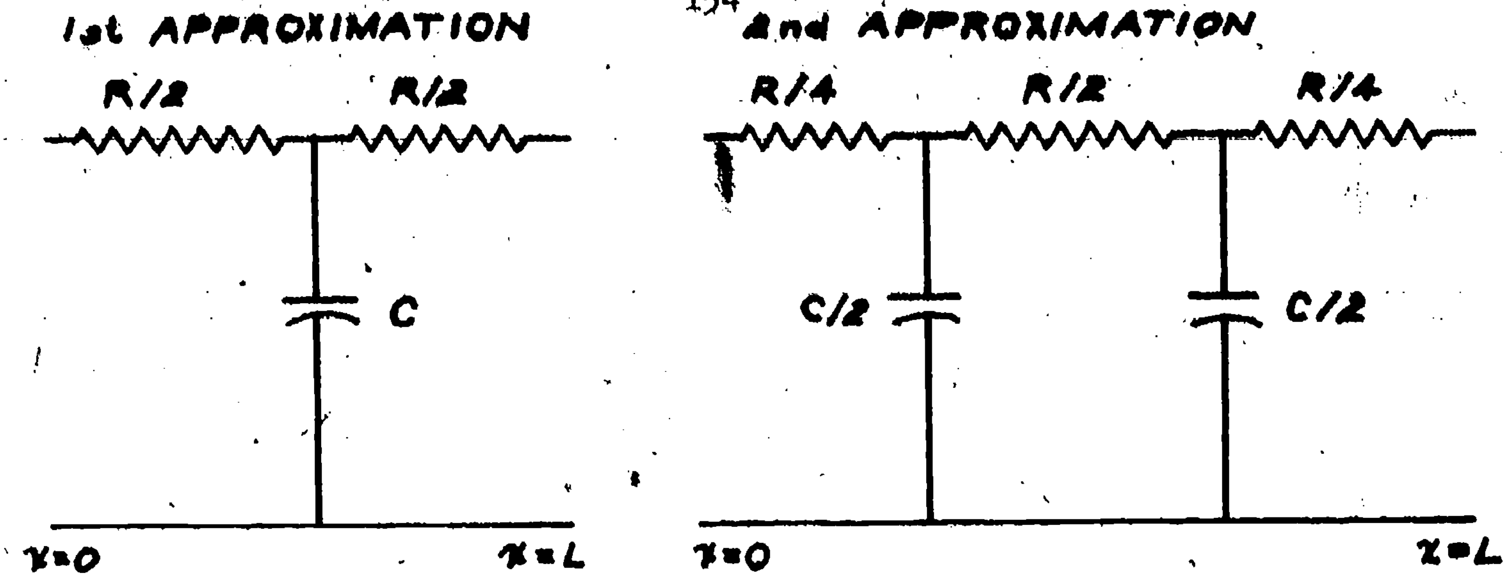


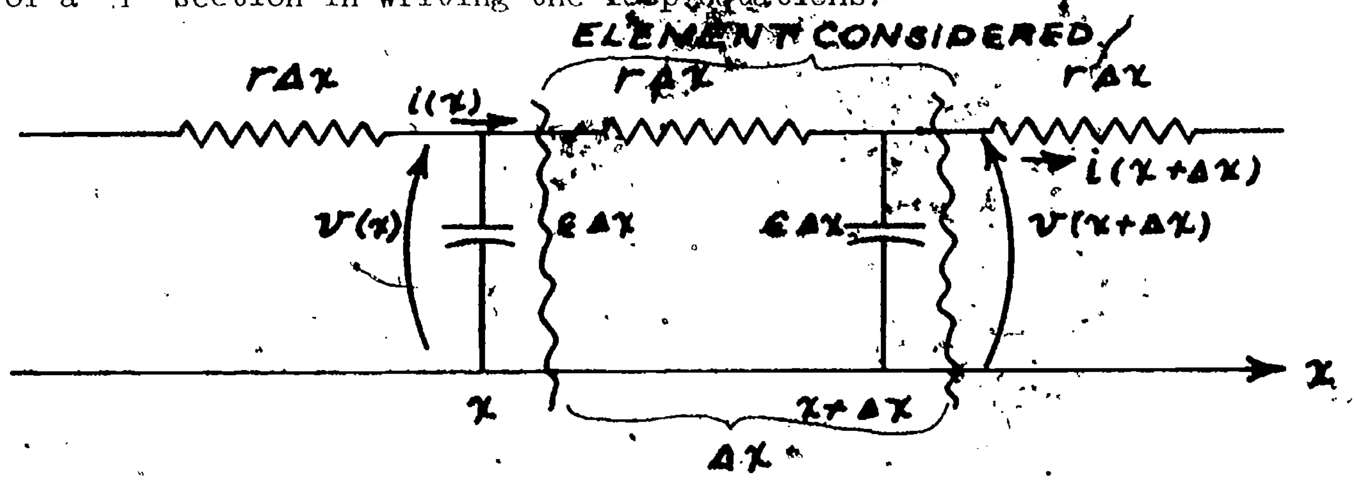
Figure 4 - 10

As the material is subdivided into thinner and thinner slabs, the accuracy increases and so does the work involved in solving the model. In the limit as the number of slabs approaches infinity, the calculated temperatures distribution approaches the actual temperature distribution to within an infinitesimal error ϵ (the fundamental idea of integral calculus). A method of approaching a desired accuracy in the temperature distribution would be to successively double the number of slabs. When the general shape of the temperature distribution curve no longer changes significantly and when the changes in temperature of the various locations of interest no longer change significantly between doublings, the required accuracy has been achieved. This is the same idea as used in determining how many terms of a mathematical series are significant.

Obviously, it will not take very many such doublings of the number of slabs until the problem becomes painfully complicated to solve. We could utilize a digital computer to perform the calculation or we could utilize the idea of allowing the number of slabs to approach infinity. The application of calculus should give us the exact answer in a single equation provided that sufficient boundary conditions are known. We explore the calculus approach for two reasons.

If we know how much work will be involved in determining the exact solution, we will be able to intelligently decide where to cut off our many-slab models. Also, the technique we shall use to solve the problem exactly (separation of variables) is a general and often used technique that is worth reviewing. The "separation of variables" technique of solving partial differential equations is also commonly used in determing the characteristics of transmission lines and waveguides, solving boundary value problems in field theory, and separating the time dependent and distance dependent portions of Schroedinger's wave equation in quantum mechanics.

We begin the exact analysis by considering the electrical analogy of figure 4 - 8. In figure 4 - 11, we consider a section of the electrical analogy and shift our origin so that an "L" shaped section need only be considered instead of a "T" section in writing the loop equations.



$r = \frac{R}{L}$ $c = \frac{C}{L}$ WHERE L = THICKNESS OF MATERIAL

v(x) = voltage at position x

i(x) = current at position x

v(x + Δx) = voltage at position x + Δx

i(x + Δx) = current at position x + Δx

Figure 4 - 11

Writing Ohm's law for the elementary resistor;

$$v(x+\Delta x) - v(x) = -r\,\Delta x\, i(x). \qquad (1$$

Writing Kirchoff's current law for the node,

$$i(x+\Delta x) - i(x) = -c\,\Delta x\,\frac{\partial v(x+\Delta x)}{\partial t}. \qquad (2$$

Note the use of the partial derivative $\frac{\partial v}{\partial t}$. The partial is indicated because the voltage v is a function of distance and time, but the current into the capacitor at position x + Δx depends only on the time rate of change of voltage at that position.

Taking the limit as Δx approaches zero (implying the number of slabs approximating the material whose total thickness is L is approaching infinity) equations 1 and 2 become:

$$\lim_{\Delta x \to 0}\left(\frac{v(x+\Delta x) - v(x)}{\Delta x}\right) = \frac{\partial v(x)}{\partial x} = -r\,i(x) \qquad (3$$

and

$$\lim_{\Delta x \to 0}\left(\frac{i(x+\Delta x) - i(x)}{\Delta x}\right) = \frac{\partial i(x)}{\partial x} = c\,\frac{\partial v(x)}{\partial t}. \qquad (4$$

Notice the $\frac{\partial v(x)}{\partial t}$ term in equation 4 results from the fact that $\frac{\partial v}{\partial t}$ (x + Δx) approaches $\frac{\partial v(x)}{\partial t}$ as Δx approaches zero.

Now that the quantities of interest in equations 3 and 4 are related at the position x instead of x and x + Δx, we can dispense with the parentheses denoting locations. Thus equations 3 and 4 may be rewritten:

$$\frac{\partial v}{\partial x} = -ri, \qquad (5$$

and

$$\frac{\partial i}{\partial x} = -c\,\frac{\partial v}{\partial t}. \qquad (6$$

To solve this pair of simultaneous partial differential equations, we first eliminate one of the variables by substitution so that we have a resulting partial differential equation of one variable. We choose to eliminate i and solve for v since if v is known as a function of x at any instant of time, i can be found from equation 5 by a simple matter of taking a derivative. To eliminate i, first take the partial derivative of equation 5 with respect to x.

$$\frac{\partial^2 v}{\partial x^2} = -r\frac{\partial i}{\partial x}. \qquad (7$$

Equation 6 may be substituted into equation 7, eliminating $\frac{\partial i}{\partial x}$ and yielding

$$\frac{\partial^2 v}{\partial x^2} = r\epsilon\frac{\partial v}{\partial t} \qquad (8$$

We attempt to solve equation 8 by a technique known as the separation of variables. v is known to be a function of both distance x and time t. If we can somehow separate v into a time variable g(t) (independent of x) and a distance variable f (x) (independent of t) we would have two "regular" differential equations (as opposed to a partial differential equation) which we know how to solve. We begin by assuming a product solution, i.e., that v(x,t) = f(x) g(t). In English, this statement reads, voltage which is a function of distance and time is equal to a function "f" of distance multiplied by another function "g" of time. We begin with a product solution because of our experience in solving similar equations in transmission lines problems. If the product solution doesn't work, we will try other mathematical operations such as v(x,t) = f(x) + g(t) or $f(x)/g(t)$ etc. We must take several partial derivatives before we can substitute the product solution into equation 8.

$$v(x,t) = f(x)g(t) \qquad (9$$

$$\frac{\partial v(x,t)}{\partial t} = f(x)\frac{\partial g(t)}{\partial t} \qquad \frac{\partial v(x,t)}{\partial x} = g(t)\frac{\partial f(x)}{\partial x}$$

$$\frac{\partial^2 v(x,t)}{\partial x^2} = g(t)\frac{\partial^2 f(x)}{\partial x^2}$$

Substituting into equation 8;

$$g(t)\frac{\partial^2 f(x)}{\partial x^2} = r\varepsilon f(x)\frac{\partial g(t)}{\partial t}.$$

Rearranging by algebra;

$$\frac{1}{f(x)}\frac{\partial^2 f(x)}{\partial x^2} = r\varepsilon \frac{1}{g(t)}\frac{\partial g(t)}{\partial t}. \qquad (10$$

In equation 10, the variables have been separated, that is, the left side of the equation is a function only of x, and the right side of the equation is a function only of t. In equation 9, a change in x results in a change in f(x). It does not change g(t) because f(x) and g(t) are assumed independent. In equation 10, the only way that a change in x and f(x) would not change g(t) would be if each side of the equation were equal to a constant, K^2. Rewriting equation 10:

$$\frac{1}{f(x)}\frac{\partial^2 f(x)}{\partial x^2} = K^2 = r\varepsilon \frac{1}{g(t)}\frac{\partial g(t)}{\partial t},$$

or

$$\frac{1}{f(x)}\frac{d^2 f(x)}{dx^2} = K^2, \qquad (11$$

and

$$r\varepsilon \frac{1}{g(t)}\frac{dg(t)}{dt} = K^2. \qquad (12$$

In equations 11 and 12, since the variables have separated resulting in two equations containing one variable each, we may use the total derivative symbols $\frac{d}{dt}$ and $\frac{d}{dx}$ instead of the partial derivative symbols $\frac{\partial}{\partial t}$ and $\frac{\partial}{\partial x}$.

Equations 11 and 12 may now be solved with the aid of a reference on differential equations (ref 3). K^2 must be a real number since $\frac{1}{f(x)}$ and $\frac{\partial^2 f(x)}{\partial x^2}$ are real numbers in a real, physical situation. Solving equation 12:

$$\frac{dg}{dt} - \frac{K^2}{r\varepsilon} g = 0$$

$$g(t) = A\, e^{(K^2/r\varepsilon)t} \qquad (13$$

The constant K^2 must be negative or zero. If K^2 were positive, g would approach infinity as time increased, thus v would also approach infinity as t increased. This is physically ridiculous, therefore;

$$K^2 \leq 0$$

We now proceed to solve equation 11, knowing that $K^2 \leq 0$.

$$\frac{d^2f}{dx^2} - K^2 f = 0$$

If $K^2 = 0$, the solution is

$$f = a x + b \qquad (14$$

where a and b are constants to be determined by the initial, final, or boundary conditions of the problem

If $K^2 < 0$ (K purely imaginary), the solution is

$$f(x) = a \sin |K| x + b \cos |K| x \qquad (15$$

where again a and b are constants to be determined by initial, final, or boundary conditions of the problem. It is now also obvious why K^2 was chosen as the constant in equations 11 and 12 instead of K. The value of K^2 must also be determined from boundary conditions and will in general be found to have many values. The temperature distribution according to equation 15 in a real problem may be expressable as a sum of sine and cosine waveforms (a Fourier series) giving an infinite number of $|K|$s and K^2s. Each K^2 will give another solution to $v(x,t) = f(x)\, g(t)$ = equation 13 multiplied by equation 14 or 15, and all such solutions are valid. Because we have a linear system (a resistor-capacitor network that doesn't contain nonlinear devices like diodes or SCRs), the principle of superposition is valid, and the total solution may be composed by summing all possible individual solutions. Expressed mathematically,

$$v(x,t) = \sum_{k^2=0}^{k^2=\infty} f_{k^2}(x)\, g_{k^2}(t) \quad .$$

This procedure is illustrated in the following example. The basic derivation we have completed is valid for all one dimensional, homogeneous material, heat flow problems. The form of the solution would be different in a cyclindrical or spherical geometry, but the method would remain the same.

Example

We wish to find the temperature distribution as a function of time in a bar of homogeneous material. Originally the bar is at 25°C as is the large heat sink at the right end of the bar.

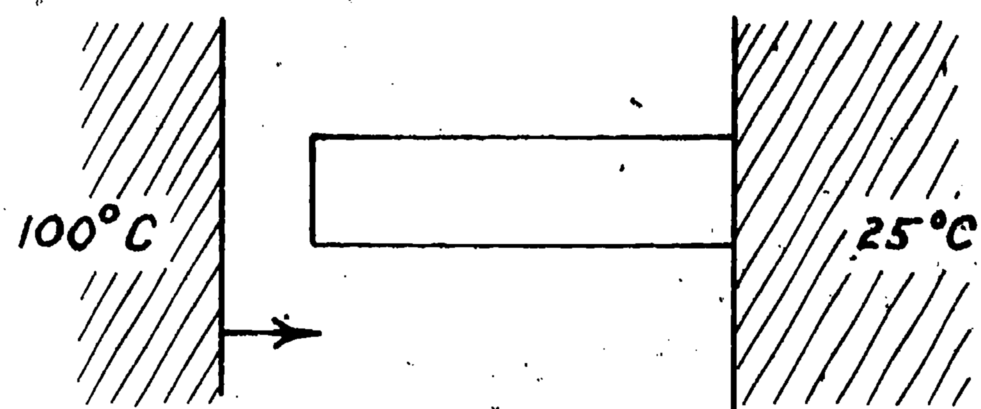


Figure 4 -12

At time t = 0, the large thermal reservoir (a large block of copper for example) is brought in contact with the left end of the bar. The temperature of the reservoir and heat-sink do not change significantly with time. If no significant amount of heat is lost from the bar due to radiation or convection (in a real problem, we would have to estimate these quantities), the heat flow is one dimensional from the reservoir to the heat-sink and our previous analysis can be used.

