RCA REVIEW

a technical journal

RADIO AND ELECTRONICS
RESEARCH • ENGINEERING

VOLUME XIX JUNE 1958 NO. 2

PRINTED IN U.S.A.

RCA REVIEW, published quarterly in March, June, September, and December by RCA Laboratories, Radio Corporation of America, Princeton, New Jersey. Entered as second class matter July 3, 1950 at the Post Office at Princeton, New Jersey, under the act of March 3, 1879. Subscription price in the United States and Canada; one year $2.00, two years $3.50, three years $4.50; in other countries: one year $2.40, two years $4.30, three years $5.70. Single copies in the United States, $.75; in other countries, $.85.

RCA REVIEW

a technical journal

RADIO AND ELECTRONICS
RESEARCH • ENGINEERING

Published quarterly by

RCA LABORATORIES

in cooperation with all subsidiaries and divisions of

RADIO CORPORATION OF AMERICA

VOLUME XIX | JUNE, 1958 | NUMBER 2

CONTENTS

RCA REVIEW is regularly abstracted and indexed by *Industrial Arts Index Science Abstracts* (I.E.E.-Brit.), *Electronic Engineering Master Index, Chemical Abstracts, Proc. I.R.E.*, and *Electronic & Radio Engineer.*

RCA REVIEW

CONTROLLED THERMONUCLEAR FUSION — PROMISE OF THE FUTURE*

By

GEORGE WARFIELD†

RCA Laboratories,
Princeton, N. J.

Editor's Note: This paper and the one which follows—"Thermonuclear Fusion—Its Meaning to the Radio and Electronics Engineer," by E. W. Herold—are review papers. Although the two papers were separately and independently prepared, they complement each other very well. Dr. Warfield's paper is an elementary introduction, and Mr. Herold's a more technical description.

The Radio Corporation of America was recently selected, along with the Allis-Chalmers Manufacturing Company, to design, fabricate, and install what is probably the world's largest experimental facility for controlled thermonuclear fusion research. This facility, known as the the "C" Stellarator, is to be used by scientists of Princeton University, under the direction of Professor Lyman Spitzer, Jr., and will be located at the James Forrestal Research Center of the University.

Summary—*Thermonuclear fusion offers attractive possibilities as the energy source of the future, provided a practical method for controlling the fusion reactions can be developed. In this paper, some of the problems involved in achieving a controlled fusion reactor are discussed in simple terms as an introduction to the paper by E. W. Herold.*

THE ENERGY PROBLEM

FOR THE past several years many groups in the United States and elsewhere throughout the world have been working vigorously to develop a practical method for harnessing for peaceful purposes the vast energy reserves tied up in thermonuclear fusion fuels existing on the earth.[1-4] This fuel has already been harnessed for destructive purposes in the form of the so-called hydrogen bomb. Indirectly, fusion fuels are already supplying practically all our energy

* Reprinted from *I.R.E. Student Quarterly*, February 1958.

† Dr. Warfield is also Associate Professor of Electrical Engineering, Princeton University.

[1] P. C. Putnam, *Energy in the Future*, D. Van Nostrand Co., Princeton, N. J., 1953.

[2] R. F. Post, "Controlled Fusion Research — An Application of the Physics of High Temperature Plasmas," *Proc. I.R.E.*, Vol. 45, p. 134, February, 1957.

[3] F. Bello, "Fusion Power: The Trail Gets Hotter," *Fortune*, Vol. 61, p. 135, July, 1957.

[4] R. F. Post, "Fusion Power," *Scientific American*, Vol. 197, p. 73, December, 1957.

needs, since all the energy that comes to the earth from the sun has its origin in the "burning" of fusion fuels on the sun. Over millions and millions of years some of this energy has been stored on the earth in coal, oil, gas, trees, etc., and it is this stored energy which we are now using.

The incentives for this work are very great indeed. A recent study[1] of the future energy requirements of the world in general and of the United States in particular, combined with the latest estimates of the *economically* recoverable reserves of the traditional capital fuels (coal, oil, and gas) resulted in some very sobering conclusions. At the present time energy is being consumed at a tremendous rate, but, of even more importance, the *increase in the rate* of consumption is itself tremendous. For example, according to Putnam[1] "In the United States one-half of all the coal ever consumed has been burned since 1920, and nearly one-half of all the oil and gas ever consumed has been burned since 1940." If this increase in the rate of consumption continues, and there is no reason to believe otherwise, within a relatively short time, of the order of fifty years or less, the reserves of the traditional fuels will not be sufficient to meet the demand.* If the standard of living in the United States is to be maintained at its present level, and if the standard of living in the rest of the world is to be improved, it is essential that new sources of relatively cheap energy be utilized.

A seemingly attractive possibility along this line is to increase the utilization of the so-called income energy sources such as windpower, hydroelectric power, photosynthesis, farm products, direct solar radiation, etc., which, incidentally, all receive their energy from the fusion process on the sun. However, studies of these sources indicate that either the amount of energy from some of the sources is minute compared to the total demand, or the cost of utilization is many times higher than the cost of present fuels, contrary to what one would expect offhand. For example, direct solar radiation would appear to be a very cheap source of energy since the sun supplies to the land areas of the earth about 6.7×10^{14} kilowatt hours each day at no cost to us. However, the density of this energy is relatively low — 125 watts per square foot — so that fairly large installations of solar converters of some form would be required to exploit this energy. In addition, since essentially no energy is supplied to any given installation at

* One can argue that all past estimates of the reserves of coal, oil, and gas have been on the pessimistic side and that these present estimates are also pessimistic. However, as Putnam points out, in planning for the future it is not wise to be too optimistic about the reserves. Before dismissing these estimates it would be well to read Putnam's complete study. As a case in point it is reported in *Time* magazine for December 2, 1957, that all the major oil companies are cutting back on their off-shore oil operations because of excessive costs.

night, some storage mechanism must be included as part of the installation. Finally, only certain places on the earth receive a dependable day-to-day supply of radiation. Efficient large power installations would have to be located in these areas, which, at the present time, are located far from those regions where the bulk of the energy is consumed. This would require an extremely costly transmission system. Consequently, even though the fuel is free, the cost of utilizing it on a large scale may be prohibitive. The indication at present is that the income fuels will not be able to fill the gap left by dwindling reserves of coal, oil, and gas.

Thus the main hope for the future is the nuclear fuels. These are of two types — fission and fusion. Fission fuels are the heavy elements, such as uranium and plutonium, whose nuclei can be split into two lighter nuclei with the release of energy, while the fusion fuels are

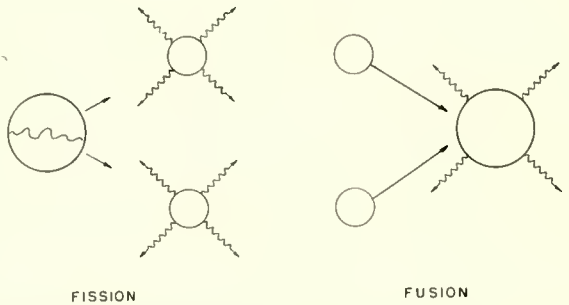


Fig. 1—Schematic representation of fission and fusion. The wavy lines represent energy released in any form, not necessarily radiation.

the light elements, such as deuterium and tritium, whose nuclei can be fused together to form heavier nuclei again with the release of energy, (Figure 1). Just why energy can be released in these processes is discussed later. Deuterium, which at the present time appears to be the most important fusion fuel, is a form of hydrogen having a nucleus made up of a proton and a neutron, while tritium is another form of hydrogen with a nucleus containing one proton and two neutrons in contrast to ordinary hydrogen whose nucleus consists of a single proton (Figure 2). A start has been made in harnessing these nuclear fuels for peaceful uses. By far the greatest effort to date has gone into the fission fuels and many power reactors have been built or are being built. Much less effort has been put into harnessing fusion fuels but this effort is now expanding quite rapidly. Since both these nuclear processes involve capital fuel, the reserves of these fuels should be of great concern. It has been estimated that the economically re-

coverable reserves of fissionable fuels represent an energy content of the order of 25 times that of the reserves of coal and 100 times that of the reserves of oil and gas. While this sounds quite adequate, it may be able to supply the energy needs for perhaps a century in the face of increasing energy demands. As for fusion fuels, one gallon of water contains about 5×10^{22} deuterium atoms which can be separated from the ordinary hydrogen quite cheaply by elecrolytic techniques. In fact, the cost of separating the deuterium in one gallon of water is about 10 cents. As shown later, the complete fusion of deuterium into ordinary helium releases about 7 million electron volts (mev) per deuterium atom. A quick calculation shows that the energy released in the complete "burning" of the deuterium in one gallon of water is roughly 3.5×10^{23} mev or 10,000 kilowatt-hours. Another rough calculation indicates that a family of five uses for all purposes —heating, lighting, cooking, car, etc.—the equivalent of about 10^5 kilowatt-hours per year. Thus 10 gallons of ordinary water should be able to supply all energy requirements of one family for a year.

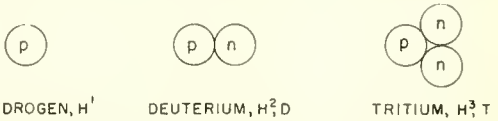


Fig. 2—Schematic representation of the three isotopes of hydrogen. In this, and all following drawings, P and N stand for proton and neutron, respectively.

When one considers the tremendous volume of water on the earth, it can be seen that the reserves of deuterium are sufficient to supply any conceivable demand for billions of years provided a method can be developed to control the fusion process. It is this vast reserve of fusion energy that furnishes one of the main incentives for the work now in progress.

The fusion process has several other inherent advantages over the fission process to add further incentives for this work. One of the most important of these is the relatively small amount of radioactive by-products of the fusion reaction compared to the fission reaction. Should a large fraction of the world's energy requirements be supplied by fission, the problem of disposing of all the radioactive by-products would be formidable indeed. To indicate just how formidable, the energy released in the explosion of an atomic bomb may be of the order of 10^8 kilowatt-hours. The total energy input in the United States in the year 1950 was roughly 10^{13} kilowatt-hours. Hence, if all the energy were supplied by fission reactors, this would be equivalent to exploding about 100,000 atomic bombs so that each year we would have to dispose

of the radioactive by-products of this many bombs. One possibility, of course, is to shoot these radioactive materials out into space, but this would be fantastically expensive. With fusion reactors, radioactive by-products are not a serious problem. In addition, the hazard of explosion for a fusion reactor is essentially nil because of the nature of the process. The fuel can be fed in as needed much as in an oil burner so that at any one time there is only a small amount of fuel in the reactor. On the other hand, since a large amount of fuel must be in a fission reactor to make the process self-sustaining, there is always the possibility of an explosion, even though remote.

From the standpoint of the electrical engineer, fusion has one very great advantage over fission. By the nature of the fusion process, all material in the reactor is completely ionized so that large electrical currents exist in the reactor. It is possible by various schemes to get electrical energy out of the reactor directly without going through the standard thermal cycles used for fission reactors. However, since a substantial fraction of the energy produced is given to uncharged neutrons, the direct conversion of all the fusion energy to electrical energy presents a challenging problem.

A final incentive for the work toward controlled fusion is that it presents a great scientific and technical challenge. In the absence of any other reasons, this would be incentive enough for many people to try to solve the problems. However, with this incentive alone the magnitude of the present effort would not be possible.

Just what is this fusion process which looks so promising? A general description of the process is given first, followed by a more detailed description of some of its features.

Basically, fusion is a process in which two light nuclei combine to form a heavier nucleus accompanied by the release of a large amount of energy. For two nuclei to combine they must first come together. Since nuclei are positively charged, there is a force of repulsion between two nuclei which tends to prevent their coming sufficiently close for fusion to occur. This repulsion is best described by saying that each nucleus is surrounded by a potential-energy barrier. Thus, if two nuclei are to get together, either the barriers must be lowered in some way, or the particles must have sufficient energy to surmount the relative barrier between them. If the barriers could be lowered, the particles could get together quite easily at ordinary temperatures. If the barriers cannot be lowered, the particles must have sufficient kinetic energy to overcome the potential barrier, and this requires very high temperatures.

The situation here is somewhat analogous to the case of two cities completely separated by a very high mountain. For example, suppose

city A is in a very deep valley surrounded completely by a very high mountain, while city B is on the other side of the mountain. For some reason, the people of these two cities would like to travel back and forth. The direct approach is to go over the mountain, but the available transportation is not powerful enough to scale the mountain. There are two courses of action open to promote a lively traffic between the two cities. Either a cut or tunnel must be constructed in the mountain between the two cities, or more powerful means of transportation be developed to carry the traffic over the mountain. If the cut or tunnel is constructed, the available cheap transportation can be used. However, it may turn out that the cost of constructing an appropriate passage may be far greater than the cost of developing, building, and operating for a reasonable period of time the more powerful transportation needed to go over the mountain. If a good, cheap method of constructing a passage is not available and is not readily foreseen, the only course of action is to construct more powerful cars. This, at present, is the situation with controlled fusion if we interpret cost as energy.

Recently, a catalyst has been found for lowering the barrier between two nuclei. This catalyst is a particle called the mu meson. The energy needed to produce one of these catalytic particles, however, is much greater than the energy released when two nuclei fuse. If this one catalytic particle could induce a sufficiently large number of fusions, more energy would be released from the system than was expended to produce the catalyst and we would realize net energy gain from the process. Unfortunately, the lifetime of a mu meson is so short that on the average it does not catalyze enough fusions to produce a net energy gain. Thus the mu meson is not a practical catalyst. It is conceivable that a practical catalyst will be found in the future. At this time, however, in the absence of a good catalyst, the second alternative — getting the particles over the barrier — must be employed to make fusion occur.

For the particles to surmount the mutual barrier, the kinetic energy associated with their relative velocities must be very large — orders of magnitude larger than the kinetic energies of practically all the particles in a group of nuclei at ordinary temperatures. Thus, to initiate the fusion reaction in a group of particles, there must be an input of energy from some external source to give some of the reacting nuclei sufficient energy to surmount the barrier and undergo fusion. Once the fusion process is started, if the energy released upon fusion is greater than the initial input, and if enough of the energy can be fed back to maintain the fusion process, it should be possible to "burn" all the fusion fuel and extract energy from the system. This is anal-

ogous to the process of burning any fuel. Some of the fuel must be raised to its ignition temperature by supplying energy to it from some external source — a lighted match for example — before the combustion process can start. If the energy released when this first bit of fuel begins to burn is greater than the energy needed to ignite it in the first place, the possibility exists of feeding some of this energy to adjacent portions of the fuel to ignite more fuel, while the excess is extracted from the system. For sustained burning to occur, it is not sufficient that the energy released when a quantity of fuel burns be greater than the ignition energy for that fuel. The geometry of the complete fuel mass must be such that all this energy does not escape from the system, but that the feedback of energy to adjacent parts of the fuel is adequate to ignite them. This should be quite clear to anyone who has tried to keep a single log burning in a large fireplace.

The average energy of a group of interacting particles may be described by the temperature of the group. For fusion to occur, the ignition temperature must be extremely high — of the order of millions of degrees centigrade. At this temperature, the average speed of the particles is enormous and the particles tend to disperse in a very short time from any region in which they are located. The more widely dispersed a group of particles becomes, the less likely it is that two particles can collide and fuse. To keep the probability of collision, and hence of fusion, reasonably high, some constraining force must be applied to the system to counteract this dispersion. The method of applying this constraining force is one of the central problems in developing fusion devices.

Constraining Forces

Known forces may be divided into two broad classes: those which act directly and those which act at a distance. Directly acting forces occur whenever a particle makes contact with the walls of a container. After contact, the particle is reflected back into the reaction region. This is the type of constraining force used in ordinary chemical reactions where the force is supplied by the walls of whatever type of vessel is used for the reaction. Solid walls are not practical containers for a fusion reactor. The highest melting point for any known solid is of the order of 3,000° centigrade. The walls of even the "best" container would have to be kept below this temperature if the reactor is to be a permanent one. The temperature of the fuel must be of the order of 100 million degrees centigrade for sustained fusion to occur. Each time a fuel atom hits the wall, it gives up energy to the wall. Thus interaction between fuel and wall results in cooling the fuel and

heating the wall. If enough particles hit the wall per unit time — this number depends on the density of fuel particles and on the temperature — the power input to the wall will be substantially greater than the power loss due to conduction; radiation, etc., so that the temperature of the wall will increase and eventually reach the melting point with disastrous results. Even if the density of the fuel atoms is not high enough to produce a significant temperature rise of the walls, the fuel will be cooled by the walls and its temperature will drop below the ignition temperature. There are three known forces which act at a distance: gravitational, electric, and magnetic. It turns out that to obtain a gravitational force large enough to keep a fusion system together, the reactor would have to be of the dimensions of a star. As a matter of interest, this is just the constraining force which keeps the fusion reaction in the sun from dispersing. For earthly devices, then, this force must be discarded. A purely electrostatic force is also not feasible, because in a reaction both electrons and positive ions must be kept away from the walls. An electrostatic field which would keep ions away from the walls would tend to move the electrons to the walls and vice versa. Fortunately, it turns out that magnetic forces are feasible and it is magnetic forces which give promise of producing controlled fusion. Quite often, one hears the expression "magnetic bottle" for controlled fusion. This is just a very descriptive way of saying the constraining mechanism is a magnetic field rather than a material wall.

One fusion device, the thermonuclear or fusion bomb, has already been developed. In the bomb, fusion takes place in a fixed amount of reacting material with the release of a tremendous amount of energy in a very short time. The rate at which energy is released is much too great for its utilization as an everyday energy source. In this device one does not have to worry about destroying the system in which the reaction takes place. Thus, any mechanism which will ensure that the reacting particles stay together long enough for a significant fraction of them to fuse can be used. Appropriate material barriers which last for a few microseconds are possible, although just how the bomb is held together has never been published.

For controlled fusion devices, one's freedom of choice of a constraining mechanism is limited. As we have just seen, the only presently feasible choice is a magnetic field. The problem, then, is to devise some configuration of magnetic field which can keep the particles away from the material walls of the devise while fusion occurs. Since magnetic fields can exert forces only on electrically charged particles, one has the additional problem of ionizing the reaction material either before or after it is introduced into the reactor. This is a relatively

minor problem since electrical engineers have been dealing with charged gas particles for many years in the form of gas discharges.

There are of course many other requirements for a controlled fusion reactor which we might just mention. The device should produce a continuous output of energy at a rate low enough that it can be utilized. This requires that the reaction material be introduced in small amounts at a continuous rate and that the released energy be extracted continuously. There is the problem of removing from the reactor the heavier nuclei after fusion — the "ash," so to speak. There is the further problem of how the released energy is to be utilized. As mentioned previously, this is a problem of particular interest to electrical engineers, since the possibility exists of converting fusion energy directly to electrical energy without any intermediate thermal stages in the process.

Inside the Atom

Now that the fusion process has been discussed in a general way, some of its aspects will be considered in more detail.

Our present picture of the nucleus is that it is composed of two types of particles — protons and neutrons — both of which are called nucleons. The proton has a positive charge of 1.6×10^{-19} coulombs and a mass of 1.00759 atomic mass units (usually written as amu) where one amu has an actual mass of 1.660×10^{-27} kilogram. The neutron is electrically neutral and has a mass of 1.00898 amus. It can be shown quite conclusively that there are no electrons in the nucleus. There is some evidence that the nucleus may contain other particles such as mesons, but the evidence for this is still sparse. All nuclear reactions of interest here can be explained quite well by assuming the simple model of the nucleus; any meson population of the nucleus can be neglected. A nucleus is specified by assigning to it an atomic number, Z, and an atomic mass number, A. The atomic number designates the number of protons in the nucleus, while the atomic mass number is the nearest integer to the mass of the nucleus expressed in atomic mass units. Symbolically, a nucleus is specified by the chemical symbol appropriate to the nucleus with the mass number as a superscript to the right and the atomic number as a subscript to the left. Thus ${}_{82}Pb^{206}$ represents a nucleus of lead whose atomic number is 82 and whose mass number is 206. The atomic number and the chemical symbol give the same information so that the subscript is redundant and can be omitted. A given chemical element may exist in several forms as far as weight is concerned so that the mass number cannot be omitted. Different forms of the same chemical element are called isotopes. Since the mass of both nucleons is close to 1 amu, the atomic

mass number A designates the total number of nucleons in the nucleus, and, since Z is the number of protons, $(A\text{-}Z)$ is the number of neutrons in the nucleus.

If the nucleus consists of neutrons and protons, we are faced with the problem of what forces hold the particles together. Since the protons are positively charged, the repulsion between the protons should tend to separate them, unless there were some negative charge to counteract this force. Since the only other particles in the nucleus are uncharged, there must be forces other than electrical holding the nucleus together. Just what these forces are is not known at present, so that rather than trying to guess about them, we deal with the energy which holds the particles together — the binding energy of the nucleus. This binding energy is tied up with the correspondence between mass and energy as given by the Einstein relation $E = mc^2$.

This equation comes from the theory of relativity. If the mass of a particle when it is at rest is m_0, the so-called rest mass, its mass when it has a speed v is

$$m = m_0/[1 - v^2/c^2]^{1/2}$$

and its momentum is mv. When a force is applied to the particle, Newton's law must be written as $F = d(mv)/dt$ rather than as $F = ma$. Suppose a force F is applied to a particle and the particle moves a distance x in the direction of the force. The work done on the particle is $\int_0^x F\,dx$ and this must equal the change in kinetic energy of the particle ΔKE. Then

$$\Delta KE = \int_0^x F\,dx = \int_0^x \frac{d}{dt}\left[\frac{m_0 v}{\left(1 - \frac{v^2}{c^2}\right)^{1/2}}\right] dx.$$

Evaluating the integral,

$$\Delta KE = m_0c^2\left[\frac{1}{\left(1 - \frac{v^2}{c^2}\right)^{1/2}} - 1\right] = mc^2 - m_0c^2 = c^2\,\Delta\, m.$$

The change in kinetic energy of a particle is just c^2 times its change in mass. Thus, mass and energy are related. A particle at rest has an

energy m_0c^2, while a moving particle has an energy mc^2, of which $(m - m_0)c^2$ is kinetic energy. If we could decrease the mass of the particle at rest, we would obtain energy. This is the origin of nuclear binding energy, and as we shall see, of the energy released in fission and fusion. The energy equivalent of one atomic mass unit at rest is

$$m_0c^2 = 1.660 \times 10^{-27} \times (3 \times 10^8)^2 = 1.49 \times 10^{-10} \text{ joules}$$
$$= 931 \text{ mev.}$$

Thus, if we could destroy just one atomic mass unit, we would release 931 mev of energy.

If the mass of any nucleus is measured with a precision mass spectrometer, it is found that the masses of all nuclei are almost but not quite an integral number of amu. Nuclear masses can be measured to a precision of 1 part in 10^5. The atomic number and the mass number of each nucleus give the constituents of that nucleus. The masses of all the constituents may be added to arrive at a calculated mass. When the calculated mass is compared with the observed mass, the observed mass is always found to be less than the calculated mass. This difference is called the mass defect and it is just this difference which supplies the binding energy of the nucleus. As an example consider the helium nucleus, He^4, which consists of two protons and two neutrons. The calculated mass is $2 \times 1.00759 + 2 \times 1.00898 = 4.03314$ amu, while the experimentally determined mass is 4.00277 amu. Thus the mass defect is 0.03037 amu which corresponds to an energy of 28.3 mev. This is the total binding energy of the He^4 nucleus. That is, when two neutrons and two protons come together to form this nucleus there is a loss in mass which is converted into energy released from the system. (Figure 3.) To put it another way, if this much energy is supplied to a He^4 nucleus, the nucleus would separate into its constituent particles. If we call this binding energy Q, the energy equation is

$$He^4 + Q \rightarrow 2H^1 + 2n^1$$

where H^1 is the symbol for a proton and n^1 is the symbol for a neutron. Such an equation can be written for any nucleus and the value of Q can be found by writing the energy equivalent of all particles on either side of the equation. Q is the number needed to balance the energy equation. If Q is positive, the nucleus is stable while if it is negative, the nucleus is unstable.

The binding energy just found for the helium nucleus, and the binding energy for all other nuclei composed of more than 1 nucleon is not the most convenient energy to use to describe nuclear binding.

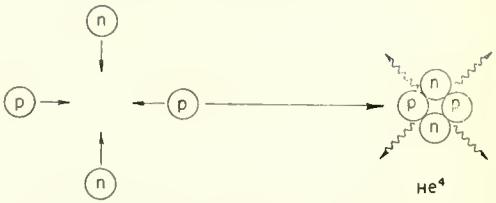


Fig. 3—Formation of He^4, with the release of energy, by combining two protons and two neutrons.

Rather, the binding energy per nucleon gives a better picture of the energies involved in different nuclei. The binding energy per nucleon is the total binding energy as found from an equation like the one above divided by the number of nucleons in the nucleus. A plot of this binding energy per nucleon as a function of the atomic number of the nucleus, gives a curve like that shown in Figure 4. There are several interesting features to this curve. Most important for this discussion, the very light elements and the very heavy elements have less binding energy per nucleon than the elements of intermediate weight. This is the basis of both fission and fusion energy sources. Thus, if a heavy element can be split into two or more lighter elements, energy should be released in the process. Or if two or more light elements with small atomic number can combine to form a heavier element, energy will also be released. The amount of energy per nucleon released in either process depends upon the difference in the binding energy per nucleon of the final and initial isotopes. One of the things that makes fusion so promising as an energy source is the relatively large difference in binding energy per nucleon between helium and deuterium.

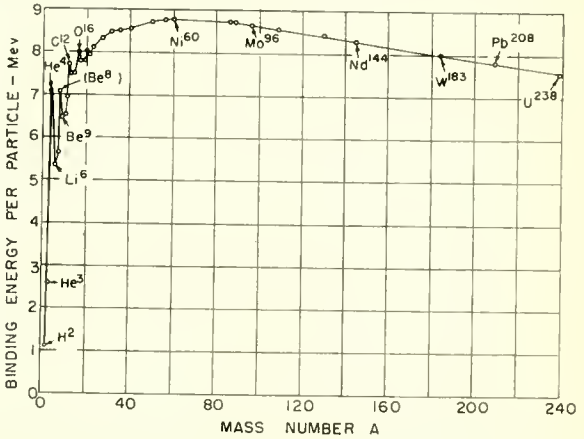


Fig. 4—Average binding energy per nucleon. (By permission: from Introduction to Modern Physics, by Richtmyer, Kennard, and Lauritsen. Copyright, 1955, McGraw-Hill Book Co., Inc.)

It may seem offhand that since the intermediate weight elements have higher binding energies than the heavier or lighter elements, we would have to supply energy to make the fission or fusion process occur. However, one must keep in mind that the binding energy is the energy needed to break up the nucleus. Conversely, then, it is the energy given off when the nucleus is formed from its constituent parts. The higher the binding energy, the more energy is released when the nucleus is formed. Notice from the curve that the range of binding energies per nucleon is from about 1 to 9 million electron volts. This should be compared with ordinary chemical processes such as the burning of coal, where the energy released is of the order of electron volts per reaction.

We can make this comparison between nuclear and chemical fuels somewhat more vivid by comparing the energy released in burning one pound of coal with the energy released in the fission of one pound of U^{235} or in the fusion of one pound of deuterium. The best coals liberate about 16,000 BTU per pound which is equivalent to approximately 10^{19} mev. One pound of any material contains 2.7×10^{26} nucleons. The average energy released per nucleon in U^{235} fission is 0.85 mev, so that one pound of U^{235} releases roughly 2×10^{26} mev during fission. This is seven orders of magnitude greater than the energy liberated by burning one pound of coal. Hence, one would have to burn 10 million pounds or 5,000 tons of coal to get the same energy as is released in the fission of one pound of U^{235}. The difference between the binding energy per nucleon of He^4 and deuterium is 6 mev. As shown later, essentially three deuterium nuclei must combine to form He^4. The total binding energy of the end product of this process, He^4, is 28.3 mev, while the total binding energy of the three deuterium fuel nuclei is 6.6 mev. In the process 21.6 mev is released, corresponding to 3.6 mev per nucleon in the fuel. This is about four times the energy per nucleon released on fission and we would have to burn 20,000 tons of coal to release as much energy as is released in "burning" one pound of deuterium. From these comparisons, the possibilities of nuclear fuels appear extremely attractive.

However, one is faced with the problem of how to split the U^{235} nucleus, or how to fuse two deuterium atoms together. Basically, this problem is one of nuclear reactions. A nuclear reaction is a process in which one particle bombards a nucleus and a new nucleus is formed usually with the emission of a second particle and some energy. If the new nucleus formed is stable, the reaction is called an artificial disintegration; if the new nucleus formed is unstable, so that it subsequently disintegrates further, the reaction is referred to as induced or artificial radioactivity. Nuclear reactions may be produced with all sorts of bombarding particles: protons, neutrons, deuterons, alpha par-

ticles, etc. (Deuteron is the name given to the deuterium nucleus.) If the incident particle is positively charged, as are all the particles just mentioned except the neutron, the particle must have a very large initial energy to produce the reaction. The reason for this, of course, is that the nucleus being bombarded repels the positive charge by ordinary coulomb repulsion. Thus to produce nuclear reactions with these charged particles, some means must be found to give them large energies. The first nuclear reactions discovered were those in which the bombarding particles were energetic alpha particles obtained from naturally radioactive isotopes. After the invention of particle accelerators, such as cyclotrons, betatrons, synchrotrons, etc., other types of charged particles could be used to induce reactions. On the other hand, since neutrons are electrically neutral, there is no barrier for neutrons around a nucleus and even very low energy neutrons may be able to induce nuclear reactions. For this reason fission, which can be produced by neutron bombardment, can occur at ordinary temperatures; while fusion, which cannot be induced by neutrons, must take place at very high temperatures, in the absence of a catalyst.

When the bombarding particle enters a nucleus two things may happen. Either the particle leaves again so that only an elastic collision has occurred, or the particle is captured by the nucleus producing a new nucleus called a compound nucleus. After a short time, the compound nucleus disintegrates to complete the reaction. If the end product is unstable, further disintegrations are possible.

As an example of a nuclear reaction leading to a stable nucleus, and to indicate the manner in which such reactions are described, consider the first nuclear reaction that was discovered. In this reaction an alpha particle bombards a nitrogen nucleus. The alpha particle is momentarily captured by the nucleus to form a new nucleus, fluorine, which then disintegrates to give an oxygen nucleus and a proton.

$$_2He^4 + {}_7N^{14} \rightarrow [{}_9F^{18}] \rightarrow {}_8O^{17} + {}_1H^1 + Q.$$

Very often the compound nucleus is not included in the equation, and the equation is written in a condensed form.

$$N^{14}\ (\alpha, p)\ O^{17}.$$

A somewhat different set of symbols is used in this second notation. Capital letters are used outside the parenthesis, with some special symbols for the lighter nuclei — D for deuterium and T for tritium. Small letters — d for deuterium, t for tritium, p for proton, n for neutron, α for alpha particle — are used inside the parenthesis. These

condensed equations are to be interpreted in the following way: the nucleus to the left of the parenthesis is bombarded by the first particle inside the parenthesis producing a compound nucleus which disintegrates to produce the nucleus to the right of the parenthesis and the second particle inside the parenthesis.

When the compound nucleus disintegrates, the energy released is shared between the two final particles — in the equation above between the O^{17} and the proton, — with the lighter particle taking on the larger share of the energy. In fact, the ratio of the energy of the two final products, found from conservation of energy and momentum is $\epsilon_1/\epsilon_2 = m_2/m_1$.

A nuclear reaction which leads to fission is as follows:

$$n^1 + U^{235} \rightarrow Xe^{140} + Sr^{94} + 2(n^1) + Q.$$

This is but one of many nuclear reactions that take place in the splitting of U^{235}. Actually when neutrons bombard U^{235} the results of the splitting are not unique but many different final nuclei are formed, all of which are quite radioactive. For this discussion it is the fusion reactions which are of interest. Some of the possible fusion reactions are listed below with the energy released in each reaction and with the cross section for reaction at 600 million degrees for some of the reactions. The significance of this cross section is discussed later. The first reaction is shown schematically in Figure 5.

$D + D \rightarrow He^3 + n + 3.25$ mev		10^{-2} barn
$D + D \rightarrow T + p + 4.00$ mev		10^{-2} barn
$D + He^3 \rightarrow He^4 + p + 18.30$ mev		2×10^{-3} barn
$D + T \rightarrow He^4 + n + 17.6$ mev		2.1 barns
$D + Li^6 \rightarrow 2He^4 + 22.4$ mev		
$p + Li^7 \rightarrow 2He^4 + 17.3$ mev		
$n + Li^6 \rightarrow T + He^4 + 4.6$ mev		

The first two reactions are both collisions between two deuterons, but with different end products. They occur with roughly equal probabilities. The energy released in these reactions is relatively small compared to that released in the other reactions. This makes reactions 3 and 4 quite interesting. The energy released in both these reactions is substantially larger than that released in the first two, and, in addition, both involve reactions between deuterons and the products of the

first two reactions. Neither of these "fuels" is available in nature. Tritium, T, another isotope of hydrogen with a nucleus containing 1 proton and 2 neutrons, is unstable and does not exist in nature, while He^3 exists in nature but with an abundance of only 10^{-4} per cent. Thus by starting with deuterium in the reactor additional fuel is produced as a by-product of the first reactions. Reactions 5 and 6 are also quite interesting because both Li^6 and Li^7 exist in nature and because the energy released is quite substantial. These two reactions are also interesting for another reason. Although we list them as fusion reac-

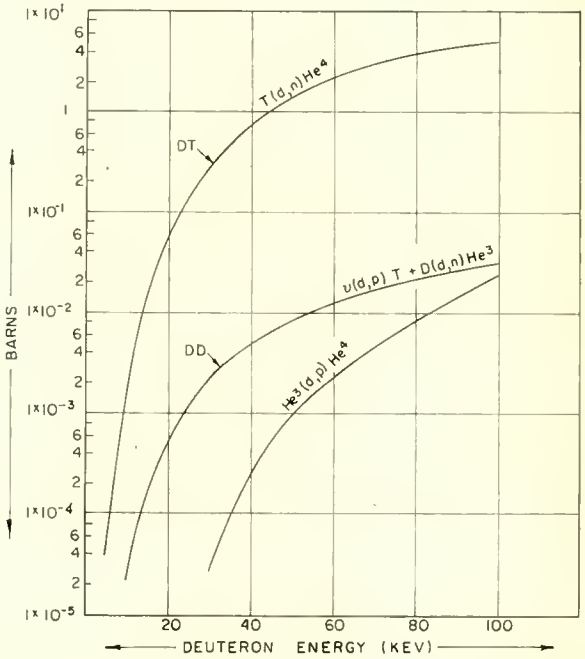


Fig. 5—Pictorial representation of the D-D reaction producing He^3 and a neutron.

tions, they are actually fission reactions since the final nuclei are lighter than the original nuclei. These are fission reactions stimulated by proton and deuteron bombardment. Since both of these particles are charged positively, these fission reactions can occur only at high temperatures so that they can safely be called thermonuclear reactions. From the list of reactions, it can be seen that at a temperature of 600 million degrees the D-T reaction is about 1,000 times more probable than the D-He^3 reaction and 100 times more probable than the D-D reaction. Thus the bulk of the energy from a deuteron reactor is produced when deuterium combines with tritium. The reactor can be made

more efficient if more tritium is introduced into the reactor than is produced by the D-D reaction itself. This makes the last reaction quite important since one of the products of the D-T reaction is a neutron. If Li^6 is introduced into the system, perhaps as some sort of lining for the reactor, these neutrons could be used to manufacture the tritium in the reactor itself.

The fact that several thermonuclear reactions are possible stimulates interest in producing a controlled fusion reactor. Consequently, it is of interest to consider the conditions under which the reactions will occur. As an analogy from chemistry, it is known that hydrogen and oxygen can unite to form water with the release of energy. However, hydrogen and oxygen can be in contact under ordinary conditions for extremely long times with very little reaction between them. Under certain conditions, for example in the presence of a spark, the reaction probability becomes extremely high. The additional information needed for fusion reactions is the probability that the reactions will occur under various conditions. This probability is described by a parameter called the effective cross section for the reaction, and this has the following meaning. Consider collisions between identical hard elastic spheres of radius r, such as billiard balls. A collision between two balls occurs whenever the separation between centers is $2r$ (Figure 6). It is convenient to think of this collision process in a different way. One of the balls may be regarded as a projectile and treated as a point, while the other ball is treated as a stationary target. A collision between the moving point and the target occurs whenever the point is moving toward the target within a cylinder whose axis is parallel to the direction of approach of the projectile and passes through the center of the target, and whose cross-sectional area is $4\pi r^2$. In reality we are assigning to the target an effective target area or collision cross section four times its actual geometric cross section. If a projectile is shot at random into a region containing many targets we can define a probability of collision, $p\ dx$, for the projectile in moving a small distance dx into the target region. This probability is the ratio of the target area seen by the projectile to the total area normal to its approach direction. Thus, for a bullet shot at random at a thin row of trees, the probability of a hit depends upon how much of the total area is occupied by trees. If only half the total area is covered by trees, the probability of a collision is one-half. If the region contains N targets per unit volume, each with a collision cross section σ, the volume dx thick with a base area A will contain $NA\ dx$ targets. If no target is shielded by another, the target area seen by the projectile is $\sigma\ NA\ dx$ and the ratio of this to the total area is $\sigma\ N\ dx$. Hence $p\ dx = \sigma N\ dx$.

From this definition one can determine the mean-free-path, $<l>$, for the collision. This, by definition, is the average distance a single particle moves between successive collisions; or it is the average distance moved by a large number of projectiles before they make their first collision. The mean-free-path for this case of hard elastic spheres turns out to be just the inverse of p

$$<l> = 1/p = 1/N\sigma$$

which depends only on the density of targets. For actual physical systems, such as nuclei, which do not behave as hard elastic spheres,

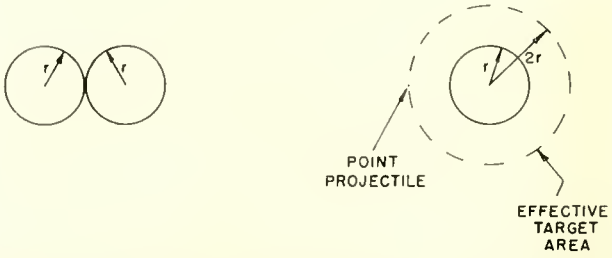


Fig. 6—Collision between hard elastic spheres. The drawing on the left represents the actual collision, while that on the right is an alternate representation with a point projectile colliding with a target having an effective target area four times its actual geometric cross section.

it is possible to determine the mean-free-path for a given type of collision† either theoretically or experimentally for a given density of targets. When this is done, it is convenient to express the relation between $<l>$ and N in the form of the equation given above. The parameter σ which must be used to put the relation in this form is called the effective cross section for that type of collision. For actual physical systems, cross sections for practically all types of processes are functions of the energy of the interacting particles. Nuclear reaction cross sections are of the order of 10^{-28} square meters. Since this is a very small number, the nuclear physicists have defined a new

† Between successive collisions of this type the particle may make many other collisions of different types. We usually define a mean-free-path for each separate type of collision, and this is the average distance moved between successive collisions of the same type. For example, we might imagine three separate groups of balls — green, red, and blue — on a billiard table. If we shoot a cue ball at random into these balls we can define a mean free path for collisions of the cue ball with red balls. Between two successive red ball collisions the cue ball may hit several blue and green balls, but these are ignored in determining the mean free path for red ball collisions. We can, of course, define a mean-free-path for two successive collisions irrespective of the type.

unit of area to describe nuclear cross sections. This unit is the "barn" which by definition is 10^{-28} square meters.

Once the cross section for a given reaction is known as a function of energy, it is a straightforward matter to determine the rate at which the reaction will proceed under given conditions. To see roughly how this is done we assume a well-collimated beam containing n projectiles per unit volume, all with the same energy, and hence with the same speed v, incident on a region containing N targets per unit volume. The number of projectiles crossing unit area per unit time into the target region is nv, and each of these projectiles has a probability of $N\sigma$ of undergoing a reaction in moving unit distance. Therefore, the number of reactions taking place in unit volume per unit time is $R = nvN\sigma$, the reaction rate. Most systems in which reactions occur do not involve well-collimated beams of projectiles, but rather the projectiles are moving in all directions with a spread in energy and speed. Under these conditions the reaction rate must be written as $R = nN \langle v\sigma \rangle$, where $\langle v\sigma \rangle$ is a weighted average of the product of v and σ. If all the reacting particles are of the same type, as in the DD (deuteron-deuteron) reaction, the reaction rate becomes $R = \frac{1}{2} n^2 \langle v\sigma \rangle$. The product of the reaction rate and the energy released per reaction is the reaction power density of the system. Unless the reaction rate and the reaction power density for a given reaction can be made sufficiently high under reasonable conditions, the reaction will be of no practical use as a power source.

One may well ask what is meant by a sufficiently high reaction power density. Some of the energy produced in any system to be used as a practical power source must leave the system to be put to use. However, we have already seen that if the burning process is to be maintained, enough of the released energy must be kept in the system to ignite more fuel. As far as the burning process itself is concerned, any energy which leaves the system may be considered a loss, even though it is put to use. If the power loss is too great, the fuel will cool and combustion will stop. The power losses, then, essentially determine what the reaction power density must be for a sustained reaction. Thus it is necessary to know the different methods by which energy can leave the system, and how the energy loss rate for each of these methods depends upon the system parameters such as temperature, fuel density, impurities, etc. A large part of the energy produced in the fusion reaction is taken on by the neutrons produced in the D-T reaction. Since neutrons are electrically neutral the magnetic bottle cannot contain them, nor can the material walls. We can keep some of this energy in the system by putting a lining of Li^6 in the reactor to produce tritium by neutron capture. Since a charged

alpha particle is produced along with the tritium, both particles can be contained by the magnetic bottle keeping the energy in the system. The main energy loss, besides that due to neutrons, is radiation produced when electrons interact with the positive ions in the system—so-called bremsstrahlung radiation. The electron is accelerated by the attractive force due to the ion, and like the accelerated electrons in a radio transmitting antenna, it gives off electromagnetic radiation. For the conditions existing in a fusion reactor this radiation is in the X-ray range. An analysis of this loss mechanism shows that the power loss from unit volume is proportional to $n_e^2 T^{\frac{1}{2}} Z^2$ where n_e is the density of electrons in the reactor, T is the temperature and Z is the atomic number of the ions with which the electrons interact. It turns out that the reaction power density also depends upon n_e^2 but its temperature dependence is exponential. Consequently, at low temperatures the power loss density is greater than the reaction power density, but as the temperature increases the reaction power density increases faster than the loss rate. At some temperature the two will be equal and this will be the minimum temperature for a sustained reaction.

One further point of practical interest is the factor Z^2 in the loss term. Since the losses increase rapidly for heavier elements in the system, a reactor should be "scrupulously clean." This imposes severe vacuum problems for the designers of reactors.

It is always emphasized that fusion can occur only at very high temperatures. It might be well to discuss briefly just what is meant by temperature for such a process. Temperature is a statistical concept, and it makes no sense to speak of the temperature of a single particle. For systems of interacting particles in equilibrium with each other and with the surroundings, the energy distribution function involves a parameter T which is the absolute temperature defined in thermodynamics. The average energy of the particles of the system can be determined from the energy distribution function, and this average energy will always involve the parameter T. Thus the average energy of a group of nuclei described by a Maxwell–Boltzmann distribution function is $\langle\epsilon\rangle = (3/2)kT$, where k is Boltzmann's constant. We can say, then, that the temperature is a measure of the average energy of the particles. When a system of particles is not in equilibrium with its surroundings, there is still an average energy per particle. Again, the average energy can be described by a parameter called temperature, but to indicate that this is not the usual thermodynamic temperature of systems in equilibrium, we call it the kinetic temperature of the system. It is convenient to define this kinetic temperature for a group of particles as $T = (2/3)\langle\epsilon\rangle/k$. This particular definition is used so that the kinetic temperature and the thermodynamic temperature

have the same numerical value when the group of particles is in equilibrium with its surroundings.†

Since energy and temperature are related, it has become customary in fusion work to use units of energy and temperature interchangeably. That is the product kT has units of energy, while the ratio ϵ/k has units of temperature. Thus, for $\epsilon = 1$ ev $= 1.6 \times 10^{-19}$ joule, $\epsilon/k = 1.6 \times 10^{-19}/1.38 \times 10^{-23} = 1.16 \times 10^{4}$ degrees Kelvin. Hence we can describe a kinetic temperature of 11,600 degrees Kelvin as a kinetic temperature of 1 ev. As a rough working figure 1 ev of kinetic temperature is equivalent to 10^{4} degrees Kelvin. From our definition of kinetic temperature, $<\epsilon> = (3/2)kT$. Consequently, the average energy of a group of particles with a kinetic temperature of 1 ev is 1.5 electron volts. In fusion one usually deals with a larger unit of kinetic temperature, the kev or thousand electron volts. As a good working figure, 1 kev of kinetic temperature is 10^{7} degrees Kelvin.

From this, it can be seen that a kinetic temperature of 10^{7} to 10^{8} degrees Kelvin should not be too difficult to attain under certain conditions. For example, in a cathode-ray tube the electrons leaving the cathode can easily be accelerated through 10 kv. Under these conditions the kinetic temperature of the electron is of the order of 10^{8} degrees Kelvin. This is a system in which particles at a very high temperature actually hit the boundaries. The boundaries, under normal conditions, do not vaporize because the number of particles in the beam is relatively small and the total power input to the face of the tube is not sufficient, in competition with the losses, to raise the temperature of the boundaries to the melting point or beyond. If the number of electrons in the beam is increased sufficiently, the spot being bombarded may melt as some people have found out to their dismay.

From the curves of reaction cross section for the D-D reaction (Figure 7), it is seen that the cross section reaches a fairly decent value at deuteron energies of 50 kev and above. It would seem then, that we need only give deuterons energies of this amount and fusions should occur. The simplest fusion reactor which comes to mind is to bombard a deuterium target with a beam of deuterium ions, or deuterons, which have been accelerated through 100 kev, something fairly easy to do. This type of reactor, however, is not feasible. In the target the deuterium exists as neutral atoms and it turns out that a very large fraction of the deuteron energy is spent in exciting and ionizing the deuterium atoms of the target. Some fusion will occur, but it can

† Other definitions of kinetic temperature are used by some people. One of the most common is $T = <\epsilon>/k$. Since, in fusion work one deals with orders of magnitude differences in temperature, the relatively small difference between the two definitions is not of much practical concern.

be shown that the energy gain in this type of reactor is of the order of 10^{-4}, not a very practical energy source.

To get around this difficulty, instead of bombarding a deuterium target we might arrange to shoot two deuteron beams, both of which have been accelerated through say 50 kv, toward each other in a long tube. Since both beams are ionized to begin with, there will be no energy loss in excitation and ionization so that all the energy is available for fusion. Unfortunately, this particular method is not very attractive. The cross section for the D-D reaction at 100 kev is 3×10^{-2} barn. The mean-free-path for the reaction is $<l> = 3 \times 10^{29}/N$ meters, where N is the number of deuterons per unit volume in either beam. If one assumes a particle density in the ion beams of $10^{25}/m^3$, the density of particles in a gas at atmospheric pressure and room temperature, the mean-free-path is 10^5 meters or about 200 miles. Thus, one would need a reactor at least this long to get any appreciable fusion. This density in a beam is extremely high. For example, an ion beam with this density accelerated through a potential of 100 kv represents an ion current density of 6×10^{12} amperes per square meter.

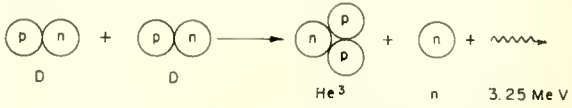


Fig. 7—Cross sections for several fusion reactions (After Richard F. Post).

This is many orders of magnitude greater than any ion current density which has ever been achieved. If we assume a more realistic ion beam current density of say 10^4 amperes per square meter, or 8 orders of magnitude less, the mean-free-path becomes 8 orders of magnitude greater than before or approximately 10^{10} miles. Since the distance from the earth to the sun is "only" 10^8 miles, this type of fusion reactor is not practical.

These long distances might be achieved on earth by making the reactor in the form of a toroid and applying a strong magnetic field normal to the plane of the toroid. However, since the two beams must go around the toroid in opposite directions, this is not practical. The magnetic field will be able to make only one of the beams go around the toroid, the other would go to the walls. Another possibility would be the use of two toroids which intersect as shown in Figure 8, with the two deuteron beams moving in the same sense in each toroid. At the intersection of the tube, the beams are moving in opposite directions, and hence bombard each other. This device also has limitations because of the density which can be put in the beams. Because of

severe space-charge problems this scheme, at the moment, does not look practical.

In all these schemes, only positive ions, in the form of deuterons, are in the system. The space charge due to these ions imposes severe limitations on the density of reacting particles which can be put into the system, and hence, on the reaction power density. There is a way to avoid the space-charge limitation. By forming the positive ions in the region in which they are to be used, the space charge of the ions can be neutralized with the electrons which are removed from the neutral atoms to form the ions. Regions in space in which there are equal densities of positive and negative charge are called plasmas. At present, it appears that any successful fusion reactor will involve a plasma. While the use of a plasma removes the space-charge limitation on the density, other factors will limit the density. However, none of

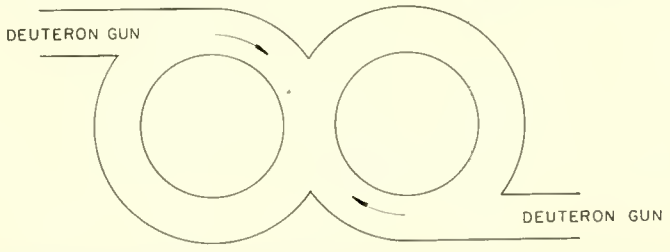


Fig. 8—Fusion reactor consisting of two intersecting toroids, with a uniform magnetic field normal to the plane of the paper. Both deuteron guns introduced deuteron beams which have been accelerated through potentials of the order of 50kv. As discussed in the text, this is not a promising form of reactor.

these other limitations on the density will be as severe as the space-charge limitation.

Once a plasma is formed in a reactor and the plasma particles are given energy by some means, a scheme must be found for keeping the energetic particles away from the walls. In principle, a magnetic field along the axis of the reactor tube can achieve this since charged particles spiral about magnetic field lines—electrons in a clockwise direction and positive ions in a counterclockwise direction as one looks along the positive direction of the field (Figure 9). The stronger the magnetic field, the tighter is the spiral for particles of a given energy.

There are several schemes for accomplishing this. Perhaps the simplest is to wrap a solenoid around a straight tube. In this scheme the problem arises as to what to do about the ends of the tube since the energetic particles must be kept away from all the walls. The most obvious solution is to bend the tube back on itself to form a toroid so that the ends are eliminated.

If this is done, however, a new difficulty arises. An analysis of the motion of a charged particle in a magnetic field shows that any external forces applied to the particle normal to the magnetic field, or any inhomogeneities in the magnetic field itself, produce a drift of the particle in a plane normal to the magnetic field. A simple example of this drift is the cycloidal motion of a charged particle in crossed electric and magnetic fields. A coil wrapped around the toroid to produce a field along the circumference of the toroid produces a stronger field along the inner circumference than along the outer circumference, so that there is an inhomogeneity in the field. Thus the charged particles would tend to drift to the walls. It still may be possible to use this scheme for a successful reactor if the time for which a particle needs to be confined before a reaction takes place can be made short compared to the time it takes a particle to drift from the center of the tube to the walls.

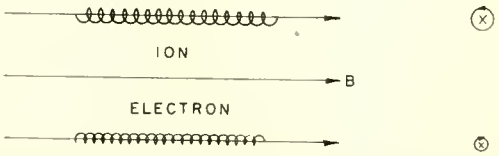


Fig. 9—Motion of charged particles about magnetic field lines in a uniform magnetic field. The drawing at the right indicates the motion as seen by an observer looking in the positive direction of the field. For equal energies, the electrons move in tighter spirals than the positive ions.

With the reactor in the form of a straight tube or in the form of a toroid, it is possible to produce a constraining magnetic field without using coils wrapped around the tube. For a long time it has been known that when a very large current exists in a plasma, the plasma tends to contract in a direction normal to the direction of the current. This contraction, called the pinch effect, is produced by the magnetic field of the current itself. The pinch effect has been achieved and reported in the literature by groups in England, Russia, and the United States. While they have all achieved this effect, they have done so for only short lengths of time. The pinch tends to be unstable in several different ways, so that once it is established it tends to destroy itself very quickly—in times of the order of microseconds. If a pinch can be stabilized it may lead to a successful fusion reactor.

These are but a few examples of the field configurations one might visualize for a reactor. Without a doubt, many more configurations have been studied or are being studied. Needless to say a successful fusion reactor has not yet been achieved. There are many complex

problems, only a few of which we have discussed in this short account, standing in the way of success. However, enough work has been done to lead many people to believe that it is only a matter of time before the problem will be solved.

Until a successful fusion reactor is achieved, and for a very long time thereafter, the whole field of thermonuclear fusion will be an exceedingly interesting and challenging field to beckon the electrical engineer.

CONTROLLED THERMONUCLEAR FUSION—ITS MEANING TO THE RADIO AND ELECTRONIC ENGINEER*†

By

Edward W. Herold

RCA Laboratories,
Princeton, N. J.

Summary—The endeavor to control thermonuclear fusion to produce useful power has been receiving wide publicity. This paper reviews the subject, first showing the need for such power and the great promise it offers to mankind. It is then shown that the most likely approach is by means of a large, electrodeless, high-temperature gas discharge using rarefied deuterium, or a deuterium–tritium mixture. Satisfactory containment and ignition, and new diagnostic tools present problems to the solution of which the radio-electronics engineer can make a considerable contribution. The rudiments of a fusion power reactor are discussed; it uses deuterium and lithium as fuels, and a deuterium–tritium discharge. The paper concludes with a discussion and evaluation of some recently published work.

Introduction

THE great interest in controlled thermonuclear fusion of heavy hydrogen, i.e., deuterium, is evidence of the great potentialities of this power source to mankind. The fundamental nuclear phenomena of fusion are found in textbooks of the 1930's, but substantial energy production did not occur until the first successful hydrogen-bomb tests of 1952. Work on controlling fusion energy for peaceful uses was largely carried on in secrecy until 1956 when an excellent survey paper,[1] and an important book,[2] were published in the U.S.A., and a series of Russian papers appeared[3] in the U.S.S.R.

* Manuscript received April 18, 1958.

† A very condensed form of this paper was presented at the IRE National Convention, March 25, 1958, and will appear in the Institute of Radio Engineers 1958 Convention Record.

[1] R. F. Post, "Controlled Fusion Research," *Rev. Mod. Phys.*, Vol. 28, p. 338, July, 1956; also *Proc. I.R.E.*, Vol. 45, pp. 134-160, February, 1957.

[2] L. Spitzer, Jr., *The Physics of Fully Ionized Gases*, Interscience Publishers, New York, N. Y., 1956.

[3] I. V. Kurchatov, "On the Possibility of Producing Thermonuclear Reactions in a Gas Discharge," *Soviet Journal of Nuclear Energy*, Vol. 1, p. 65, 1956. This was the first of a series of 5 papers by different authors, all in the same issue. All 5 papers are reprinted, in English, in *Journal of Nuclear Energy*, Vol. 4, p. 193-234, February, 1957.

Since then there have been many published papers, and the 1958 Geneva Conference on the Peaceful Uses of Atomic Energy will undoubtedly bring forth much more work throughout the world.

The present paper is a review of the major problems of controlled thermonuclear fusion, with particular emphasis on the part the radio and electronic engineer can play in helping with their solution. For a more rigorous scientific treatment of the fundamentals, the reader is referred to the aforementioned references, particularly (1) and (2), and the literature references quoted therein. Data and elementary formulas have been freely drawn from these sources without repeating the derivations. The evaluations and interpretations, however, are those of the author.

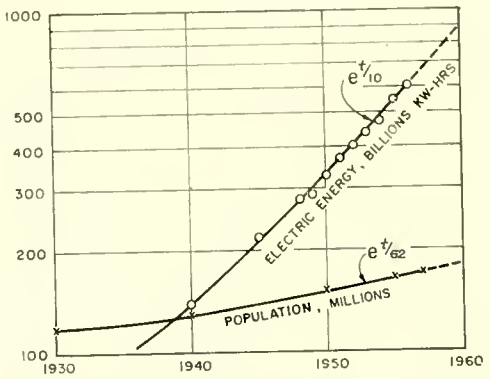


Fig. 1—Annual statistics of the U. S. A. Taken from the 1958 *World Almanac*.

The Need for Fusion Power

Many other demonstrations have been made that our energy-producing fuels are rapidly depleting.[4] Only one simple illustration will be used here, shown in Figure 1. It is seen that the population trend, and the increase in use of electric energy in the U.S.A., appear to be exponential, but the electric energy curve rises at 6 times the rate of that of population. These curves will not necessarily continue with the exponents shown, but they illustrate a basic trend in our civilization. This trend is toward such a rapid increase in our per capita energy requirements as to be alarming. If the energy curve were to continue for 100 years, with its time constant of 10 years, our energy requirement would rise to e^{10}, which is 22,000 times the present rate!

[4] P. C. Putnam, *Energy in the Future*, D. Van Nostrand Co., Princeton, N. J., 1953.

Nuclear power is the only known solution, and it will play a dominant role in less than 100 years. Unfortunately, uranium is limited in supply, and has highly dangerous by-products which may yet be fatal to our survival. There are no such disadvantages to fusion of heavy hydrogen, or deuterium as we shall call it. Furthermore, the sun proves that controlled thermonuclear power is practical, useful, reliable and permanent. It seems probable that fusion power will be achieved on earth, though in quite different form from that of the sun, in the laboratory within 5 to 10 years, and in practical application within 50 years; this is in time to supplant conventional sources and fission reactors both, well before they are exhausted of fuel.

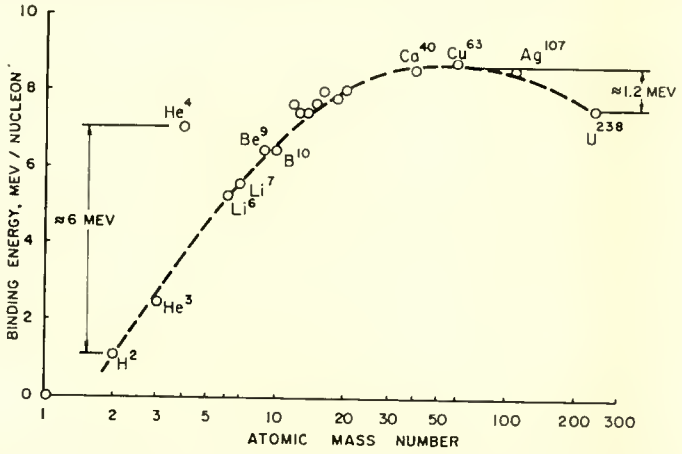


Fig. 2—Nuclear binding energy curve of the stable elements, per nucleon. Energy is released whenever nuclei of elements low on the curve are used to produce elements higher up on the curve.

Nuclear Reactions

One needs very little knowledge of nuclear physics to understand the fusion process, because the latter is so closely analogous to burning of chemical fuels. In both cases, one needs to ignite the fuel and, for each combining event, a known energy is produced together with a relatively stable end product. Figure 2 shows a nuclear binding energy curve for the stable elements.[5] The binding energy is the energy per nucleon needed to dissociate an original nucleus into its component nucleons—protons and neutrons. Thus the highest binding energy materials are the most stable, and the greatest energy is produced when they are formed. Any nuclei, which either combine or split to

[5] The curve was prepared from nuclear mass data published in *American Institute of Physics Handbook*, Section 8, Nuclear Physics, McGraw-Hill, New York, 1957.

form new nuclei of higher binding energy, can potentially release energy according to the differences shown on this curve.

A very remarkable nucleus is that of helium-4, also known as the alpha particle. Its nucleus of two protons and two neutrons is so stable, and yet so simple, that it is an outstanding end product for fusion of all the light elements. In the sun, hydrogen eventually fuses to helium. On earth, it is much simpler to use deuterium, since it is also very plentiful, and it already contains a neutron, one of the building blocks needed to produce helium. Also plentiful and an interesting candidate is lithium, of which we shall speak later since it appears likely to be a fuel, along with deuterium, in early fusion reactors.

Just as we write chemical reactions in formulas, nuclear fusion formulas are also used. The four reactions of greatest interest with deuterium are

$$D + D \rightarrow (He^3 + 0.8\ \text{mev}) + (n + 2.45\ \text{mev}), \tag{1}$$

$$D + D \rightarrow (T + 1.0\ \text{mev}) + (p + 3.0\ \text{mev}), \tag{2}$$

$$D + T \rightarrow (He^4 + 3.6\ \text{mev}) + (n + 14.1\ \text{mev}), \tag{3}$$

$$D + He^3 \rightarrow (He^4 + 3.7\ \text{mev}) + (p + 14.6\ \text{mev}), \tag{4}$$

where energies are given in millions of electron volts (mev). Measurements indicate that the first two reactions take place with about equal probability, while Reaction (3) is about a hundred times more probable than (1) and (2). The last reaction is least likely, having about a tenth of the crossection of Reactions (1) and (2). All four reactions are needed for complete fusion of deuterium to helium, which is seen by adding them; this gives a balance

$$6D \rightarrow 2He^4 + 2n + 2p + 43.2\ \text{mev}. \tag{5}$$

Thus the average energy produced is 7.2 mev per deuteron, or 3.6 mev per nucleon.

Since the D-T reaction is so much more probable, and leads to the lowest ignition temperatures, it is tempting to use it. However, tritium is not found naturally; it is also radioactive and expensive to produce. A way around this is to use another nuclear reaction,

$$n + Li^6 \rightarrow T + He^4 + 4.6\ \text{mev}. \tag{6}$$

It will be noted that Reaction (3) uses up a triton, generates a neutron; Reaction (6) makes use of this neutron to react with the relatively plentiful isotope-6 of lithium to regenerate the triton. Thus, adding (3) and (5), it is possible to obtain a closed-cycle balance

reactor. It must also be realized that deuterons bombarding a solid or liquid target will also not cause many fusions, because they will lose almost all their energy in heating the solid or liquid.

We conclude that the presently practical way to achieve fusion power is by the heating of all the gas particles until they have an average energy of 10 kev or more. They must then be held together long enough, and in a relatively fixed volume, so they can eventually react. This is known as a thermonuclear reaction. With a Maxwellian distribution, an average energy of 10 kev corresponds to a temperature of about 10^{8}°K. At such temperatures, attainable only in gases, the gas is fully ionized and forms a plasma (free electrons and ions with equal charge density). Furthermore, material walls, or electrodes, would immediately cool such a gas.

Evidently, a fusion reactor is what physicists have called an "electrodeless" gas discharge tube, and the ignition process is most conveniently performed by electric induction or some such means. It is these factors which make fusion power of such direct interest to the radio and electronic engineer.

To find out how much power can be produced, if the hot gas could be held together, we must find the reaction rate. For a deuterium fuel, the reaction rate requires taking the cross section, σ, of Figure 3 and multiplying by the particle velocity, v, and averaging over the Maxwellian velocity distribution. This has been done at Los Alamos and published by Post.[1] For a deuterium gas of ion density n_i, which in this case is also the deuteron density, n_D, one gets a reaction rate[1] of

$$R_{DD} = \frac{1}{2} n_D^2 \, (\overline{\sigma v})_{DD}. \tag{8}$$

If it is assumed that Reactions (1) and (2) are equally likely, but that each triton produced is immediately "burned" in Reaction (3), it is found that the average reaction produces 12.6 mev, 34 per cent of which is in charged helium and hydrogen nuclei, and 66 per cent in neutrons. Thus, one gets

$$\text{D-D Power Density} = \frac{1}{2} n_D^2 \, (\overline{\sigma v})_{DD} \times 12.6 \text{ mev}. \tag{9}$$

A curve of the all-deuterium power density, plotted against temperature, is shown in Figure 4, using a typical particle density $n_i = n_D = 10^{15}/\text{cm}^3$. The dotted curve just below is the portion which is in charged particles.

For the deuterium–tritium reaction, the corresponding formula[1] is

$$\text{D-T Power Density} = n_D\, n_T\, (\overline{\sigma v})_{DT} \times 17.6 \text{ mev.} \qquad (10)$$

In this case, the energy is about 20 per cent in charged particles and 80 per cent in neutrons. Figure 4 shows curves for the case of equal parts tritium and deuterium and a density of

$$n_i = (n_T + n_D) = 10^{15}/\text{cm}^3.$$

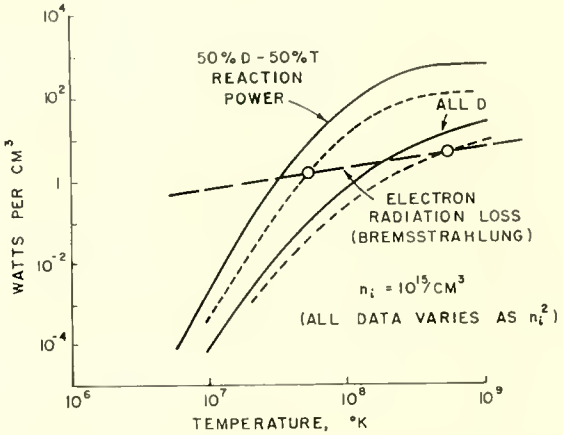


Fig. 4—Calculated total fusion reaction power density (solid curves) from a thermonuclear plasma of density 10^{15} ions/cm^3. The dotted curves give the part of the power retained by charged particles. The ignition temperatures are the intersection of the dotted curve with the radiation-loss curve (dashed line).

Since the power density varies as n_i^2, the data of Figure 4 is readily converted to other densities.

Now that we understand the thermonuclear process as one of a very hot, fully-ionized gas, we must consider any limiting processes which might cause more energy to leave the reaction than is created in it, so as to cool it. Normal black-body radiation from a gas is practically negligible and, if we continue to assume perfect containment, and no electrode system, losses to the walls and electrodes are not a factor. The one loss which is basic is the X-ray radiation from the plasma electrons which comes about whenever their velocity is altered by a so-called "free–free" collision with an ion. The electron energy losses in the Maxwellian gas are random and the radiation

covers a broad energy spectrum, from the radio and infrared to the X-ray region. These radiations are called bremsstrahlung and their power density is given theoretically by[2]

$$\text{Radiation Power Density} = 0.15 \times 10^{-33}\, n_e\, n_i\, Z^2\, T^{1/2} \text{ watts/cm}^3, \quad (11)$$

where n_e and n_i are the electron and ion densities in particles per cubic centimeter, Z is the charge on each ion (this is unity for deuterium and tritium), and T is the plasma temperature in °K. Since this power loss also varies as the square of the particle density (since $n_e \sim n_i$), it is seen that the fusion power produced, and the radiation power lost, vary in the same way with density. The temperature variation is quite different, however, as is seen in Figure 4. The hydrogenic radiation loss was increased by 30 per cent (to include the effect of a small proportion of helium and heavier impurity elements).

As seen later, confinement of a thermonuclear gas is most easily done by magnetic fields, so that only charged particles are held, and the neutrons escape. Thus, the ignition temperatures can be found from Figure 4 as the point of intersection of the dotted curves (charged-particle power) with the radiation-loss curve. For the pure deuterium reaction, 5×10^8°K is required before the charged-particle power exceeds the radiation loss. For the *D-T* case, on the other hand, the required temperature is ten times lower, or 5×10^7°K. Actual operating temperatures must exceed these values by enough to overcome other losses. For the *D-T* reaction, 10^8°K is a suitable temperature. It produces a fusion power of 60 watts/cm^3, which compares with fission reactors and other high-power generating equipment.

Now let us examine the plasma density. Although the ignition temperature does not change with plasma density, the useful range of density is still rather narrow. If the density is lowered to 10^{13}/cm^3, the power generated at 10^8°K drops all the way down to 6 milliwatts/cm^3, which is far too small to be useful. If the density is raised to 10^{17}/cm^3, one gets 0.6 megawatt/cm^3, which is uncontrollably large. Atmospheric density is over 10^{19}/cm^3; if this density were used for a fusion reaction, an explosive would result. Thus the most suitable density centers around 10^{15}/cm^3, differing at most by an order of magnitude higher or lower. At room temperature, this density gives a pressure of about 0.03 mm of mercury, i.e., 30 microns, but at thermonuclear temperatures, the pressure is very much higher. This is one of the problems of containment which will be discussed.

Ordinary gas discharges, for example a fluorescent lamp, also contain about the same density of gas molecules, but most of the gas is not ionized so the *plasma* density seldom exceeds 10^{12}/cm^3. Their

electron plasma temperatures are only about 10^{4}°K, and so the thermonuclear plasma is a very different medium, having 1,000 times the density and 10,000 times the temperature.

Containment

Thus far, the ground covered seems firm and well understood. It is now necessary to discuss two areas, containment and ignition, which are poorly understood. As a result, there will be less to say about them, and yet they are the problems to which the radio-electronics engineer can make the greatest contributions.

Containment of a gas is ordinarily by a vessel of some solid material. However, as we have seen, this is impossible for the thermonuclear fusion gas because of the cooling effect of such a vessel. A containing force which acts at a distance is required, i.e., either gravitation, or electric, or magnetic fields. Gravitation requires such a large mass that it is suited only for the sun and stars. Static electric fields tend to act in opposite directions on ions and electrons and so are ruled out. Most present ideas on containment involve magnetic fields, of a quasi-static nature.

The containment problem is difficult and complex chiefly because we are dealing with a high-temperature Maxwellian-distribution gas plasma which has remarkable properties.[2] Perhaps the simplest summary is to describe such a plasma as like a nearly perfect conductor made of an elastic, self-expanding sort of material. In the gas, collisions are so infrequent that there are few losses corresponding to friction; thus the Q is high, and violent oscillations and instabilities can occur. The science of such plasmas in magnetic fields is called magnetohydrodynamics,[9] and it is still in its infancy. It is not practical to discuss the fine points of the science in this paper; however some of the better-known properties of plasmas will be outlined.