Stating the initial, final, and boundary conditions:

initially, T = 25°C throughout the bar; finally, in steady state the temperature is linearly distributed over the length of the bar as shown in figure 4 - 13.

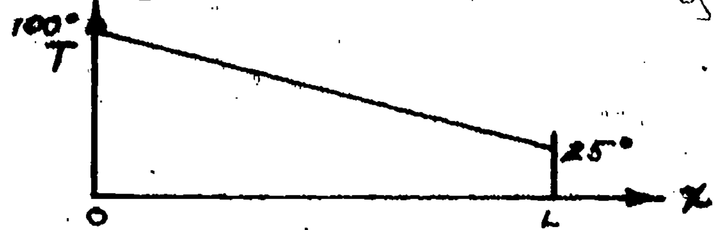


Figure 4 - 13

This distribution follows as a direct result of the steady-state heat flow equation in figure 4 - 1.

$$H = -kA\frac{dT}{dx} \qquad H = \text{CONST.}, \text{ STEADY STATE}$$

$$\therefore dT = -\frac{H}{kA}dx \quad \Rightarrow T = -\frac{H}{kA}x + B$$

$$\text{AT } x = 0, T = 100°$$

$$\text{AT } x = L, T = 25°$$

$$\therefore 25 = -\frac{H}{kA}L + 100°$$

$$\therefore \frac{H}{kA} = \frac{75}{L}.$$

Next we form the products f(x) g(t)

$$v(x,t) = (a_0 x + b_0)A_0 e^0 + \sum_{K^2<0}^{K^2=-\infty} A_{K^2}(a_{K^2}\sin|K|x + b_{K^2}\cos|K|x)e^{K^2/r_2 t}$$

For the final steady state, the sum terms disappear because the $e^{(K^2/r_2)t}$ terms approach zero (remember K^2 is negative) as t approaches infinity. We can now determine the constants of the first term ($K^2 = 0$) of the series using the final conditions,

$$T(x,\infty) = C_0 x + d_0 \quad (A_0 a_0 = C_0, A_0 b_0 = d_0),$$

$$= -\frac{75}{L}x + 100.$$

Initially, T (x,t) was equal to 25°C throughout the bar. Therefore

$$25 = -\frac{75}{L}x + 100 + \sum_{K^2<0}^{K^2=-\infty} A_{K^2}(a_{K^2}\sin|K|x + b_{K^2}\cos|K|x)e^0$$

or, lumping A_{K^2} times a_{K^2} and A_{K^2} times b_{K^2} together, at t = 0

$$\frac{75}{L}x - 75 = \sum_{K^2<0}^{K^2=-\infty}(C_{K^2}\sin|K|x + d_{K^2}\cos|K|x).$$

We use the standard technique of defining a temperature cyclic in x and which uses the left side of the above equation as part of the cycle. Then the right side of the equation becomes a Fourier series. The answer will be valid only over that portion of the cycle that the left side of the equation defines the temperature.

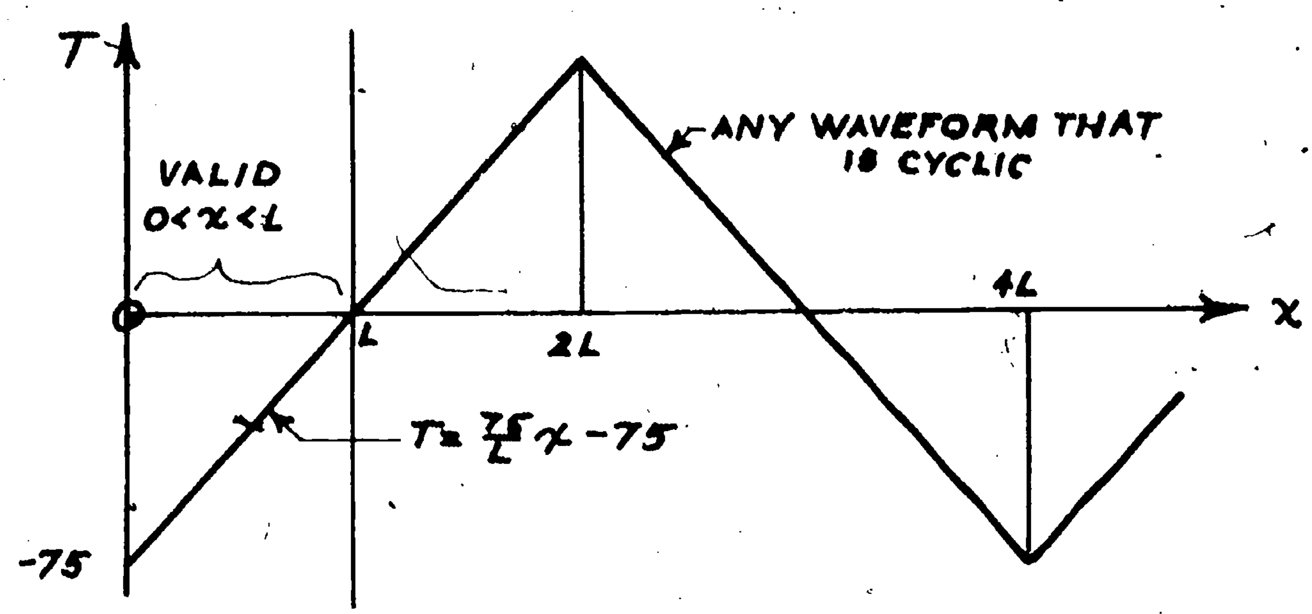


Figure 4 - 14

Note that $|K| = 0$ has already been used. Therefore the waveform we choose to define a cycle $(L < x < 4L)$ must have a zero average value so that no dc level will appear in the Fourier series. The waveform chosen has a symmetry such that no sine terms appear in its Fourier series. Also, we can now determine the possible values of K^2.

When $x = 4L$, the first cosine argument is 2π.

$$K = \frac{2\pi}{4L} = \frac{\pi}{2L} \text{ for the first term}$$

and

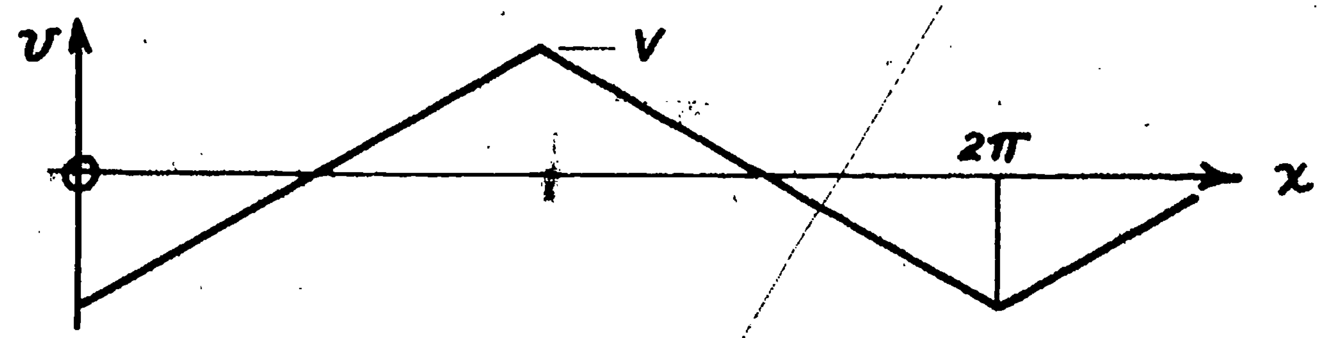
$$K^2 = \frac{\pi^2}{4L^2}$$

For the second term of the series

$$|K| = \text{twice as much as the first term} = \pi/L$$

$$K^2 = \frac{\pi^2}{L^2} \qquad \text{etc.}$$

The Fourier series for the waveform shown in figure 4 - 15 is well known.



$$v = -\frac{8V}{\pi^2}\left(\cos x + \frac{1}{9}\cos 3x + \frac{1}{25}\cos 5x + \frac{1}{49}\cos 7x + \cdots\right)$$

Figure 4 - 15

Translating this solution into terms of our problem gives

$$\sum_{K^2<0}^{K^2=-\infty} \left(c_{K^2} \sin|K|x + d_{K^2}\cos|K|x\right) =$$

$$-\frac{8(75)}{\pi^2}\left(\cos\left(\frac{\pi}{2L}\right)x + \frac{1}{9}\cos\left(\frac{3\pi}{2L}\right)x + \frac{1}{25}\cos\left(\frac{5\pi}{2L}\right)x + \frac{1}{49}\cos\left(\frac{7\pi}{2L}\right)x + \cdots\right).$$

The values of K^2 are:

$\lvert K\rvert$	K^2
$\frac{\pi}{2L}$	$-\frac{\pi^2}{4L^2}$
$\frac{3\pi}{2L}$	$-\frac{9\pi^2}{4L^2}$
$\frac{5\pi}{2L}$	$-\frac{25\pi^2}{4L^2}$

etc.

We have yet to determine the thermal equivalents of $r\Omega$.

$r = \frac{R}{L}$ is analogous to $\frac{\theta}{L}$.

$$\mathcal{C} = \frac{C_{elec.}}{L} \text{ is analogous to } \frac{C_{therm}}{L}$$

$$r\mathcal{C} = \theta C/L^2.$$

Substituting into the full expression for T (x, t) =

$$T(x,t) = -\frac{75}{L}x + 100$$
$$-\frac{8(75)}{\pi^2}\Big[\left(\cos\frac{\pi x}{2L}\right)e^{-\frac{\pi^2}{4\theta c}t} + \frac{1}{9}\left(\cos\frac{3\pi x}{2L}\right)e^{-\frac{9\pi^2}{4\theta c}t}$$
$$+\frac{1}{25}\left(\cos\frac{5\pi x}{2L}\right)e^{-\frac{25\pi^2}{4\theta c}t} + \frac{1}{49}\left(\cos\frac{7\pi x}{2L}\right)e^{-\frac{49\pi^2}{4\theta c}t} + \dots\Big].$$

Figure 4 - 16 shows the temperature distribution in the bar at various times according to the above equation.

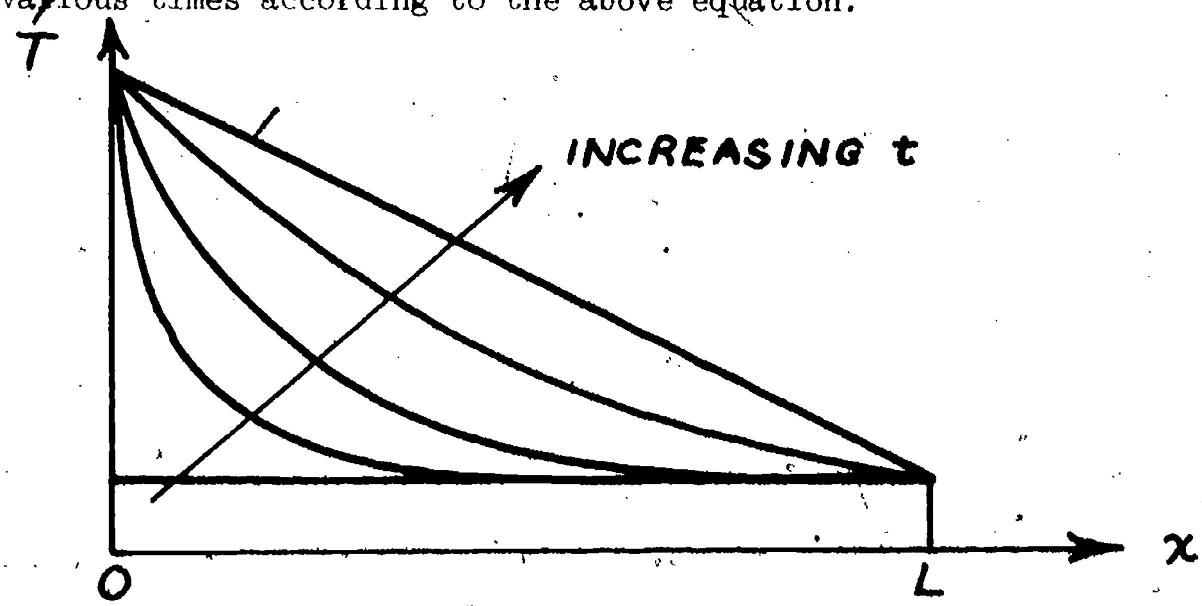


Figure 4 - 16.

It must be remembered that the solution of the previous example is only "exact" in the sense that its accuracy is limited only by the accuracy of the model (one directional heat flow, no radiation or convection, perfect heat source and sink), provided a sufficient number of terms of the solution are considered. The method of analysis chosen should also depend upon whether or not the more laborious model will in fact give more accurate answers.