It is easy for the radio engineer to understand containment of *low-density* electric charges by magnetic fields because this occurs in every magnetically focussed electron tube, and in magnetrons when they are beyond cutoff. These simple concepts provide a start to understanding high-density plasma containment. Figure 5 shows the simple case of an isolated ion or electron in a magnetic field, as taken from Spitzer.[2] Such charges experience force at right angles to the magnetic field and to the perpendicular component of their own velocity. Only velocity components parallel to the field are unaffected. Thus the particles, whether they be electrons or ions, are held to the magnetic field lines,

[9] T. G. Cowling, *Magnetohydrodynamics*, Interscience Publishers, New York, N. Y., 1957.

rotating in circles around them, as shown at the left of the figure, and free to move back and forth along field lines. A large magnetic field gives small circles and relatively "tight" containment. When there are other fields present at right angles to the magnetic field, such as an electric field, or a gradient of a magnetic field, the particles drift across field lines and are no longer contained, as illustrated in the right-hand portions of Figure 5. This phenomenon is the first basic problem of containment, since there is no way known of establishing a completely uniform magnetic field with no gradients or electric forces. However, the drift velocity is inversely proportional to the magnetic field, so that the "tight" containment of a large magnetic field continues to be relatively effective.[2]

With a plasma, the problem is more complex because plasmas are good conductors, and any electric current in the plasma sets up its own field and disturbing force. Yet, as we have already suggested, a most

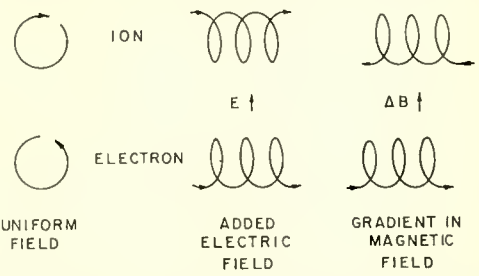


Fig. 5—Containment of isolated charged particles in a magnetic field normal to the plane of the drawing. Electric fields, and magnetic field gradients cause a drift across the field lines.

convenient heating means is electromagnetic induction, in which a plasma field is a requirement. Also, electron–ion collisions in a plasma lead to a sudden energy change and a random-walk diffusion process, whereby particles from a high pressure region diffuse to a low-pressure region. It can be shown[2] that the diffusion velocity is inversely proportional to the square of the magnetic field, so that again a strong field leads to much better containment. The bugbear of all containment problems is the effect of the plasma dynamic instabilities. These are the result of the other forces, and of the finite wave-propagation phenomena in a plasma, and are rather poorly understood.

To understand plasma containment quantitatively, one must consider some of the chief properties of a fully ionized dense plasma. First of all, there is charge neutrality, because even a slight charge unbalance leads to extremely high electric fields and currents. For a deuterium or tritium plasma, this is equivalent to saying that ion and

electron densities are equal, and any separation of charge tends to be suppressed. A second important characteristic is the very low electrical resistance of hot plasmas. Unlike metals, in which lattice collisions cause electrical resistance, a plasma has no lattice; the hotter it gets, the faster move the charges and the lower the resistance. Thus the plasma resistivity of a hydrogenic gas is given by a very simple inverse temperature relationship, which is approximately

$$\text{Resistivity} \approx 7 \times 10^4 \, T_e^{-3/2} \text{ ohm-cms}, \tag{12}$$

where T_e is the electron temperature in °K. Particularly to be noted is the approximate independence of resistivity with plasma density. This is because, if the charge carrier density changes, the effect of collisions changes in proportion, leaving the resistivity unaltered, to the first order. A more exact treatment is found in Spitzer.[2] Very low plasma resistances result at the thermonuclear temperatures, e.g., at 10^8°K the resistance is 4 per cent of that of copper at room temperature; at 10^9°K, it drops to .13 per cent. In a magnetic field, the resistivity along the field lines remain the same, but the perpendicular resistivity is twice as large.[2]

The pressure of a fully ionized plasma, as with all gases, is a simple function of the temperature and is given by the kinetic-energy density. For an electron density in particles/cm^3, n_e, an ion density, n_i, and equal electron and ion temperatures,

$$\text{Pressure} = (n_e + n_i)\, kT. \tag{13}$$

If k, Boltzmann's constant, is in practical units (1.37×10^{-23} watt-sec/°K) and T in °K, the energy-density kinetic pressure is in units of watt-sec/cm^3 (1 watt-sec/cm^3 = 10^7 dynes/cm^2 $\approx$ 10 atmospheres). This quantity also represents the energy needed to heat the gas. A density of $n_e = n_i = 10^{15}$/cm^3 at a temperature of 10^8°K requires 2.7 watt-seconds/cm^3, which represents about 27 atmospheres pressure.

In a magnetic field, any changing fields sets up strong counteracting currents in the plasma tending to keep the field out, as in a perfect conductor. If the field is set up first, and the plasma then heated, the field is trapped inside. The plasma is, therefore, diamagnetic, a property also associated with a superconductor. Barring the previously discussed loss-of-containment effects, the plasma particles tend to follow the magnetic field lines and revolve around them. This has the consequence that, if the field is contracted, the plasma contracts with it, and we are able to move the plasma by moving the field. But strangest of all is the reciprocal property; if the plasma contracts, the

field must contract with it. The plasma is like a nearly perfect conductor which is free to change its shape or size to conform to any static or dynamic forces. The intriguing possibilities in the science of magnetohydrodynamics are indeed challenging.

A final basic property of the plasma is the magnetic field needed for containment. It has been pointed out that high fields lead to reduced drift and diffusion velocities and so are advantageous. However, one must also impose, as a minimum requirement, that the magnetic containment force exceed the basic plasma pressure. One may reason that the force on a contained plasma at its boundary is given, as with all conductors, by the vector product of current and magnetic field. Since the current flow is always such as to resist a change in field, the current and the magnetic field must be proportional to each other. Thus the containing force is proportional to the square of the magnetic field, or to the square of the current. This will be seen more clearly below, in considering the pinch effect. Since this containing force must exceed the plasma pressure, the minimum relation for containment[1] is

$$\frac{B^2}{8\pi} \geqq (n_e + n_i)\, kT, \tag{14}$$

where B is the magnetic field in gausses when the pressure term is expressed in dynes/cm^2. For $n_e = n_i = 10^{15}$, and a temperature of 10^{8}°K, it was seen that the kinetic pressure was 2.7 watt-sec/cm^3 $= 2.7 \times 10^7$ dynes/cm^2, and therefore the containment field must exceed 26,000 gausses.

With magnetic containment, the maximum radius of gyration of randomly oriented individual charged particles is proportional to the square root of the mass and the square root of the energy, and inversely proportional to the magnetic field. For a 10 kev particle, typical of a plasma temperature around 10^{8}°K, the maximum radii, r_g, are

10 kev electron $r_g \approx 330/B$ cms,

10 kev deuteron $r_g \approx 2 \times 10^4/B$ cms,

10 kev triton $r_g \approx 2.6 \times 10^4/B$ cms,

10 kev alpha particle $r_g \approx 2 \times 10^4/B$ cms.

For $B = 26{,}000$ gausses, these correspond to radii of 0.01 centimeter for electrons, 0.8 centimeter for deuterons and alpha particles, and 1.0 centimeter for tritons. Of course many particles have higher ener-

gies; the 3.6 mev alpha particle from the D-T reaction is an extreme, and will have a radius of 15 centimeters in a 26,000-gauss field.

The Pinch Effect

Clearly, if containment requires a magnetic field of such high values, the cheapest way to get the field is to use the plasma itself as the conductor. Its resistance can be lower than copper, and any ohmic loss contributes to heating it up, which must be done anyway. This is the tremendous advantage of a method of containment known as the pinch effect.[1] Most published work on thermonuclear plasmas throughout the world is based on this effect. (An exception is the "Stellarator" which uses other containment means.)

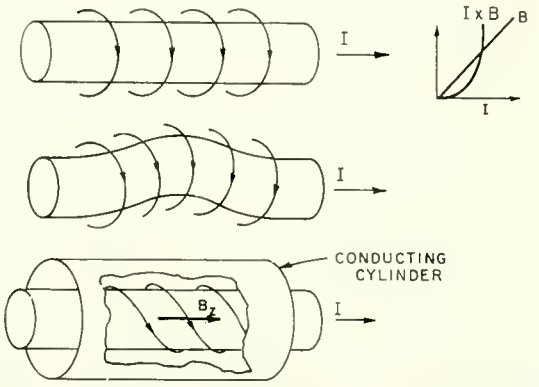


Fig. 6—The pinch effect and its stabilization by a longitudinal field and a conducting cylinder. The inward radial containment depends on $I \times B$ and hence on I^2 and B^2, since I and B are proportional.

Figure 6 (top) shows a tube of plasma with heavy currents along it. This produces a circular magnetic field as shown, and a radial inward force proportional to $I \times B$. Since B increases as I, the force increases as I^2 or B^2 and, above some definite value, can exceed the plasma pressure. When this happens, the plasma contracts and its particle density and internal pressure increase. The ohmic heating of the current, together with the adiabatic compression, increase the internal temperature and pressure even more. However, the process is a dynamic one and does not reach an equilibrium. Furthermore, since in a thermonuclear case the current, I, must be an induced one, it cannot be sustained indefinitely in the same direction. At best, then, the pinch effect is a transient condition, and those working with it hope for a long enough containment to allow heating and fusion to take place. The end effects have not been mentioned; however, it is obvious that a torus can be used so as to make an essentially endless plasma.

There are many instabilities in a pinched plasma, of which the kink instability shown in Figure 6 (center) is one of the simplest to understand. It is easily seen that, should the plasma bend at any time, the field, B, becomes stronger at the concave side and, since the force varies as B^2, the concavity is accentuated and the kink gets worse until the pinch "explodes," so to speak. It is now generally known that this type of instability, and other types also, can be eliminated by two simple means. The first is to introduce a relatively weak longitudinal magnetic field, shown as B_z in the bottom of Figure 6. This introduces an initial spiral-like field configuration instead of the circular one of the simple pinch. When the pinch contracts, the field lines are trapped within it and contract also. Eventually, at small diameters, the small initial B_z is increased ten or more fold and acts like a strong longitudinal field, confining the plasma much as does the focussing magnetic field of a traveling-wave tube or klystron. A second stabilizing means

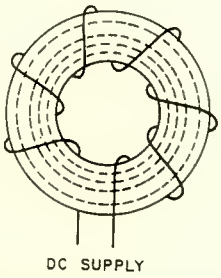


Fig. 7—A toroidal magnetic field, by itself, produces a radial gradient which leads to loss of containment.

is a conducting cylindrical wall around the plasma. By eddy-current reactions, any tendency for the plasma to move toward or away from the wall is resisted. Although these two stabilizing means are now employed in all pinch work, they have not solved all the problems, and pinch containment is still far short of expected performance.

Longitudinal-Field Containment

If the pinch improves by adding a small longitudinal field, it certainly should be possible to go all the way and provide a strong longitudinal field and, by brute force, contain the plasma. Numerically, fields of the order of more than 26,000 gausses are necessary, as seen from the containment relation of Equation (14), and containment improves as the field increases. Unfortunately, a simple toroid, as shown in Figure 7 will not do. Inherently, the field lines at smaller radii are shorter, and the effective longitudinal magnetic field is in-

versely proportional to the radius. Thus, the simple scheme fails because it has built into it a gradient which causes the particles to drift across the containment field lines.

Solutions to this problem lie in the direction of variations of tube and field geometry but are beyond the scope of this paper. Needless to say, no solution has solved all the problems, and the electron physicist, electron tube engineer, and radio engineer can all contribute new concepts to this field of gas discharge technology.

Ignition and Heating

Of all the mysteries of thermonuclear plasmas, heating and ignition are among the foremost. Excellent theoretical understanding of certain electrical properties under equilibrium conditions are completely upset by the fact that heating is a nonequilibrium process. To understand some of the difficulties, consider heating by passage of direct current through a plasma. Because electrons travel 60 times as fast as deuterium ions, they carry most of the current, and this form of heating is effective mainly on electrons, tending to raise their temperatures but not that of ions, except after many thermalizing collisions. Furthermore, as the temperatures rise, the resistance goes down and the I^2R loss decreases, since induced currents are normally limited by the plasma inductance. Thus such heating appears to be self-limiting. If a mechanism is found which heats the ions but not the electrons, the cooling effect of the electrons will be of concern.

In radio-frequency heating, peculiar skin effects occur in a plasma with long mean-free-paths and magnetic constraints on the particle paths. The radio engineer will be interested in the so-called cyclotron frequencies of charged particles in a magnetic field.[1] Quantitatively, these are (in cycles per second),

$$\text{Electron cyclotron frequency} \approx 2.9 \times 10^6 B,$$

$$\text{Deuteron cyclotron frequency} \approx 770B,$$

$$\text{Triton cyclotron frequency} \approx 510B,$$

$$\text{Alpha particle cyclotron frequency} \approx 770B.$$

For $B = 26{,}000$ gausses, these correspond to 75,000 megacycles for electrons, 20 megacycles for deuterons and alpha particles, and 13 megacycles for tritons.

It is the writer's conclusion that, at the present time, the greatest contribution can be made by the radio-electronic engineer to thermonuclear fusion by attention to this problem of heating and ignition.

High-power radio frequencies, which are adapted to the plasma characteristics, and to the ion and electron dynamics, offer great hope, both for ignition, and for help in the containment problem. If containment time is very short, very high power radio frequency may supply the rapid heating necessary to achieve useful results.

Plasma Diagnostics

Some of the study tools used in thermonuclear fusion work are listed in Table I. In a pure deuterium gas, at thermonuclear temperatures, there is no optical spectrum; optical spectra are useful only when undesired impurities are present. The X-ray diagnosis is chiefly useful for finding out whether stray high-energy electrons are striking

Table I—Plasma Diagnostics

Optical Spectra	
	Purity and composition of gas.
	Ion temperature (from Doppler broadening).
X-Rays	
	Electron confinement.
Electric and Magnetic Data	
	Electron temperature (from resistivity).
	Density × temperature (from magnetic field data).
	Current distribution (from magnetic field data).
Microwaves	
	Electron density.
Neutron Data	
	Origin (from isotropy and from variation n_e).
	Ion temperature(if thermonuclear).

metal walls, since the broad spectra of bremsstrahlung give little useful data. Electric and magnetic data are obvious but difficult since probes are contraindicated. A most useful tool is microwave measurements of both noise and transmission phase shift.[10] This work should be done at or near the so-called plasma frequency of the electrons which is

$$f_e = 9000 \sqrt{n_e},$$

where n_e is the electron density. For the typical case of $n_e = 10^{15}/\text{cm}^3$, the frequency will lie at about 3×10^5 megacycles or 1 millimeter wave-

[10] M. A. Heald, "Microwave Measurements in Controlled Fusion Research," Institute of Radio Engineers 1958 Convention Record, to be published.

length. There is a great need for improved millimeter-wave sources and technology; much work is presently limited to low plasma densities, because of deficiencies in this area.

The most useful of all diagnostic means will not be available until success is near at hand. This is the analysis of copious neutron output. A true thermonuclear discharge will have a completely isotropic energy and quantity distribution of neutrons around any given plasma volume element. Most other fusion processes, such as would be caused by localized fields, will not show this isotropy and can be detected as a spurious effect. The neutron yield, under thermonuclear conditions, will be a good measure of ion temperature. Recently reported British and Los Alamos work did not yield enough data to assure isotropy, so the temperatures calculated are suspect.

Rudiments of a Fusion Reactor

From information derived above, certain specifications for a reasonable fusion power reactor can be set forth. First, we need a gas discharge, preferably of deuterium and tritium in a 50-50 mixture. At room temperature, this will be at about 30 microns pressure, corresponding to a density of $10^{15}/cm^3$. This gas must be contained by magnetic means, such as the stabilized pinch effect, or by an externally applied magnetic field, and must be heated to about $10^{8}°K$, probably by electrical induction processes. Assuming these are all possible, Figure 8 shows the reactor in schematic form. The gaseous fuel is deuterium, the "ash" is helium. The energetic neutrons, since they are uncharged, escape and strike a moderator and absorber containing lithium-6. As we saw, the neutron–lithium reaction produces tritium, which can then be fed back into the reactor. The initial D-D reaction soon builds up as a D-T reaction and then consumes both lithium and deuterium as fuels. The heat absorbed is used in the customary way, with the Carnot cycle as a bottleneck.

In Table II, there are some calculations on this hypothetical reaction. Of particular interest is the mean collision time and the mean ion path, of 20 seconds and 12,000 miles respectively. Fortunately, so much energy is given out per collision that a working reactor is possible at a fraction of these values. However, the data shows the major requirement, i.e., long containment and a re-entrant gas discharge to provide a very long effective length.

Table III gives the calculated energy and approximate power densities for the hypothetical reactor. The energy needed to heat the plasma is 3 watt-seconds/cm^3; if this must be supplied rapidly, high power is needed, for example heating in one millisecond requires 3 kw/cm^3.

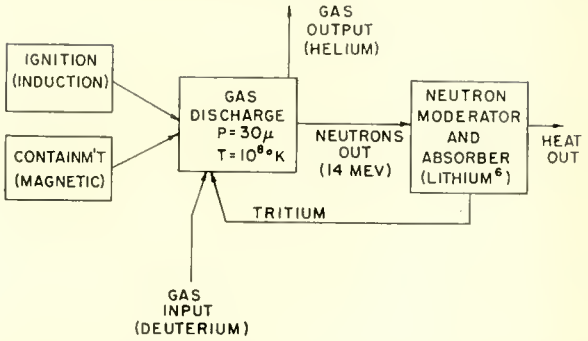


Fig. 8—The rudiments of a deuterium–tritium fusion reactor at 10^{8}°K, and $n_i = 10^{15}/\text{cm}^3$. Ignition requires 3 watt-seconds/cm³, containment requires a field of over 26,000 gausses; a neutron power output of 50 watts/cm³ is achieved.

The bremsstrahlung radiation is 2 watts/cm³, which is comfortably exceeded by the power generated in charged helium ions; namely, 10 watts/cm³. Thus, there is a margin of 8 watts/cm³ to provide continuous heating of cold gas coming in and to take care of other losses. However, since these charged alpha particles have about 3.6 mev energy, their containment for a period sufficiently long to provide reignition is not a negligible problem. The potentially useful output, then, is 50 watts/cm³ in the form of 14 mev neutrons. When these are moderated, and caused to react with lithium, additional heat is generated corresponding to 4.6 mev per neutron. The neutron flux is about 3×10^{13} per second for each cubic centimeter of the discharge, so that radiation damage to all inner parts of the reactor is to be reckoned with.

It may be erroneously assumed that the reactor could be of any size. Although the bremsstrahlung loss is independent of size, almost

Table II—Reaction at 10^{8}°K (50% Deuterium + 50% Tritium, $n_e = n_i = 10^{15}$ cm,$^{-3}$ Initial Pressure $\approx 3 \times 10^{-2}$ mm Hg)

Quantity	Symbols	Approximate Value
Reaction rate parameter	$\overline{\sigma v}$	10^{-16} cm³/sec
Mean collision time	$(n_i \overline{\sigma v})^{-1}$	20 sec
Mean ion path	$(n_i \bar{\sigma})^{-1}$	12,000 miles
Pressure (P)	$(n_e + n_i)\ kT$	27 atmosphere
Minimum containment field	$B \geqq (8\pi P)^{1/2}$	> 26,000 gausses

every other loss depends on a favorable volume-to-surface ratio, and fusion reactors will normally have a large discharge chamber, probably some hundred or more cubic feet. Scaling considerations can be deduced by assuming a fixed shape, any one of whose dimensions is D. The power for the containment field is given by I^2R where I is the necessary current and R is the resistance. The magnetic field, B, and the current are proportional according to[1]

$$I^2 \sim (BD)^2,$$

and the resistance varies as

$$R \sim \rho D^{-1},$$

Table III—Energy and Power Densities, Reaction at 10^8°K (50% Deuterium +50% Tritium, $n_e = n_i = 10^{15}$ cm,$^{-3}$ Initial Pressure $\approx 3 \times 10^{-2}$ mm Hg)

Quantity	Symbols	Approximate Value
Heating energy	$(n_i + n_e)\ kT$	3 watt-sec/cm^3
Bremsstrahlung	$Kn_e n_i Z^2 T^{1/2}$	2 watt/cm^3
Charged ions	$(n_i^2/4)\ \overline{\sigma v} \times 3.6$ mev	10 watt/cm^3
Neutrons	$(n_i^2/4)\ \overline{\sigma v} \times 14.1$ mev	50 watt/cm^3

where ρ is the resistivity of the conductor used (in the pinch containment, this is the plasma resistivity). Thus we see that

$$\text{Power for Magnetic Field} \sim I^2R \sim B^2D. \tag{15}$$

As seen from Equation (10), the total reactor power is proportional to the square of plasma density and the volume. Equation (14) showed that the minimum value for the square of the magnetic field is proportional to plasma density. Thus

$$\text{Reactor Power} \sim n_i^2\, D^3 \sim B^4\, D^3. \tag{16}$$

If we wish our reactor power to be as much as possible above the power needed for the magnetic field, the ratio of Equation (16) to (15) should be as large as possible, i.e.

$$\frac{\text{Reactor Power}}{\text{Power for Magnetic Field}} \sim B^2\, D^2 \tag{17}$$

should be large. Thus, large fields and large dimensions are favorable, and we may compare the merits of two reactor designs on the basis of $B^2 D^2$.

There are other power losses besides those for the magnetic field; in particular there are those due to heated (unburned) ions and electrons which diffuse away, or are swept away with the helium ash, or are lost because of other failures in containment. Depending on the assumptions made, merit factors for these other losses can be derived; qualitatively they show similar results to those of Equation (17), i.e., a large size, and a large field, are favorable.

Major Future Problems

Table IV lists a few of the major future problems, and the areas in which the electronics industry can contribute. In containment, one would like a major invention to reduce the power required for a steady magnetic field of perhaps 10^4 to 10^5 gausses over a volume of many

Table IV—Major Future Problems

- Containment
  - Steady magnetic fields at lower power
  - A-C methods
- Heating of plasma by induction (r-f methods)
- Neutron-resistant materials
- Conversion of fusion power to electricity
  - The charged particles
  - The neutrons
- Diagnostics
  - MM-wave sources to measure n_e
  - New means to measure ion temperature
  - New means to measure internal fields
- Transmission of power

cubic feet. Even better would be a containment method which used the hydrodynamics of the plasma as a key to the use of alternating-field containment. Initial ignition is a major unsolved problem, requiring many megawatts to get the plasma fired up in a short enough time. Although this is needed only initially, like a pilot light on an oil burner or gas range, it is likely to be operated continuously to avoid shut-down.

Also needed are radiation-resistant materials as well as ideas on

the direct conversion of fusion power to electricity. Particularly challenging would be some method which does this by utilizing the neutron directly, preferably without going through the heat cycle. Plasma diagnostics is a fringe area, but of importance to the radio engineer, since the best techniques will undoubtedly come out of microwave technology and physical electronics. Finally, just a word on transmission of power. A fusion reactor, such as the rudimentary one discussed above, will produce a total power of several megawatts per cubic foot of plasma. Since the cost, size, and weight of fuel are all negligible, the economic factors favor large installations, perhaps exceeding the largest now known for electric power. A typical example might be a million kilowatts from a single generator, using a little over ½ pound of fuel (lithium and deuterium combined) per day. The major cost of useful power may be the transmission system. Much will have to be done by the electrical engineering profession to improve wire-line transmission. It is possible, however, that the era of fusion power may make this power so cheap and plentiful that we can afford to explore completely new transmission methods, perhaps using induction fields over short distances, or electromagnetic radiation, over longer distances. This day, should it come, may see our earth less dependent on the sun whose inefficient but powerful radiant transmission system has been serving us so well for the past billion years.

Evaluation of Recent Work

Early in 1958, publicity was given to U. S. and British results on some stabilized-pinch deuterium plasma experiments in which neutrons were produced.[11–18] Although these neutrons were without question the

---

[11] *New York Times*, pp. 1 and 8-10, January 25, 1958.

[12] P. C. Thonemann, et al, "Production of High Temperatures and Nuclear Reactions in a Gas Discharge," *Nature*, Vol. 181, p. 217, January 25, 1958.

[13] L. Spitzer, Jr., "Cooperative Phenomena in Hot Plasmas," *Nature*, Vol. 181, p. 221, January 25, 1958.

[14] N. L. Allen, et al, "A Stabilized High-Current Toroidal Discharge Producing High-Temperatures," *Nature*, Vol. 181, p. 22, January 25, 1958.

[15] L. C. Burkhardt, R. H. Lovberg, and J. A. Phillips, "Magnetic Measurement of Plasma Confinement in a Partially Stabilized Linear Pinch," *Nature*, Vol. 181, p. 224, January 25, 1958.

[16] D. C. Hagerman and J. W. Mather, "Neutron Production in a High-Power Pinch Apparatus," *Nature*, Vol. 181, p. 226, January 25, 1958.

[17] L. C. Burkhardt and R. H. Lovberg, "New Confinement Phenomena and Neutron Production in a Linear Stabilized Pinch," *Nature*, Vol. 181, p. 228, January 25, 1958.

[18] J. Honsaker, et al, "Neutrons from a Stabilized Toroidal Pinch," *Nature*, Vol. 181, p. 231, January 25, 1958.

result of deuterium fusion, it is by no means certain that they were all of thermonuclear origin, i.e., that they came from a gas plasma completely free of selective acceleration of particular groups of ions. As background, it must be remembered that early Russian[19] and U. S.[20] pinch discharges gave substantial neutron production. Kurchatov reported that the time of maximum neutron yield indicated that the neutrons were probably the result of deuterons selectively accelerated by one of the pinch contractions, rather than a thermonuclear process. In the U. S. work, a definite anistropy of neutron energy was found, again indicating non-thermonuclear fusions. Such processes will not

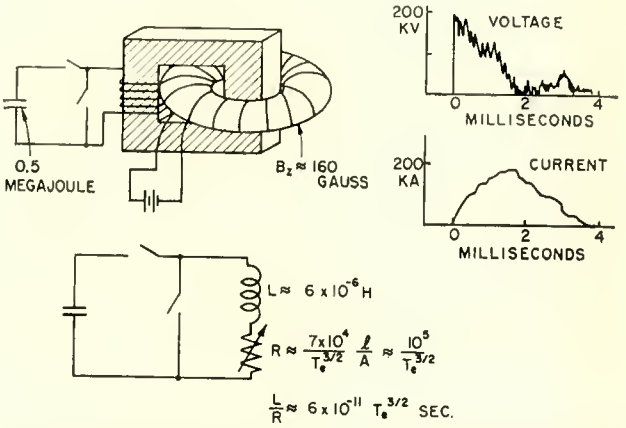


Fig. 9—The British "Zeta" experiment, as reported January 25, 1958. The apparatus and an equivalent circuit are shown, as well as the discharge current and voltage.

produce a favorable power balance; in this respect they resemble the straightforward experiment of measuring cross section with a deuteron beam against a gas target.[8]

Figure 9 is a redrawn diagram of the British "Zeta" pinch machine,[12] which received the greatest attention in the January release. A 3-foot diameter aluminum tube, 30 feet long, toroidal in shape, was filled with deuterium gas at very low pressure and a weak longitudinal magnetic field applied. This discharge was "pinched" by using the plasma as the one-turn secondary of a transformer whose primary is excited by closing a switch to a capacitor bank. At the moment when

---

[19] I. V. Kurchatov, "Russian Thermonuclear Experiments," *Nucleonics*, Vol. 14, p. 36-43 and p. 123, June, 1956.

[20] O. A. Anderson, et al, "Neutron Production in Linear Deuterium Pinches," *Phys. Rev.*, Vol. 109, p. 612, January 15, 1958.

the energy is transferred to inductive storage in the plasma, i.e., at the voltage zero, a shorting switch is closed and the plasma current allowed to decay. Ideally, the decay time should depend on the L/R time constant which depends chiefly on the plasma resistivity. Assuming a "pinched" plasma diameter of 12 inches, the inductance is easily found from a standard formula[21] and is about 6 microhenries. Using the theoretical plasma formula, Equation (12) above, the resistivity is approximately $10^5\, T_e^{-3/2}$, so that the time constant is

$$\frac{L}{R} \approx 6 \times 10^{-11}\, T_e^{3/2} \text{ seconds.}$$

If the electron temperature averaged around $10^{6\circ}$K, L/R would be about 60 milliseconds. The experiments showed a time decay some 20 times

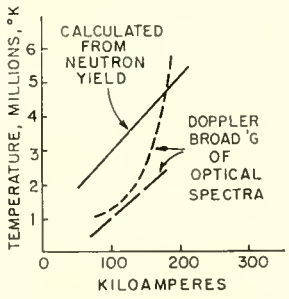


Fig. 10—Ion temperatures of the "Zeta" discharge, assuming thermonuclear neutrons. The optical spectra used were those of impurities, O-V, and N-IV.

shorter, i.e., 3 milliseconds (Figure 9). One suspects, therefore, that the electron temperatures were not so high, on the average, perhaps because of loss of containment.

Spitzer[13] calculated that, *even if the electron temperature had been high* in the Zeta machine, the 3-4 millisecond elapsed time was not long enough for the electrons to come to equilibrium with the ions and heat them. However, the British neutron data (if thermonuclear conditions are assumed) indicated that ion temperatures perhaps as high as $5 \times 10^{6\circ}$K were attained, as shown in Figure 10. Doppler line-broadening of one of two impurity gases also indicated such a temperature. It would appear that the electron temperatures are well below this figure, and the ion-heating mechanism is unexplained. Also unexplained are the relatively high neutron yields at low currents, and the absence of the anticipated cooling effect of impurity gases, when they were

[21] F. F. Terman, *Radio Engineer's Handbook*, First Edition, McGraw-Hill Book Co., New York, 1943, p. 52.

introduced in small percentages. But, above all, conclusive data on neutron energy isotropy could not be obtained, so the thermonuclear nature of the yield remains in doubt.[22]

In comparing the Zeta work with that of other toroidal pinch machines[14,18] a calculation shows the expected L/R time constant varies as the square of the torus size. One expects the ratio of time constants, for the three experiments called Zeta,[7] Sceptre III,[14] and Perhapsatron[18] would then be as 1000:100:2. The published data shows excellent agreement, with milliseconds in the Zeta, tenths of milliseconds in the Sceptre III, and microseconds in the Perhapsatron. All three gave neutrons corresponding to thermonuclear temperatures ranging from 3 to 5×10^6°K. One may justly conclude that the three gave equally meaningful data, differing only in time scale, as determined by the dimensions. All three gave temperatures and containment times which are equally far below those desired.

Conclusion

Controlled thermonuclear fusion in plasma discharges offers hope of a revolutionary new energy source. However, major containment and heating problems are an obstacle, and achieved temperature and containment times are factors of a hundred below objectives. It is probable that these obstacles will be overcome, in the laboratory, within 5 to 10 years, but practical utilization may require a period of several decades beyond this. The radio and electronic engineer has an opportunity to make major contributions, perhaps even to shorten the time appreciably.

---

[22] Added in proof: *The New York Times* of May 17, 1958, on page 5, reported that the British scientists had found anisotropy in neutron yield and concluded that the neutron origin was not thermonuclear.

COMMERCIAL AIRBORNE WEATHER RADAR*

By

A. W. Vose and F. V. Wilson

RCA Defense Electronic Products,
Los Angeles, Cal.

Summary—*This paper discusses factors relating to the detection and penetration of atmospheric precipitation at microwavelengths. Technical design features of both 5.6-centimeter weather penetration and 3.2-centimeter weather avoidance radars used in present-day commercial aircraft are discussed.*

Introduction

ELECTRONIC equipment is essential in the airline industry in both communication and navigation. Commercial transport aircraft, as well as private and business aircraft, depend daily on its use to maintain ground contact along the airways for safe and expeditious flight. ADF, VOR, HF, Loran, and many other equipments have long since proved their worth in the vast electronics network that covers the airways the world over.

In spite of these time-proven aids, unpredictable weather conditions along the airways have long been a serious hazard to the pilot trying to maintain a safe and on-schedule flight to his destination. Under such conditions, ground contact provides only general weather information usually derived some hours before or, at the best, relayed reports from preceding aircraft flying through the same area. The decision whether to go around or through a storm front will, in any case, affect the operation of the airline since a wrong decision could result in passenger discomfort, schedule delays, and even structural damage to the aircraft. The display of weather echoes by radar was discovered as a by-product of military equipments designed for bombing and navigation purposes near the start of World War II. Since local storm conditions aloft can change radically in a matter of minutes, or even seconds, it was apparent that these devices could be used to provide continuous up-to-the-minute indication of changing weather conditions during flight.

However, it was not until 1952 that firm operational requirements were established by the Air Transport Association. The most critical of these requirements was the one pertaining to attenuation of the radar beam by heavy rain, namely,

* Manuscript received February 10, 1958.

"The equipment must be capable of penetrating and displaying at short range rainfall rates of 60 millimeters per hour to a depth of 15 miles."

Today, major airlines throughout the world use airborne radar to locate turbulent areas in thunderstorms, and pilots rely on it to guide them safely through storm areas with a minimum amount of lost time and passenger discomfort. The AVQ-10 and AVQ-50 are two such radars; the first, being designed for airline use in weather penetration and avoidance, uses an operating wavelength of 5.6 centimeters, and the second, designed for business and private aircraft use in weather avoidance, uses an operating wavelength of 3.2 centimeters.

Rainfall Detection

The basic problem in radar storm detection is the absorption and scattering of a plane electromagnetic wave by a sphere. This problem was first studied by Mie[1] who analyzed the absorption and scattering of light from gold particles suspended in a liquid. Application to the theory of the detection of precipitation was first made by Ryde.[2] He predicted the back-scatter and attenuation by meteorological phenomena, and, in almost every respect, the results obtained in later experiments have confirmed the predictions.

The radar signal from precipitation is not steady as is that from a point target. The received power at any instant is made up of the resultant of the individual signals from a very large number of particles contained in a volume of space determined by the angular and range resolution of the radar set. The exact arrangement in space of the precipitation particles is being continually changed by the turbulent motion of the air, resulting in a random spatial distribution and so-called incoherent scattering. Precipitation echoes would be expected to fluctuate because of the changing interference effects among the randomly moving scatterers. Indeed, the rapid fluctuation of meteorological echoes is so striking as to provide one of the easiest ways to identify them when viewed on an A-scope. Airborne weather radars however usually employ PPI type indicators, and other characteristics of the precipitation echo are therefore used to identify them. We shall not deal further in this section with target identification and scope interpretation, but will proceed to define and discuss the radar equations applicable to rainfall detection and penetration,

[1] G. Mie, "Beitroge zur Optik truber Medien," *Ann. de Physik*, Vol. 25, p. 377, March, 1908.