Transient Thermal Impedance

It is useful to devise another model of the thermal system that is valid only for a particular power dissipated versus time curve. Such a model has the advantage of extreme simplicity. For example, an SCR may be used as a "static switch" in which a current is turned on for a short period of time. Assume the current waveform and the power dissipated in the SCR are as shown in figure 4 - 17a and b. The SCR junction temperature is given as a function of time in figure 4 - 17c. The transient thermal impedance "θ_c" is commonly defined as the ratio of the maximum temperature rise to the power dissipated or

$$P = \theta_c (T_{max} - T_{ambient}).$$

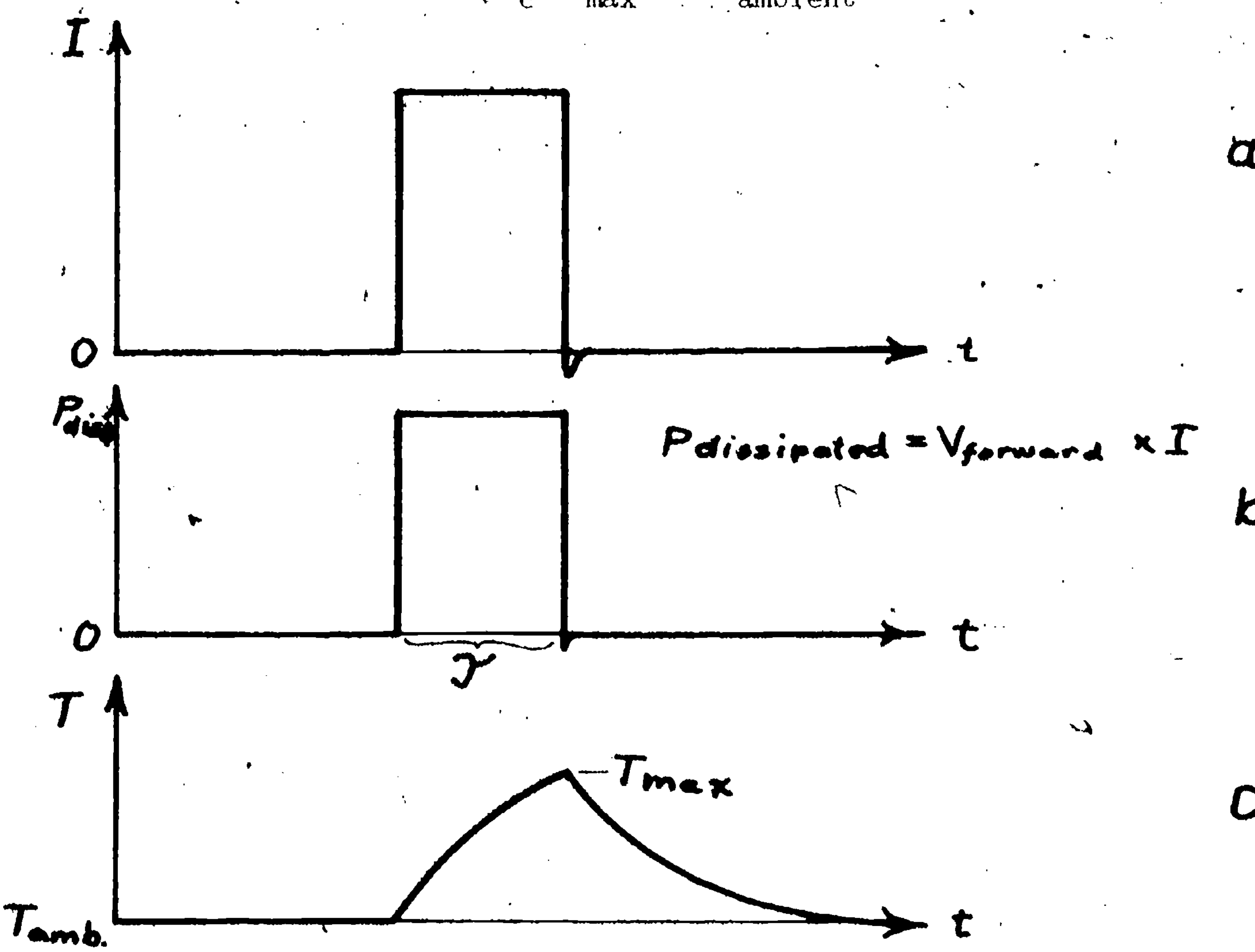


Figure 4 - 17

The transient thermal impedance defined in this way changes as a function of the time "τ" the power pulse is applied. If τ is very short, θ_c is small. As τ becomes long, θ_c approaches the steady-state thermal resistance. Figure 4 - 18 shows a typical plot of θ_c vs τ for a square pulse of power dissipated in an SCR. Because we still have a linear system, the change in temperature is proportional to the amplitude of the power applied in the same time interval so that to a first approximation θ is not dependent upon P.

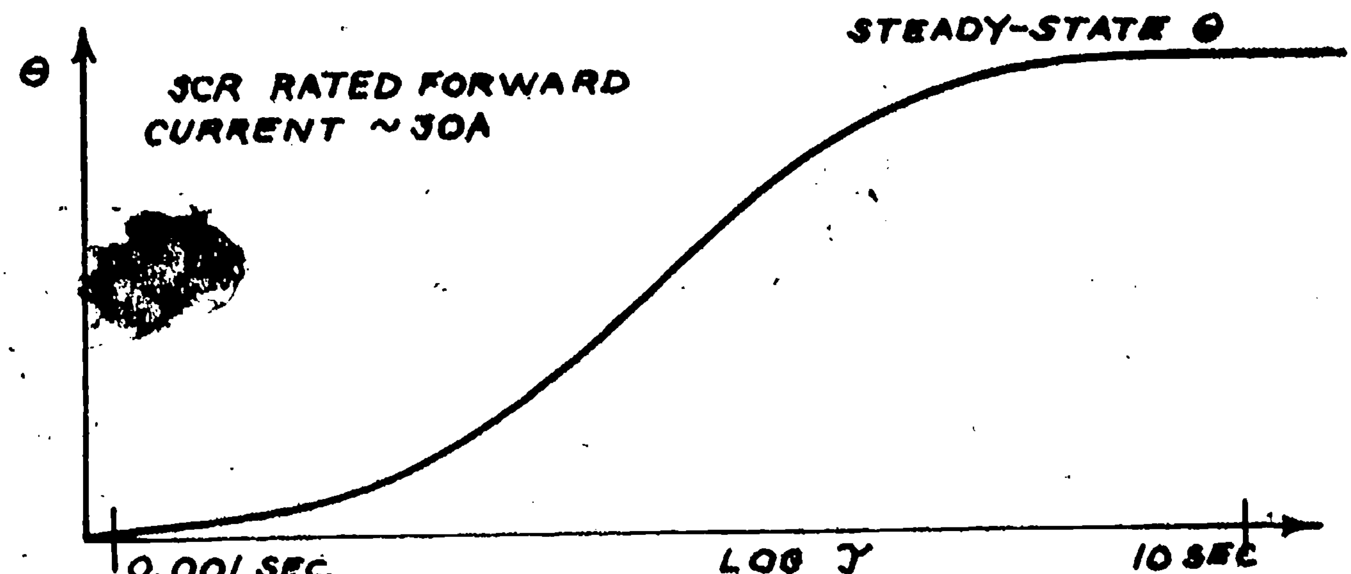


Figure 4 - 18

Summary

In this chapter, we have reviewed two basic thermal properties of materials, namely, thermal conductivity and heat capacity. Thermal resistance was introduced as a convenience in solving heat flow problems related to SCR heat sinking. In addition, on the basis of empirical evidence, it was found that heat sinks which cool by processes of convection and radiation could also be said to have a thermal resistance.

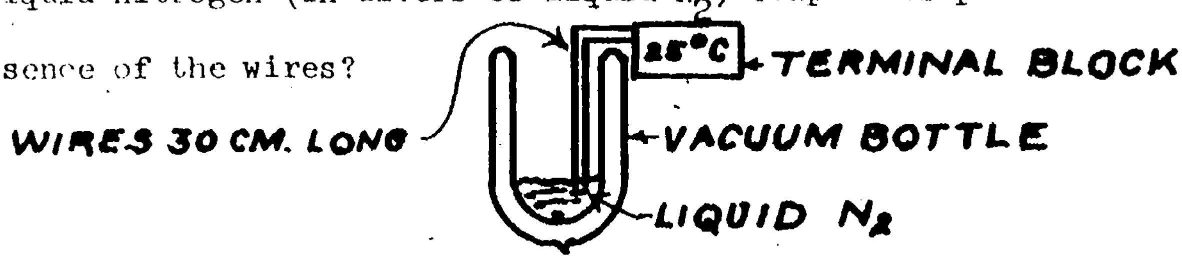
Thermal capacitance became a useful concept in the consideration of transient temperature distributions in thermal conduction problems. A material conducting heat was modeled by superimposing the models for steady-state heat conductivity and thermal capacitance (called a "lumped parameter model" because it was composed of blocks of "ideal" materials). It was then argued that the accuracy of the transient model increased as the thickness of the material being modeled decreased. Accordingly, thick materials were subdivided into larger numbers of thin slabs. The electrical analogy (developed alongside as an aid to gaining insight into the thermal behavior of materials) to the thermal model suggested a standard mathematical approach in which the thickness of each slab could be made to approach zero and an exact transient solution to transient heat conduction problems could be found. An example one dimension heat flow by conduction problem was presented to indicate the amount of effort involved in calculating an "exact" solution. It was intimated that in a three dimensional heat flow problem, the difficulties and effort involved would drastically increase. This required effort gave rise to the notion of transient thermal impedance, a simple empirical relation between power dissipated in the device and temperature rise for a specific dissipation waveform.

This chapter has introduced two more kinds of modeling. Modeling by analogy, as exemplified by the electrical analogues of the thermal properties, and superposition of models of different linear phenomena, as exemplified by the development of the lumped parameter model of transient heat flow. It must be remembered that superposition is only valid in linear systems (in which output is proportional to input).

Exercises

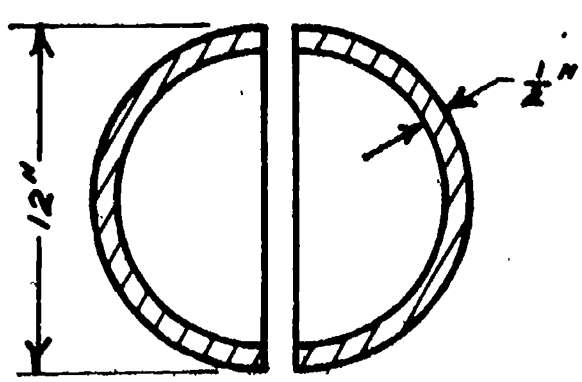
1. Two blocks of aluminum, A and B, are thermally insulated from their surroundings and each other. Block A weighs 5 times as much as block B. Block A is at a temperature of 100°C while block B is at a temperature of 20°C. If the two blocks are brought together (touch) while they are still insulated from their surroundings, what will be the steady-state temperature of block A?

2. Two electrical leads (#14 copper wire) extend into a vacuum Dewar (thermos bottle) as shown in the figure below. The upper ends of the copper wires are at room temperature. The lower ends of the wire are in liquid nitrogen. How much liquid nitrogen (in liters of liquid N_2) evaporates per minute due to the presence of the wires?



Problem 1

In the design of electronic instrumentation for use in the deep ocean, it is frequently necessary to protect the electronics from the high pressures occurring in the ocean depths (5,000 - 15,000 psi). One exceptionally strong instrument package for these purposes is a hollow glass sphere. The sphere is composed of two hollow hemispheres. The high pressures force the two hemispheres together, slightly deforming the glass into a perfect seal.



glass thickness = ½ inch

sphere diameter (outside measurement) = 12 inches

If the electronics package inside the sphere dissipates 100 watts, and the outer surface of the sphere is at the ambient temperature of the deep ocean ($0^{o}C$) what is the temperature at the inside surface of the sphere? Assume the heat flow is uniform over the surface of the sphere. What is the maximum thermal gradient (in $^{o}C/cm$) in the glass?

If the thermal gradient in the glass becomes too large, the thermal stresses caused by the thermal gradient will cause the glass to crack. Of course, the glass can sustain much larger thermal gradients when it is under pressure, but there still is likely to be some maximum allowable value. Recall that in transient effects, very large thermal gradients can exist before steady state is established. If the electronics inside the glass sphere is brought up to power slowly, the transient thermal gradients can be minimized. How slowly should the electronics in the sphere be brought up to full power so as not to put undue thermal stress on the glass sphere? (Please answer in terms of seconds). Explain your reasoning!

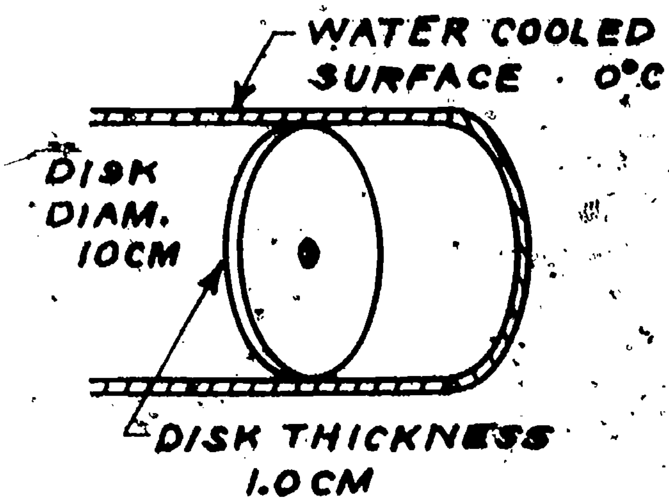
Thermal conductivity of glass = 0.002 calories/sec through a plate
1 cm thick of area 1 cm^2 with a temperature difference of $1^{o}C$
across the plate

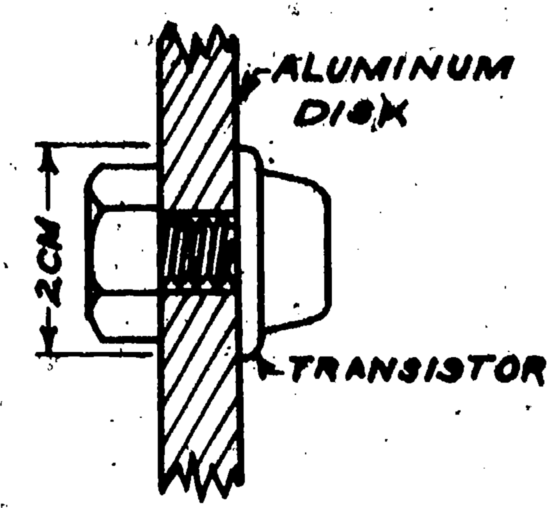
Specific heat of glass = 0.117 calories/gram

Density of glass = 4.0 grams/cm^3 (rf. Handbook of Chem. & Phys.).

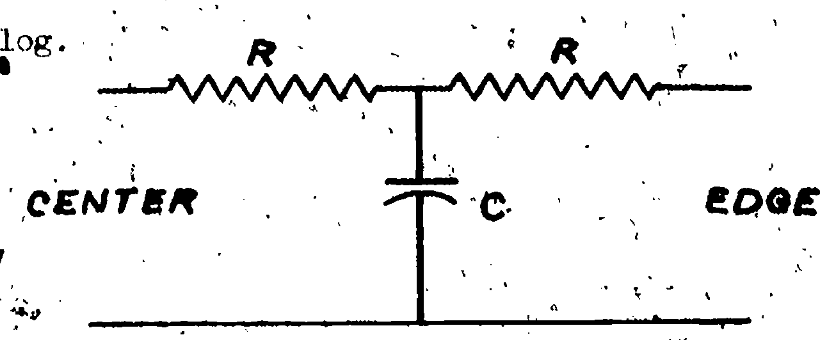
Problem 2

A power transistor (a heat source) is located at the center of an aluminum disk. The outer edge of the disk is in intimate contact with a salt-water cooled surface such that temperature measurements at the edge of the disk show a constant temperature of $0^{o}C$ (within experimental accuracy of $\pm 1/2^{o}C$) for the thermal power range under consideration. The transistor is mounted tightly on the aluminum disk and has an "effective" diameter of 2 cm. That is, the plate "can be said" to have a uniform temperature in a 2 cm circle directly under the transistor.





As a first try at analyzing the thermal properties of this system, we shall represent the thermal characteristics of the disk by our familiar resistor-capacitor analog.



Knowing that:

thermal conductivity of aluminum

$= 0.50 \frac{\text{cal}}{\text{sec}}$ through a plate 1 cm thick with areas 1 sq. cm for a temperature difference of $1^{\circ}C$

specific heat of aluminum

$= 0.2185$ cal/gram$^{\circ}$C

density of aluminum = 2.74 grams/cc

and that 1 cal = 4.186 joules

1) Calculate the appropriate values of R and C for our analog.

2) Calculate a new R and C if the thickness of the disk is only 0.50 cm.

3) Assuming a junction-to-case thermal impedance for the transistor of $1^{\circ}C$/watt and a case to disk impedance of $0.4^{\circ}C$/watt, find the steady-state temperature of the junction if the transistor dissipates 20 watts.