[2] J. W. Ryde, "The Attenuation of Centimetre Radio Waves and Echo Intensities Resulting from Atmospheric Phenomena," *Jour. Inst. Elec. Engrs.*, Vol. 93, Part 3A, p. 101, 1946.

and to illustrate the use of these equations in determining weather radar design parameters.

The equation of rainfall detection is conveniently derived from the standard radar equation as follows: assuming a conventional pulsed-type radar,

$$P_r = \frac{P_t G}{4\pi r^2} \frac{\sigma}{4\pi r^2} A_e , \tag{1}$$

where P_r is the average received power,
P_t is the transmitted peak power,
G is the antenna gain factor,
σ is the radar scattering cross section of all the precipitation particles contributing to the return echo,
A_e is the effective antenna receiving area,
r is the radar target range.

Any self consistent set of units may be used.

In the interest of obtaining a more general and useful equation for weather detection purposes, the following additional radar and meteorological terms are defined:

τ = time duration of transmitted pulse,
c = velocity of light,
h = pulse length in space ($h = c\tau$),
ϕ = vertical radar beam width between half-power points,
θ = horizontal radar beam width between half-power points,
V = volume of space illuminated by the pulse which can contribute to the instantaneous received rain echo,
$\sum_i n_i\sigma_i$ = scattering cross section per unit volume of space due to n_i particles of cross section σ_i each within this unit volume.

Now, assuming that the precipitation completely fills the beam, the volume, V, of scatterers that can contribute to the instantaneous received echo may be shown[3] to be approximately

$$V = \theta\phi r^2 \frac{\tau c}{2} , \tag{2}$$

and since $\sum_i n_i\sigma_i$ is defined to be the scattering cross section per unit

[3] D. E. Kerr, "Propagation of Short Radio Waves," *MIT Radiation Laboratory Series*, Vol. 13, p. 588, 1951.

volume, it follows that if the precipitation is uniform within the beam, then the total scattering cross section, σ, can be written as

$$\sigma = V \sum_i n_i \sigma_i = \theta \phi r^2 \frac{\tau c}{2} \sum_i n_i \sigma_i . \tag{3}$$

Substitution of Equation (3) into Equation (1) gives

$$P_r = \frac{P_t G}{4\pi r^2} \cdot \frac{\theta \phi}{4\pi} \cdot \frac{\tau c}{2} \sum_i n_i \sigma_i A_e . \tag{4}$$

Also it may be shown[3] that, approximately,

$$\frac{\theta \phi}{4\pi} = \frac{1}{G} . \tag{5}$$

Inserting Equation (5) into Equation (4) gives

$$P_r = \frac{P_t}{4\pi r^2} \cdot \frac{\tau c}{2} \cdot \sum_i n_i \sigma_i A_e , \tag{6}$$

and to account for atmospheric and rain attenuation there must be added a factor $10^{-0.2\left[\int_R^r K_R dr + \int_0^r K_a dr\right]}$ where K_R is the rain attenuation in decibels per unit distance (one way), and K_a is the atmospheric attenuation in decibels per unit distance (one way).

The equation for storm detection, including the effect of attenuation, and allowing that only a fraction (ψ) of the beam may be filled is

$$P_r = \frac{P_t}{4\pi r^2} \frac{\tau c}{2} \sum_i n_i \sigma_i A_e \psi \, 10^{-0.2\left[\int_R^r K_R dr + \int_0^r K_a dr\right]} \tag{7}$$

Clearly, there are several assumptions implicit in Equation (7), and it is perhaps worth while to point them out along with some precautions in use of this equation.

(1) The beam widths θ and ϕ are conventionally defined here as the half-power beam widths. When using these widths it must be understood that some power is actually present beyond these angles; however, for most purposes this definition is probably sufficient.

(2) The supposition of uniform precipitation throughout the beam is generally invalid in reality. However the complexity of the precipitation distribution is generally such as to make this assumption necessary in order to arrive at a satisfactory answer.

(3) When considering cases where the precipitation partially fills the beam, it should be remembered that a storm of given size will occupy only a fractional part of the beam solid angle, and this fraction will vary inversely as the square of range.

(4) When using Equation (7), certain loss factors inherent in equipment design or necessitated by installation should be included. Examples are wave guide attenuation and radome losses.

(5) P_r, the lower limit of power which the receiver can detect, is limited by the receiver sensitivity. In microwave regions this limit is not set by atmospheric static, but rather by the electron or thermal "noise" which originates in the receiver components. If the power for the signal which may just be distinguished from noise is used in Equation (7), then the equation may be solved for maximum range at which a given storm may be detected. It should be remembered that P_r is the average returned signal power, and it should be emphasized that detection of the fluctuating return signal in noise is a statistical process, dependent on both the signal-to-noise power ratio, and the nature of the detection apparatus employed. For a PPI type of display a signal to noise power ratio of 3 decibels is usually considered adequate for reliable detection. The receiver noise power, N, with which the signal must compete, is readily determined[4] from the receiver bandwidth β, the over-all receiver noise figure $\overline{NF}$ and the thermal equilibrium temperature, T, as

$$N = \overline{NF}\, kT\beta, \tag{8}$$

where k is Boltzmann's constant (1.38×10^{-23} joules/degree). The minimum detectable signal $P_{r\min}$ is conveniently expressed as

$$P_{r\min} = \left(\frac{P_r}{N} \right)_{\substack{\min \\ \text{det.}}} \overline{NF}\, kT\beta. \tag{9}$$

[4] L. N. Ridenour, "Radar System Engineering," *MIT Radiation Laboratory Series*, Vol. 1, p. 28, 1947.

Before Equation (7) can be applied to storm detection, it will be necessary to define the scattering cross section and the attenuation as a function of rainfall rate and wavelength. The fundamental laws of scattering developed by Mie[1] and Rayleigh[5], and later applied by Ryde[2] to the special case of scattering and diffraction by rain drops, assert that a spherical particle, which is small relative to the wave length of radiation falling upon it, scatters the energy in proportion to the sixth power of its diameter. Reflectivity per unit volume turns out to be a function of ND^6, where N is the number of drops per unit volume, and D is their diameter. However, when the drops become large compared to the wave length, ($D > \lambda/10$) "Rayleigh scattering" no longer applies, and the exact dependence of scattering cross section upon λ is a complex function. The composition of the particle affects its dielectric properties, and therefore its efficiency as a scatterer; water spheres, for example, will scatter about five times as much radiation as an equivalent size ice sphere. Shape and aspect are also important for nonspherical particles. If a rain drop is spherical its back-scatter cross section will be approximately

$$\sigma = \frac{\pi^5 D^6}{\lambda^4} \left(\frac{m^2 - 1}{m^2 + 2} \right)^2, \tag{10}$$

where D = drop diameter

λ = wave length

m = complex refractive index of water.

If the rainfall intensity is known, and one assumes a drop-size distribution in accordance with that found by Laws and Parsons[6], it is possible to compute the reflectivity per unit volume. Distributions for rain at the ground have been determined in various parts of the world by many investigators, and the reflectivity per unit volume computed and related to the rainfall rate. One method is to find the sum of the sixth powers of the drop diameters per unit volume and substitute this value into Equation (10). Marshall and Hitschfeld[7] report the following expression

$$\sum D^6 = 2.0\, R^{1.6}\, 10^{-10}\ \text{cm}^6\ \text{cm}^{-3}, \tag{11}$$

[5] Lord, Rayleigh, "On the Light from the Sky, Its Polarization and Colour," *Phil. Mag.*, Vol. 41, p. 107, p. 274, 1871.

[6] J. O. Laws and D. A. Parsons, "The Relation of Raindrop-Size to Intensity," *Trans. Amer. Geophys. Un.*, Vol. 24, p. 453, 1943.

[7] J. S. Marshall and W. Hitschfeld, "Calculated Sensitivity of Airborne Weather Radars," *Air Force Surveys in Geophysics*, Air Force Cambridge Research Center, No. 23, December, 1952.

where D = drop diameter,

R = rainfall rate in millimeters per hour.

Combining Equations (10) and (11) and inserting into Equation (7) gives

$$P_r = \frac{P_t}{4\pi r^2} \frac{\tau c}{2} \frac{\pi^5}{\lambda^4} (.93)\,(2.0\, R^{1.6} \times 10^{-10})\, A_e \psi 10^{-0.2\left[\int_R^r K_R dr + \int_0^r K_a dr\right]} \tag{12}$$

where all units of length are in centimeters and $[(m^2 - 1)/(m^2 + 2)]^2$ has been set equal to 0.93 (a value nearly independent of wavelength). Numerical evaluation of the constants gives

$$P_r = 2.26 \times 10^{-9} \frac{P_t}{r^2} \frac{\tau c}{\lambda^4} R^{1.6} A_e \psi F\, 10^{-0.2\left[\int_R^r K_R dr + \int_0^r K_a dr\right]} \tag{13}$$

According to Marshall and Hitschfeld, radar rain attenuation may be accurately expressed by $K_R = K'R^{\alpha}$ = rain attenuation in decibels per mile (one way) where K' and α are functions of wavelength and R is in millimeters per hour. For those frequencies most generally considered suitable for airborne weather radar applications (X and C band), Reference (7) gives data from which the following table was constructed:

Table I

λ	3.2	5.7
K'	0.0144	0.0047
α	1.3	1.1

The attenuation by the atmosphere is chiefly due to oxygen and water vapor, and varies with temperature, pressure, and humidity. This attenuation is relatively small even at sea level, and decreases with altitude so that only representative figures corresponding closely to standard temperature, pressure, and water vapor content need be used. Table II shows the values of atmospheric attenuation in decibels per mile (one way) used in Reference (7) for X and C band.

Table II

λ	3.2 cm	5.7 cm
K_a	0.027 db/mi	0.018 db/mi

One factor of considerable interest to the theory of rain detection remains to be discussed before Equation (13) is used to predict performance. For reasons which are to date unknown, the most accurate experiments performed to confirm Equation (13) have quite consistently yielded a return power approximately 4 to 7 decibels below the predicted values for both rain and snow. Following the notation of Marshall and Hitschfeld[7] there has been inserted into Equation (13) the factor F, to account for the discrepancy between theory and experiment. Numerical choice of F is open to the discretion of the user between the limits shown above. However, in the interest of maintaining high penetration and detection capabilities throughout the service life of the radar, a conservative selection of 6 to 7 decibels would be the logical choice for design considerations.

Penetration

Penetration is the ability of a radar to look through extended rainfall areas and display the meteorological precipitation patterns without distortion. An airborne weather radar designed for navigation through storm areas must be able to display the weather picture to the user with sufficient penetration range to insure that nearby precipitation does not mask or distort more distant precipitation. In this matter wavelength becomes an important consideration. Decreasing the wavelength would increase both the angular resolution and the detectability of a given storm were it not for an accompanying increase in the attenuation by rain. The attenuation works against the resolution by introducing distortion, nearby precipitation patterns casting shadows on more distant patterns. It works powerfully against the sensitivity, since at shorter wavelengths the detectability of a target rain can depend more on the intervening rain than it does on the distance to the target rain. On the other hand, weather avoidance radars, if used principally to avoid weather penetration, would not normally suffer in the same degree from attenuation due to intervening rain. Used under these conditions the shorter wavelengths offer distinct advantages in increased sensitivity and resolution, and, as was done in AVQ-50, the increased resolution can be traded for smaller and lighter equipment while retaining the angular resolution of a longer wavelength system employing a larger antenna system.

A comparison of X and C band detection and penetration capabilities for typical airborne weather radars is shown in Figure 1. The ordinate shows, for curves X and C, the amount of allowable intervening rain attenuation in decibels for detection of a 10 millimeters per hour target rain at the ranges shown on the abscissa. The break

in the two curves indicates the approximate ranges at which a target rain cloud of 10 square nautical miles will just fill the radar beam. For ranges below these points the received power varies as the inverse square of the range, and beyond the break points varies either as the

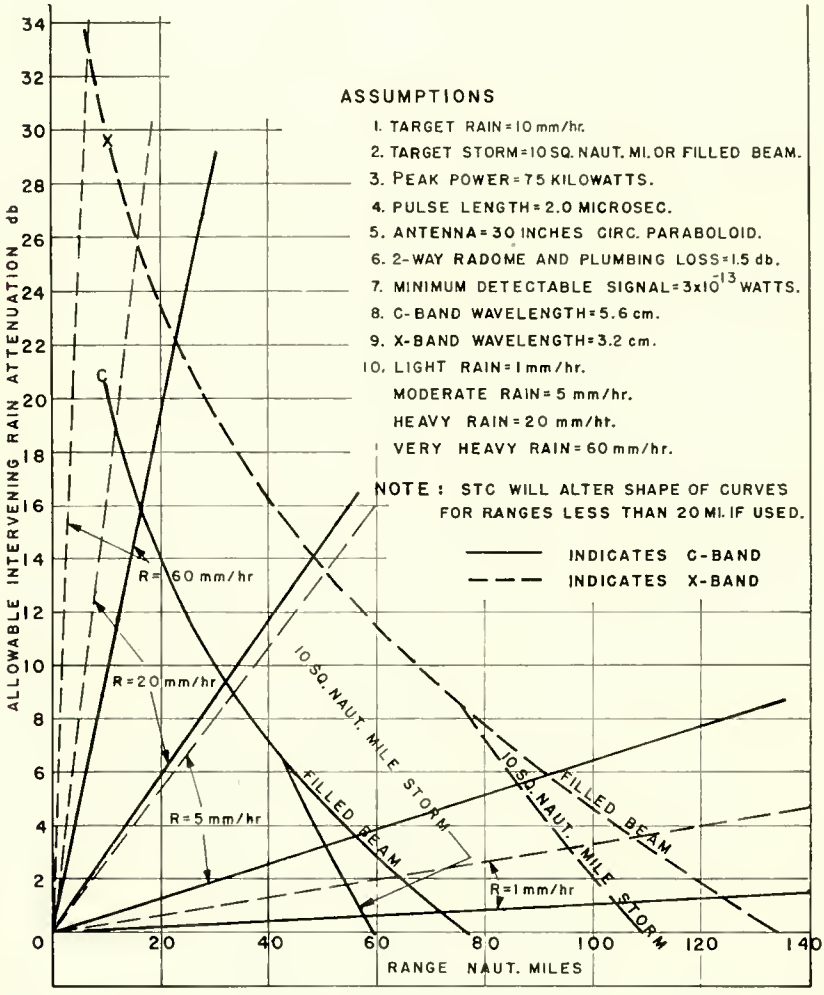


Fig. 1—Weather radar storm detection and penetration capabilities for X-band and C-band radars.

inverse square or fourth power depending on whether or not the beam is filled. The vertical separation between the X and C curves indicate that the sensitivity of the X-band set, in absence of rain attenuation, is everywhere superior to the C-band set. This situation is drastically

changed by the presence of moderate to heavy intervening rain. The straight lines originating from the origin indicate the rain attenuation versus range for both X and C band for typically light, moderate, heavy, and very heavy rainfall. The intersections of the rain attenuation lines with the allowable attenuation curves define the limits of penetration when all the intervening space between the radar and the target rain is filled with precipitation. Thus, for example, when looking through 60 millimeters per hour rain, the C-band set can detect a 10 millimeters per hour rain at approximately 16½ miles, while the X-band performance under similar conditions is reduced to about 7 miles. The foregoing example assumes that sensitivity-time control (STC) is not employed. Since this is not representative of normal operation, it can be expected that further range reduction and distortion will be apparent in the X band set. Numerous other examples of penetration ranges for various combinations of intervening rain may be determined by finding the integrated rain attenuation and correlating this with the appropriate allowable rain attenuation curve. The important difference to note between the two wave lengths is the very marked increase in X band attenuation with rainfall rate.

Another design consideration of importance in achieving adequate penetration capabilities at short range is the sensitivity time control that preserves the proper echo signal levels with range, and allows the use of iso-echo technique to display rainfall contours and gradients. Sensitivity time controls of various forms of circuitry are well known and have been used with radars for a number of years. The importance of STC and iso-echo, and some of their design features as used in the AVQ-10 and AVQ-50, are discussed later.

Magnetron Development

With 5.6 centimeters established by airline requirements as the optimum wavelength for weather detection and penetration, a suitable power source was required. As a result, the RCA-6521 "weather radar" magnetron was developed. This tube fills the need for an extremely reliable, long-life magnetron where size and weight are of secondary importance. Figure 2 shows the 6521 magnetron.

The tube design engineers believed that the greatest deterrent to good magnetron life was the operation of the cathode at excessive current density. In the design of the 6521, the interaction space was scaled from the RCA-4J50, a 3 centimeter magnetron. Since the number of resonators was not changed, the 6521 has larger anode and cathode diameters, which, in turn, permit lower cathode-current and anode dissipation density. Figure 3 shows a cutaway view of the 6521.

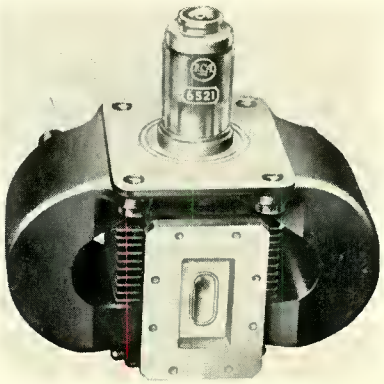


Fig. 2—6521 C-band magnetron.

The 6521 uses a vane-type anode which contributes to a simple rugged unit. Cooling is accomplished by forced-air flow over the external anode radiator fins. Because the air-flow system might fail during some critical period of radar operation, the anode structure is rather massive. This larger size insures that a magnetron "catastrophic failure" will be unlikely, even if the flow of cooling air should be lacking for average flight durations. The cathode support structure is also massive to minimize the deteriorating effects of electron "back-bombardment" heating of the cathode. Three getters, integrally mounted in the cathode assembly, maintain a very high vacuum during

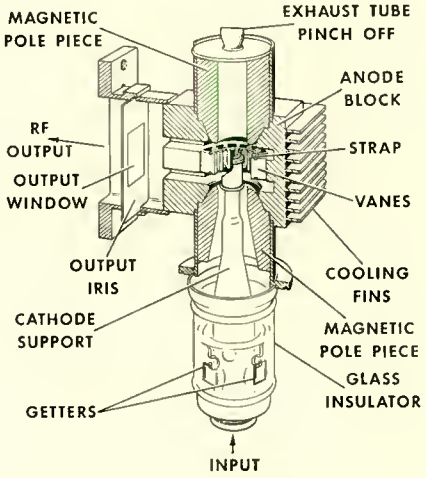


Fig. 3—Cutaway view of the 6521 magnetron.

both shelf storage and operation of the 6521. This aspect is of importance in a field-stocked, long-life tube, because spare-tube shelf-storage time may be considerable. A double-helical heater is used to prevent electromechanical resonances and heater modulation. The heater design is also important to long life and greater reliability of performance.

Good magnetron mode stability is important to the AVQ-10 Weather Radar System. If the magnetron should operate in an improper mode during a pulse, no echo information could possibly be presented on the pilot's PPI scope, because the magnetron output would be at the wrong frequency. As a result, the pilot would observe a "missing-line" from what otherwise might be a solid mass of storm-area information. Although the amount of target information missed would be quite small, the sporadic lack of lines on the PPI screen could be disturbing

Table III — Typical Operating Conditions for the 6521

Heater Voltage	9.5 volts
Peak Anode Voltage	15 kilovolts
Peak Anode Current	13.5 amperes
Pulse Repetition Rate	400 pps
Duty Factor	0.0008
Pulse Width	2 μ sec
Frequency	5400 $\pm$20 mc
R-F Bandwidth (Maximum)	1.5 mc
Peak Power Output	85 kilowatts
Pulling Figure	5.5 mc
Missing Pulses (Maximum)	0.25 percent
Efficiency	50 percent
Voltage-Standing-Wave Ratio (Maximum)	1.5
Thermal Factor	0.15 mc/°C

to the pilot. It was found that the mode stability of the 6521 was greatly influenced by the relative electrical position of the radar duplexer. As a result, experiments were performed to determine the best position for the duplexer, i.e., the position in which the unwanted modes have the least tendency to build up oscillations. Because the electrical characteristics of the 6521 are substantially the same from tube to tube, incorporation of the optimum duplexer position in the radar equipment insures best mode stability in the original 6521 supplied with the equipment, and also in subsequent replacement tubes.

The results of coordinated design and application engineering, together with close liaison with the equipment designer, were highly profitable. In daily airline use, 6521 tubes normally operate for more than 2,000 hours under severe conditions before failure. Measurements have been made of variations during life of several of the more

important parameters of the 6521. It is interesting to note that mode stability remained essentially constant throughout life, the over-all change in frequency being less than 1.5 megacycles. The tubes were still in satisfactory condition when removed from test after more than 2,000 hours.

The final result was a packaged 12-pound tube that delivers a peak pulse power output of 85 kilowatts at a frequency of 5,400 megacycles. Table III gives typical operating conditions for the 6521. The Rieke

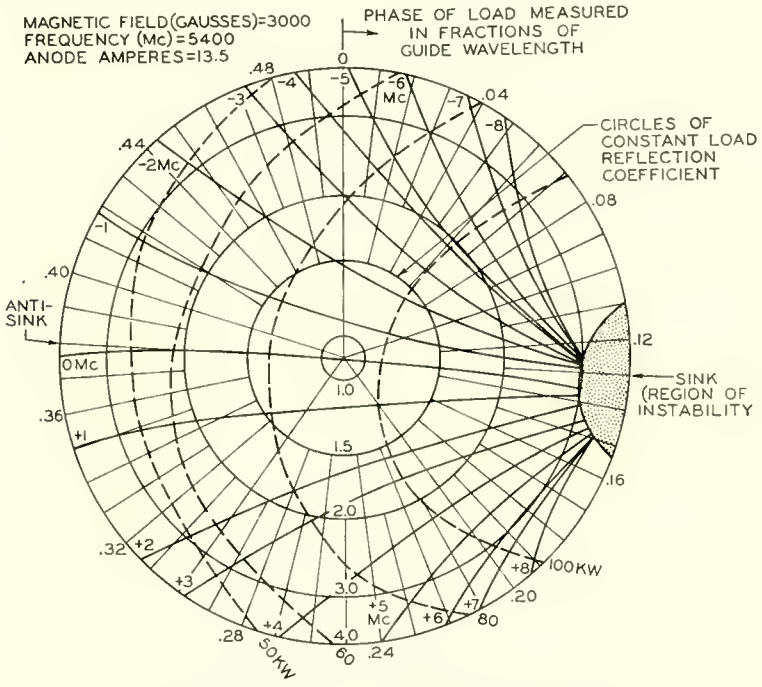


Fig. 4—Rieke diagram for the 6521 magnetron.

diagram shown in Figure 4 relates power and frequency as the load is varied with constant input conditions. Very extensive testing has been done with regard to the voltage-standing-wave ratio into which the tube operates. Although operation is specified to be within the circle representing a mismatch of 1.5 to 1, no appreciable deterioration of the weather information mapped by the system is evident when the tube operates into a very substantial reflective load. This feature is of great importance in airline service because reliability of safety equipment under unusual conditions is absolutely imperative.

Equipment Design Considerations

Airline requirements for weather radar make necessary new design concepts not encountered in previous airborne radar equipment. Reliability is of primary importance. However, the restrictions imposed on size and weight by aircraft space limitations make necessary design compromises to meet this most important of all requirements. As a

Table IV — General Characteristics of the AVQ-10 and AVQ-50

	AVQ-10	*AVQ-50*
System Complement	5 major units	4 major units (plus controls kit)
Over-all Weight	125 pounds	50 pounds
Primary Power	850 VA at 400 cycles 30 watts at 27.5 VDC	500 VA at 400 cycles
Type Presentation	360° PPI	80° PPI
Operating Altitude	16,000 feet (antenna 45,000 ft.)	16,000 feet (antenna 40,000 ft.)
Operating Frequency	5400 ± 30 megacycles	9375 ± 40 megacycles
Power Output	75 kw peak	40 kw peak
Pulse Width	1.80 microseconds	1.80 microseconds
Repetition Rate	400 cycles (Synchronized to line frequency)	400 cycles (Synchronized to line frequency)
Ranges	20-50-150 nautical miles	20-80 nautical miles
Antenna Stabilization	line-of-sight, pitch and roll	not provided
Antenna Scan Rate	15 rpm, continuous	85 "looks" per minute, sector
Antenna Beam Width	7° pencil beam with 22" reflector	7° pencil beam with 12" reflector
PPI Indicator	5" normal P7 or Display Storage Tube	3" or 5" normal P7 display
System Noise Figure	12 decibels	12.5 decibels
Receiver Bandwidth	1 megacycle at 3-db points	1 megacycle at 3-db points
Iso and STC	Iso-Contour and STC circuitry provided	Iso-Contour and STC circuitry provided

result, many of the "frills" found in previous equipment have been eliminated by the use of simplified circuits using a bare minimum of vacuum tubes. Vacuum tubes have long been recognized by the airlines as the greatest single source of equipment failure in operational service. Results obtained have been most gratifying, and weather

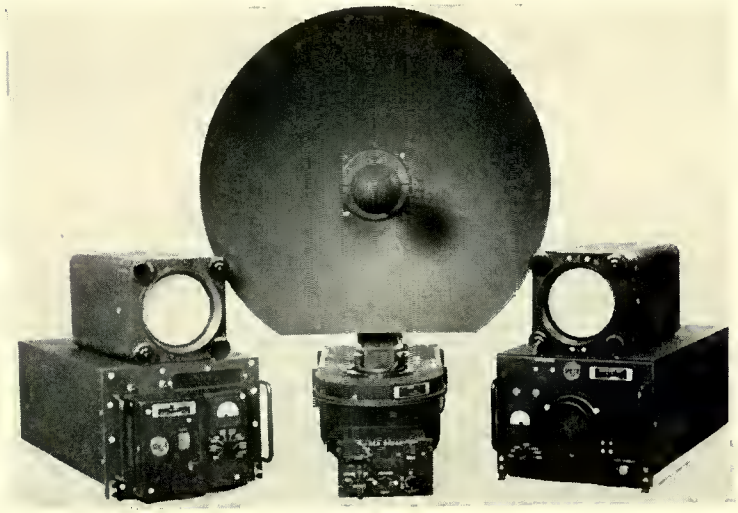


Fig. 5—Complete AVQ-10 weather radar system (2-indicator installation).

radar has already proved to be one of the most reliable pieces of electronic equipment installed in commercial aircraft today. Table IV gives a comparison of the general characteristics of the AVQ-10 and AVQ-50 Weather Radars. Figures 5-7 are photographs of the AVQ-10 equipment.

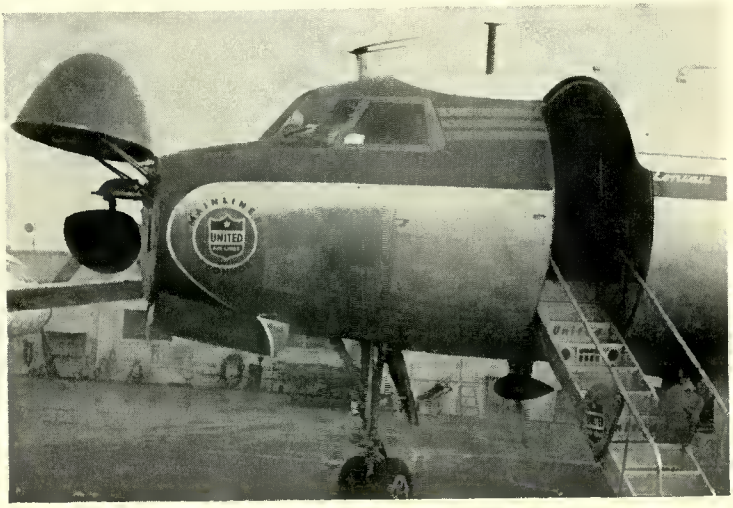


Fig. 6—Typical AVQ-10 antenna installation in a Convair 340. (Courtesy United Airlines.)

Iso-Echo Contour Circuits

All commercial airborne weather radars use some form of iso-echo contour operation. The purpose of iso-echo is to provide the radar operator with a better means of determining the high-turbulence areas of a storm front than that provided by a normal PPI display.

Past experience with airborne weather radar has shown a definite correlation between turbulence and sharp-edged radar echoes, i.e., sharp rainfall gradient (sharp edged echoes) is associated with severe turbulence. While an experienced operator can usually distinguish a less-dangerous storm target by its soft, fuzzy, appearance on the

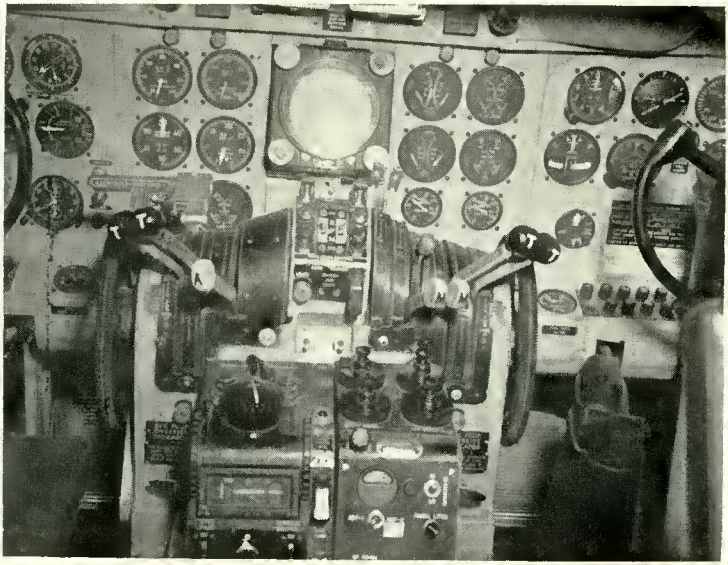


Fig. 7—Typical AVQ-10 indicator installation in a Convair 340. (Courtesy United Airlines.)

cathode-ray-tube screen, the "contour" function provides a more definite means of selection. This is done by showing two contours, one corresponding to the minimum discernible signal level, and the other corresponding to a predetermined signal amplitude above this level. The first contour is defined by the outline of the target on the cathode-ray tube while the second corresponds to the edge of the "hole" left when signals above the predetermined level are inverted. The separation between contours indicates the rainfall gradient, the less the separation the greater the rainfall gradient. Figure 8 illustrates the use of "contour" operation on a typical storm display.

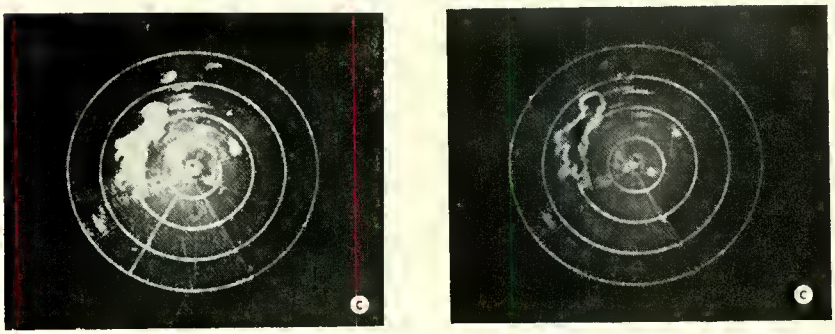


Fig. 8—Storm display on AVQ-10 indicator showing normal and contour presentation. (Courtesy United Airlines.)

The AVQ-10 iso-echo circuit consists simply of an additional amplifier chain included in the regular video circuits in such a way that only video above a predetermined level is accepted and amplified by the iso-echo channel. These video peaks are then inverted and mixed in opposite phase with the video in the normal channel. The net result is that video peaks (return from heavy rainfall) above the predetermined level buck out the normal video at that particular point and produce less signal to the PPI tube resulting in "black" holes.

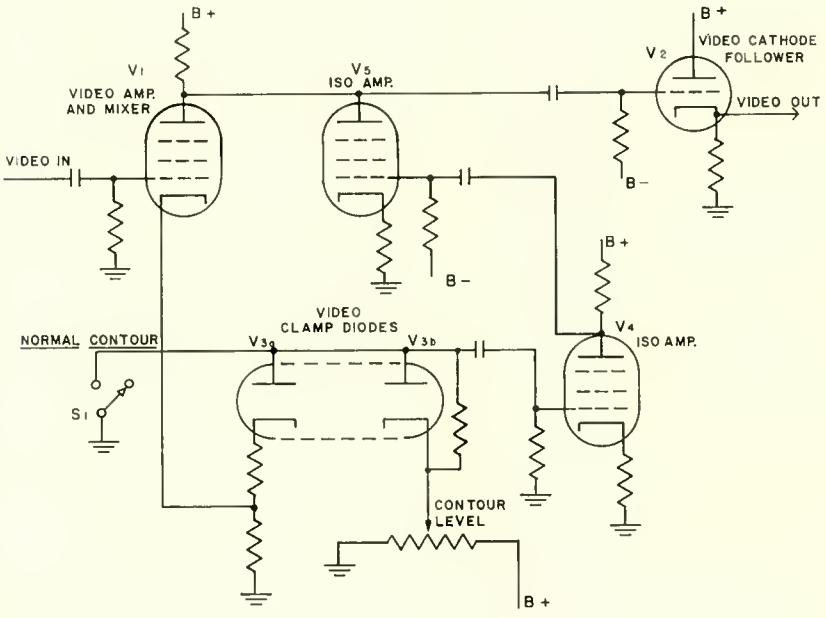


Fig. 9—Simplied diagram of AVQ-10 Iso-Contour circuit.

Figure 9 shows a simplified drawing of a typical iso-echo contour circuit as used in the AVQ-10. V_1 and V_2 make up the normal video channel. V_{3a}, V_{3b}, V_4, and V_5 make up the iso-echo channel. With S_1 in the "contour" position, the iso-echo video channel becomes operative and feeds to the mixer stage V_1 any signals above the level determined by the bias on V_{3a}. This bias, or "contour level" control, is set so that all signals representing undesirable turbulence or rainfall gradient can be detected on the PPI display. Experimental data derived from numerous flight tests involving the type of storms encountered over the western plains of the United States shows the optimum setting to be in the signal region 15 to 20 decibels above the minimum detectable.

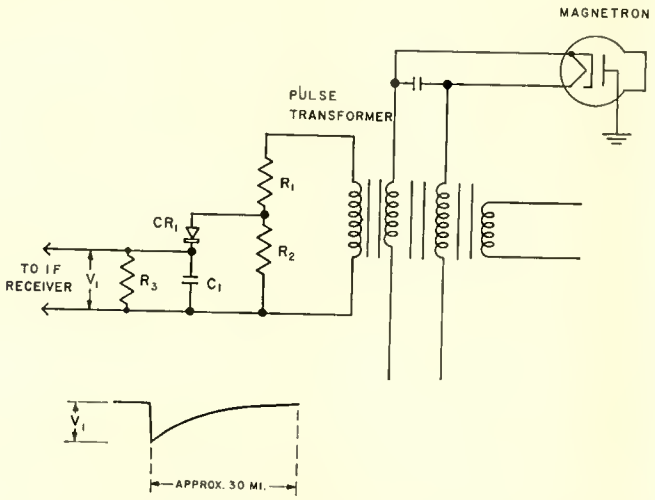


Fig. 10—Simplified diagram of AVQ-50 STC circuit.