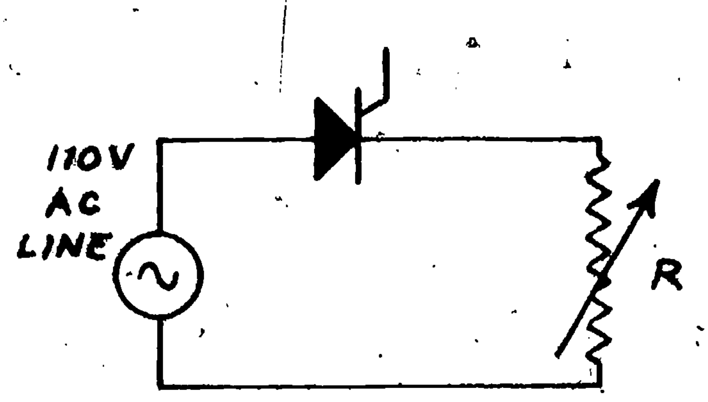
4) Find the temperature distribution in the disk to within $\pm 1/2^{\circ}C$ 1 second after the transistor is "turned on."

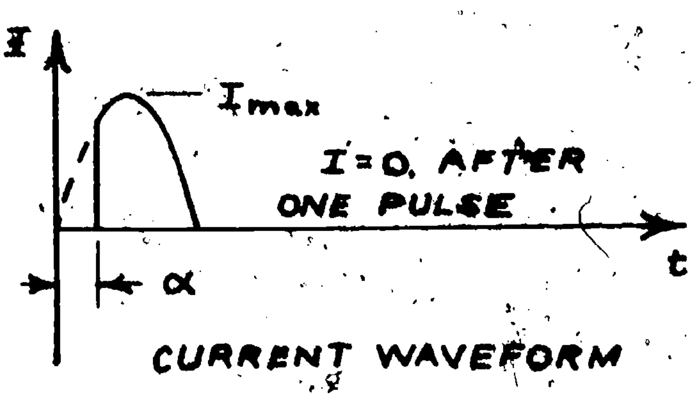
Laboratory Problem 1

We have a number (about a dozen) of 35 amp, 500 volt SCR's mounted on heat-sinks for student use in the laboratory. Several of these SCR's have failed recently for unknown reasons. A cursory inspection of the SCR's shows that they are not mounted on their heat-sinks in accordance with the manufacturer's specifications. We must decide whether remounting the SCR's is worth our trouble. Determine the maximum ratings of the SCR as presently attached to its heat-sink. Please do not destroy any more SCRs.

Laboratory Problem 2

A 35 amp, 500 volt SCR is to be used in a single pulse generating circuit is shown in the following figure. In order that we can relate the maximum current I to the firing angle, determine the transient thermal impedance of the SCR-heat-sink system for $0^{\circ} < \alpha < 90^{\circ}$.





References

1) Robert Murray, Jr., Ed., Silicon Controlled Rectifier Designers' Handbook, Westinghouse Electric Corp., Youngwood, Pa., 1963.

This reference gives a brief review of the thermal properties materials and "cooling power curves" for typical heat-sinks for both natural and forced air convection. Thermal impedance data for stud-mounted rectifiers are also presented.

2) F. W. Gutzwiller, Ed., SCR Manual, 4th Ed., General Electric Corp., Syracuse, N.Y., 1967.

This reference lists methods of mounting SCR's other than stud mounting, and contains a brief section on heat-sink fin design.

3) Dwight, Herbert Bristol, Tables of Integrals and Other Mathematical Data, 3rd Ed., Macmillan Co., New York, 1957.

This reference is one of many excellent tables of integrals and solutions to differential equations.

The following articles are available from the Wakefield Engineering Inc. Company, Wakefiled, Massachusetts in addition to the referenced journals.

4) Wayne Goldman, "An Introduction to the Art of Heat Sinking", Electronic Packaging and Production, July, 1966.

5) Wayne Goldman, "9 Ways to Improve Heat Sink Performance", Electronic Products, Oct., 1966.

6) Wayne Goldman, "Torque and Thermal Resistance", Electronics, Sept. 7, 1964.

Chapter 5 A Free-wheeling Diode DC Motor Drive

Introduction

In this chapter, we consider a simple but efficient motor controller. The methods used in determining the steady-state behavior of the circuit are the same as used in Chapter 3, but the analysis is slightly more complicated by the motor and mechanical load properties. The calculation of a "turn-on" or speed-change transient is considered for the case of a non-linear load characteristic. The non-linear system equations are solved by an iteration technique, and the model of "quasi-steady-state" in which some system quantities can be assumed to be in steady-state while other quantities are considered as undergoing a transient is introduced.

Free-wheeling Diode Circuit

The circuit under consideration is shown in figure 5-1. This circuit is frequently referred to as a "chopper circuit" because the action of the SCRs is to "chop" the direct voltage into a series of voltage pulses.

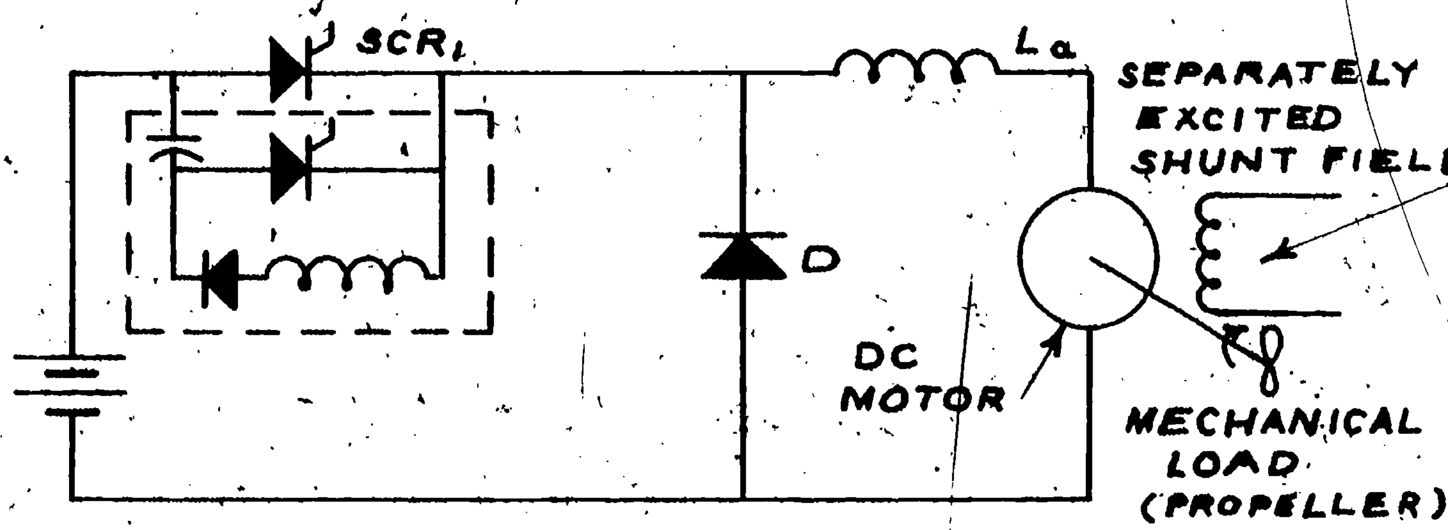


Figure 5 - 1

That portion of the circuit enclosed in dashed lines can be recognized as the turn-off circuit for SCR_1. The basic idea of the circuit is that SCR_1 acts as a simple switch. If the switch is closed for a long time, the motor will reach the maximum steady-state speed determined by the battery voltage, the motor characteristics and the mechanical load characteristics. If the switch remains open, the motor doesn't turn the propeller. If the switch is alternately open and closed in a cyclic manner, the motor will run at some speed in between zero and maximum.

The circuit is basically an efficient circuit since except for winding resistance and the forward conducting characteristics of the SCR and diode, there are no dissipative elements in the circuit. The inductance L_a (including the armature inductance of the motor) maintains the armature current through diode D when SCR_1 is not conducting. Thus the motor torque (proportional to the armature current) is smooth in time rather than pulsating, a very desirable feature. This circuit is most often used with a series wound motor rather than a shunt wound motor. In such a case, L_a, the series field inductance is sufficiently large that an additional external inductor is not usually necessary. However, the shunt motor is slightly easier to analyze in the circuit, and so the series motor problem is left as a home problem at the end of the chapter.

Modeling the Circuit #1 - Steady-state

The SCR_1 and its turn-off circuit is modeled as a simple switch as shown in figure 5 - 2. In actual problems it may be necessary to consider whether or not the capacitor delivers significant energy to the motor during the discharge part of its cycle, however, this additional consideration adds difficulty without being particularly instructive. Therefore, for the sake of brevity (the calculation is not all that hard) we shall assume the turn-off circuit has no significant effect on the operation of the circuit other than to turn-off

SCR_1. Likewise, we neglect the forward conducting voltage drops across the SCR and diode in all the models of this problem, and we assume no reverse or blocking leakage current.

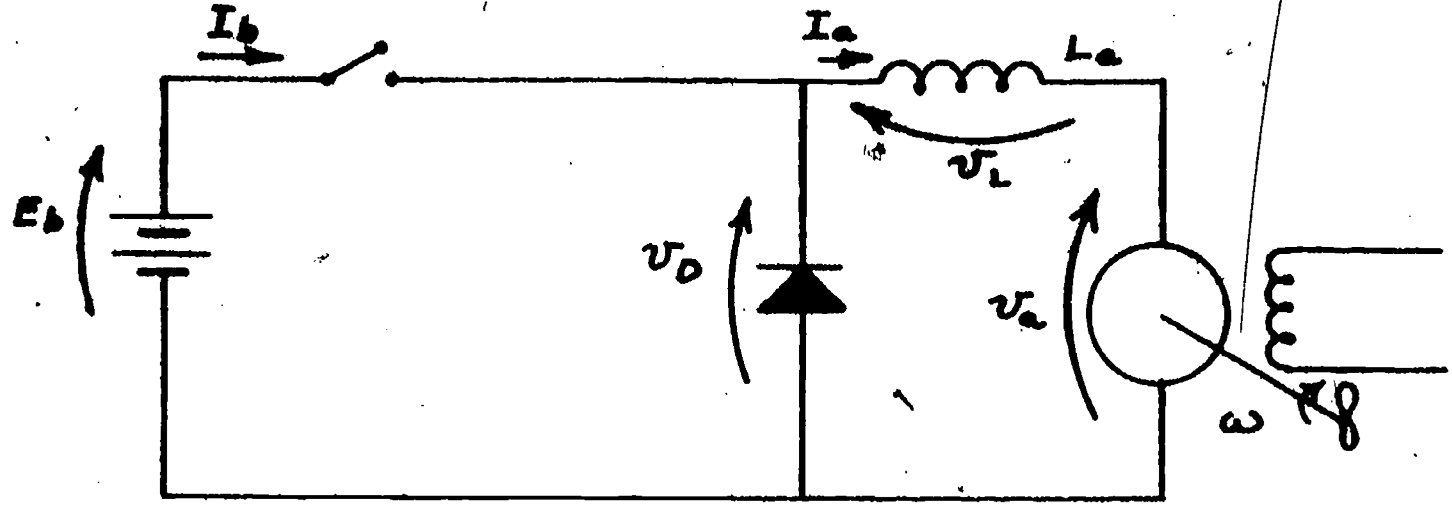


Figure 5 - 2

As a first try at analyzing the circuit, we make some assumptions known to be incorrect, but helpful. Assume L_a is so large that I_b can be considered constant and assume the inductor winding resistance and the motor armature resistance are negligibly small. These assumptions will be relaxed in succeeding models. Also, if the switch is opened and closed frequently enough, the inertia of the motor rotor and propeller will tend to keep ω, the angular frequency of the shaft rotation, constant. The motor armature voltage "E_a", which is equal to the terminal voltage v_a if the brush drop and armature winding resistance are negligible, is proportional to the field flux multiplied by the angular frequency of rotation, i.e.

$$v_a = k\phi\omega \qquad (1)$$

where k = constant of proportionality. Assume that the flux for the shunt machine is constant, thus neglecting armature reaction and assuming a constant field current. We also know the armature current is proportional to the motor torque multiplied by the flux.

$$T = k'\phi I_a \qquad (2)$$

where T = motor torque.

Note that k' is not necessarily equal to k because in linearizing the non-linear motor characteristics, neglecting losses, and neglecting armature reaction (calling ϕ constant), we have neglected motor characteristics that are not negligible. Accordingly, k and k' are determined by a "best fit" approximation to empirical motor data. Usually, k and k' differ by only a few percent for common motors exceeding a few horsepower.

The motor load is indicated as a propeller. To a first approximation, we then expect the torque required to drive the propeller to be proportional to the square of the angular velocity "ω", that is,

$$T = \alpha \omega^2 \tag{3}$$

where α = proportionality constant.

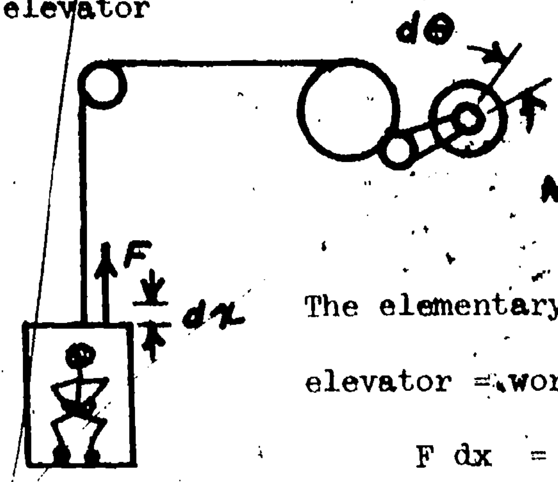
If the motor is properly "matched" to the propeller using a mechanical matching device (gearbox), the motor will deliver its rated torque at the rated motor speed so that

$$\alpha = \frac{T_{rated}}{\omega^2_{rated}}.$$

Digression-Common Motor Loads

The steady-state torque-speed characteristics of many common motor loads are easily derivable from the elementary principles of mechanics. Provided the losses of the load are negligible, it is usually possible to avoid an analysis of the load mechanism in detail and consider only the basic intent of the load device. Such estimates are more valid for large power capability systems (above 15 hp.) than for low power capability systems (fractional horsepower systems) because at low power levels, it is usually not practical to spend the extra effort and money required to manufacture a "low-loss" device. A series of elementary examples of torque-speed characteristic estimating follows:

The elevator



The elementary work done in raising the elevator = work delivered by motor.

$$F\,dx = T\,d\theta$$

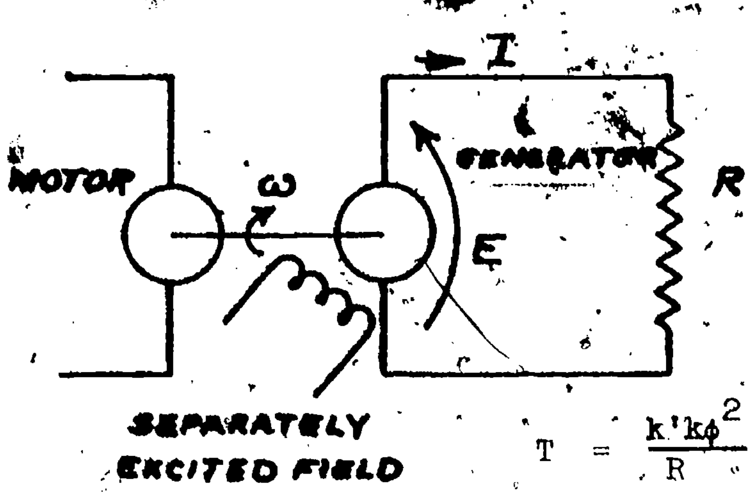
Power "P" = rate of doing work

$$P = F\frac{dx}{dt} = F\,v = T\frac{d\theta}{dt} = T\omega$$

$T = F(\frac{v}{\omega})$ ← a constant depending on the gear ratio and drive winch size.

$\underline{T = \alpha F}$, a constant

The shunt-excited generator



Generator equations

$$E = k\phi\omega$$

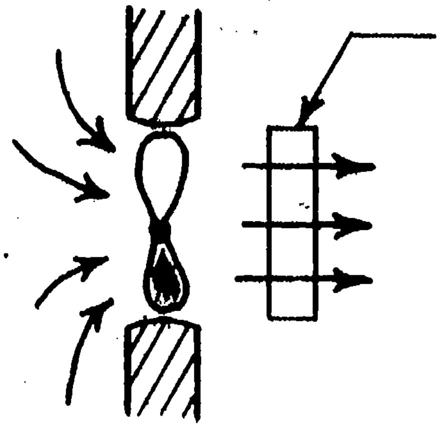
$$T = k'\phi I$$

$$\frac{E}{I} = R = \frac{k\phi\omega}{T/k'\phi}$$

$$T = \frac{k'k\phi^2}{R}\,\omega$$

$\underline{T = \alpha\omega}$, a linear relation

The propeller or fan



Consider a control volume of fluid having a negligible velocity when entering the propeller and having a high velocity "v" when exiting the propeller.

Change in kinetic energy "ΔKE" of control volume $= \frac{1}{2} mv^2$

v^2 is proportional to ω^2

$\therefore dKE = C\omega^2$

The power "P" delivered by the propeller is the rate of change of energy with time

$P = \frac{\Delta KE}{\Delta t}$ where Δt is the time required to change the energy of the control volume.

Δt is inversely proportional to ω since more control volumes per unit time pass through the fan as ω is increased.

Thus $P = \frac{C\omega^2}{C'/\omega} = \alpha\,\omega^3.$

Since $P = T\omega = \alpha\,\omega^3,$

$\underline{\underline{T = \alpha\,\omega^2}}$, a quadratic relation.

We make our first try at finding the steady-state speed, current, and voltage using the same "averaging over the cycle" technique used in the inverter circuits, thereby eliminating the inductor in our equations.

$$v_D = v_L + v_a \qquad (4)$$

$$\frac{1}{2\pi}\int_0^{2\pi} v_D \, d\theta = \frac{1}{2\pi}\cancelto{0}{\int_0^{2\pi} v_L \, d\theta} + \frac{1}{2\pi}\int_0^{2\pi} v_a \, d\theta \qquad (5)$$

Since v_a is proportional to speed (equation 1), and we have assumed a constant speed; v_a is constant.

Therefore,

$$\frac{1}{2\pi}\int_0^{2\pi} v_a \, d\theta = v_a$$

The v_D term is only slightly more difficult to evaluate. Referring back to figure 5 - 2: when the switch is closed, $v_D = E_b$; when the switch is open, I_a flows through the diode which must be conducting and $v_D = 0$. Define θ_{on} (figure 5 - 3) as the interval the switch is closed, θ_{off} as the interval the switch is open, and $g = \frac{\theta_{on}}{\theta_{off} + \theta_{on}}$ as the fraction of the time the switch is closed.

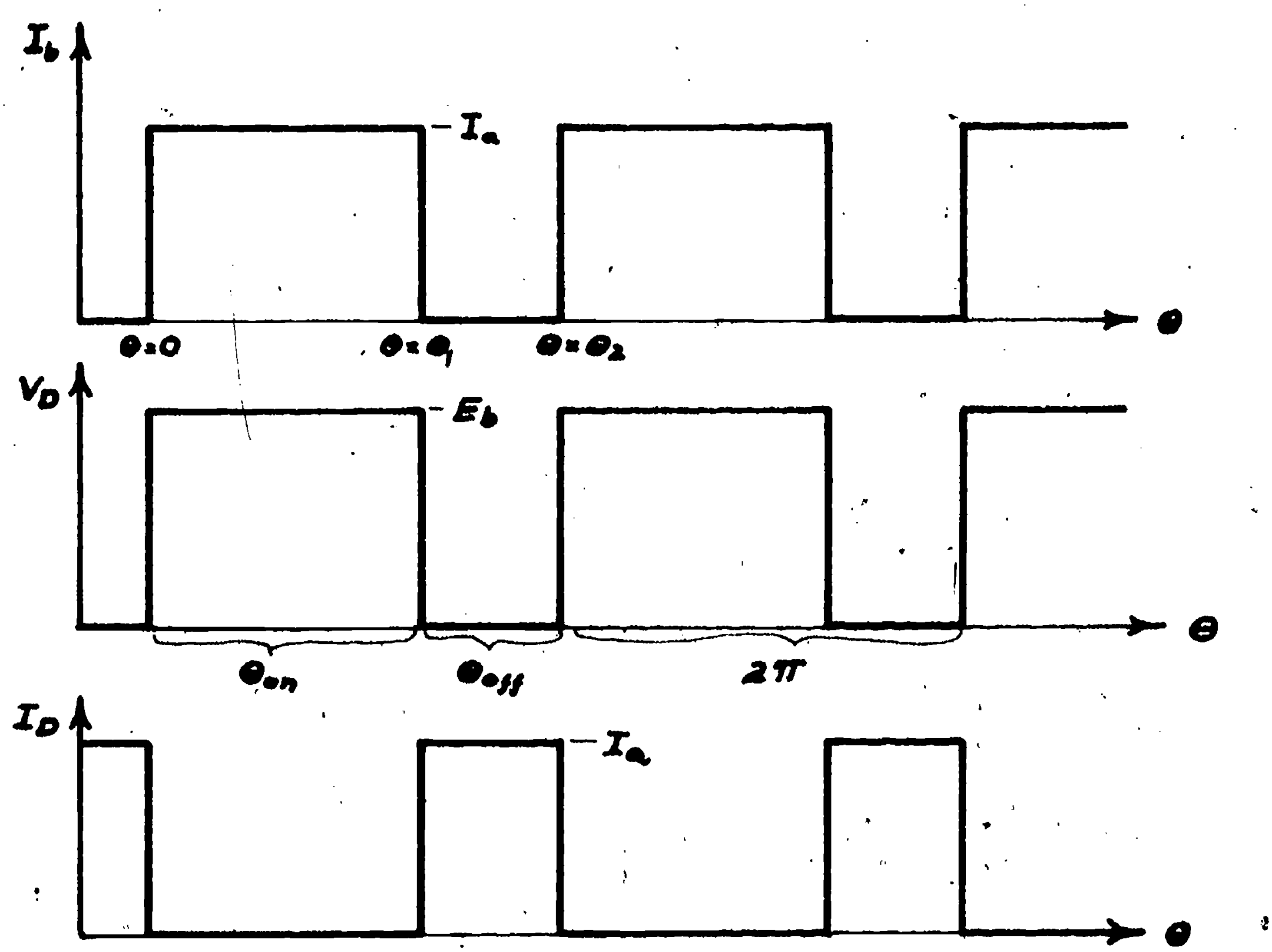


Figure 5 - 3

It is clear from figure 5 - 3 that

$$\frac{1}{2\pi}\int_0^{2\pi} v_D \, d\theta = \frac{1}{2\pi}\int_0^{\theta_1} E_b \, d\theta + \frac{1}{2\pi}\int_{\theta_1}^{2\pi} 0 \, d\theta$$

$$= \frac{1}{2\pi} E_b \theta_{on} .$$

Since

$$\theta_{on} + \theta_{off} = 2\pi ,$$

$$\frac{1}{2\pi}\int_0^{2\pi} v_D \, d\theta = \frac{1}{\theta_{on}+\theta_{off}} E_b \theta_{on} = g E_b$$

Substituting into equation 5 yields

$$g E_b = v_a \qquad (6)$$

Knowing the armature voltage v_a, we find the speed using equation 1.

$$\omega = \frac{v_a}{k\phi} = \frac{g E_b}{k\phi} \qquad (7)$$

Knowing ω allows us to determine the value of I_a using the motor torque equation (equation 2) and the load characteristic (equation 3).

$$T = k'\phi I_a = \alpha \omega^2 ,$$

$$I_a = \frac{\alpha \omega^2}{k'\phi} \quad OR \quad I_a = \frac{\alpha g^2 E_b^2}{k' k^2 \phi^3} \qquad (8)$$

Modeling the Circuit #2 - Steady-state

As the first "refinement" of the calculation we choose to include the effects of the winding resistance of the inductor and the motor armature resistance. The motor brush voltage drop and the SCR and diode forward voltage drops could also be included at this point. We redraw the circuit model as shown in figure 5 - 4. Note that the motor armature resistance has been included with the winding resistance in one lumped "R" just as the armature inductance was included in "L".

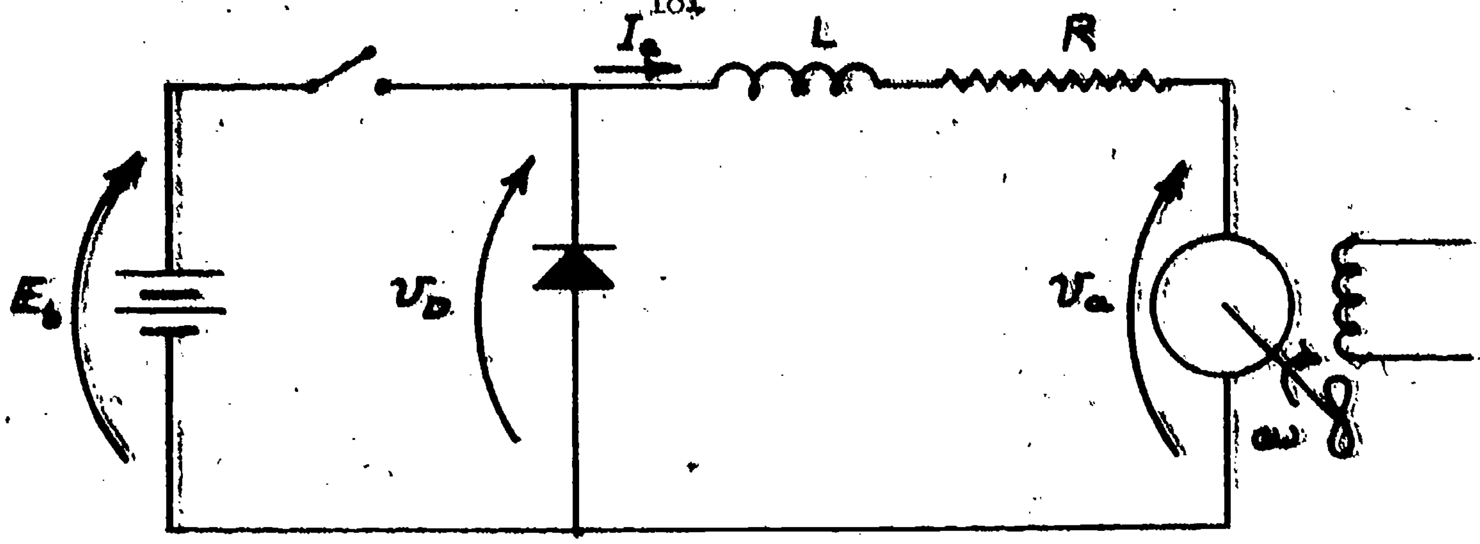


Figure 5 - 4

We retain the assumptions of constant current I_a and constant speed ω. We could rewrite Kirchoff's voltage law around the diode-inductor-resistor-motor loop and average as before, arriving at new (and slightly more complicated) general expressions for v_a, I_a and ω. If we were interested in v_a, ω, and I_a for only a few specific values of g, another approach would be well worth considering.

We have already calculated a first approximation of v_a, ω, and I_a using model #1. We could make a next approximation by using I_a as previously calculated to determine the $I_a R$ voltage drop (a constant because I_a is constant). Subtracting this drop from the average value of v_d yields a new value of v_a.

$$\frac{1}{2\pi}\int_0^{2\pi} v_D \, d\theta - I_a R = g E_b - I_a R = v_{a1}.$$

The new value of v_a (now noted as v_{a1}) can be used to determine a new ω_1 and then a new I_{a1}. These in turn can be used again to recalculate again new values of v_a, ω, and I_a. When the variables no longer change significantly, the answer has been found. Such calculations are frequently performed in a tabular manner for convenience:

FROM MODEL #1

v_{a0}	ω_0	I_{a0}	$I_{a0}R$
$v_{a1} = v_{D\,avg} - I_{a0}R$	$\omega_1 = \frac{v_{a1}}{k\phi}$	$I_{a1} = \frac{\alpha\omega_1^2}{k'\phi}$	$I_{a1}R$
$v_{a2} = v_{D\,avg} - I_{a1}R$	$\omega_2 = \frac{v_{a2}}{k\phi}$	$I_{a2} = \frac{\alpha\omega_2^2}{k'\phi}$	$I_{a2}R$
$v_{a3} = \cdots$	$\cdots$	$\cdots$	$\cdots$

This method of calculation is called an "iterated" calculation and is a frequently used method in digital computer calculations. However, in many practical problems, sufficient accuracy may be attained after only one or two iterations; and if only a few values are required, the iterated calculation may be a significant "short cut" even for hand calculations compared to solving the more complicated problem in general terms.

Modeling the Circuit #3 - Steady-state

As the next refinement we shall let I_a vary within the cycle provided I_a does not go to zero (although it may approach zero arbitrarily closely). In this way we avoid having to break the time interval into more than two "pieces", corresponding to the switch open and closed. We retain the assumption of constant speed (due to the "large" inertia of the motor-propeller system). We first question the validity of our previous analysis.

Returning to the circuit models 1 and 2, since speed is constant, the determination of v_a (table) has the same validity. Since speed is proportional to v_a (equation 1), the determination of ω is also still valid (equation 7). However, the torque relations (equations 2 and 3) are only valid if inertia is ignored, and we have postulated that ω is constant only because of inertia, I_a varying. We must re-examine the torque relations.

Define J = moment of inertia of the armature, gearbox (if any), and propeller. Lumping the moment of inertia of the armature in with the propeller allows us to define a motor torque T as applied to the armature by the electromagnetic field interactions in the motor. Thus the torque relations become (refer to figure 5 - 5).

$$T = k'\phi I_a \quad \text{as before,}$$

and

$$T = J\frac{d\omega}{dt} + \alpha\omega^2. \qquad (9)$$

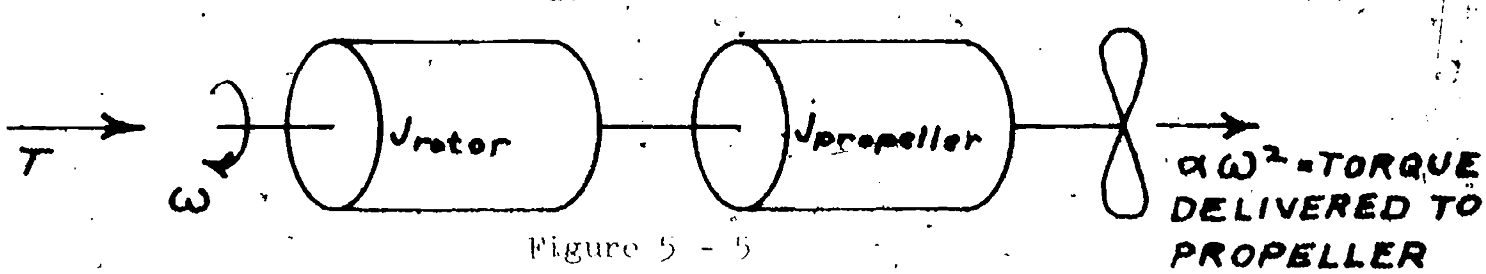


Figure 5 - 5

We can eliminate the torque T from the equations yielding

$$k'\phi I_a = J\frac{d\omega}{dt} + \alpha\omega^2$$

Note that if we average over a cycle, the $J\frac{d\omega}{dt}$ term must disappear, since averaging over a cycle

$$\frac{1}{\tau}\int_0^\tau J\frac{d\omega}{dt}dt = \frac{1}{\tau}J\int_{\omega_0}^{\omega_\tau}d\omega = \frac{1}{\tau}J(\omega_\tau - \omega_0),$$

and $\omega_\tau - \omega_0$ must be equal to zero in a steady-state cyclic process.