Sensitivity Time Control

Weather radars using iso-echo circuits of the type discussed above must also use some form of sensitivity time control. Since the iso-echo circuit operates on difference in signal amplitude, means must be provided to assure that this amplitude difference is produced entirely by rainfall rate change within the storm area, and is essentially independent of range to the target. The conventional STC circuit uses at least one and sometimes two vacuum tubes. The STC circuit shown in Figure 10 is used in the AVQ-50. It illustrates one of the several opportunities taken in this equipment to eliminate vacuum tubes and improve equipment reliability. An additional winding on the mag-

netron pulse transformer provides a negative pulse voltage of low impedance and high peak current capability to charge capacitor C_1 through the low forward resistance of crystal diode CR_1. At the completion of the modulator pulse, C_1 discharges exponentially through R_3 to produce the desired STC time constant. R_1 and R_2 are selected to give the desired initial "depth of drive."

A-C Resonant Charging Modulator

A-C charging modulators date back to the very early history of radar, being used primarily in high-power rotary-gap modulators of low recurrence frequency. With the advent of multiple pulse widths

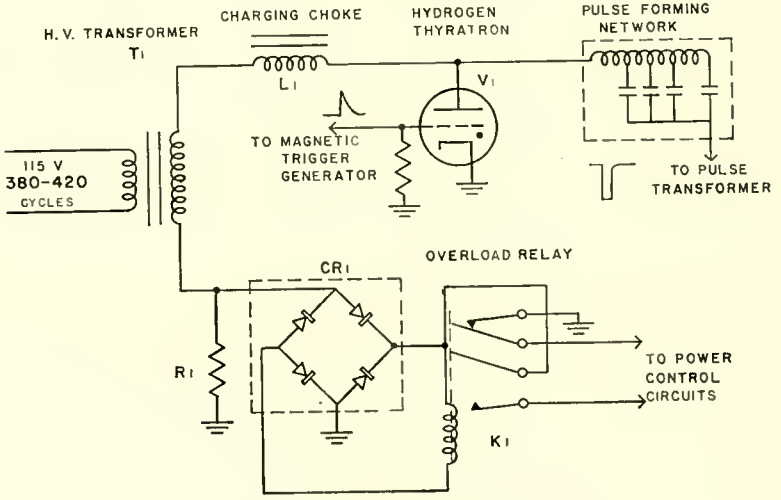


Fig. 11—A-C resonant charging modulator.

and repetition rates in latter day equipment, this type of pulser saw little service. However, the requirements of commercial weather radar permit a single pulse width and a single repetition rate. Also, the repetition rate may be an integral multiple of the line frequency. This makes a-c resonant charging again practical with resulting circuit simplicity and improved reliability. A single-tube a-c resonant charging modulator has been used in both the AVQ/10 and AVQ/50 weather radar systems. Figure 11 shows a simplified circuit diagram. The charging circuit, consisting essentially of charging inductor L_1 and the pulse-forming network (PFN) capacitance, is tuned to resonance at the a-c line frequency of 400 cycles. Because of the resonant condition the PFN voltage reaches a maximum value when the impressed sinusoidal voltage passes through zero. If the trigger voltage

to the thyratron grid is in proper phase relationship, the modulator output pulse will always occur at this zero point and stable circuit operation will be maintained.

One disadvantage of the a-c resonant charging circuit is that the voltage across the PFN continues to build up if the thyratron misses one or several pulses. To prevent damage being incurred as a result of such failure, a protective circuit consisting of R_1, CR_1, and K_1 has been provided. This circuit acts to remove the input voltage to the high voltage transformer T_1 if the charging current becomes too great. Since variation in the phase relationship between the thyratron trigger and the resonant charging wave will also affect stability of operation, special means have been taken to counteract phase variations.

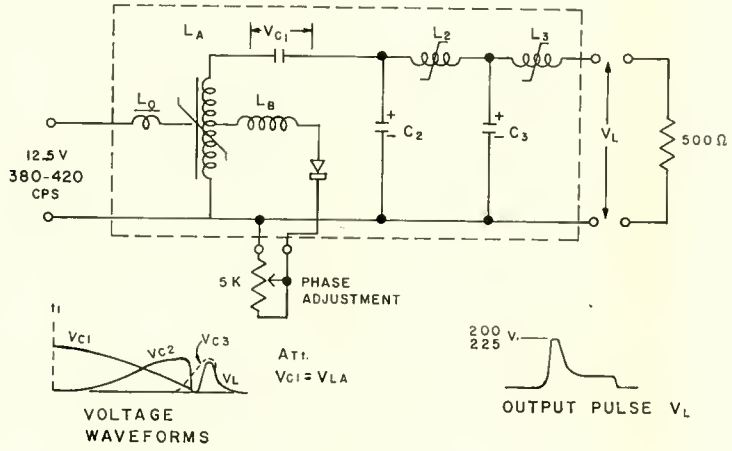


Fig. 12—Magnetic trigger generator.

Magnetic Trigger Generator

Circuit constants of the a-c resonant charging modulator, by the nature of its operation, can be optimized for only one frequency, the resonant frequency. Although the modulator used in the AVQ-10 and AVQ-50 is generally thought of as a 400-cycle unit, in actual use it must operate over a frequency range of 380 to 420 cycles, the normal allowable frequency tolerance for aircraft inverters. Operation over this frequency range will produce phase changes in the resonant charging circuit which can alter the operation. Switching no longer occurs at the zero point of the sinusoidal input voltage, and unstable operation will result. To counteract this effect (as well as eliminate vacuum tubes for improved reliability), a magnetic trigger generator

is used in the AVQ-10. Since the magnetic trigger generator also operates on the a-c charging principle, variation in line frequency produces essentially the same effect in its operation as in the modulator. Proper phase relationship between the trigger pulse and the PFN peak voltage is always maintained to produce stable circuit operation.

A circuit diagram of the magnetic trigger generator is shown in Figure 12. Initially L_A, L_2, and L_3 are assumed saturated. Resonant charging of the resulting series circuit consisting of L_0 and C_1 causes the voltage V_{C1} to rise to a value approximately equal to the peak value appearing across the autotransformer, L_A. The charge is then transferred to C_2, causing the voltage at C_2, V_{C2} to rise. At a certain value of voltage, the saturable reactor L_2, changes from a very high impedance to essentially zero impedance allowing a rapid transfer of energy from C_2 to C_3. This causes the voltage at C_3, V_{C3}, to rise rapidly to a value where L_3 saturates, and the output pulse appears across the load. Since the transferred charge is the same in each stage, by proper design of L_2 and L_3 the output pulse is formed. The bias arrangement enables single-phase operation, i.e., one output pulse per cycle of input. The phase adjustment allows a 30° to 50° control of the point of pulse formation near the end of a cycle of input.

Acknowledgment

The authors wish to acknowledge the assistance given by Willis F. Beltz and Robert W. Kissinger in providing the material on the RCA 6521 magnetron.

THE USE OF VERTICAL POLARIZATION TO SOLVE UHF TELEVISION "GHOSTING" PROBLEMS IN A SHADOWED VALLEY*

By

Donald W. Peterson

RCA Laboratories,
Princeton, N. J.

Summary—UHF television broadcasting in hilly or mountainous terrain suffers from "ghosting" or multi-path reception in shadowed valleys. An experimental study of the ghosting problem has led to the conclusion that vertically polarized co-channel amplifier boosters and passive boosters will solve the problem of serving shadowed valleys. New service can be rendered without denying any existing service resulting from creation of a "mush" area.

Introduction

IN UHF television broadcasting, much difficulty has been experienced due to the "shadowing" effects which are present in hilly areas. Areas relatively close to the transmitter may receive no satisfactory signal if they lie in the shadow of a hill. In such cases, "satellites" may be employed to supply signals to deficient areas.

The two kinds of satellites in use are the co-channel booster and the frequency-changing repeater. Thus far there have been very few co-channel booster installations.[1] A serious drawback has been the existence of "mush" areas where direct and boosted signals interfere with each other to produce a strong ghost in the picture. On the other hand, frequency-changing repeaters avoid the mush area problem only to experience other difficulties. For one thing, the identity and symbolism of the station channel number are lost. Thus reception of channel 40 via channel 72 has much less than the desired impact on viewers. Furthermore, since FCC regulations require that repeaters operate on the upper UHF channels, the high frequency of repeaters in itself is a handicap because of propagation, line loss, and converter effectiveness problems.

The typical shadowed area is usually an area of ghost or multipath reception problems. Such ghosting may be observed by the simple

* Manuscript received March 31, 1958.

[1] There are a number of papers on boosters in *I.R.E. Transactions on Broadcast Transmission Systems*, March, 1955.

expedient of driving through a shadowed valley with a unidirectional receiving antenna and a UHF television receiver. This procedure, however, will not be very informative beyond demonstrating the existence of severe ghosting from multipath propagation. Ordinary receiving antennas with front-to-back ratios up to 20 decibels are not capable of sufficient directional discrimination to determine the directions from which the several signals are coming. Thus, rotation of the receiving antenna may show completely unrecognizable pictures apparently coming from all directions.

Pulse Observation of Multipath Effects

It has been shown by pulse studies in the field that shadowed-valley reception can commonly be accounted for by two rays. The direct-ray pulse, which comes over the shadowing hill, often is undistorted. The principal alternate-path-ray pulse frequently comes from behind the receiving antenna. There may actually be many rays coming from behind, but even in the case of a single wooded-ridge reflector there will be great reflected-pulse distortion. It may be said in general that reflected pulses from wooded hills are always lengthened and distorted to a degree that indicates that pictures received by reflection from the hills will be useless. On the other hand, single bare-hill reflections may be useful.

Picture Observation

It is now possible to explain the ghosting observed with a unidirectional antenna in a shadowed valley. The signal from behind the receiving antenna usually comes from a hill behind the antenna. The strongest reflected signal comes from the upper part of the hill. Despite reflection loss of 20 or more decibels, the reflected signal may be greater than the direct-ray signal by an amount in excess of the front-to-back ratio of the receiving antenna. When this happens, the multipath signals appear to come from all directions insofar as can be determined by picture observation while rotating the receiving antenna.

Ghost Reduction

Although in some instances horizontally polarized co-channel boosters have been successfully used to get signals into shadowed areas, the installation of a booster generally leads to a "mush" area where the space median values of direct and boosted signals are the same. This produces a ghosting effect since the two path lengths will not, in general, be equal. Discrimination between the two signals may sometimes

be achieved by utilizing the directivity of a standard receiving antenna through proper location and orientation. If receiving antennas of improved design were developed, this approach might be quite fruitful, particularly in those cases where the interfering signal is coming from *behind* the antenna. Generally, relatively little attention is given to the *vertical* patterns of receiving antennas. Since the strongest reflected signals may be expected from high on hills behind the antenna, this may be a clue to ghost reduction.

It was felt, however, that the use of vertical polarization[2,3] might offer a simpler and more effective solution.

Field Experiment

With a view toward learning how to cope with severe ghosting in shadowed valleys, a field experiment was undertaken. The WINR, Binghamton, New York Television broadcasting system[4] furnished an ideal situation for study in Chenango Valley, Figure 1. The valley is only a few miles from the broadcasting antenna. There is ghosting ranging from moderate to severe. WINR is a channel 40 station (626-632 megacycles).

A unidirectional vertically polarized transmitting antenna was set up in a location overlooking the valley; this site is labeled "640 MC Transmitter" in Figure 1. The antenna and a view of the valley are shown in Figure 2. The antenna was fed with a 10-watt CW signal at 640 megacycles, about 8 megacycles above the 631.75-megacycle of the sound carrier of WINR. The antenna was a 24-foot colinear array with a flat reflector; the half-power beam width was 3.7° in elevation and 58° in azimuth, and the gain over a half-wave dipole was 20 decibels. It was aimed down into the valley by tilting for the purpose of illuminating the populated area and so that very little of the radiated power could escape the valley. Thus there were available in the valley both the WINR horizontally polarized signal which was shadowed, and the 640-megacycle vertically polarized unshadowed signal. The frequencies were close enough so that reflections from the hills beyond the valley would not show significant frequency variation.

The receiving antenna was a cylindrical parabola reflector with two colinear dipoles at the focus; the pattern in the rear 180° was down

---

[2] J. T. Dixon and J. M. Taff, "Polarization Discrimination in Television Broadcasting," *FCC Report 4.3.10*, March 14, 1958.

[3] K. H. Kappelhoff, "Advantages Gained by Considering Polarization in the Planning of VHF and UHF Broadcasting Services," *NBS Report 5538*, September 15, 1954.

[4] Owned and operated by the Gannett Company.

at least 25 decibels. The antenna was arranged in such a way that it could be tilted in the principal elevation plane, rotated in azimuth, and rotated in polarization.

In order to learn where the potential mush area lay, a number of

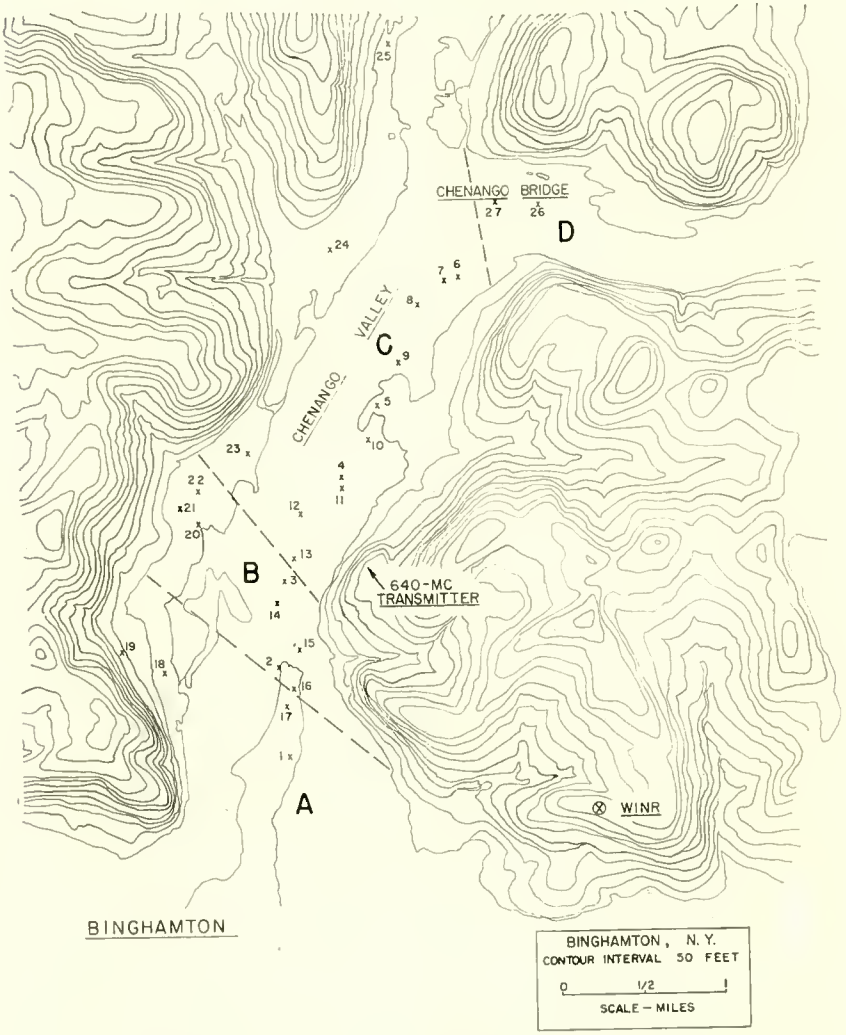


Fig. 1—Contour map of the Chenango Valley area of Binghamton, N. Y.

receiving locations were chosen. These extended from well within the area at the mouth of the valley, where there was high-quality direct reception from WINR, to a distance of 4 miles up the valley. Four measurements were made with a field intensity meter at each location.

The terminal voltage at the input to the meter was measured for both the WINR sound channel and the 640-megacycle signal. Each signal was measured with the receiving antenna properly oriented for the polarization of the particular signal, and then a second measurement was made with the receiving antenna polarization oriented so as to minimize the signal. The data obtained is given in Table I. In addition to the field-intensity measurements, visual observations of WINR pictures were made at each location on a portable receiver. Examination of the data suggests what might be expected from the use of a booster.

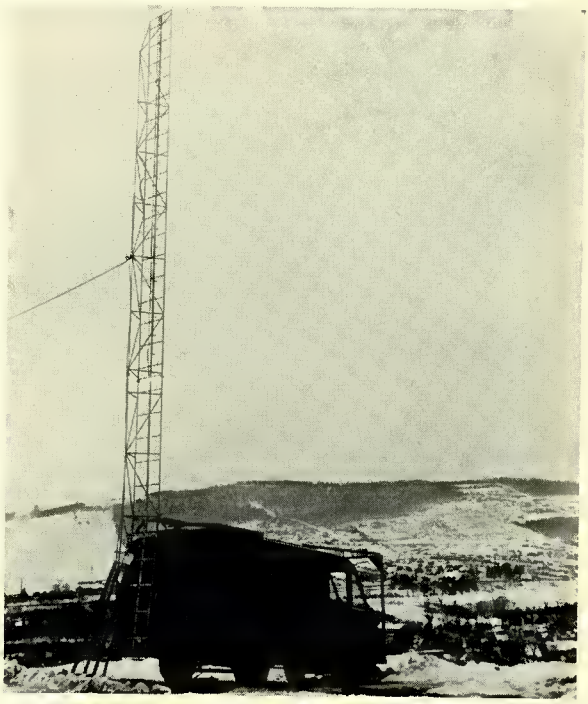


Fig. 2—Transmitting location and valley illuminated by 640-megacycle transmitter.

Referring to Figure 1, the area designated A receives strong signals directly from WINR; these are free of harmful ghosting. There are the usual short delay effects always found in cluttered surroundings but received pictures are excellent virtually everywhere in the area. In the area designated B, ghosting on the WINR signal is of such magnitude as to result in somewhat degraded pictures. By the standards which prevail in the field, the pictures are useful but are con-

siderably deteriorated by ghosting. Direct reception in area C is very poor. Pictures are almost everywhere distorted beyond recognition as a result of multi-path signals. The signal is weak enough to produce snow in the picture even in the best installation. Direct reception of useful pictures in area D is non-existent.

In area A a booster signal would be so far below the direct signal as to create no new problem regardless of booster polarization.

In area B the data shows that with horizontal polarization of the booster this would be a mush area. On the average, in this area, the horizontally polarized direct and vertically polarized booster signals

Table I — Field Intensity Meter Relative Terminal Voltage

	WINR		Booster	
Location	*Horizontal*	*Vertical*	*Horizontal*	*Vertical*
1	1000	800	76	110
2	130	17	51	340
3	350	27	61	1300
4	370	150	130	2200
5	280	57	200	1100
6	450	42	23	370
7	540	75	5	210
8	330	100	16	230
9	190	35	100	940
10	72	25	37	330
11	68	18	17	100
12	110	13	70	4400
13	140	40	280	760
14	560	38	37	270
15	115	16	7	40
16	140	15	44	140
17	1000	150	4	48
18	5400	350	16	98
19	8800	900	52	620
20	1400	190	180	1300
21	1700	260	40	410
22	680	250	100	960
23	330	50	320	2700
24	170	45	150	1200
25	9400	800	60	540
26	320	170	6	15
27	280	35	33	280

would be equal. In this area a booster would provide the desired good quality signal in most cases. It was found that the undesired horizontally polarized direct signal could be reduced a median value of 17 decibels by polarization discrimination. The lowest value of polarization discrimination observed in area B was 8.7 decibels, and it should be noted that in this case, in location 22, there was ample additional discrimination because of the differences in signal levels to bring the actual discrimination to 11.7 decibels.

It has been learned, by picture observation under laboratory conditions, that a ghost signal at —20 decibels is hardly discernible. Even a

—10 decibel ghost is not serious by comparison with ghosts commonly tolerated in the field.

It is apparent from a study of the data that a mush area would exist in Chenango Valley if horizontal polarization were used with a co-channel booster. This area could be minimized by careful design and location of the booster re-radiating antenna. However, it is necessary to contend with three practical difficulties. (1) The booster antenna patterns determine where potential mush areas will be created. It may be difficult or impossible to tailor the pattern to the topography. (2) The fact that the clutter of houses and trees usually is present at receiving locations of interest means almost completely independent randomness of field strength distributions from the two signal sources in areas of equal median field strength. (3) Conveniently elevated locations for booster antennas will not, in general, exist right where they are desired. Therefore, mush areas are likely to occur in horizontally polarized booster installations. The extra degree of freedom afforded by a vertically polarized booster makes boosters relatively easy to engineer.

It should be pointed out that the Chenango Valley experiment was deliberately set up so as to present a sizable mush-area. Because of the location chosen for the 640-megacycle antenna, the receiving antenna was aimed in essentially the same direction for both transmissions in a considerable part of the valley. It is believed that no situation can exist where installation of a vertically polarized booster will deny reception to existing receiving installations, although it may be necessary to reorient or relocate some antennas in the area where received pictures are of border-line quality. New installations would, of course, be polarized for the best signal.

Receiving Antenna Installation

The use of vertically polarized boosters requires nothing new of the serviceman who installs the receiving antenna. He is already accustomed to searching for favorable receiving locations in shadowed areas and carefully orienting the antenna. The booster signal will be of much higher quality than the direct signal in the shadowed area. In the border area where useful signals of both polarizations are present, either signal can be used in the rare instance when both are equal by the simple experimental choice of polarization by the antenna installer.

Passive Boosters

In certain instances passive boosters or boosters without amplifiers will be useful. The choice between passive boosters and amplifier

boosters is a matter of economics. Consider a case where a shadowing hill one mile distant is high enough to be on the main beam of a station which radiates a megawatt. The field intensity will be

$$F\ (\text{mv/m}) = \frac{137.6\ \sqrt{P}\ (\text{kw})}{d\ (\text{miles})},$$

$$= 137.6\ \sqrt{1000} = 4{,}360\ \text{mv/meter}.$$

A 630-megacycle uni-directional receiving antenna with a power gain of 1000 referred to a half-wave dipole, which will be roughly 20 feet square, will make available for re-radiation nearly 1.5 watts. For this kind of application a paraboloid receiving antenna may be prohibitively expensive, but a flat reflector surfaced with a suitable commercial fencing and mounting an array of, for example, zigzag elements is a possibility. Such a passive booster can serve an area of many square miles at the foot of the shadowing hill.

Conclusions

It has been shown by experiment that mush areas caused by co-channel UHF television boosters can be eliminated by the use of vertical polarization. The study also indicates that passive boosters would be useful in special situations.

Acknowledgment

The author gratefully acknowledges the cooperation of Richard Blackburn and Gino Riciardelli of the Gannett Company in setting up and conducting the field experiment.

TRACING DISTORTION IN STEREOPHONIC DISC RECORDING*

By

M. S. Corrington and T. Murakami

RCA Victor Television Division,
Camden, N. J.

Summary—*Tracing distortion in the 45°-45° and the vertical–lateral systems has been calculated to compare the two systems. Analysis shows that with an ideal pickup there is no cross modulation between the two channels in the 45°-45° system of recording if the groove angle is 90°. The intermodulation and harmonic distortion within each of the channels in the 45°-45° system is the same as that obtained in the normal vertically cut record. In the vertical–lateral system there is cross modulation from the lateral channel to the vertical channel and the amount has been calculated. The intermodulation and harmonic distortion in each channel has also been calculated. Curves of the amounts of distortion, crosstalk, and cross modulation for various groove velocities, stylus radii, and recording velocities are given for both systems.*

Introduction

ALTHOUGH the theory of tracing distortion in vertically and laterally cut records has been known for many years, it will be reviewed briefly to show the method of computation and the modifications required for stereophonic recordings. In previous work, the tracing distortion was obtained by a harmonic analysis of the curve traced by the center of a spherical stylus. The amplitudes of the harmonic frequencies were found by (1) finding the coordinates of the curve and making a numerical harmonic analysis[1] or (2) by expanding an explicit expression for the curve in a power series.[2,3] In either event, the amount of work necessary was prodigious when more than one frequency was involved. When the record groove is modulated laterally and vertically simultaneously, the amount of labor required to compute the various distortion components, using a desk

* Manuscript received March 25, 1958.

[1] J. A. Pierce and F. V. Hunt, "On Distortion in Sound Reproduction from Phonograph Records," *Jour. Acoust. Soc. Amer.*, Vol. 10, p. 14, July, 1938.

[2] W. D. Lewis and F. V. Hunt, "A Theory of Tracing Distortion in Sound Reproduction from Phonograph Records," *Jour. Acoust. Soc. Amer.*, Vol. 12, p. 348, January, 1941.

[3] M. S. Corrington, "Tracing Distortion in Phonograph Records," *RCA Review*, Vol. X, p. 241, June, 1949.

calculator, would be too great to attempt. In view of this, it was decided to make the computations by programming an electronic digital computer to use method (1).

No Crosstalk Between Channels

In the 45°-45° stereophonic record system with a groove angle of 90° as shown by Figure 1, the outer wall of the groove is modulated by moving it parallel to itself with amplitude proportional to the signal in the right-hand channel. The inner groove wall is moved in a similar manner in accord with the signal in the left-hand channel. If identical signals are fed into both channels, the phasing is chosen so the groove motion is lateral only.

The pickup is designed with two output circuits and the axes of the elements are arranged so each circuit responds to the motion of one groove wall only. If adequate care is taken in the pickup design there

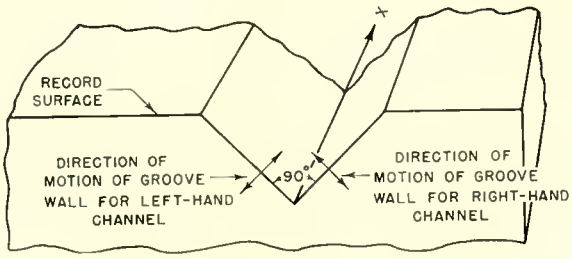


Fig. 1—Groove wall motions for 45°-45° system.

will be no crosstalk between channels since the groove walls move independently.

Equations for Stylus Motion

Assume that a cosine wave is recorded in one channel. A cross section of the wall of the groove, looking in a direction parallel to the elements of the cylindrical surface, will be as shown by the solid curve of Figure 2. It is assumed that the curvature of the stylus is always greater than that of the groove wall, so there is only one point of tangency (x_1, y_1).

The equation of the center of the sphere (ξ_1, η_1) using the coordinate system shown is

$$\xi_1 = x_1 + r \sin\theta, \tag{1}$$

$$\eta_1 = y_1 + r \cos\theta, \tag{2}$$

where $y_1 = a \cos kx_1$ and $k = 2\pi/\lambda$. The angle θ is defined by

$$\tan \theta = -\left.\frac{dy}{dx}\right|_{x=x_1} = -y_1', \tag{3}$$

so that $\sin \theta = -y_1'/[1 + (y_1')^2]^{1/2}$, and $\cos \theta = 1/[1 + (y_1')^2]^{1/2}$.

When the curve traced by a spherical stylus of radius r is an arbitrary function $y = f(x)$, the coordinates of the center of the sphere can be written

$$\xi_1 = x_1 - \frac{r\, y_1'}{\sqrt{1 + (y_1')^2}} \tag{4}$$

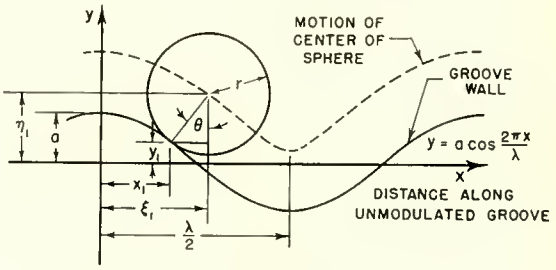


Fig. 2—Spherical stylus tracing a cosine wave.

$$\eta_1 = y_1 + \frac{r}{\sqrt{1 + (y_1')^2}} \tag{5}$$

where $$y_1' = -\left.\frac{df}{dx}\right|_{x=x_1}.$$

The coordinates of the center of the sphere for the cosine wave $y = a \cos kx$ become

$$\xi_1 = x_1 + \frac{r\, a\, k \sin k\, x_1}{\sqrt{1 + a^2k^2 \sin^2 k\, x_1}}, \tag{6}$$

$$\eta_1 = a \cos k\, x_1 + \frac{r}{\sqrt{1 + a^2k^2 \sin^2 k\, x_1}}. \tag{7}$$

If Equations (6) and (7) are each multiplied by k and the substitutions

$$\begin{aligned} A &= k\,a & H &= \eta_1 k \\ R &= k\,r & \Xi &= \xi_1 k \\ X &= k\,x_1 \end{aligned} \tag{8}$$

are made, the normalized form of Equations (6) and (7) are

$$\Xi = X + \frac{R\,A \sin X}{(1 + A^2 \sin^2 X)^{1/2}}, \tag{9}$$

$$H = A \cos X + \frac{R}{(1 + A^2 \sin^2 X)^{1/2}}. \tag{10}$$

Solution of Equations for Stylus Motion

To obtain a table of corresponding values of Ξ and H, Equations (9) and (10) must be solved simultaneously by eliminating X, the normalized distance down the unmodulated groove. H should be tabulated at equidistant steps in Ξ to simplify the numerical harmonic analysis.

Equation (9) is solved for X for each value of Ξ by Newton's method[4] of successive approximations. If X_1 is an approximate value of the desired root, Newton's formula for a more accurate value, X_2 for the root of the equation $f(X) = 0$ is

$$X_2 = X_1 - \frac{f(X_1)}{f'(X_1)}, \tag{11}$$

where X_1 is the first approximation and $f'(X_1)$ is the value of the first derivative at X_1. This formula can be used successively until the desired degree of accuracy is obtained. When X has been found, Equation (10) is used to find the value of H corresponding to the assumed value of Ξ. The result is a table of values of H for uniformly spaced values of Ξ. The increment in Ξ is determined by how rapidly the modulation varies and by the accuracy required.

Harmonic Analysis of Stylus Motion

From these values of Ξ and H, the Fourier series representation

[4] E. P. Adams, "Smithsonian Mathematical Formulae and Tables of Elliptic Functions," Smithsonian Institution, Washington, D. C., 1939, p. 7.

$$H = \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \cos n\,\Xi, \tag{12}$$

where

$$A_n = \frac{2}{\pi} \int_0^{\pi} H \cos n\Xi \, d\Xi$$

$$n = 0, 1, 2, \ldots . \tag{13}$$

can be obtained by the Newton-Cotes formulas[5] for numerical integration. The evaluation of Equation (13) for the harmonic amplitudes was programmed so that the entire calculation was done by the computer. Although Equations (9) and (10) were derived on the basis of a single tone, it is obvious that the analysis can be extended to multitone operation by using Equations (4) and (5).

Analysis of Vertical–Lateral System

Cross-Modulation Obtained

In the vertical–lateral system of stereorecording, the cutting stylus is modulated in the vertical and lateral directions simultaneously. The lateral channel produces a second-harmonic term in the vertical channel because of the pinch-effect. The tones in the vertical and lateral channels beat together to produce sum and difference frequencies in the lateral channel.

Equations for Stylus Motion

The vertical–lateral system can be analyzed in the following manner: Let the vertical and lateral modulation of the cutting stylus be given by $F_1(x)$ and $F_2(x)$ respectively, where x is the distance down the unmodulated groove. The vertical and lateral motions may be resolved into components $f_1(x)$ and $f_2(x)$ respectively perpendicular to the groove walls OB and OA as shown in Figure 3. In terms of $f_1(x)$ and $f_2(x)$, the vertical and lateral displacements become

$$F_1(x) = \frac{1}{\sqrt{2}} [f_1(x) + f_2(x)], \tag{14}$$

[5] National Bureau of Standards, "Mathematical Tables Project, Tables of Lagrangian Interpolation Coefficients," *Columbia University Press*, New York, N. Y., 1944, p. xxxii.

and

$$F_2(x) = \frac{1}{\sqrt{2}} [-f_1(x) + f_2(x)]. \tag{15}$$

Solving Equations (14) and (15) for $f_1(x)$ and $f_2(x)$ gives

$$f_1(x) = \frac{1}{\sqrt{2}} [F_1(x) - F_2(x)], \tag{16}$$

$$f_2(x) = \frac{1}{\sqrt{2}} [F_1(x) + F_2(x)]. \tag{17}$$

By use of Equations (4) and (5), the distance from the center of

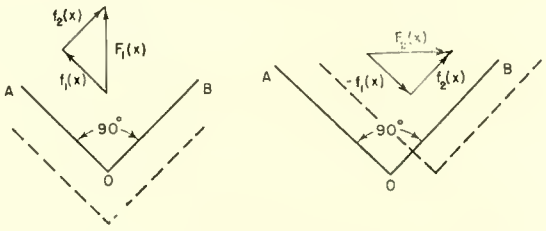


Fig. 3—Vertical and lateral motion of cutting stylus.

the sphere to the unmodulated sidewalls OB and OA may be found as a function of the distance x, along the unmodulated groove.

Single Tone in Each Channel

Let the cutting stylus be modulated by the tones

$$F_1(x) = a_1 \cos k_1 x, \tag{18}$$

and

$$F_2(x) = a_2 \cos k_2 x \tag{19}$$

in the vertical and lateral directions respectively, where a_1 and a_2 are the amplitudes of motion and k_1 and k_2 are the angular frequencies of the tones. Using Equations (16) and (17), the equations for the sidewall motion are

$$y_1 = f_1(x) = \frac{1}{\sqrt{2}} [a_1 \cos k_1 x - a_2 \cos k_2 x], \tag{20}$$

and

$$y_2 = f_2(x) = \frac{1}{\sqrt{2}} [a_1 \cos k_1 x + a_2 \cos k_2 x]. \tag{21}$$

The coordinates for the center of the sphere are obtained by substituting Equations (20) and (21) into Equations (4) and (5). After normalization, the equations of the stylus motion are given by

$$\Xi_1 = X_1 - \frac{RW_1}{\sqrt{1 + W_1^2}}, \tag{22}$$

$$H_1 = A \cos X_1 - \beta A \cos \alpha X_1 + \frac{R}{\sqrt{1 + W_1^2}}, \tag{23}$$

$$\Xi_2 = X_2 + \frac{RW_2}{\sqrt{1 + W_2^2}}, \tag{24}$$

$$H_2 = A \cos X_2 + \beta A \cos \alpha X_2 + \frac{R}{\sqrt{1 + W_2^2}}, \tag{25}$$

where

$$W_1 = - A \sin X_1 + \alpha\beta A \sin \alpha X_1 ,$$

$$W_2 = + A \sin X_2 + \alpha\beta A \sin \alpha X_2,$$

$$\alpha = k_2/k_1 , \quad A = \frac{a_1 k_1}{\sqrt{2}}, \quad X_1 = k_1 x_1 , \quad \Xi_1 = k_1 \xi_1 , \quad H_1 = k_1 \eta_1 , \quad \beta = a_2/a_1 ,$$

$R = k_1 r$, $X_2 = k_1 x_2$, $\Xi_2 = k_1 \xi_2$, $H_2 = k_1 \eta_2$. x_1 is the distance to the point of contact on groove wall OB, and x_2 is the distance to the point of contact on groove wall OA.