Therefore,

$$\frac{1}{\tau}\int_0^\tau k'\phi I_a\,dt = \frac{1}{\tau}\int_0^\tau J\frac{d\omega}{dt}dt + \frac{1}{\tau}\int_0^\tau \alpha\omega^2 dt$$

$$k'\phi I_{a\,avg} = \alpha\omega^2$$

or

$$I_{a\,avg} = \frac{\alpha\omega^2}{k'\phi}. \qquad (10)$$

which is very similar to equation 8.

Digression

Note that we could not have argued that $J \frac{d\omega}{dt}$ is negligible because ω is constant. ω is constant because J is so large, therefore $J \frac{d\omega}{dt}$ may be significant despite $\frac{d\omega}{dt}$ being "small". Also, if $J \frac{d\omega}{dt}$ were negligible, I_a would be constant which contradicts the assumption of I_a varying. Since we are using this argument to determine I_a, assuming $J \frac{d\omega}{dt}$ negligible would be a contradictory and therefore invalid assumption.

We have shown that model #1 retains its validity if the average value of armature current $I_{a\ avg}$ is substituted for the previously constant value of I_a. Furthermore, since

$$\frac{1}{\tau} \int_0^\tau v_R \, dt = R \frac{1}{\tau} \int_0^\tau I_a \, dt,$$

that is,

$$R I_{avg} = v_{R\ avg},$$ all the arguments of model #2 are also valid. Thus we begin model #3 knowing

$$v_a, \ \omega, \text{ and } I_{a\ avg}.$$

We must yet find the instantaneous values of I_a.

Writing Kirchoff's voltage low around the diode-inductor-resistor-motor loop of figure 5 - 4 yields

$$v_D = L \frac{d I_a}{dt} + R I_a + v_a. \qquad (11)$$

Equation 11 is really two equations:

switch open, $v_D = 0$

$$L \frac{d I_a}{dt} + R I_a + v_a = 0 \qquad (12)$$

where v_a is known from model #2,

and

Switch closed, $v_D = E_b$

$$L \frac{dI_a}{dt} + R I_a - (E_b - v_a) = 0 \qquad (13)$$

where E_b will be larger than v_a, still known from model #2.

We still also have

$$I_{a\ avg} = \text{number known from model \#2 for the chosen value of } g \qquad (14)$$

We must find a way to "solve" equations 12 and 13. As we shall see, equation 14 will provide a numerical check of the solution of I_a.

The solution of equation 12 is

$$I_a = \frac{v_a}{R} + C e^{-\frac{R}{L} t}$$

where C is a constant yet to be determined.

The solution of equation 13 is

$$I_a = \frac{-(E_b - V_a)}{R} + C' e^{-\frac{R}{L} t}$$

where C' is another constant yet to be determined.

Changing the variable t into θ (refer to figure 5 - 3) knowing that $\frac{\theta}{2\pi} = \frac{t}{\tau}$ where τ = time the switch is open plus the time the switch is closed, we have:

switch open, $\theta_1 < \theta < 2\pi$

$$I_a = \frac{v_a}{R} + C e^{-\frac{\tau}{2\pi}\frac{R}{L}\theta} \qquad (15)$$

and

switch closed, $0 < \theta < \theta_1$

$$I_a = -\frac{(E_b - v_a)}{R} + C' e^{-\frac{\tau}{2\pi}\frac{R}{L}\theta} \qquad (16)$$

The inductor prevents I_a from changing instantaneously, therefore I_a is a continuous function of time. At θ_1, when the switch opens we must have

$$I_a \text{ (equation 15)} = I_a \text{ (equation 16) at } \theta_1 \qquad (17)$$

Also, because we are in the steady-state of a cyclic process,

$$I_a(\theta) = I_a(\theta + 2\pi)$$

or,

$$I_a(\theta = 0, \text{eqn. } 16) = I_a(\theta = 2\pi, \text{eqn. } 15) \qquad (18)$$

Solving these equations, we first substitute equations 15 and 16 into equation 17.

$$\frac{V_a}{R} + C e^{-\frac{\gamma}{2\pi}\frac{R}{L}\theta_1} = -\frac{(E_b - V_a)}{R} + C' e^{-\frac{\gamma}{2\pi}\frac{R}{L}\theta_1}$$

or, rearranging the terms by algebra,

$$\frac{E_b}{R} e^{+\frac{\gamma}{2\pi}\frac{R}{L}\theta_1} = C' - C. \qquad (19)$$

Next we substitute equations 15 and 16 into equation 18.

$$-\frac{(E_b - V_a)}{R} + C' = \frac{V_a}{R} + C e^{-\gamma R/L}$$

or, rearranging the terms,

$$\frac{E_b}{R} = C' - C e^{-\frac{\gamma R}{L}} \qquad (20)$$

Equations 19 and 20 can be solved for C and C' yielding

$$C = -\frac{E_b}{R}\frac{(e^{\frac{\gamma}{2\pi}\frac{R}{L}\theta_1} - 1)}{(1 - e^{-\gamma R/L})} \qquad (21)$$

and

$$C' = \frac{E_b}{R}\left(1 + \frac{(e^{\frac{\gamma}{2\pi}\frac{R}{L}\theta_1} - 1)}{(e^{\gamma R/L} - 1)}\right). \qquad (22)$$

The values of C and C' may be substituted back into equations 15 and 16. We have found I_a as a function of time. For a given time duration of a cycle τ and a value of the conduction angle $\theta_{on} = \theta_1$, we could plot I_a as a function of θ as in figure 5 - 6. The average value of I_a as calculated by averaging I_a from equations 15 and 16 should be compared with the average value of I_a (equation 14) or determined from model #2, checking the calculation.

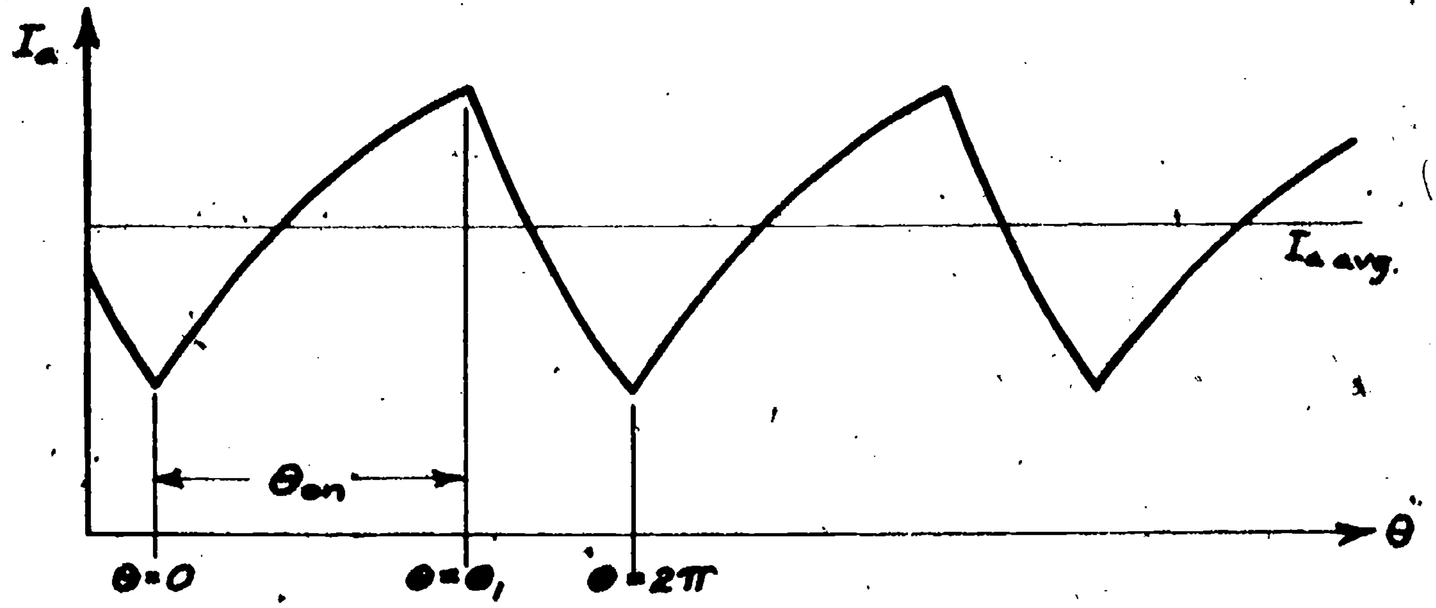


Figure 5 - 6

Modeling the Circuit #4 - Steady-state

As the next and final refinement of the steady-state model that we shall consider, we allow the motor angular speed ω to vary within the switching cycle. Again, we question the validity of our previous models. Returning to model #1 which neglects all resistances, but letting ω vary, we see that the average voltage across the motor armature is still known.

$$v_D \text{ average} = v_a \text{ average}$$

Referring to equation 1, that is

$$v_a = k\phi\omega$$

v_a is no longer constant since ω varies.

However, knowing the average armature voltage allows us to determine the average speed:

$$\frac{1}{2\pi}\int_0^{2\pi} v_a \, d\theta = v_{a\,average} = \frac{k\phi}{2\pi}\int_0^{2\pi}\omega \, d\theta = k\phi\,\omega_{average}$$

Thus

$$\omega_{average} = \frac{v_{a\,average}}{k\phi} = \frac{qE_b}{k\phi}.$$

However, in determining the current, we have some difficulty, from equation 9, that is,

$$k'\phi I_a = \alpha\omega^2 + J\frac{d\omega}{dt}$$

we see the current depends on ω^2. If we try to average this relation, we have

$$k'\phi \frac{1}{2\pi}\int_0^{2\pi} I_a\, d\theta = k'\phi I_{a\,average}$$
$$= \alpha \frac{1}{2\pi}\int_0^{2\pi} \omega^2 d\theta + J\frac{1}{2\pi}\int_0^{2\pi} d\omega \quad (\rightarrow 0)$$
$$= \alpha(\omega_{rms})^2$$

Thus the average current I_a is proportional to the mean-square of the speed ω. There is no way to determine the rms or the mean square of ω if we only know the average value of ω. The relationship between these quantities depends on the specific waveform involved (recall that for a sine wave, $v_{rms} = .707\ v_{peak}$, $v_{aug} = o$; for a full wave rectified sine wave, $v_{rms} = .707\ v_{peak}$, $v_{aug} = .636 v_{peak}$).

Because the average current cannot be determined, the methods used in model #3 cannot be used. There is no way to "adjust" v_a and ω unless the average current is known. We shall have to apply the techniques used in model #3, namely: write the system differential equations, impose the conditions of continuity of current and/or speed, require the variables to have the same values at the beginning and end of a steady-state cycle, and then solve the resulting equations at the appropriate instants of time (or θ).

As a matter of convenience, we rewrite the system equations:

$$v_a = k\phi\omega \qquad (1)$$

$$T = k'\phi I_a \qquad (2)$$

$$T = J\frac{d\omega}{dt} + \alpha\omega^2 \qquad (9)$$

switch open

$$L\frac{dI_a}{dt} + RI_a + v_a = 0 \qquad (12)$$

switch closed

$$L\frac{dI_a}{dt} + RI_a - (E_b - v_a) = 0 \qquad (13)$$

I_a is continuous at switching times $\theta = \theta_1$

and $\theta = 0, 2\pi$ (23)

I_a has the same value at $\theta = 0$, $\theta = 2\pi$ (24)

and/or

ω is continuous at switching times $\theta = \theta_1$,

and $\theta = 0, 2\pi$ (25)

ω has the same value at $\theta = 0$, $\theta = 2\pi$. (26)

We choose to solve for I_a as a function of time. Eliminating the torque T in equations 2 and 9,

$$k'\phi I_a = J\frac{d\omega}{dt} + \alpha\omega^2. \quad (27)$$

Using equation 1 to eliminate v_a from equations 12 and 13 gives:

switch open

$$L\frac{dI_a}{dt} + RI_a + k\phi\omega = 0, \text{ and} \quad (28)$$

switch closed

$$L\frac{dI_a}{dt} + RI_a - (E_b - k\phi\omega) = 0. \quad (29)$$

If we now substitute equation 27 into equations 28 and 29 in order to eliminate ω, we would come up with a set of fairly complicated non-linear (squared terms) differential equations. For example, simultaneous solution of equations 27 and 28 yields:

$$L\frac{d^2I_a}{dt^2} + R\frac{dI_a}{dt} - \frac{\alpha L^2}{Jk\phi}\left(\frac{dI_a}{dt}\right)^2 - \frac{\alpha}{Jk\phi}I_a\frac{dI_a}{dt} + k'\phi I_a - \frac{\alpha R^2}{Jk\phi}I_a^2 - \frac{[illegible]}{Jk\phi} = 0 \quad (30)$$

Such complicated differential equations are frequently solved on an analog computer although a digital computer could also be used. Obviously equation 29 will yield an even more complicated expression. The analog computer can simulate both expressions (30 and 29 and 27 solved together) with the appropriate switching functions. The solution, even using the computer, will be time and effort consuming. It is worth noting that if we had approached this problem by writing down the system equations and solving as indicated in model #4, the relatively "easy" solutions of models #1, 2, and 3 would not have been discovered except by a great expenditure of time and effort, even though those models may be valid in the particular physical situation being considered.

The Starting Transient - Cyclic Iteration

Assume the motor is to be started by cyclicly opening and closing the switch (still representing the SCR circuit as a switch as shown in figure 5 - 2). The starting current transient is important because the peak current determines the peak torque applied to the motor rotor (whose windings could be damaged by excessive torque) and because the motor circuit protection equipment (such as overcurrent relays) must be set to accommodate the peak current. Large armature currents could also damage the motor commutator or if of sufficiently long duration, cause damaging local heating in various circuit elements. Therefore we choose to analyze the starting current transient $I_a(t)$.

Experience in starting DC motors tells us that the circuit resistance plays a non-negligible role in limiting the starting current. Hence our analysis must consider the circuit resistance, the variation of I_a, and indirectly, the variation of ω. We might be tempted to write the system equations and solve them starting from the initial conditions $I_a = 0$, $\omega = 0$. Recall from model #4 that the resulting differential equation is very complicated and now the problem is even worse. In model #4, we had only to determine I_a during a single steady-state cycle. In the present transient problem, we must find I_a during <u>each</u> cycle of a transient lasting possibly many thousands of switching cycles. While it is possible that such a detailed calculation is necessary in certain specific cases, we consider some of the situations in which the transient solution can be simplified.