Since the normalized coordinate of the center of the sphere measured along the unmodulated groove is the same for both sidewalls, $\Xi_1 = \Xi_2$

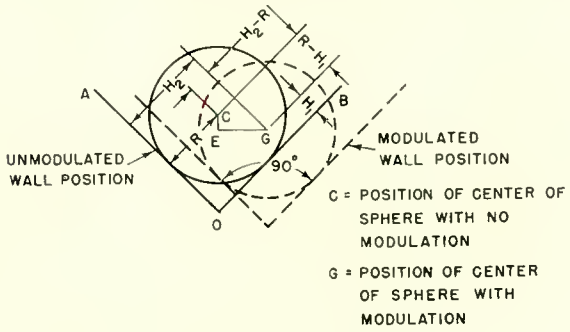


Fig. 4—Coordinates of movement of sphere.

$= \Xi$. H_1 and H_2 are therefore each expressible as a function of Ξ. $H_1(\Xi)$ is found by solving Equation (22) for X_1 for each successive assumed value of Ξ by Newton's method and substituting this value of X_1 into Equation (23). $H_2(\Xi)$ is found in the same manner using Equations (24) and (25).

Stylus Displacement

Figure 4 shows the position of the groove in the modulated and unmodulated states. In this figure H_1 represents the distance of the center of the sphere from the unmodulated right-hand groove wall, OB, at any displaced position. The distance of the center of the sphere from the other unmodulated sidewall, OA, is given by H_2. Let the quantities D_1 and D_2 denote the normalized vertical and horizontal displacements of the center of the sphere. In Figure 4 these distances are given by $D_1 = CE$ and $D_2 = EG$, respectively. When expressed in terms of H_1 and H_2 with the aid of Figure 5, the vertical and horizontal displacements are

$$D_1 = \frac{H_1 + H_2 - 2R}{\sqrt{2}}, \tag{26}$$

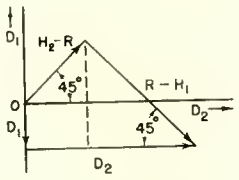


Fig. 5—Relation between H_1, H_2, D_1 and D_2.

$$D_2 = \frac{H_2 - H_1}{\sqrt{2}}. \tag{27}$$

A harmonic analysis of Equations (26) and (27) by the method previously described can be made to find the harmonic and crosstalk distortion in the two channels. To solve for the harmonic coefficients, the integration indicated by Equation (13) must extend over π radians of the difference frequency $k_1 - k_2$, which may be several cycles of each signal frequency.

The analysis described above can be extended to cases where there is more than one tone in each channel. The computation time will increase rapidly, since the integration interval must be one half cycle of the lowest possible beat frequency of the tones involved.

Curves of Calculated Distortion

To obtain curves of harmonic distortion in terms of the tangential groove velocity, v, and the recording velocity, u, the normalized amplitude, A, and radius, R, in Equations (9) and (10) must be expressed in terms of these quantities. Since the tangential groove velocity for a frequency f is given by $v = \lambda f$ and the amplitude a for a recording velocity u is $a = u/2\pi f$, the normalized amplitude is

$$A = ka = \frac{2\pi a}{\lambda} = \frac{u}{v}. \tag{28}$$

The normalized radius is then

$$R = kr = \frac{2\pi r}{\lambda} = \frac{2\pi f r}{v}. \tag{29}$$

Harmonic Distortion in Vertical and Lateral Systems

Using the proper values of A and R in Equations (9) and (10), the coordinates of the stylus motion and its harmonic amplitudes have been found for a recording frequency of 400 cycles per second. The percent second harmonic amplitude is shown in Figures 6(a), 6(b), and 7(a) for recording velocities of 7, 14, and 22 centimeters per second for three different stylus radii. These curves correspond to the tracing distortion obtained in each channel of the 45°-45° system using an amplitude-sensitive pickup. For a velocity sensitive pickup, the second harmonic amplitude would be multiplied by 2, the third by 3, etc. The same curves also apply to the vertical channel of the vertical–lateral

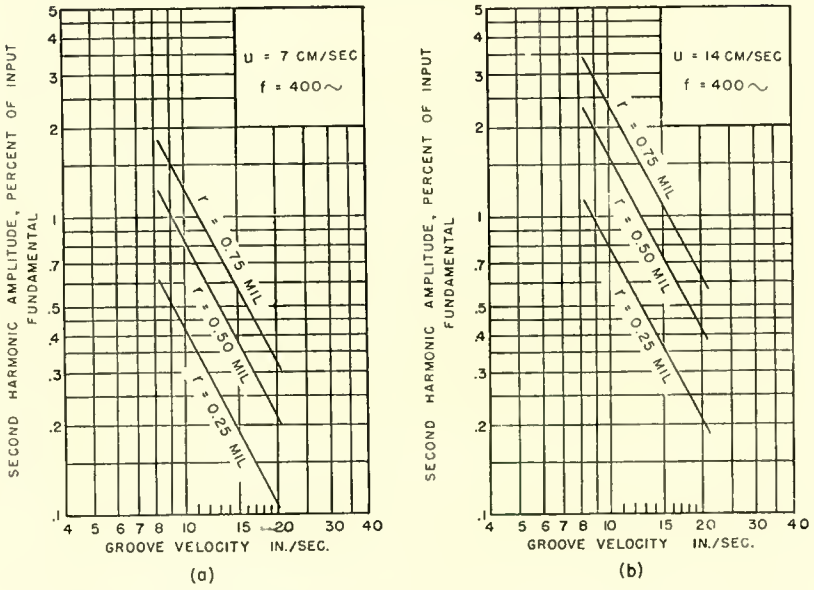


Fig. 6—Second-harmonic amplitude for vertical cut recording.

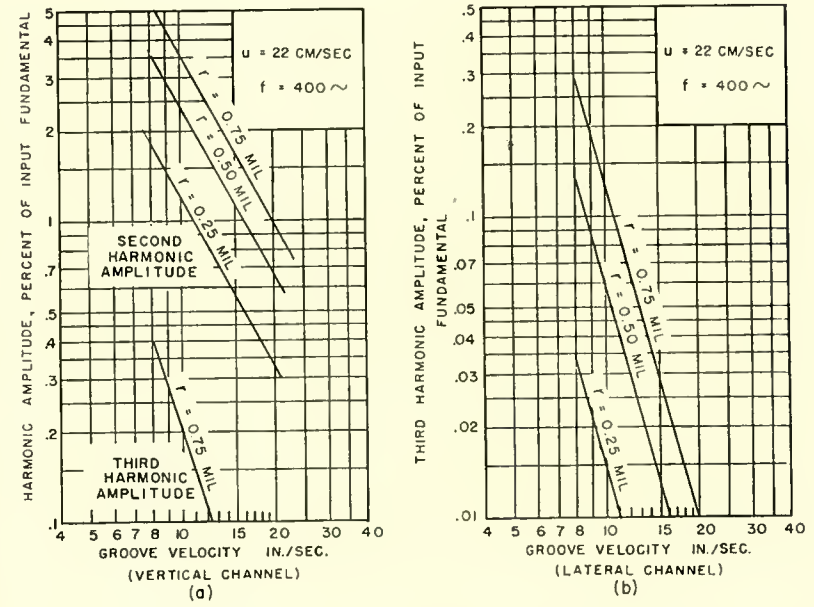


Fig. 7—Harmonic distortion in vertical and lateral channels.

system when the crosstalk from the lateral channel is less than 0.1 per cent. Higher-order harmonics which are not shown in these figures have amplitudes which are in the order of or less than 0.1 per cent. With a recording velocity $u = 22$ centimeters per second, the magnitude of the third harmonic component is more than 20 decibels down from the amplitude of the second harmonic component on an amplitude basis.

The harmonic distortion in a laterally cut record, recorded at a velocity of 22 centimeters per second, is shown in Figure 7(b) for comparison purposes. In this case the only significant term is the third-harmonic component which is more than 20 to 1 down in amplitude from the second harmonic component in the corresponding vertical recording.

Intermodulation Distortion in Vertical and Lateral Channels

The intermodulation distortion with two tones of frequencies 400 and 4,000 cycles in one channel has been plotted in Figures 8(a) and 8(b) for the vertical and lateral channels.

For a vertically cut record the intermodulation, I, has been defined as

$$I = \frac{\text{amplitude of 3,600-cycle tone} + \text{amplitude of 4,400-cycle tone}}{\text{amplitude of 4,000-cycle tone}}. \quad (30)$$

For a laterally cut record the intermodulation is

$$I = \frac{\text{amplitude of 3,200-cycle tone} + \text{amplitude of 4,800-cycle tone}}{\text{amplitude of 4,000-cycle tone}}. \quad (31)$$

A recording velocity ratio of $u_1/u_2 = 4$ has been assumed for the 400- and 4,000-cycle tones so that the amplitude of the 4,000-cycle tone for a constant velocity recording is 1/40 of that for the 400-cycle tone.

It is noted that the intermodulation in the vertical channel is more than 10 times that in the lateral channel for the same recording conditions. The intermodulation limitations are the same in the vertical–lateral system as in the 45°-45° system since each uses a vertical channel. To decrease the intermodulation distortion it is necessary to reduce the recording level or the stylus radius.

A close approximation to the intermodulation is given by the formulas

$$I = \frac{800\,\pi\,u_1 r}{v^2} \text{ (vertical recording)}, \quad (32)$$

$$I = \frac{800\,\pi^2 u_1^2 r^2}{v^4} \text{ (lateral recording)}, \tag{33}$$

where

I = per cent intermodulation,

u_1 = recording velocity in inches per second (400 cycles),

r = stylus radius in mils,

v = groove velocity in inches per second.

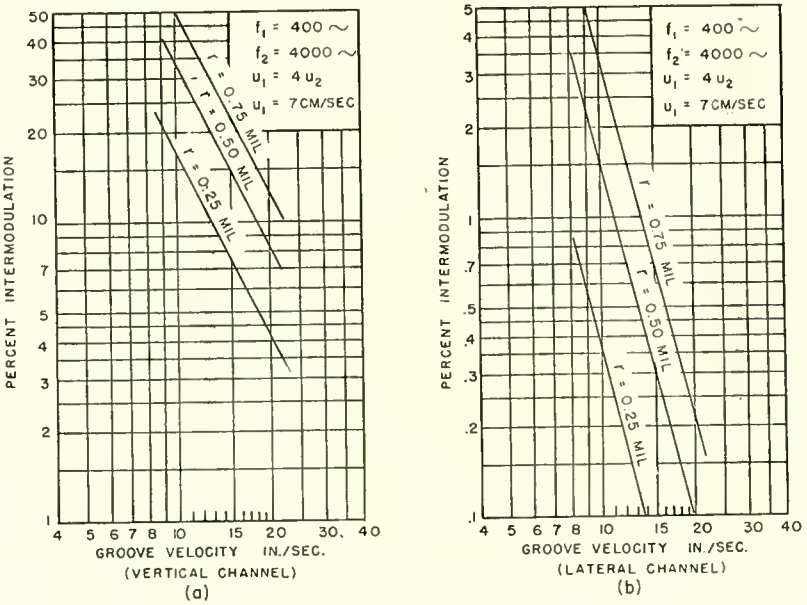


Fig. 8—Intermodulation distortion in vertical and lateral channels.

Equations (32) and (33) are accurate to within 5 per cent at 75 per cent intermodulation and within 1 per cent at 25 per cent intermodulation. These formulas are based on the relation between the harmonic amplitudes and the intermodulation distortion, and on the formulas derived by Corrington.[3]

Combination Tone Amplitudes for Vertical–Lateral System

Combination tones include all possible frequencies, harmonic, sum, and difference, etc., that are produced in a given channel. The curves

shown in Figures 9(a) and (b) were calculated with equal amplitude sine waves of frequencies 400 cycles and 300 cycles recorded in the vertical and lateral channels respectively. Combination tone amplitudes in the vertical channel include the 600-cycle pinch-effect crosstalk from the lateral channel and the 800-cycle second-harmonic distortion generated within the channel. Two recording amplitudes were chosen, namely, 7 and 14 centimeters per second; the stylus radius was 0.75 mil.

In the lateral channel the significant distortion frequencies are the sum and difference frequencies of 100 and 700 cycles per second. All other distortion terms are less than 0.1 per cent. These distortion components are seen to be larger than the harmonic distortion or the pinch-effect component in the vertical channel.

The curves of Figures 10 and 11 show the variation of the combination-tone amplitudes when the radius of the stylus is made 0.50 mil and 0.25 mil, respectively.

Relation Between Groove Velocity and Record Diameter

The groove velocity was given as the abscissa of the preceding curves so that they would be independent of the particular choice of revolutions per minute. Figure 12 can be used to find the record diameter corresponding to a given groove velocity.

Conclusions

Each wall of the groove of the 45°-45° system is equivalent to a channel with vertical recording. Since this is a single-sided system there will be second- and third-harmonic distortion. If the pickup is properly designed, there will be no cross modulation between channels. Curves are given for the harmonic and intermodulation distortion in either channel for various levels and stylus radii. The second harmonic component is the dominant distortion term and is approximately directly proportional to the stylus radius, the recording velocity and inversely proportional to the square of the groove velocity.

In either the 45°-45° or the vertical channel of the vertical–lateral system, the percentage distortion is the same; therefore the distortion limitations are the same for either system when the channels are considered individually.

To keep the distortion as low as in lateral recordings, the level should be lower than that presently used for single-channel lateral recordings, and the stylus radius should be less than the 1 mil commonly used. There is a limit to the allowable reduction in the recording level since the signal-to-noise ratio is also reduced as the level is decreased.

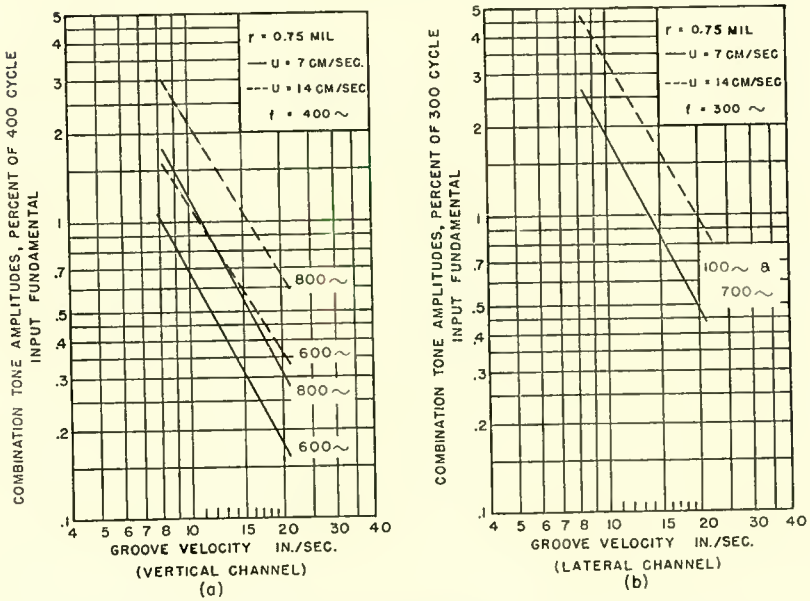


Fig. 9—Combination tone amplitudes for vertical–lateral system (r = 0.75 mil).

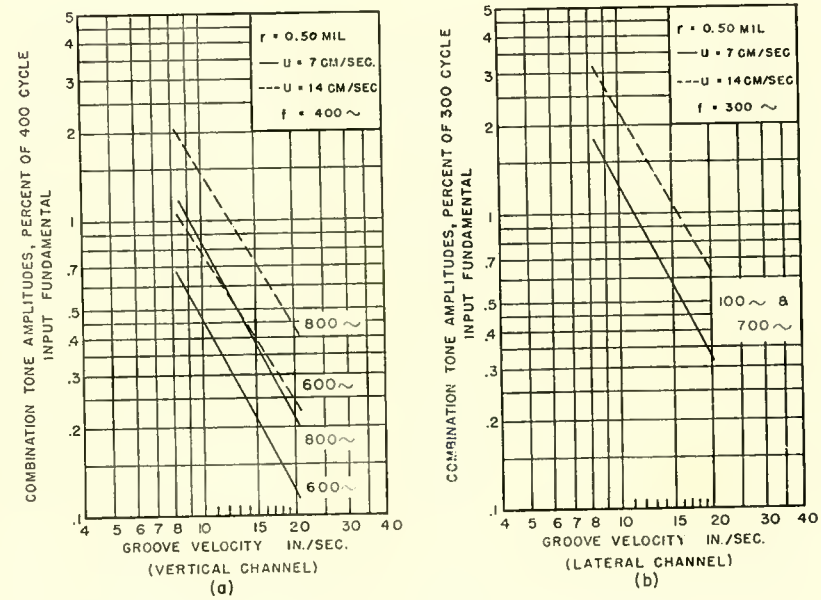


Fig. 10—Combination tone amplitudes for vertical–lateral system (r = 0.50 mil).

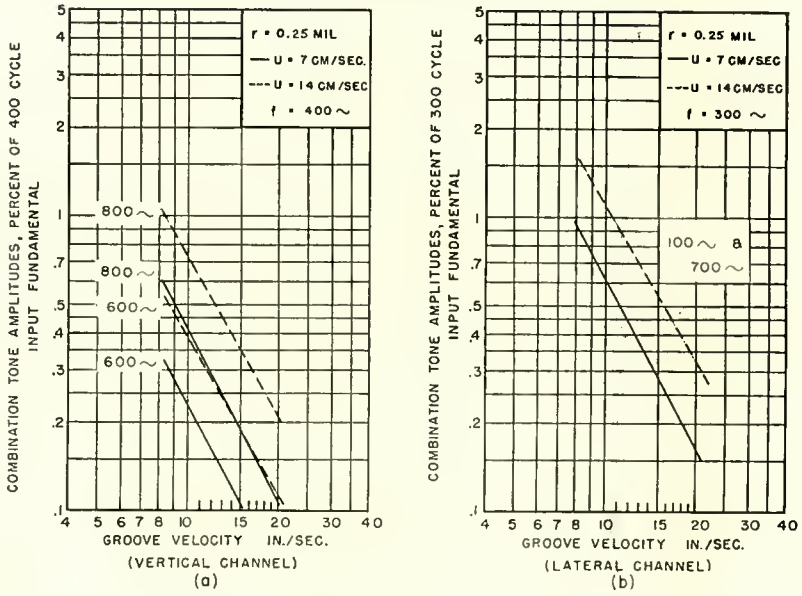


Fig. 11—Combination tone amplitudes for vertical-lateral system ($r = 0.25$ mil).

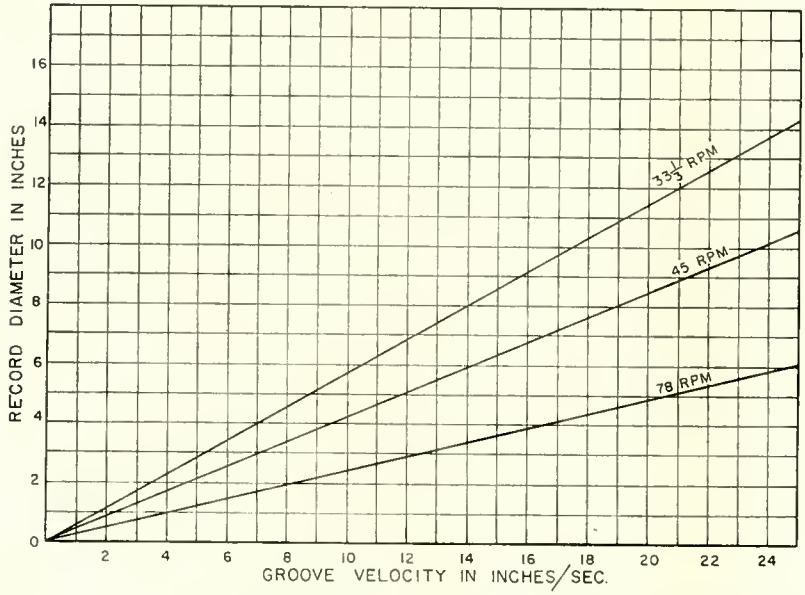


Fig. 12—Relation between record diameter and groove velocity.

When the vertical–lateral system is used, the distortion will be different in the two channels. The second harmonic of the tone in the lateral channel will appear in the vertical channel because of pinch effect. The vertical channel will also contain second and third harmonics of the modulation in the vertical channel.

The tones in the vertical and lateral channels will beat together and produce sum and difference tones in the lateral channel. These sum and difference amplitudes in the lateral channel are larger than the pinch-effect term produced in the vertical channel. There will also be third-harmonic distortion in the lateral channel; however, this harmonic will be quite small in amplitude compared to the second harmonic in the vertical channel.

No attempt has been made to set maximum allowable limits for the distortion in a stereophonic system. This should be done after extensive listening tests where the levels and stylus radii are varied systematically. On the basis of a limited number of listening tests made to date, it appears possible that the present limits for single-channel systems are unnecessarily low for a pleasing stereophonic system.

Acknowledgment

The program for the computer was prepared by R. F. Kolar. Without his skill, persistence, and efficient use of the computer, it would have been difficult to study so many cases.

A NEW HIGH-TRANSCONDUCTANCE ELECTRON GUN FOR KINESCOPES*

By

James W. Schwartz†

Summary—*This paper deals with a new kinescope electron gun. It employs a single aperture and two mesh-covered aperture grids in the modulator section. The first mesh is operated at a positive potential and acts as a space-charge grid. Modulation is applied to the second mesh grid or to the cathode. Present designs exhibit peak currents of 800 to 1,600 microamperes, require two to five volts drive, and produce resolution essentially equivalent to conventional commercial guns.*

Introduction

CONVENTIONAL kinescope electron guns exhibit a transconductance of approximately 20 micromhos. Many vacuum-tube amplifier structures produce a transconductance of 100 to 500 times this figure. It is not possible to apply receiving tube modulator geometry directly to the control of electron-gun current. This is because of the additional requirement that the electron stream must form a dense narrow-angle beam which can be imaged to a small spot.

Nicoll[1] has reported an experimental gun geometry which employs a fine metal mesh on the cathode side of a conventional control grid. It requires only five volts of drive and produces peak currents as high as 800 microamperes. Close tolerance requirements, critical design considerations, and limited resolution have restricted the application of this gun.

The present paper describes a new type of electron gun which employs two mesh-covered apertures. The first mesh grid is operated at a positive potential and acts as a space-charge grid. Modulation is usually applied to the second mesh of the double-mesh-grid gun. Present designs exhibit peak currents of 800 to 1,600 microamperes, require two to five volts of drive, and produce resolution essentially equivalent to that of conventional commercial kinescope guns. Critical mechanical construction problems are eliminated.

---

* Manuscript received April 10, 1958.

† Formerly, RCA Laboratories, Princeton, N. J.; now with Kaiser Aircraft and Electronics Company, Oakland, Calif.

[1] F. H. Nicoll, "Experimental High-Transconductance Gun for Cathode Ray Tubes," *RCA Review*, Vol. 16, p. 612, December, 1955.

Electrical Characteristics

Conventional kinescope guns are basically similar to ordinary triode or tetrode amplifier tubes. A negative grid located between the cathode and a positive acceleration electrode controls the electron flow by altering the electric field near the cathode.

The double-mesh-grid gun is basically similar in operation to a space-charge tetrode amplifier. A space-charge-grid amplifier tube is geometrically similar to an ordinary tetrode, but operates on an entirely different principle. The tube employs a small positive accelerating potential on the (space-charge) grid adjacent to the cathode. The second grid is used for control of the plate current and is operated at a small negative potential. A dense virtual cathode forms in the region

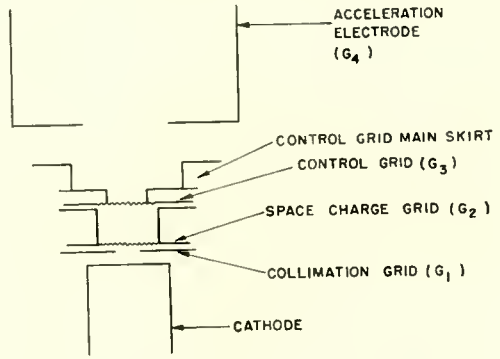


Fig. 1—Double-mesh-grid electron gun.

between the two grids, resulting in very high transconductance. Space-charge-grid tubes have had only limited application.* This is due to the fact that more linear characteristics are obtained with conventional structures, and that the space-charge-grid current usually exceeds the screen-grid current of a conventional tube resulting in poorer cathode emission utilization.

Figure 1 shows the geometric configuration employed in the double-mesh-grid gun. An aperture disk, G_1, is operated at or below cathode potential and masks the cathode. The space-charge grid, G_2, operates at a positive potential producing a collimated beam of electrons. Some of the beam is collected by the space-charge grid, but the majority of the electrons pass through to a retarding field between G_2 and G_3. If

* Recently this type of tube has become popular for 12-volt operation in automobile radios.

the control grid is positive, most of the electrons will pass through the mesh openings in it. Some electrons, of course, are also intercepted by the control grid, G_3. Those that pass through the G_3 mesh are focused to a crossover by the field formed by the cooperation of the accelerating electrode, G_4, and the grid skirt attached to G_3. The crossover may then be imaged upon the screen by the use of a conventional magnetic or electrostatic final lens.

A typical set of control characteristics is shown in Figure 2. It should be noted that the required drive range is less than 5 volts. The characteristic curves indicate that the space-charge-grid current,

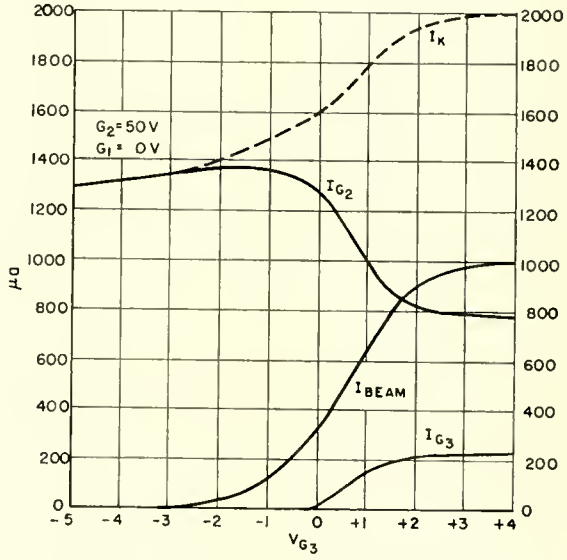


Fig. 2—Typical grid control characteristics.

I_{G2}, increases slowly as the control-grid voltage is increased in the region below beam cutoff. As the beam current increases in the region near zero bias on the control grid, G_3, the space-charge-grid current decreases abruptly. Below beam cutoff, electrons passing through the space-charge grid are reflected by the control grid and either return to the cathode or are collected by the space-charge grid. The abrupt decrease in space-charge-grid current is accompanied by the onset of electron transmission by the control grid as its voltage is raised in the region near zero bias. At a few volts positive bias on the control grid, no electrons are returned to the cathode or the G_3 side of the space-charge-grid mesh. The beam current saturates at a small posi-

tive control-grid voltage causing a break point in the characteristic curve. Figure 2 also shows the change in cathode current due to this effect.

The control grid can intercept electrons from the beam when it is at positive bias. The proportion of the beam current lost to the control grid is usually slightly less than one would expect from consideration of the percentage of mesh open area. This is due to a slight focusing action which tends to crowd electrons into the mesh openings.

Reduction of the space-charge-grid voltage reduces the accelerating field at the cathode and results in lower cathode emission. All of the

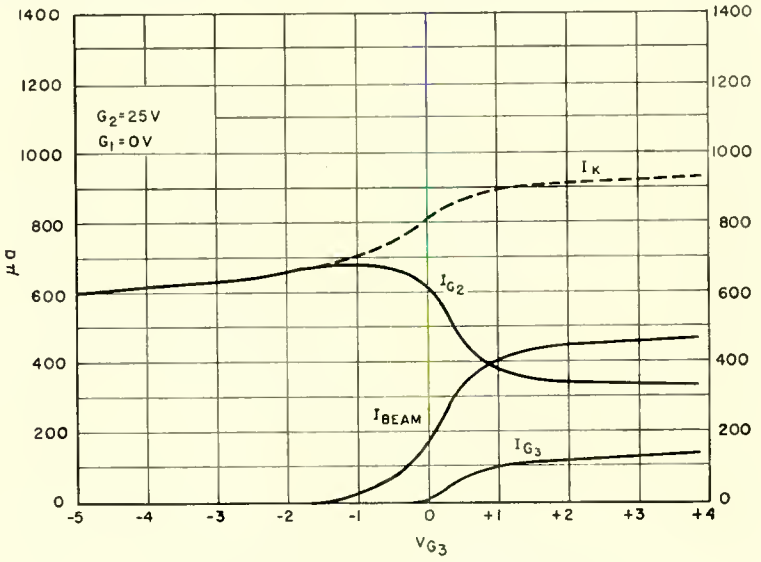


Fig. 3—Typical grid control characteristics.

tube currents are attenuated in approximately the same proportion when the cathode current is reduced in this manner. Figure 3 illustrates the characteristic curves for the same tube used to obtain the data of Figure 2. In Figure 3, however, the space-charge grid is operated at 25 volts instead of 50 volts. It will be noted that a small shift in cutoff and break-point voltage accompany the reduction in electrode currents.

Figure 4 is a logarithmic plot of beam current versus grid-drive voltage. It is interesting to note that the curve appears to be composed of two essentially straight line segments. The useful drive range is clearly defined by the break point in the curve. The slope of the curve

before the break point represents the γ of the electron gun. It is approximately 2.1. Changes in various electrode voltages have some effect upon γ. In some instances the logarithmic drive curve exhibits a gradual change in slope throughout the useful drive range.

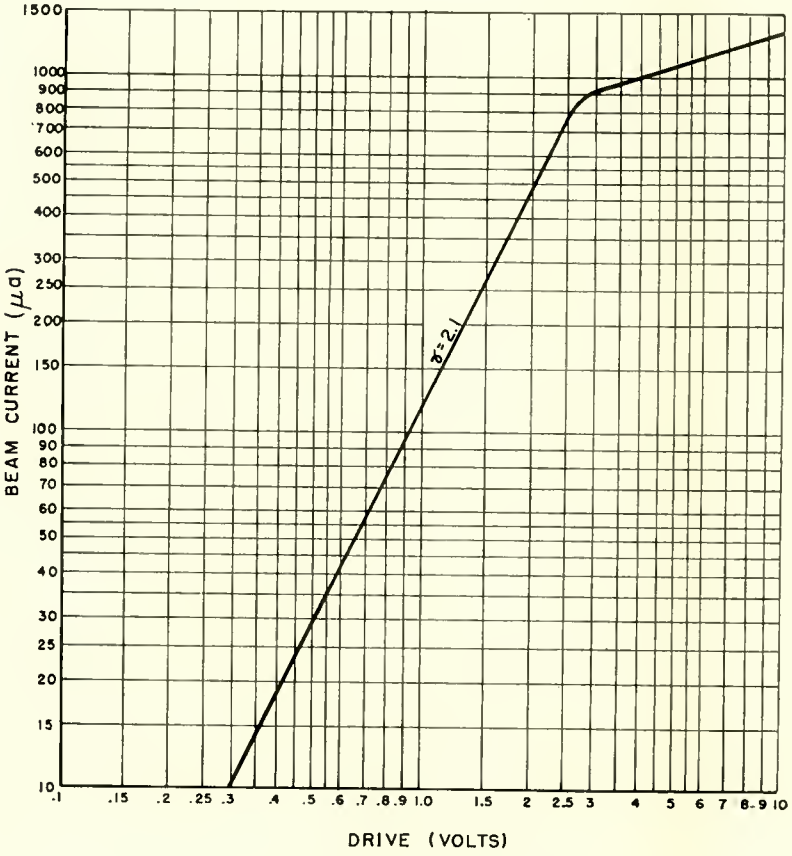


Fig. 4—Log transfer characteristics.

Extraneous magnetic fields appear to have no more pronounced effects on the operation of the double-mesh gun than on conventional types.

Cathode Loading

The cathode loading conditions in the new gun are somewhat different than those usually encountered in electron guns. In conventional kinescope guns the peak current density, which occurs at the

center of the cathode, is approximately 2.5 times larger than the average density over the area from which the emission is drawn. Due to intercepted grid currents in the new gun, the cathode current is about twice as large for the same maximum beam current. The emission density, however, is somewhat more uniform and in some designs may be spread over a considerably larger area. In the new gun, the net cathode current varies less than 2:1 with modulation. This is in contrast to the infinite ratio existing for conventional guns.

The destructive effect of high-speed positive ions upon the cathode may not be as important in the operation of the double-mesh gun. This is because of the more uniformly distributed emission in the new gun and because the meshes partially shield the cathode from bombarding particles. At this time it is not known whether the different loading conditions in the double-mesh gun are more favorable or less favorable in promoting long cathode life. Several experimental tubes have been operated satisfactorily for more than 2,000 hours.

Resolution

The crossover-forming electron optical fields of the double-mesh gun are basically different from those of conventional guns. In the latter, as the control grid is driven close to zero bias, the strong beam-focusing field rapidly deteriorates. This causes the large and abrupt increase in spot size at high beam currents, which is commonly known as "blooming." Figure 5 shows the crossover-forming field of a typical double-mesh gun.* It is obvious that small potential variations encountered over the drive range of the control grid, G_3, have very little effect upon the strong focusing lens in the region beyond G_3. This eliminates the "blooming" present in conventional guns. The spot size, in fact, changes only slightly throughout the range of beam current control.

Each mesh opening in the G_3 mesh serves as a small lens. The strength of the lens depends upon the field strengths on both sides of the mesh. The square of the transverse velocity imparted to an electron passing close to the edge of a mesh opening depends proportionately upon the field strength and the diameter of the mesh opening. The effect of mesh-induced transverse velocities is equivalent to the effect of transverse thermal velocity components at the cathode of a conven-

* The radial scale of Figure 5 is expanded five times more than the axial scale. This aids in illustrating the cross-over details. In such a scale transformation, however, angles are distorted. This causes the trajectories to appear to have a steeper slope than the normal to the equipotential of greatest curvature. Such an effect, of course, is impossible.

tional gun. They increase the size of the crossover and consequently limit resolution. Experimental magnetic focus double-mesh guns employing 500 openings per inch G_3 mesh exhibit resolution equivalent to that found in commercial kinescope guns. The resolution of guns employing 200 mesh, however, is degraded a moderate amount.

Gun-Design Parameters

A large number of geometrical parameters affect double-mesh-grid gun performance. Among these are the size and thickness of the various apertures and their relative spacings, the mesh sizes, locations

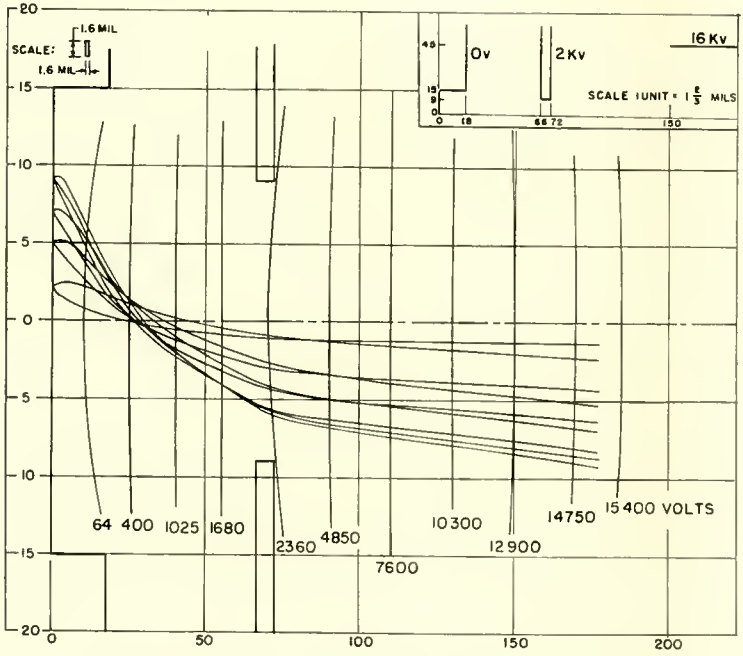


Fig. 5—Electron trajectories in crossover region.

and their relative alignments. A complete treatment of the effect of each variable will not be given here. Reference will be made, however, to some of the more interesting parameters.

An increase in the cathode-to-G_1 spacing reduces the electric field at the cathode and consequently reduces the electrode and beam currents. It also tends to make the break point in the beam-current versus G_3-voltage less abrupt and to increase the ratio of G_2 current

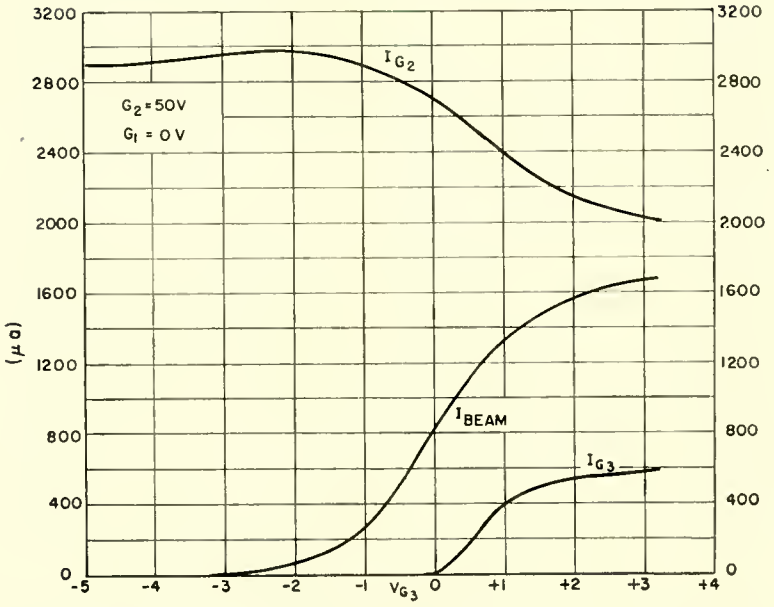


Fig. 6—Typical grid control characteristics for gun with increased G_1 aperture diameter.

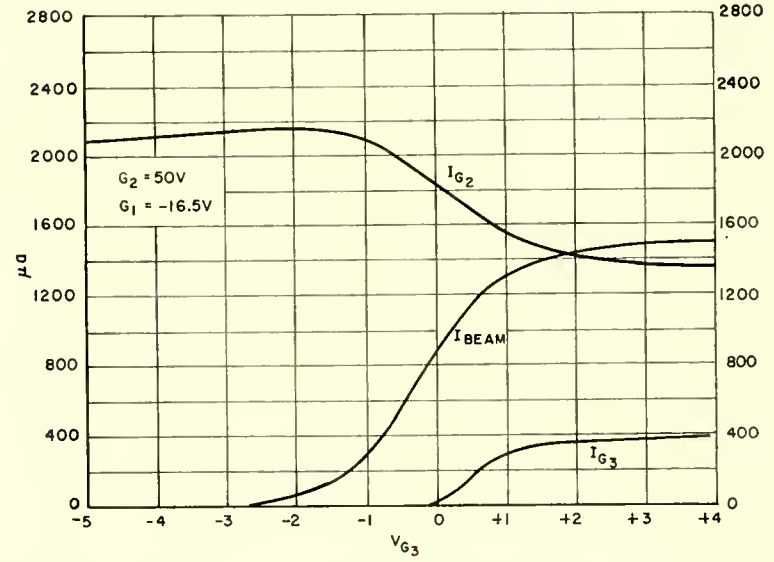


Fig. 7—Typical grid control characteristics for gun with increased G_1 aperture diameter.

to beam current. Similar effects are observed in poorly activated tubes, or in tubes purposely operated with reduced cathode temperature.

The size of the G_1 aperture is a particularly interesting parameter. If a large opening is employed, emission takes place from a larger cathode area. However, much of the increased current does not enter the beam when G_1 is operated at cathode potential. Instead, the G_2 and G_3 currents are increased. If a negative potential is applied to G_1, it is possible to decrease the G_2 and G_3 currents, while only slightly decreasing the beam current. Further increase in the negative bias applied to G_1 ultimately reduces the beam current as well. Figures 6, 7, and 8 show the characteristic curves for a gun constructed with a .050-inch instead of a .035-inch diameter G_1 aperture. Comparison of

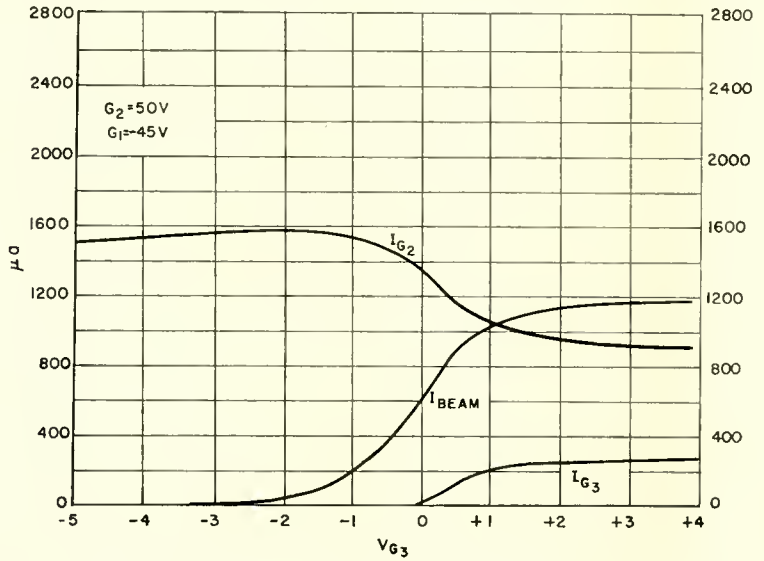


Fig. 8—Typical grid control characteristics for gun with increased G_1 aperture diameter.

these curves with those of Figure 2 reveals a sizable increase in maximum beam current at the same G_2 accelerating potential.

The operation of a G_1 aperture with negative bias and an enlarged aperture may be explained as follows: An increase in the G_1 hole size tends to allow emission from a larger cathode area. This produces a stream of electrons which is too large to pass freely through the G_3 opening. Instead, much of the stream is collected by the solid periphery of G_3 or is deflected to various parts of the G_2 electrode structure. If

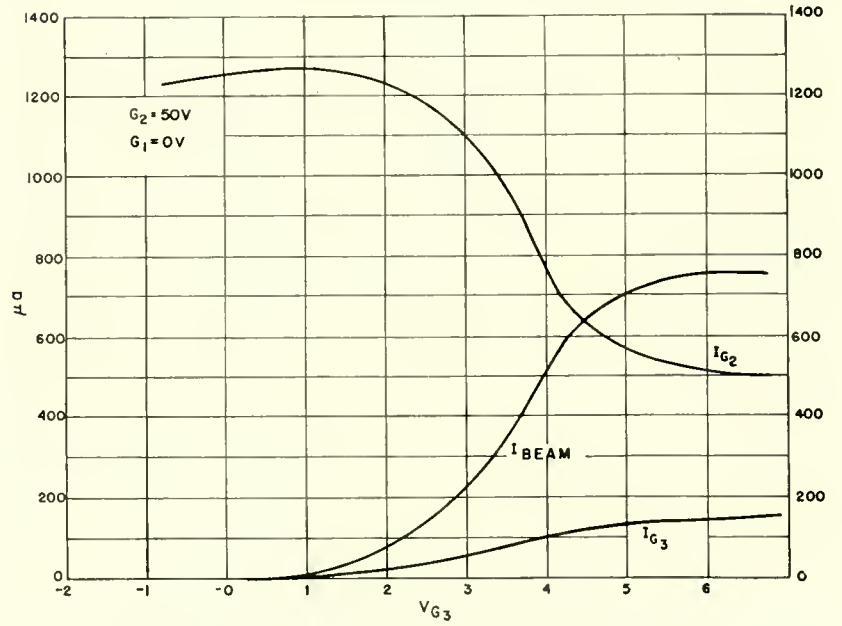


Fig. 9—Typical grid control characteristics for gun with increased G_2 thickness.

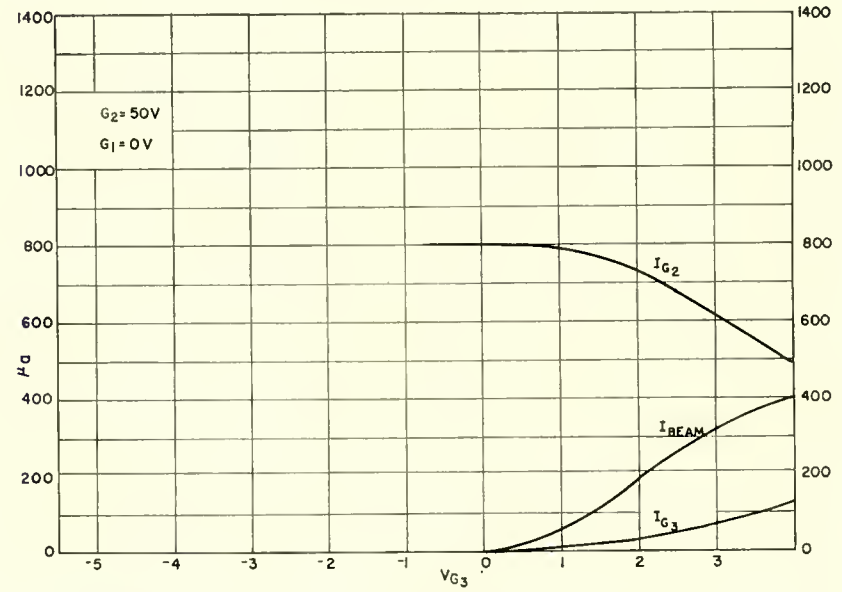


Fig. 10—Typical grid control characteristics for gun with parallel aligned meshes oriented for minimum optical transmission.

a negative potential is applied to G_1, however, a focusing field is created in the cathode–G_1 region. The stream of electrons is focused down so that relatively fewer electrons are collected by the edges of G_3 and by G_2. The cathode-current density is also also reduced by the application of negative bias to G_1. In addition, some reduction in emission area is effected. At sufficiently negative bias, the latter two effects can significantly reduce the emitted current and cancel the gain in beam current. Normally, however, a substantial beam current gain may be obtained with the enlarged G_1 hole geometry and proper G_1 bias.

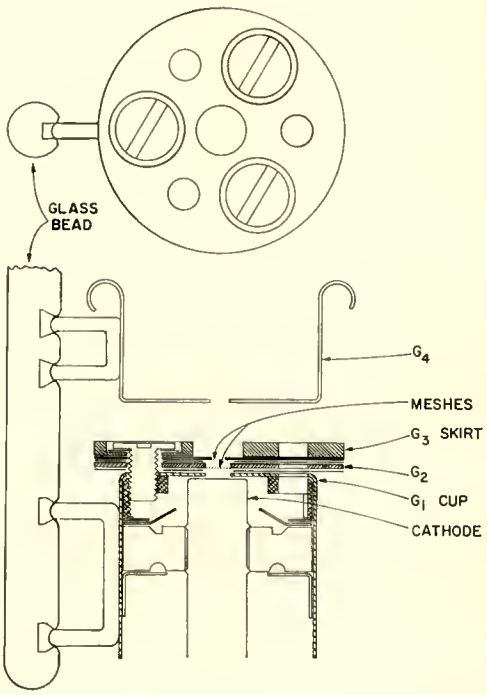


Fig. 11—Experimental construction details.

Increases in the G_2 thickness move all of the control characteristics (electrode currents versus V_{G3}) toward the positive region of the graph. This is illustrated in a comparison of Figure 9 with Figure 2. The curves of Figure 9 were taken with a gun wherein G_2 was .020 inch thick instead of .015 inch.

The relative orientation of the two meshes has some effect upon the characteristic curves. The most favorable operation appears to

result from 45° crossing of equal-pitch meshes. Other orientations produce somewhat differently shaped characteristic curves. Figure 10, for example, shows the operation curves of a gun built with parallel aligned meshes oriented for minimum optical transmission.

None of the gun parameters are particularly critical. In many cases, spacings, hole sizes, and thicknesses may be varied over a large range without significantly affecting gun performance.

Gun Construction

The experimental double-mesh guns employ a stacked disk construction in the modulator section. Mica insulators are used for separators. Figure 11 illustrates the geometry of a typical experimental gun.

The control-grid capacitance is 46 micromicrofarads when this geometry is used. Mechanical construction modifications allow it to be reduced to less than 15 micromicrofarads. The cathode has a capacitance of 6 micromicrofarads to all other electrodes.

Single-Mesh Guns

It is possible to use a G_2 aperture without a mesh covering and yet retain performance similar to that of the double-mesh gun. A few experimental guns of this type have been built. Careful attention to the electron optical design is required to prevent excessive loss of electrons to the G_2 structure. The best design that was investigated exhibited similar efficiency, lower currents, and required higher G_2 voltages and somewhat higher drive voltages than the double-mesh types. The performance of this gun is probably not indicative of that which might be obtained from an optimum design single-mesh gun of this type.

A single-mesh gun may also be made by replacing the G_3 mesh with a thin single aperture disk. In general, this produces a higher drive voltage requirement — typically two to three times more. Guns of this type produce a moderate increase in peak beam current and picture resolution.

Acknowledgment

The author is indebted to Paul Messineo, Ray Shankweiler, and John Mirsh for their invaluable aid in the construction and evaluation of the experimental double-mesh-grid guns.

THERMIONIC ENERGY CONVERTER*

By

K. G. Hernqvist, M. Kanefsky, and F. H. Norman

RCA Laboratories,
Princeton, N. J.

Summary—*The thermionic energy converter is a device for direct conversion of heat into electrical energy. It is a thermionic diode having a high-work-function cathode and a low-work-function anode. The operational characteristics of this device are analyzed. Experimental studies of a thermionic energy converter having a tungsten cathode are described. Cesium vapor introduced into the tube serves two purposes. Cesium atoms ionized at the cathode surface cause space-charge neutralization in the interelectrode space. Cesium condensed at the surface of the cold nickel anode reduces the anode work function. The output voltage is 2.5 volts. Conversion efficiencies of approximately 10 per cent have been measured. Finally, the materials problems and possible applications of thermionic energy converters are discussed.*

Introduction

IT IS well known[1,2] that direct conversion of heat to electricity can be achieved by means of a thermionic diode. Thermal energy supplied to heat the cathode is directly converted into electrical energy which may be utilized in the circuit connected between the anode and the cathode. For high conversion efficiency in such a device, it is necessary to satisfy the following conditions: (1) The difference between the work function of the cathode and the work function of the anode must be high in order to take advantage of the contact difference of potential. (2) The space-charge effects in the interelectrode space must be overcome.

It is the purpose of this paper to describe a converter tube in which these two conditions are met by using cesium vapor in the tube.

Theory of Operation

The operation of the thermionic energy converter is most easily understood if reference is made to the potential-energy diagram for electrons shown in Figure 1. In Figure 1, ϕ_C is the work function of

* Manuscript received April 10, 1958.

[1] L. Malter, E. O. Johnson, and W. M. Webster, "Studies of Externally Heated Hot Cathode Arcs, Part I—Modes of the Discharge," *RCA Review*, Vol. XII, p. 415, September, 1951.

[2] H. Moss, "Thermionic Diodes As Energy Converters," *Jour. Electronics*, Vol. 2, p. 305, January, 1957.

the cathode and ϕ_A is the work function of the anode. When the external circuit between the cathode and anode is open and there is no net charge between the cathode and anode, there is no field between cathode and anode and the potential is constant as shown. By supplying heat to the cathode, some electrons are raised on a potential energy scale from the Fermi level of the cathode to the top of the potential barrier at the cathode surface. Some of these electrons are emitted from the cathode surface and flow to the anode provided space-charge effects in the interelectrode space can be overcome. When captured by the anode, the electrons fall down on a potential energy scale by an amount equal to the work function of the anode, ϕ_A volts. This energy appears as heat in the anode. The remaining potential energy of the electrons when they have reached the Fermi level of the anode, is the contact difference of potential, $\phi_C - \phi_A$. This energy is delivered directly as electrical energy to the circuit connected between the anode and the cathode.

Electrically, the converter acts as a constant-current generator capable of a maximum output voltage equal to $\phi_C - \phi_A$. A theoretical volt–ampere characteristic is shown in Figure 2. Again assuming complete space-charge cancellation in the interelectrode space, the volt–ampere characteristic for negative anode voltages may be written as

$$I_A = I_C \qquad \text{for } 0 \leq |V_A| \leq \phi_C - \phi_A,$$

$$I_A = I_C \exp\left\{\left(\frac{-e}{kT}\right)\Big(|V_A| - [\phi_C - \phi_A]\Big)\right\} \text{ for } |V_A| > \phi_C - \phi_A, \quad (1)$$

where

I_A = anode current,

I_C = saturated cathode emission current,

V_A = anode voltage,

e = electronic charge,

k = Boltzmann's constant,

$\frac{e}{k} = 11600\ ^\circ\text{K} \times (\text{volt})^{-1}$,

T = cathode temperature in °K.

Under the condition $\phi_C - \phi_A \geq kT/e$, the maximum conversion efficiency occurs at an output voltage $= \phi_C - \phi_A$, corresponding to a load impedance of $(\phi_C - \phi_A)/I_C$ ohms. Assuming $\phi_C - \phi_A \gg kT/e$, the maximum conversion efficiency may be written as

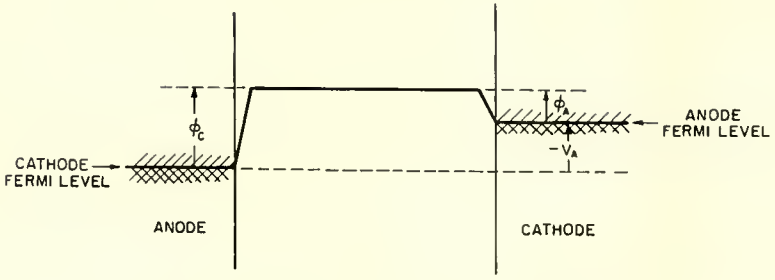


Fig. 1—Potential-energy diagram for electrons.

$$\eta \cong \frac{I_A\,(\phi_C - \phi_A)}{I_A\phi_C + P_L}, \tag{2}$$

where P_L is the sum of all the power losses of the system other than the heat delivered to the anode, $I_A\phi_A$. At high cathode temperatures the dominating loss factor will be the radiation loss. Then $P_L = \epsilon_t\,\sigma T^4\,A_r$, where ϵ_t is the thermal emissivity of the cathode, σ is the Stefan–Boltzmann constant, and A_r is the radiating cathode area. Furthermore the saturated cathode emission current is $I_C = A_e B T^2 \exp\,[-e\,\phi_C/kT]$, where A_e is the electron emitting cathode area, and B is the emission constant. Under the assumption that $I_A = I_C$, the efficiency is

$$\eta \cong \frac{\phi_C - \phi_A}{\phi_C + \dfrac{\epsilon_t\sigma}{B}\dfrac{A_r}{A_e}T^2 \exp\,[e\phi_C/kT]}. \tag{3}$$

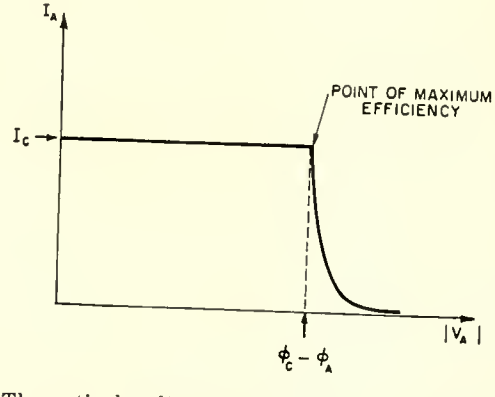


Fig. 2—Theoretical volt-ampere characteristic of thermionic energy converter.

Methods for Space-Charge Reduction

As stated in the previous section, space-charge cancellation in the interelectrode space is necessary for high conversion efficiency. Several methods for reduction of space-charge effects have been suggested or tried in the past. These are listed in Table I. The most straightforward method is that of reducing the cathode-to-anode spacing. Moss[2] has analyzed this principle and shown that it is not practical because spacings as low as 0.01 millimeter are needed for reasonable efficiencies. A second method[3] is that of increasing the particle velocity in the interelectrode space. It is based on the principle of magnetron operation under cutoff conditions. Figure 3 shows an embodiment of the device having the cathode and anode side by side and opposite to

Table I — Methods of Space-Charge Reduction for Thermionic Energy Converter

	References	Remarks
Short Cathode-to-Anode Distance	2	Requires cathode-to-anode spacings $\leqq$ 0.01mm
High Particle Velocity	3	Requires high-voltage supply and magnetic field
Ball-of-fire Discharge	1, 4	Energy for ion generation $\approx$ 2 volts/ampere
Plasmatron Principle	2, 5	Energy for ion generation $\approx$ 1 volt/ampere
Resonance Ionization	6	Energy for ion generation $\approx$ 1/100 volt/ampere

an accelerating electrode. A high voltage is applied to the accelerating electrode, but the electrons are forced to move to the anode by the perpendicular magnetic field. This method requires a high voltage source and a magnet which are not power consuming. However at high emission current densities, impracticably high voltages and magnetic fields may be required.

In a class by themselves are methods using space-charge neutralization by positive ions in the interelectrode space. Malter, Johnson, and Webster[1] obtained a negative arc drop in an externally heated hot-cathode arc operated in the "ball-of-fire" mode. Here a small region in the interelectrode space having a high potential is formed. Ions formed in this region neutralize the negative space charge in the surrounding dark plasma. An electrode potential exceeding the ioniza-

[3] H. Nelson, unpublished work. The authors are indebted to Mr. Nelson for permission to refer to this work performed at RCA in 1937.

tion potential of the gas is required for starting the discharge. From an energy balance calculation[4] it can be shown that a voltage drop approximately equal to $2\ kT_e/e + P_{ex}/I_A$ is required to maintain the discharge. Here T_e is the electron temperature and P_{ex} is the power lost in radiation. This voltage drop may amount to several volts, by which the output voltage is lowered. The conversion efficiency may then be written

$$\eta = \frac{\phi_C - \phi_A - P_{ex}/I_A - 2\,kT_e/e}{\phi_C + P_L/I_A}. \tag{4}$$

Although this method of operation may give a negative anode voltage (by using a tungsten cathode and a barium covered nickel anode),

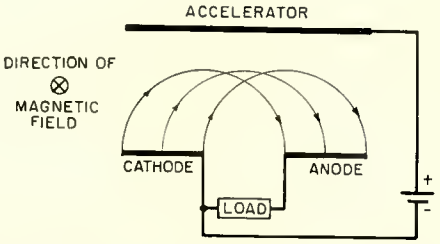


Fig. 3—Schematic diagram showing operation of thermionic energy converter using magnetic deflection.

Equation (4) shows that this process of ion generation is quite inefficient.

A more efficient method of ion generation is by the injection of ions from an auxiliary discharge.[2] This is the principle of operation used in the plasmatron.[5] Results obtained in plasmatron studies indicate that the current in the auxiliary discharge is of the order of 1/100 of the main discharge current. The auxiliary discharge voltage was about 50 volts. In addition, the auxiliary discharge required a cathode heater power of the order of 50 watts per ampere. This adds up to about 1 watt per ampere of main discharge current. This energy must be delivered in the form of electrical energy to the system and could in principle be taken from the output power of the converter. Compared to the ball-of-fire mode of operation, this method of ion generation would be more efficient by a factor of at least 2.

---

[4] K. G. Hernqvist, "Discharge Mechanism of Mercury Pool Arcs," *Jour. Appl. Phys.*, Vol. 27, p. 1226, October, 1956.

[5] E. O. Johnson and W. M. Webster, "The Plasmatron, A Continuously Controllable Gas-Discharge Developmental Tube," *Proc. I.R.E.*, Vol. 40, p. 645, June, 1952.

The third method of ion generation, which is the one used in the energy converter described in this paper, is based on the phenomenon of resonance ionization.[6] A gas whose ionization potential, V_i, is lower than the work function of the cathode, ϕ_C, is introduced into the tube. When the condition $\phi_C > V_i$ is satisfied, neutral atoms striking the hot cathode will be ionized and injected as positive ions into the interelectrode space. In this way a plasma beam which extends from the cathode to the anode is formed. For complete neutralization, the ion density must equal the electron density, or, for a plane parallel geometry,

$$\frac{I_p}{\bar{C}_p} = \frac{I_A}{\bar{C}_e}, \tag{5}$$

where I_p is the ion current, and $\bar{C}_p$ and $\bar{C}_e$ are the average Maxwellian ion and electron velocities, respectively. Assuming the ion and electron temperatures to be equal, Equation (5) may be written

$$I_p = I_A \sqrt{\frac{m_e}{m_p}}, \tag{6}$$

where m_e/m_p is the electron-to-ion mass ratio. Neglecting electron–ion recombination in the interelectrode space, the efficiency becomes

$$\eta = \frac{(\phi_C - \phi_A)\left(1 - \sqrt{\frac{m_e}{m_p}}\right)}{\phi_C + P_L/I_A}. \tag{7}$$

The power lost in ion generation is $(\phi_C - \phi_A)\sqrt{m_e/m_p}$, which may be of the order of a few hundredths of a watt per ampere. This is one or two orders of magnitude better than the previously discussed ion-generation methods. The required gas pressure may be computed as follows: The number of atoms striking the cathode surface per unit time is (1/4) $A_c N_g \bar{C}$ where N_g is the gas density and $\bar{C}$ is the average Maxwellian particle velocity. Assuming an ion yield of the order of unity, the ion current becomes

[6] A. V. Engel and M. Steenbeck, *Elektrische Gasentladungen*, Julius Springer Verlag, Berlin, Germany, 1932, p. 130.

$$I_p = -\frac{e}{4} A_c N_g \bar{C}. \tag{8}$$

Combining Equations (5) and (8) yields

$$N_g = 4 \frac{\bar{C}_p}{e \bar{C} \bar{C}_c} \frac{I_A}{A_c}, \tag{9}$$

or, expressed in terms of pressure p in mm of Hg,

$$p = 4.12 \times 10^{-6} \sqrt{T_g}\, j_A, \tag{10}$$

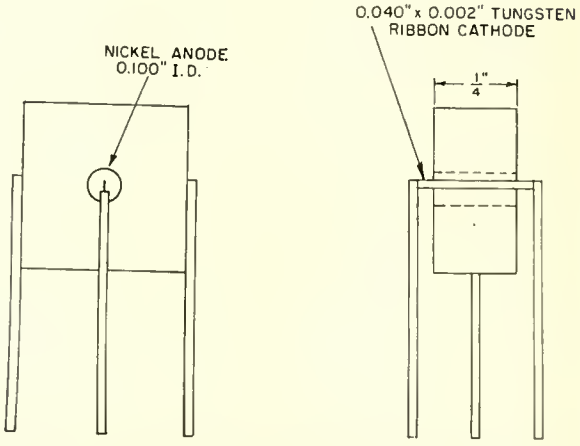


Fig. 4—Schematic drawing showing electrode structure of thermionic energy converter.

where T_g = the gas temperature in °K and j_A is the electron current density in amperes per square centimeter.

Experimental Studies of a Thermionic Energy Converter

Description of tube and measurement procedure

The experimental tube is shown in Figures 4 and 5. It is a coaxial diode having a directly heated tungsten ribbon cathode and a nickel anode. The tube contains cesium vapor. A cesium reservoir is introduced into the tube after tube processing by the reduction of cesium dichromate. The tube contains a sufficient amount of cesium to yield the vapor pressure in equilibrium with liquid cesium for the range of ambient temperatures used.

As noted earlier, the cesium in the tube serves two purposes. (1) It provides an ion source due to resonance ionization at the cathode. Since $\phi_C = 4.52$ volts for tungsten and $V_i = 3.89$ volts for cesium, the condition for resonance ionization $\phi_C > V_i$ is satisfied. (2) Cesium condenses on the cold anode and provides a coverage which reduces the anode work function to a value approximately equal to the work function of cesium. Thus $\phi_A = 1.81$ volts. For sufficient coverage it

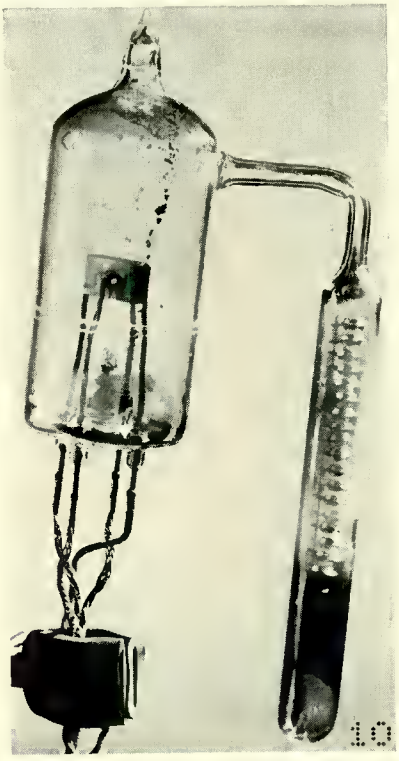


Fig. 5—Experimental energy converter.

is essential that the anode be kept near the ambient temperature.

The vapor pressure of cesium in the tube is determined by the ambient temperature (temperature of the tube walls). Equation (10) therefore sets a lower limit on the ambient temperature for a certain electron current density. From vapor pressure data on cesium,[7] Equation (1) has been plotted in Figure 6, which also shows the electron mean free path. It is seen that for current densities of interest (up

[7] J. B. Taylor and I. Langmuir, "Vapor Pressure of Caesium by the Positive Ion Method," *Phys. Rev.*, Vol. 51, p. 753, May, 1937.

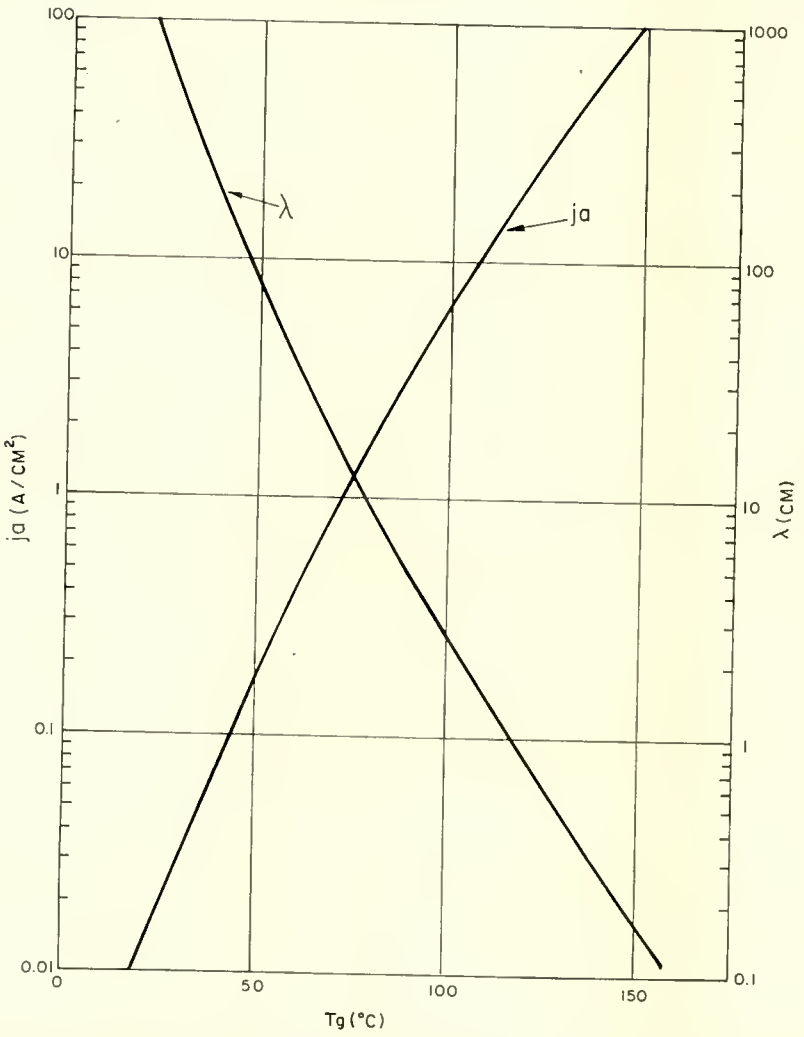


Fig. 6—Electron mean free path, λ, and maximum neutralized current density, j_A, in cesium vapor as a function of ambient temperature T_g in °C.

to 10 amps/cm²) ambient temperatures up to 110°C are required. This was achieved experimentally by placing the tube in an oven. Once the optimum temperature is reached, the anode current is quite insensitive to temperature, until at higher temperatures the current decreases due to back scattering of electrons. Although the cathode-to-anode distance is not critical, it is desirable to choose this distance shorter than the mean free path in order to reduce back scattering. In the experimental

tube the average cathode-to-anode distance was about 1.2 millimeters.

The circuit used for performance studies of the energy converter is shown schematically in Figure 7. Energy is supplied to the cathode in the form of joule heat. It is essential that there be no voltage gradient along the cathode during measurements of anode current and voltage. This is achieved by heating the filament by a half-wave rectified 400-cycle a-c current. The anode current and anode voltage are measured during the off part of the a-c cycle by displaying the volt–ampere characteristic on an oscilloscope. The grid of the oscilloscope is modulated to give high brightness only during the off part of the

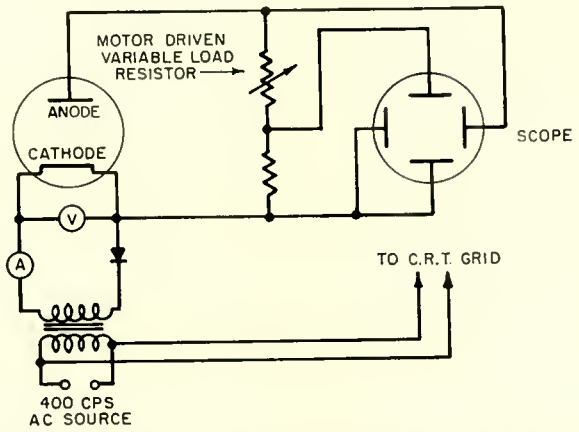


Fig. 7—Schematic diagram showing automatic display of volt-ampere characteristic for thermionic energy converter.

a-c cycle. The volt–ampere characteristic is displayed by varying the load resistor from zero to a high value.