Suppose ω could be considered constant throughout a switching cycle due to the inertia of the rotor and propeller (as in model #3). Such an assumption would be valid in the case that the change in ω during a cycle and from cycle to cycle is small compared to the value of ω. Clearly, such an approximation is questionable at the starting instant when $\omega = 0$, but the accuracy of a solution based on such an assumption grows as ω increases. Assuming ω constant during a cycle, we rewrite the system equations in differential (instead of derivative) form.

Equation 27 becomes

$$\frac{(k'\phi I_a - \alpha\omega^2)\Delta t}{J} = \Delta\omega \qquad (31)$$

Equation 28 becomes

$$-\frac{(RI_a + k\phi\omega)\Delta t_{open}}{L} = \Delta I_{a\ open} \qquad (32)$$

Equation 29 becomes

$$\frac{(E_b - k\phi\omega - RI_a)\Delta t_{closed}}{L} = \Delta I_{a\ closed} \qquad (33)$$

Figure 5 - 7 shows that

$$\Delta t = \Delta t_{open} + \Delta t_{closed}$$

and

$$\Delta I_a = \Delta I_{a\ open} + \Delta I_{a\ closed}$$

Note that in writing equations 31 and 32 we have tacitly assumed that the circuit time constnat L/R is so large that the exponential variation of armature current with time in a cycle can be approximated by a straight line as shown in figure 5 - 7. If this is not the case, the differential equations of model #3 (constant ω) can be solved to give a more accurate $\Delta I_{a\ open}$ and $\Delta I_{a\ closed}$

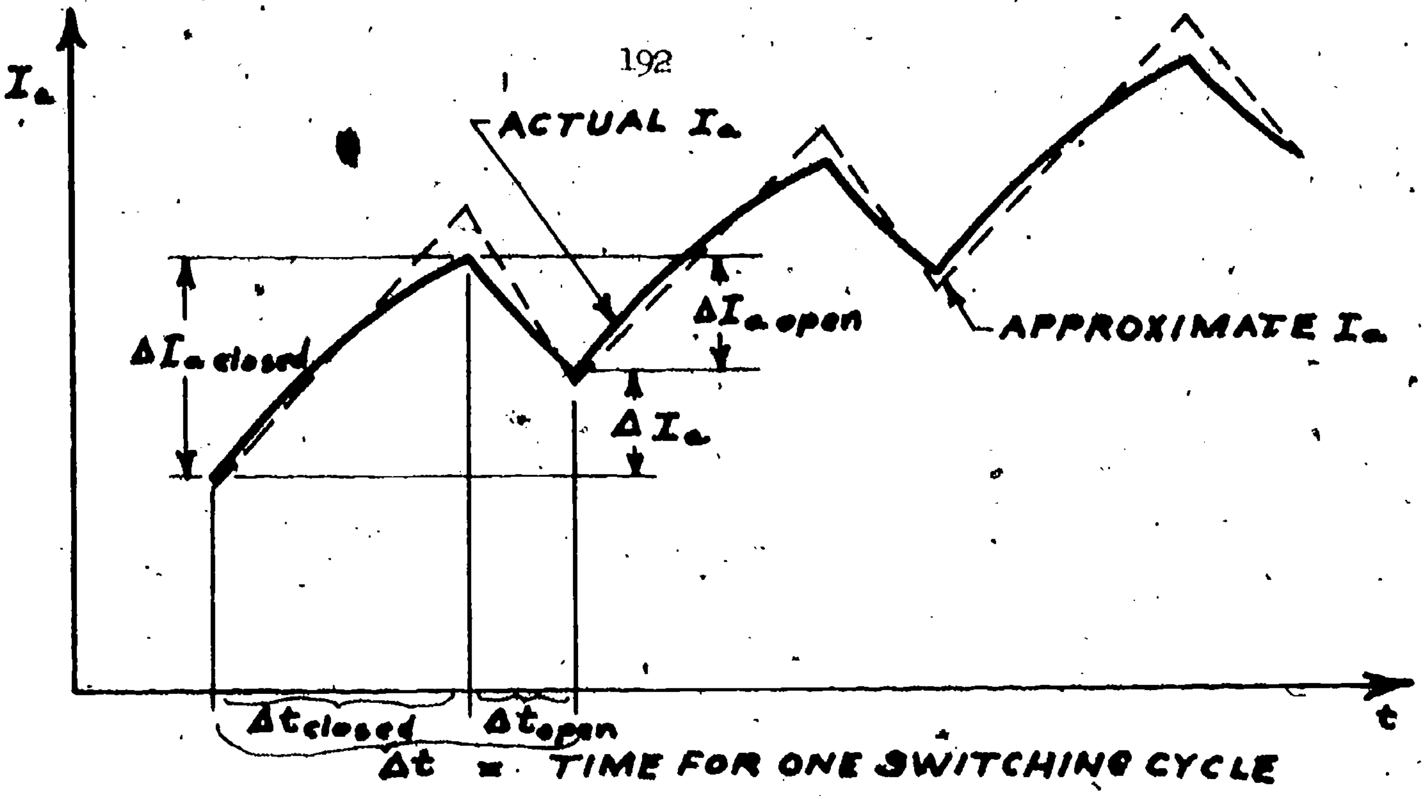


Figure 5 - 7

The solution to the problem can be approximated by an iterated solution of equations 33, 32, and 31 as follows

At $t = 0$, $I_a = 0$, $\omega = 0$,

Solve equation 33 for ΔI_a closed

Next solve equation 32 for ΔI_a open still using the same value of I_a and ω as used in solving equation 33.

Find $\Delta I_a = \Delta I_a$ closed $+$ ΔI_a open

a negative number

Using $I_a = I_a$ previous cycle (0 for first time) $+ \Delta I_a$, solve equation 31 for $\Delta\omega$.

Define a new $\omega = \omega$ previous cycle $+ \Delta\omega$ for the next cycle and repeat the process until ΔI_a and $\Delta\omega$ approach zero.

Such a procedure is particularly suitable for digital computer calculation. The iteration procedure must be repeated until steady-state I_a and ω are reached, possibly involving thousands of cycles. Thus assuming ω constant during a cycle has made the problem much easier to solve although still not tractable for hand calculation.

The Starting Transient - Quasi-steady-state

In the case that the time required for the motor to reach steady-state speed is very long compared to L/R and the switching time Δt (note, we can't speak of a mechanical time constant because of the nonlinear dependence of torque on ω^2), we can make a further simplification. We assume that the electrical circuit is at the steady-state behavior the circuit would have for a particular value of ω. The model in which the steady-state is considered to vary slowly in time is said to be in a "quasi-steady-state". In the quasi-steady-state, we can again average the voltage across the inductor. Referring to figure 5 - 4,

$$v_D = L\frac{dI_a}{dt} + I_a R + v_a,$$

or since $v_a = k\phi\omega$

$$v_D = L\frac{dI_a}{dt} + I_a R + k\phi\omega .$$

Averaging,

$$\frac{1}{2\pi}\int_{2\pi} v_D\, d\theta = \frac{1}{2\pi}\int_{2\pi} L\frac{dI_a}{dt}\, d\theta + \frac{1}{2\pi}\int_{2\pi} I_a R\, d\theta + \frac{1}{2\pi}\int_{2\pi} k\phi\omega\, d\theta .$$

Since ω is "constant",

$$v_{D\,avg} = gE_b = 0 + RI_{a\,avg.} + k\phi\omega,$$

or

$$gE_b = RI_{a\,avg.} + k\phi\omega . \qquad (34)$$

Averaging the mechanical torque equation,

$$k'\phi I_a = J\frac{d\omega}{dt} + \alpha\omega^2 , \qquad (27)$$

$$\frac{1}{2\pi}\int_{2\pi} k'\phi I_a\, d\theta = \frac{1}{2\pi}\int_{2\pi} J\frac{d\omega}{dt}\, d\theta + \frac{1}{2\pi}\int_{2\pi}\alpha\omega^2\, d\theta .$$

Since ω varies only slowly within a cycle, ω^2 is again assumed constant and

$$k'\phi I_{a\,avg.} = J\frac{d\omega}{dt}_{avg.} + \alpha\omega^2_{avg.} \qquad (35)$$

Although we could take equations 34 and 35 and solve them by an iterative technique using one iteration per switching cycle, we choose a different and faster technique. The average for one cycle of a steady-state variable is the same as the average over two, three, or any number of cycles. Therefore we define a new time interval ΔT, long compared to L/R but short compared to the time required for the transient to end. ΔT must also contain an integer number of switching cycles, but that number may be hundreds of switching cycles. We then solve the following equations by iteration

$$R I_{a\,avg.} = g E_b - k \phi \omega \qquad (34)$$

$$\Delta \omega = \frac{(k' \phi I_{a\,avg} - \alpha \omega^2_{avg}) \Delta T}{J} \qquad (36)$$

For example, at $t = 0, \omega = 0$

Solve 34 for $I_{a\,avg}$,

then solve 36 for $\Delta\omega$,

and repeat.

A table provides a convenient way to carry out the computation

ΔT	$t = \Sigma \Delta T$	$\Delta\omega$	$\omega = \Sigma \Delta\omega$	$I_{a\,avg.}$
ΔT_0	ΔT_0	—	0	$\frac{gE_b}{R}$
ΔT_1	$\Delta T_0 + \Delta T_1$	$\frac{k'\phi g E_b}{RJ} = \Delta\omega_1$	$\Delta\omega_1$	$\frac{gE_b}{R} - k\phi\Delta\omega_1 = I_{a1}$
ΔT_2	$\Delta T_0 + \Delta T_1 + \Delta T_2$	$\frac{k'\phi I_{a1} - \alpha\Delta\omega_1^2}{J} = \Delta\omega_2$	$\Delta\omega_1 + \Delta\omega_2$	$\frac{gE_b}{R} - k\phi(\Delta\omega_1 + \Delta\omega_2) = I_{a2}$
etc.	. . .	. . .	. . .	. . .

Such a system may require only 5 or 10 iterations to achieve the desired accuracy and is thus amenable to hand calculation. The current waveform from such a calculation may have the form shown in figure 5 - 8.

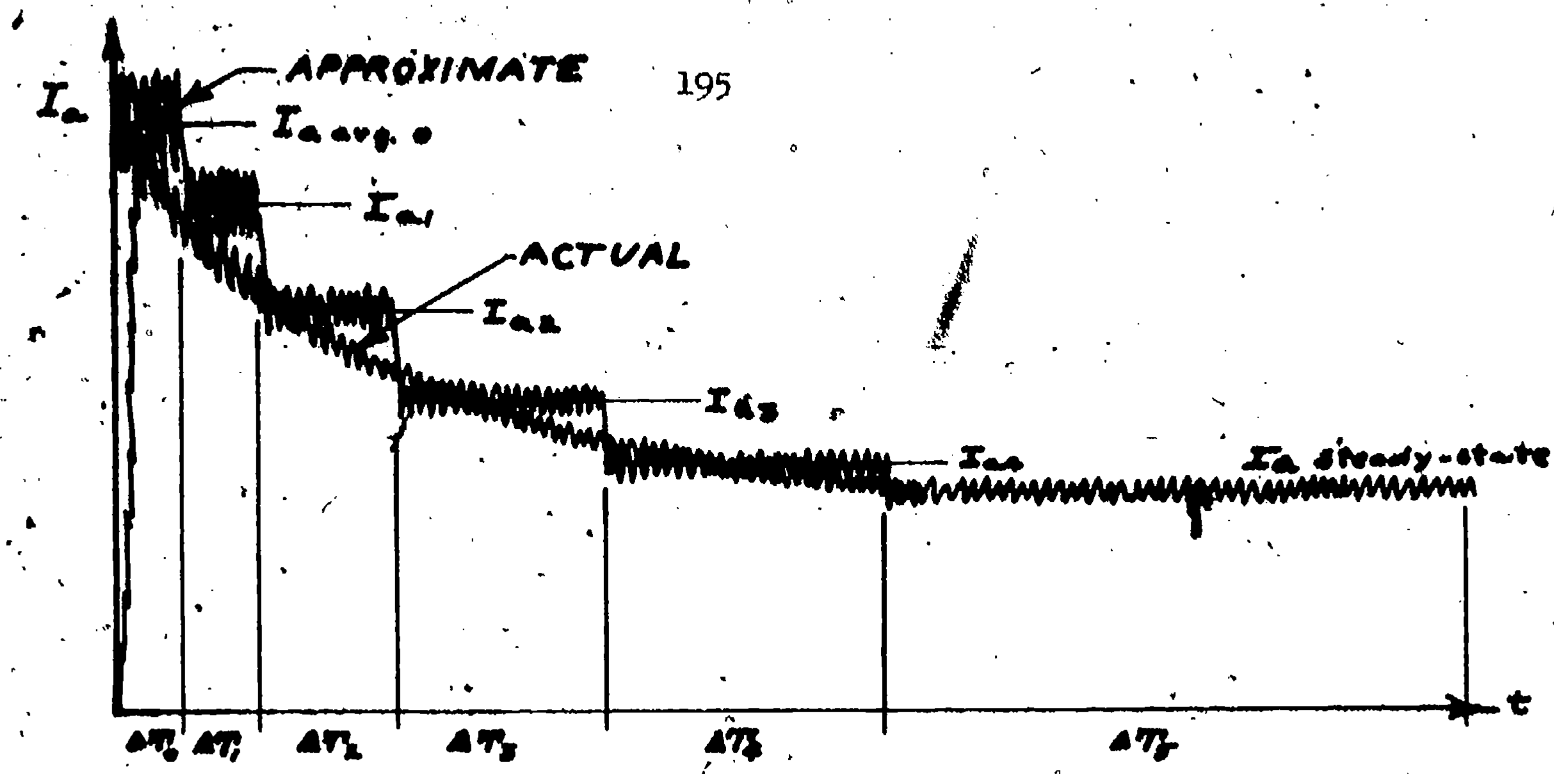


Figure 5 - 8

The size of the current "wiggles" could be estimated by taking the known average values of I_a and ω and substituting them into equation 32 to estimate $\Delta I_{a\ open}$. If the physical parameters such as the size of L, R, J, and α permit such assumptions, we have simplified the transient calculations by a factor of many thousands.

Summary

In this chapter, we modeled a nonlinear system in a manner similar to that of chapters 1 and 3 in order to find the steady-state behavior. As the model was refined step by step the difficulty in solving the problem drastically increased step by step, once again demonstrating the disadvantages of analyzing a problem starting from the detailed system equations. If we had begun with a very refined model (such as model #4) we might never have been able to solve a possibly simple problem. A fairly detailed analysis of the steady-state was presented using the basic idea of a steady-state cyclic process, i.e., $f(\theta) = f(\theta + 2\pi)$. Thus we were able to "skip" to the steady-state instead of following the variables through a starting transient to get to the steady-state.

In considering the turn-on transient of the system, while everything appeared significant at first, we were able to use the steady-state models as a logical framework which we retraced to simplify the problem. Thus the steady-state refinements were not wasted, but on the contrary were extremely helpful in determining the methods of simplifying the problem in addition to providing the final values for the variables undergoing a transient. The steady-state models and the cyclic process idea allowed us to make an easy step in our considerations to a differant kind of model, the quasi-steady-state model, of the circuit behavior. In turn, the quasi-steady-state model shortened enormously the amount of calculations required to solve the nonlinear differential equations associated with the system. Although in some specific problems, the simplified solutions may not be valid and there may be no escape from the tedious and time consuming solution of the detailed system equations, the procedure used here (steady-state model → refined steady-state models → detailed transient model → simplified transient model) is a powerful general technique used in solving transient problems.

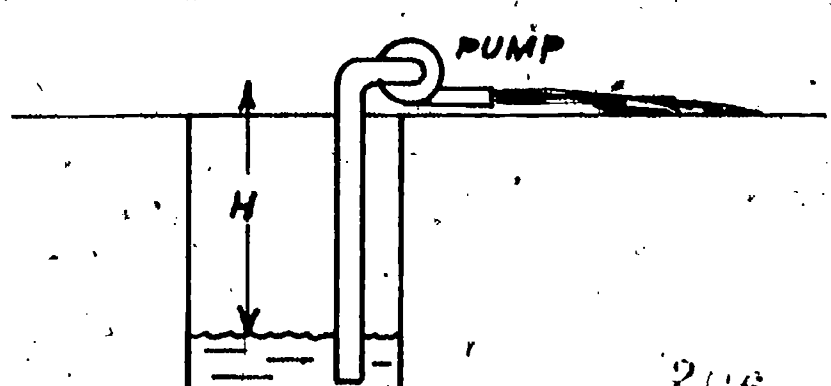
Exercises

(1). A centrifugal water pump is used to drain a pit containing water. Determine the steady-state torque-speed characteristic of the pump knowing:

(a) H = 20 feet

(b) pump efficiency = 85%

(c) the pump delivers 20 gal of water/min when operated at 1800 rpm

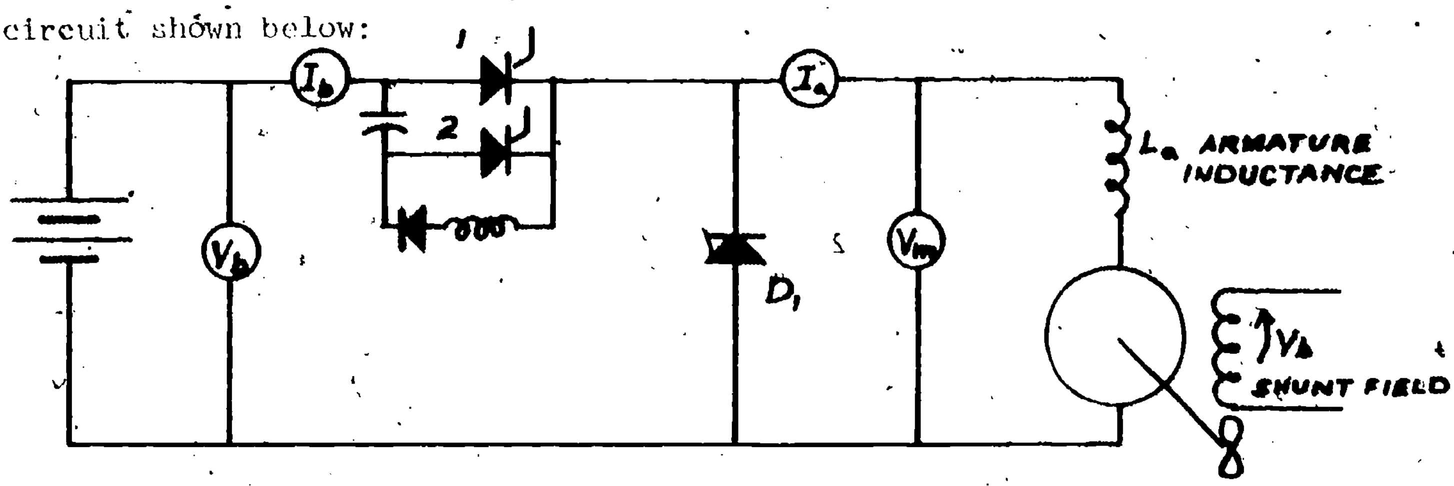


(2) For the problem considered in this chapter, determine the equations or equations to be solved by iterative means yielding I_a and ω as functions of time for the case that L is so large that it limits the current and speed transient rather than the system's mechancial inertia.

Problem 1

A deep sea submersible is driven by a direct current motor (battery power supply) directly coupled to the main drive propeller. The battery voltage is 200 volts. The motor is rated at 10 hp, 200 volts, 900 rpm. The motor is shunt field excited by direct connection to the battery. The propeller has a torque speed characteristic such that $T_{prop} = a\omega^2$ where ω is the angular velocity of the propeller.

An SCR system is chosen to regulate the propeller speed. Clearly, resistor controls would be wasteful of battery energy and submersible cruising time valued around $2,000/hr (since the battery would need more frequent charging). The circuit shown below:

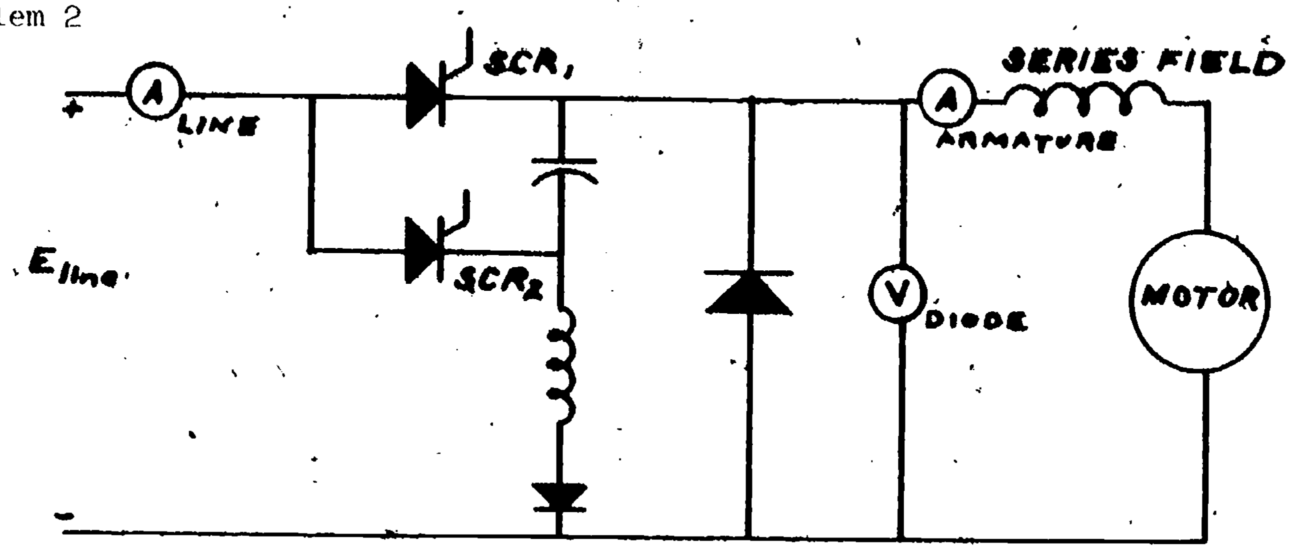


The SCR's are triggered such that SCR_1 is "on" for 2 msec, then SCR_2 is triggered and SCR_1 is triggered 1 msec later (on - 2 msec, off 1 msec).

Calculate the speed of the propeller, I_a, I_b, V_b, V_m assuming the meters are ordinary D'Arsonval meters. (they read "average")

As a first cut at the problem, you may assume ω is constant

Problem 2



E = 200 volts

SCR_1 triggered 1 msec after SCR_2

SCR_2 triggered 2 msec after SCR_1

Motor nameplate

200 volts, 10 hp,

1800 rpm, efficiency = 85% at rated load,

50^{o}C rise continuous service

During a disasterous rainstorm and flood, the motor drives a centrifugal pump used to drain a leaking cellar. The pump will wear out if the water intake should go to zero. Therefore the pump speed must be matched to the rate that water leaks into the cellar. Note, with a centrifugal pump, this rate of discharge is proportional to the speed. Since the kinetic energy associated with the water flow is proportional to the discharge velocity squared, the torque "T" is proportional to ω^2. The pump requires 10 hp at 1800 rpm.

Some measured constants of the system are:

$R_{\text{series fld + arm}} = 0.20\,\Omega$

$L_{\text{series fld}} = 0.2$ henry

$J = 0.25$ kg m^2

(1) In the steady-state, find the average values of I_{line}, I_{armature}, V_{diode}, and the speed ω (in rpm).

(2) Assuming the SCR's and the Diode can handle the current, determine $I_a(t)$ and $\omega(t)$ for the starting transient, i.e., at $t = 0$, $\omega = 0$, $E = 200$ volts, and the SCR's switch as before. Outline your method in detail and carry it through.

(3) Could the starting current amplitude be reduced by leaving SCR_2 turned on for 2 msec instead of only one?

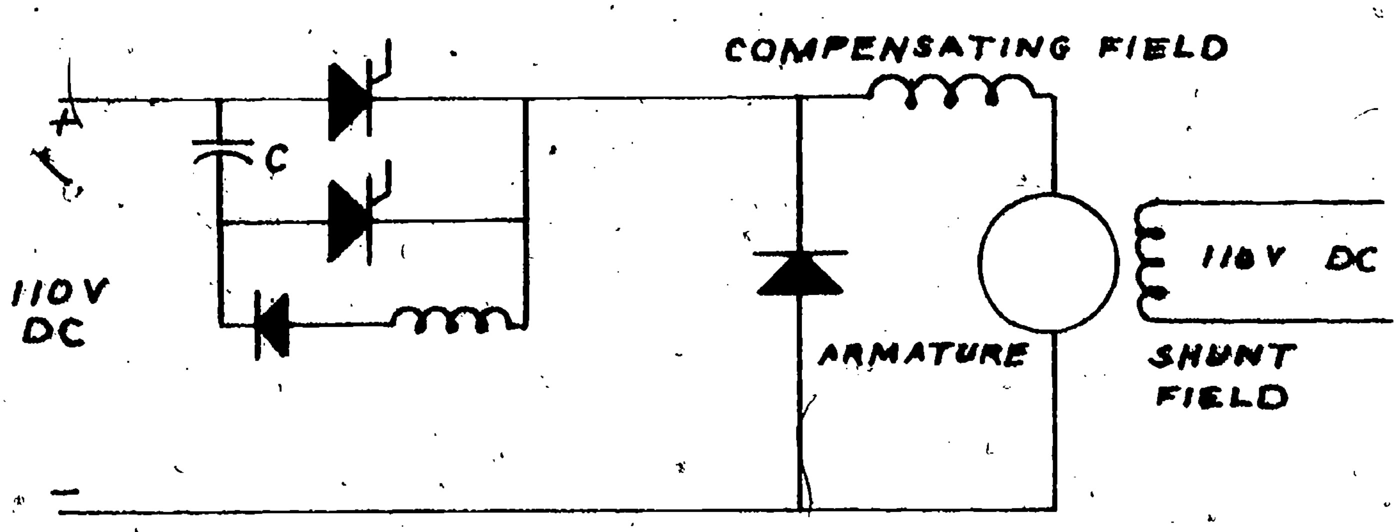
Laboratory Problem 1

There are a number of compound wound DC motors in the laboratory. Using the circuit below as a speed controller, you are to determine:

a) the torque-speed characteristics of the motor for all possible constant speed "settings" of the controller.

b) the time required for the system to reach steady state under no-load conditions.

Since the machines in the laboratory will only be available to you for a short time, it is suggested that you may not have sufficient time to construct the circuit and make the required measurements directly.



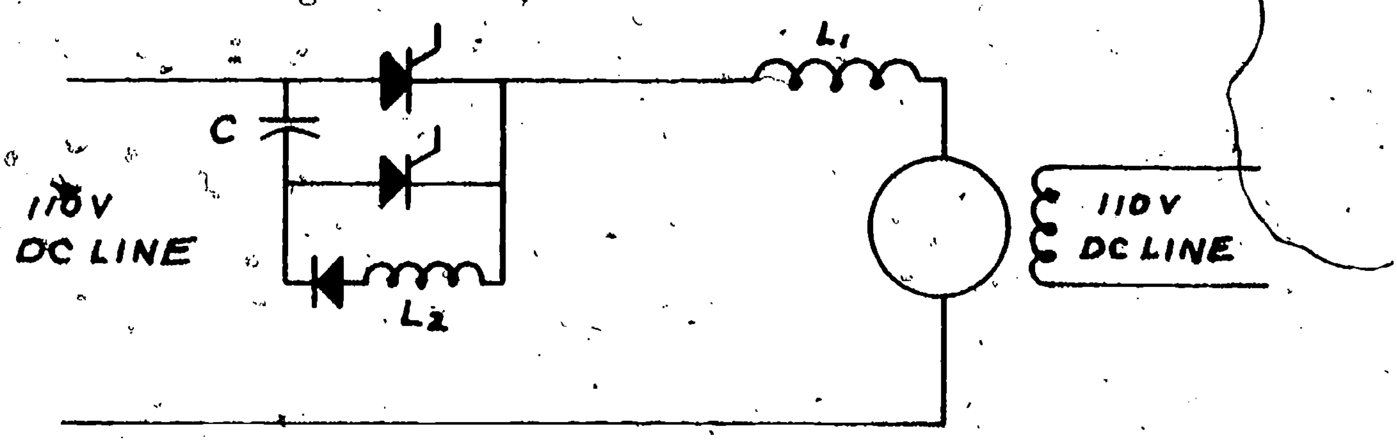
Laboratory Problem 2

The circuit below is used to control the speed of a shunt-wound DC motor. The motor is mechanically connected to a matched mechanical load whose torque is proportional to speed. You are required to determine

(a) the voltage and current ratings of the SCR's and diode

(b) the size of C, L_1, and L_2

(c) the basic scheme for a logic circuit to trigger the SCR's. The logic circuit does not have to be designed in detail, but the components must be specified sufficiently and realistically enough that the circuit could be designed and built on the basis of your specification. The logic circuit should have provision for starting the motor under load.



References

(1) Nielsen, Kaj L., Methods in Numerical Analysis, Macmillen Co., New York, 1956.

This reference outlines many useful techniques for solving systems of equations and differential equations by numerical techniques, adaptable for hand or computer analysis.

(2) Fitzgerald, A. E. and Kingsley, Charles Jr., Electric Machines, 1st Ed., McGraw-Hill, New York, 1952.

This is a general text regarding the behavior of AC and DC machines. The first edition is particularly good in describing and calculating the nonlinear properties of rotating machines as well as the underlying physics of operation.

Appendix

The following is a suggested rough guide for a time schedule of presentation of material, allowing sufficient time for detailed discussion of one "problem" and one laboratory problem" per chapter in a small class (15-25 students).

Chapter	1	3 weeks
	2	3 weeks
	3	2 weeks
	4	2 weeks
	5	3 weeks

In the case of smaller classes, a longer term, or particularly able students, further topics may be considered. The following are offered as suggestions.

6) The Parallel Capacitor Commutated Inverter with an Inductive Load

This problem, discussed in some detail in Principles of Inverter Circuits by Bedford and Hoft, is particularly suitable for modeling on an analog or hybrid computer.

7) The problem of producing an approximate sine wave by summing the outputs of several square-wave inverters with differing amplitudes and phases but the same repetition rate is a useful example of designing a circuit (choosing the phases and amplitudes) on the basis of a simplified model of operation. The problem can be generalized into designing the 3 phase, harmonic neutralized inverter. Such problems are most instructive if the third and fifth harmonics are absent from the desired output sine wave. Again, Principles of Inverter Circuits by Bedford and Hoft is an excellent beginning reference for the problem. An additional helpful

reference is Kernick, Roof, and Heinrich," Static Inverter with Neutralization of Harmonics", AIEE Transactions, Pt. II, Vol. 81 (May, 1962), pp. 59-68.