The temperature of the cathode was measured with an optical pyrometer and cross checked with the data on current density.[8] A further temperature check was obtained from resistance measurements on the filament.

Performance measurements on the thermionic energy converter

A typical volt–ampere characteristic is shown in Figure 8. It is seen to follow the predicted characteristic (Equation (11)) quite well. The anode current is constant up to about 2.5 volts, which is approximately equal to the contact difference of potential $\phi_C - \phi_A = 2.7$

[8] H. A. Jones and I. Langmuir, "Characteristics of Tungsten Filaments as Functions of Temperature," *General Electric Review*, Vol. 30, p. 310, June, 1927.

volts. Application of an external positive voltage to the anode does not appreciably change the maximum current. This shows that the condition $I_A = I_C$ is very nearly fulfilled.

The ambient temperatures necessary for space-charge neutralization were experimentally determined. Figure 9 is a plot of emission current versus ambient temperature with the emission current density as parameter. Current saturation occurred at slightly higher temperatures than predicted by Equation (10) (Figure 6). This is believed to be due to the fact that a slight voltage gradient existed along the filament as a result of the part of the discharge current which had to flow to one end of the filament. This voltage gradient causes a transverse flow of ions and corresponds to an ion drainage. A falling off of the anode current because of back scattering was not noticeable for a fairly large temperature range. This indicates that the ambient temperature is not a critical parameter.

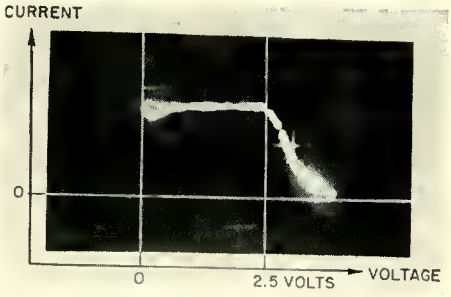


Fig. 8—Photograph showing experimental volt–ampere characteristic.

Measurements of the maximum conversion efficiency (at 2.5-volt anode voltage) as a function of temperature were made. The results are shown in Table II and Figure 10. The theoretical relation

$$\eta = \frac{2.5}{4.52 + P_{HL}/I_C} \tag{11}$$

is shown in Table III and also in Figure 10, using published[8] values of P_{HL}/I_C, and assuming $A_r = A_c$. As seen from Figure 10, agreement between theory and experiment is quite good. Measurements could not be extended beyond $\eta = 10.5$ per cent because of current instabilities. The reason for these instabilities is believed to be the presence of a voltage gradient along the cathode as a result of discharge current flowing through the filament. This voltage gradent causes an appreciable disturbance in the ion flow. This difficulty may be eliminated by the use of a heavier filament.

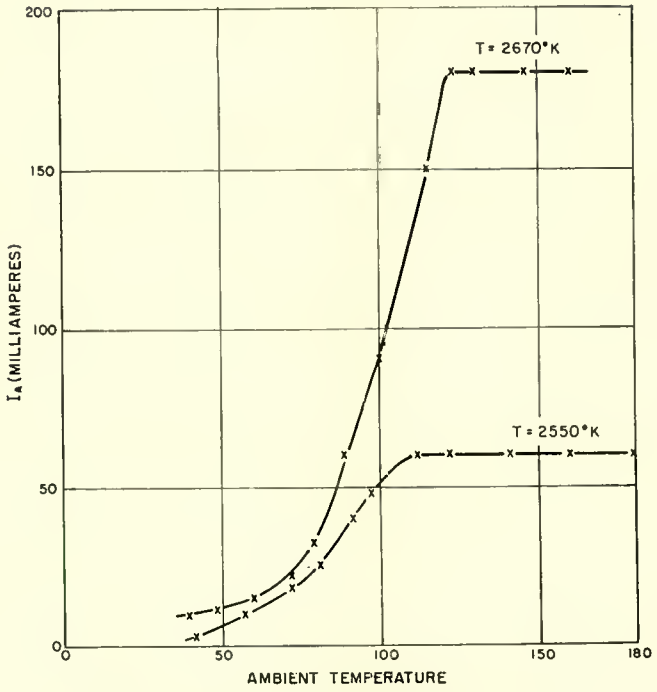


Fig. 9—Diagram showing anode current I_A as a function of ambient temperature in °C.

Tube life

As seen in Table III, the efficiency increases with cathode temperature. However, the evaporation rate of cathode material also increases with temperature. Table III shows published[8] values of the evaporation rate of tungsten given in millimeters of thickness of cathode material evaporated in 1000 hours. Clearly, a compromise must be

Table II—Efficiency Measurements on Thermionic Energy Converter

T °K	Heater Power watts	V_A volts	I_A amperes	η %
2350	8	2.5	0.009	.28
2550	11	2.5	0.06	1.4
2670	13.5	2.5	0.18	3.3
2750	15.4	2.5	0.30	4.9
2800	17.5	2.5	0.50	7.1
2860	22	2.5	0.35	8.5
2910	24	2.5	1.00	10.4

made between the desired efficiency and the life of the converter. Exactly how much material can be evaporated before the operation of the tube is seriously disturbed is difficult to estimate. Experience with

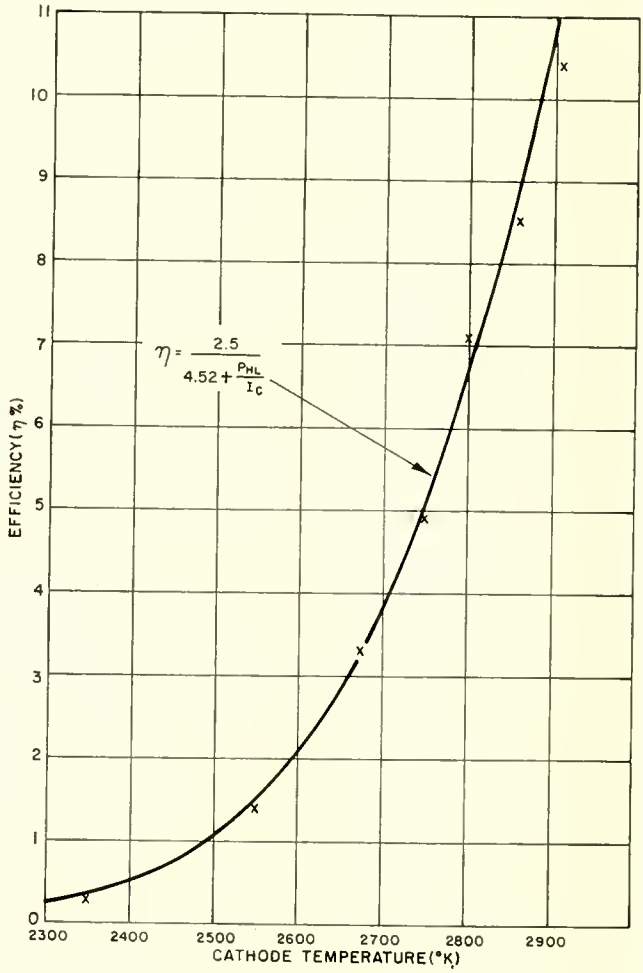


Fig. 10—Diagram showing conversion efficiency as a function of cathode temperature. Solid line shows theoretical relation. Crosses indicated experimental results.

the present tube indicates that the anode work function is not seriously affected by the tungsten deposit. An ultimate evaluation can only be done by actual life tests of a converter in its specific application. The conventional technique of lowering the evaporation rate by introducing

a buffer gas[9] is believed not to be successful in this case, because the accompanying back scattering of electrons would lower the efficiency.

Materials Problems of the Thermionic Energy Converter

As discussed in the previous section, operation of the thermionic energy converter at high conversion efficiencies using a tungsten cathode is incompatible with long tube life. Also, the high cathode temperature required for high efficiencies limits the application of the converter. As seen from Equation (3), high conversion efficiencies at

Table III—Thecretical Efficiency and Cathode Evaporation Rate for Thermionic Energy Converter Using Tungsten Cathode

T °K	P_{HL}/I_C watts per ampere	j_A A/cm^2	$\eta = \frac{2.5}{4.52 + P_{HL}/I_C}$ %	Evaporation Rate mm per 1000 hours
2300	1158	0.041	0.215	0.00015
2400	498	0.116	0.5	0.0008
2500	235	0.298	1.05	0.0038
2600	117	0.716	2.06	0.0164
2700	61.2	1.63	3.81	0.0592
2800	33.2	3.54	6.63	0.209
2900	18.9	7.31	10.7	0.642
3000	11.3	14.2	15.8	1.81
3100	7.03	26.5	21.7	4.96
3200	4.47	47.8	27.9	12.4
3300	2.9	84.4	33.8	29.9
3400	1.97	142	38.6	66.6

lower temperatures call for a lower anode work function and/or a less power consuming emitter. Of particular importance is the later requirement. For a visualization of the cathode problem, assume an emitter having all properties the same as tungsten except for the work function. Figure 11 is a plot of Equation (3) assuming $\phi_A = 1.8$ volts and $A_r = A_o$, and with the cathode temperature, T, as a parameter. It is seen that for each temperature the efficiency η has a maximum for a certain value of the cathode work function, ϕ_C. The temperature for which the efficiency curve has a maximum at $\phi_C = 4.52$ volts is prohibitively high for tungsten. A cathode material having a lower work

[9] G. R. Fonda, "Evaporation of Tungsten Under Various Pressures of Argon," *Phys. Rev.*, Vol. 31, p. 260, February, 1928.

function[10] than tungsten would allow operation at high efficiencies and lower temperatures. Possible materials are: thoriated tungsten or a dispenser type cathode (L-cathode).

Application of Thermionic Energy Converters

The high temperature requirements of the converter using a tungsten cathode appears to limit its applications to solar energy applications. Temperatures up to 3000°K are readily available in

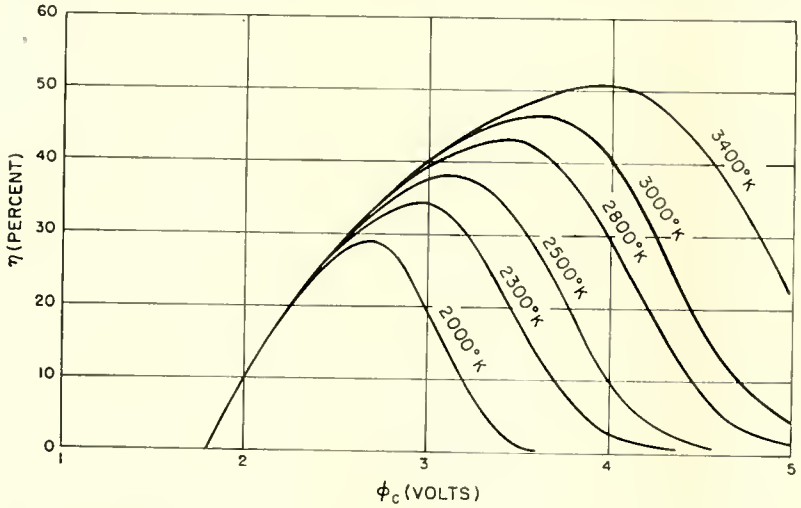


Fig. 11—Conversion efficiency, η, as a function of cathode work function, ϕ_c, with cathode temperature, T, as a parameter.

solar furnaces using parabolic mirrors for energy concentration. Because a single converter tube has a high power handling capability, the thermionic energy converter may compare favourably with conventional thermoelectric devices. For solar energy applications the absorption coefficient of the cathode must be considered. The efficiency calculations used in this paper refer to actual power delivered to the cathode.

Applications of the converter to other power sources (heat from a flame or nuclear energy) will presumably require operation at much lower temperatures. This may require lower work functions both for the cathode and the anode.

---

[10] Studies by W. Knauer of RCA Laboratories have shown that resonance ionization of cesium occurs with low work function matrix-type cathodes.

EFFECT OF COLLECTOR POTENTIAL ON THE EFFICIENCY OF TRAVELING-WAVE TUBES*

By

H. J. Wolkstein

RCA Electron Tube Division,
Harrison, N. J.

Summary—*This paper describes methods which can be used to increase the over-all efficiencies of traveling-wave tubes. These methods permit the collector potential to be depressed substantially below the synchronous helix voltage, and provide corresponding reductions in collector-element dissipation without appreciable degradation of r-f performance. Although these conditions are accompanied by secondary-electron emission which increases helix dissipation, the effects of such emission can be minimized by the use of high-resistance collector coatings, hollow suppressor electrodes, magnetic shielding of the collector, and other means.*

This paper also describes a method which can be used to estimate the minimum potential to which an axially symmetrical collector electrode can be depressed without reduction of beam current. This method introduces the concept of "beam-flow perveance," and is applicable to both Brillouin-flow and parallel-flow electron beams. The minimum collector potentials determined by this method are compared with the measured values obtained in several developmental traveling-wave tubes.

Introduction

THE gain–bandwidth products of traveling-wave amplifier tubes are extremely high compared with those of other types of amplifiers; small-signal gains of more than 30 decibels over 2 to 1 frequency ranges in the S and X bands are easily achieved. The excellent gain and bandwidth characteristics of these tubes have received so much developmental attention that they have virtually overshadowed some of the less desirable features, particularly the low collector efficiencies, which are in the order of 10 per cent or less. Although this low efficiency is of no consequence in low-power traveling-wave tubes, it becomes an increasingly important consideration in higher-power types because the attendant d-c losses in the collector increase the cost, size, and complexity of power supplies, blowers, solenoids, and associated equipment. Although this low collector efficiency is not an inherent limitation, it has not received either the consideration or the developmental effort given other characteristics of traveling-wave tubes.

* Manuscript received January 28, 1958.

This paper describes some of the methods pursued in the attempt to increase the collector efficiency of traveling-wave tubes, and the results obtained. In these methods the d-c collector potential is reduced to the value which provides the lowest collector dissipation possible without appreciable reduction of the beam current and r-f power output.

Collector-Potential Depression

To increase the efficiency of a traveling-wave tube by depression of its collector potential presents many interesting problems, all of which seem to manifest themselves in increased helix dissipation.

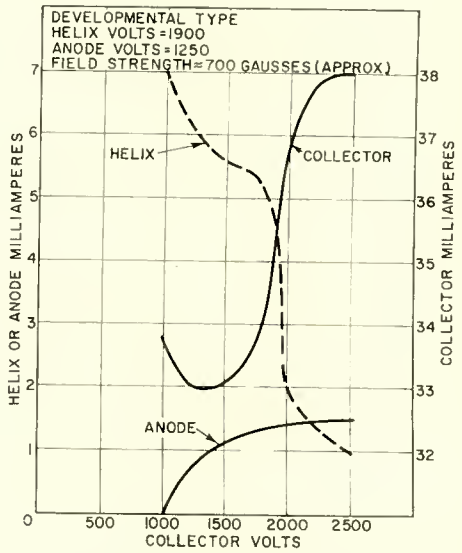


Fig. 1—Current characteristics of a developmental traveling-wave tube with depressed collector potential.

Extensive measurements were made on a number of traveling-wave tubes to determine the maximum amount the collector potential could be reduced without exceeding helix-dissipation limits in the absence of r-f input, i.e., with no beam bunching. Figure 1 shows the currents intercepted by the helix, collector, and anode as functions of collector potential. Depression of the collector potential below the value of helix voltage produces a sharp rise in helix current and a corresponding decrease in collector current. The out-of-phase relationship between helix current and collector current in the vicinity of the crossover point ($E_{coll} = E_{helix}$) strongly suggests the emission of secondary

electrons by the collector. This hypothesis is strengthened by the fact that the reduction of the helix voltage delays the inception of the sharp rise in helix current as the collector potential is decreased (see Figure 2). Unfortunately, a substantial reduction of helix potential to permit a correspondingly greater reduction of the collector potential cannot be tolerated for r-f operation because the helix voltage must have a relatively high minimum value to provide the required synchronous interaction between the electron beam and the growing electromagnetic wave. Furthermore, the helix is normally operated at the highest possible d-c potential in order to impart maximum kinetic energy to the electron beam, a portion of which is extracted in the form of r-f power output of the tube.

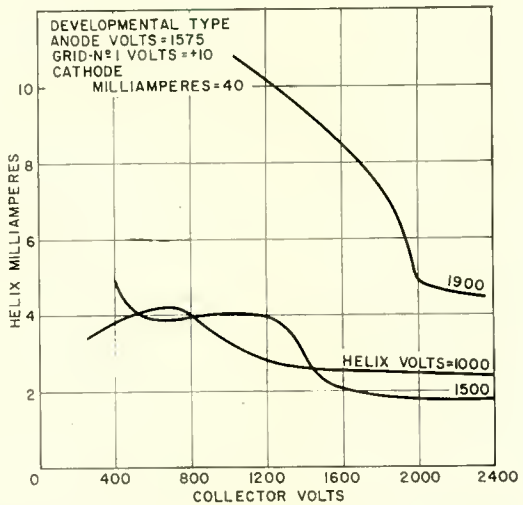


Fig. 2—Helix voltage–current characteristics as related to depressed collector potential.

Inhibition of Secondary Emission

There are a number of ways to reduce the effects of the secondary emission which occurs at depressed collector potentials. One is the use of a ring-shaped repeller electrode between collector and helix, as shown in Figure 3a. This repeller, when connected to an independent adjustable source of voltage, provides a means for the establishment of a retarding field for secondary electrons. The current intercepted by the helix as a function of collector potential and repeller-electrode voltage is shown in Figure 3b. These curves indicate that the retarding field of the repeller can be made effective over a wide range of collector

potential. For example, with the helix at a fixed potential of 1,900 volts and with 700 volts on the repeller, adequate suppression of secondaries is achieved to below 800 volts on the collector. Note, however, that there is no effective secondary suppression when the repeller is tied to the collector.

Drastic suppression of secondary electrons has been achieved by magnetic shielding of the bombarded portion of the collector. Three types of shielding employed in developmental tubes are shown in Figure 4. This shielding terminates the magnetic flux short of the

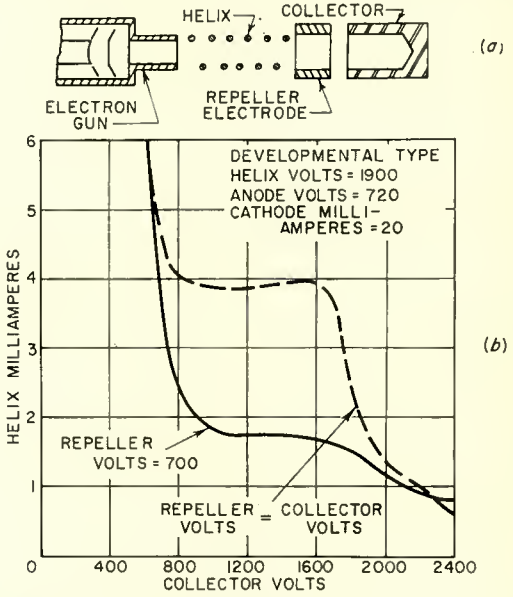


Fig. 3—The effects of a repeller electrode on depressed-collector characteristics.

bombarded region; as a result, a high percentage of the secondary electrons launched by the collector are not in confined flow and are not returned to the region of the helix. This principle was successfully applied in the periodic focusing magnets used by Siekanowicz and Sterzer.[1] The last pole piece of their structure (shown in Figure 4c) is terminated short of the active collector region. The number of released secondary electrons trapped could undoubtedly be increased by the use of Faraday electrostatic shielding in conjunction with magnetic flux termination.

[1] W. W. Siekanowicz and F. Sterzer, "A Developmental Wide-Band, 100 Watt, 20 DB, S-Band Traveling-Wave Amplifier Utilizing Periodic Permanent Magnets," *Proc. I.R.E.*, Vol. 44, p. 55, January, 1956.

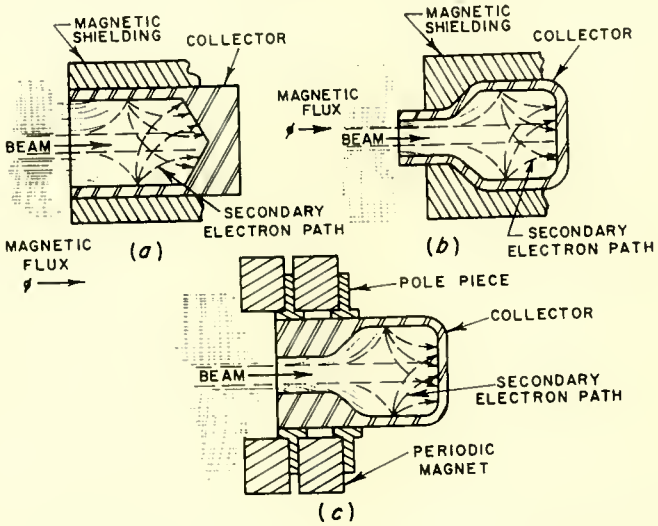


Fig. 4—Various collector configurations to inhibit secondary electrons.

Curves showing the magnitude of the current intercepted by the helix at depressed collector voltage with and without the type of magnetic shielding described above are shown in Figure 5. The curve for the magnetically shielded collector shows that in the absence of beam bunching, the collector potential may be as little as one third

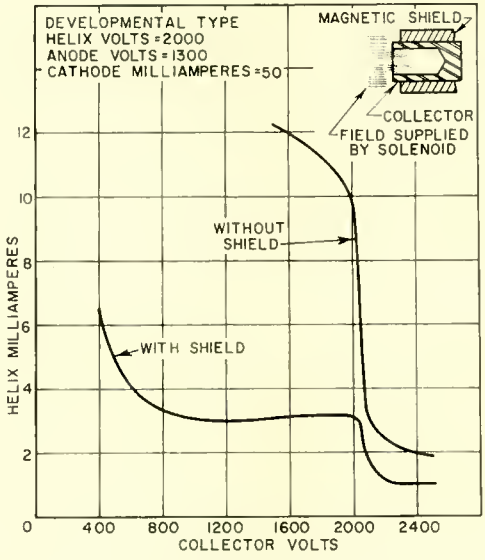


Fig. 5—The effects of magnetic shielding on helix-current characteristics.

of the synchronous helix voltage. The marked improvement obtained by the use of magnetic shielding is evident.

Decreased helix dissipation can also be achieved by a reduction in the secondary-emission ratio of the collector. Such a reduction has been achieved by proper pretreatment of the collector-electrode surface. Figure 6 (from Spangenberg[2]) shows the secondary-emission characteristics of various metals as functions of beam-impact potential. Most of the traveling-wave tubes tested employed copper collector elements for thermal reasons. According to Figure 6 and other published data, the maximum ratio of secondary to primary electrons for vacuum-deposited copper at an impact potential of approximately 600 volts is

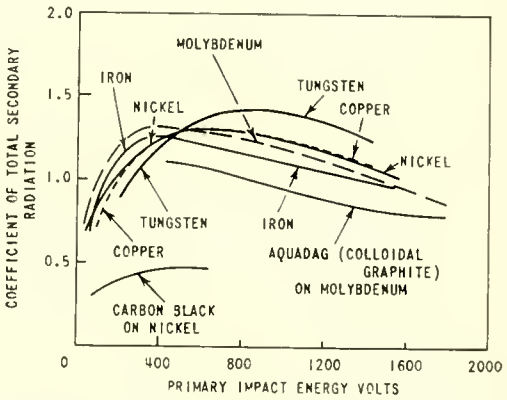


Fig. 6—Secondary electron characteristics of various elements.

approximately 1.28. The presence of superficial oxide layers or other contaminants, however, can increase this ratio significantly, causing greater helix dissipation with depressed collector potential. In some of the tests conducted, carbonized and lampblack-coated collectors having low secondary-emission ratios were used, and gave better results than vacuum-deposited copper.

Beam-Flow Discontinuity

Drastic reduction of secondary emission in a developmental traveling-wave tube by the methods described above allowed the collector potential of the tube to be depressed to approximately 1/6 to 1/3 the helix voltage, at beam-current magnitudes ranging from 10 to 40

[2] K. R. Spangenberg, *Vacuum Tubes*, McGraw-Hill Book Co., Inc., New York, N.Y., 1949, p. 50.

milliamperes. It was found, however, that for each level of beam current there was a critical value of the hollow-collector voltage below which limiting of beam-current flow occurred and resulted in excessive helix dissipation. The effects of this phenomenon in a traveling-wave tube in Brillouin flow are shown in Figure 7. Note that the point of beam-current-flow limitation for a given collector potential manifests itself by a reduction in cathode current and an abrupt increase in helix current. Furthermore, flow discontinuity takes place at succes-

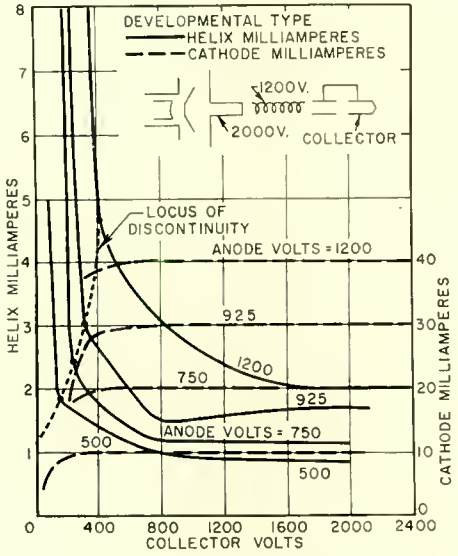


Fig. 7—Experimental observation of helix-current discontinuity with depressed collector potential.

sively higher collector potentials as the cathode current is increased. Studies by Pierce[3] and Smith[4] suggest that this beam-flow discontinuity determines the maximum current that can be sustained in a magnetically confined, axially symmetrical beam. These studies also suggest that the discontinuity is due to the combined effects of voltage gradients which depress the potential at the center of the beam to low values (see Figure 8). The equations for Brillouin-flow and parallel-flow beams given below indicate the minimum potential to which a hollow collector electrode can be depressed with a fixed magnitude of beam

[3] J. R. Pierce, *Theory and Design of Electron Beams*, D. VanNostrand Co., Inc., Princeton, N.J., 1949, pp. 153-159.

[4] L. P. Smith and P. L. Hartman, "The Formation and Maintenance of Electron and Ion Beams," *Jour. Appl. Phys.*, Vol. 11, p. 220, March, 1940.

current, assuming a beam of uniform current density and the application of the proper axial magnetic flux.

For Brillouin flow

$$V_b = I_m^{2/3} (1{,}157 + 1{,}544 \ln b/a) ; \quad (1)$$

when $b/a = 1$,

$$I_m = 25.4 \times 10^{-6} V_a^{3/2}. \quad (2)$$

For parallel flow

$$V_b = I_m^{2/3} (984 + 1{,}262 \ln b/a) ; \quad (3)$$

when $b/a = 1$,

$$I_m = 32.4 \times 10^{-6} V_a^{3/2}, \quad (4)$$

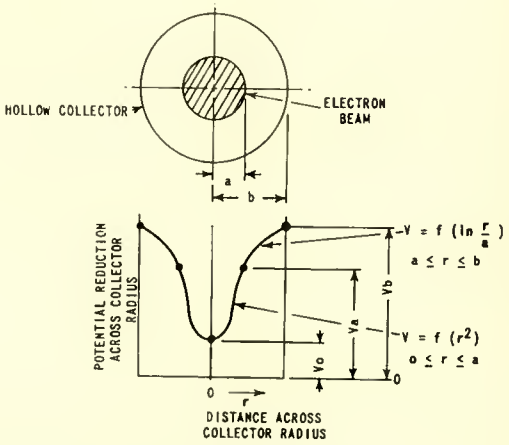


Fig. 8—Space-charge potential reduction effects for Brillouin and parallel-flow electron beams.

where V_b is the potential of the hollow electrode, in volts,
I_m is the maximum possible beam current, in amperes,
V_a is the potential at the beam boundary, in volts.

The derivations of these equations are given in the Appendix.

The solution in each case was derived from the equation for the maximum current for a given beam boundary voltage. The solution was then applied to the coaxial system comprising a hollow collector electrode partially filled by an electron beam. As indicated by the equations, a decrease in the potential of the hollow electrode, V_b, beyond the critical value results in an abrupt reduction of beam flow and reflection of the beam back into the helix region.

It is interesting to note that the constant coefficient of Equations (1), (2), (3), and (4) has the dimensions of amp/volt$^{3/2}$, and, therefore, is dimensionally identical to the coefficient of the equation for the perveance (K) of a space-charge-limited diode. This identity suggests that there is a value of "beam-flow perveance" for the maximum current that can be sustained at a given collector voltage, beam diameter, and collector diameter.[5]

Maximum beam-flow perveance is realized when the ratio of collector diameter to beam diameter is unity, because at this ratio there is no logarithmic voltage drop between the beam boundary and the collector. A ratio of unity, therefore, minimizes the collector potential required to sustain a given value of beam current; this performance is consistent with greater efficiency. In the application of this principle, however, the effects of r-f beam bunching on the diameter of the beam must be considered.

The collector efficiency of a traveling-wave tube is also limited by the very phenomenon which produces amplification, that is, the interaction between a wave on the helix and the coupled electron beam. The interchange of energy between wave and beam and the resulting increase in wave amplitude produces an electron velocity "spread" within the beam. The greatest reduction in velocity takes place at the periphery of the beam, where the beam is most closely coupled to the growing wave. This phenomenon reduces both the beam voltage and the space-charge voltage within the beam, and prohibits depression of the collector potential below the net beam voltage corresponding to the maximum reduction in the electron velocity. This observation suggests that the collector efficiency can be increased by the use of a velocity-sorting electron collector. This paper, however, considers only the effects of a reduction in space-charge potential.

The collector efficiency (η) at the point of beam-flow discontinuity is related to the collector dissipation in the manner indicated by Equation (5a).

$$\eta = \frac{P_{RF}}{P_C + P_{RF}}, \tag{5a}$$

$$\eta = \frac{I_K V_H \eta_e}{\dfrac{(I_K)^{5/3}}{K^{2/3}} + I_K V_H \eta_e}, \tag{5b}$$

[5] The term "Beam Flow Perveance," which is used herein is intended to define the current/voltage 3/2 relationship of a drift area.

$$\eta = \frac{1}{\frac{1}{\eta_e}\frac{V_C}{V_H} + 1}. \qquad (5c)$$

where P_{RF} is the r-f output power,
P_C is the power dissipated by the collector,
V_C is the collector voltage,
V_H is the helix voltage,
I_K is the cathode current, and
η_e is the approximate electronic efficiency.

Because the kinetic energy of the beam is reduced by an amount equivalent to the energy extracted, the collector voltage required when r-f input is applied is greater than that required to sustain a fixed value of beam current in the absence of bunching. If the approximate electronic efficiency, η_e, the synchronous helix potential, V_H, and the flow perveance, K, are known, the maximum collector efficiency is readily determined from Equation (5c) which expresses collector efficiency in terms of the ratio of minimum tolerable collector potential to synchronous-helix voltage. Collector efficiency as a function of this ratio for electronic efficiencies of 5, 10, and 20 per cent is plotted in Figure 9. These curves suggest that in the absence of beam bunching a collector efficiency of 50 per cent is attainable if the collector can be depressed to one-fifth the synchronous helix voltage. An electronic efficiency of 20 per cent is assumed for this condition.

The efficiency can be increased still further by reduction of the effects of space charge along the beam radius. The degree to which the collector potential can be depressed, and, consequently, the efficiency increased, is limited by the potential at the center of the beam. As shown previously in the discussion of the completely beam-filled collector, the potential at the center of the beam is determined by the collector voltage, the beam current, and the parabolic potential drop across the beam. Figure 10 shows one type of modified collector utilized to reduce space charge within the beam, thus permitting greater depression of the collector voltage. The modified collector employs a symmetrical axially located electrode and magnetic shielding of the type previously discussed. This electrode establishes an area having the same potential as the collector, but which is not affected by reductions of space-charge potential. Curves showing the effects of this modified collector on the current intercepted by the helix, as a function of collector voltage are also given in Figure 10. These curves clearly

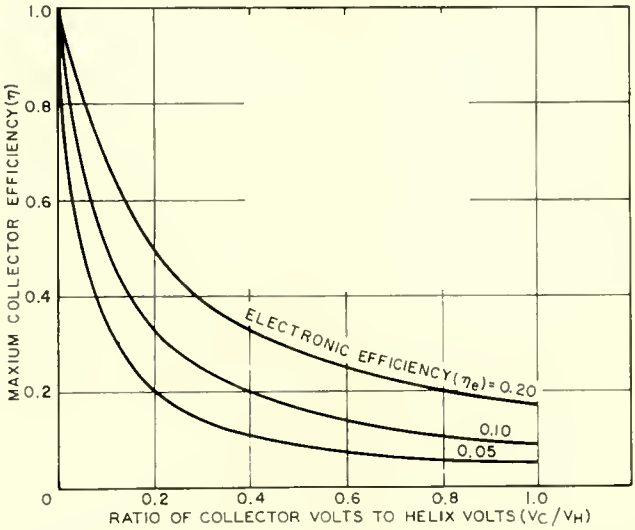


Fig. 9—An estimate of the maximum collector efficiency related to both the electronic efficiency and the ratio of collector to helix potentials.

indicate the effectiveness of the central electrode in delaying the abrupt rise in helix-intercepted current which indicates a drop in collector current. The addition of this central electrode, therefore, should make it possible to achieve high collector efficiency during r-f operation.

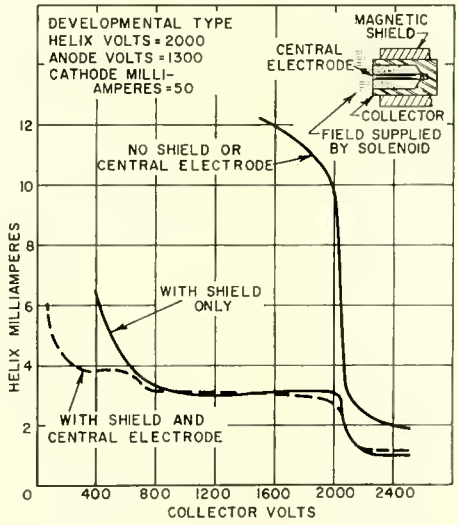


Fig. 10—The effects of a central electrode on helix-current characteristics.

Figure 11 shows the collector efficiency obtained in a developmental traveling-wave tube employing only partial shielding of the collector. The collector employed in this tube had a b/a ratio greater than three, and did not contain any provision for reducing the effect of space charge. Note that the highest collector efficiency obtained (33 per cent) is three times as great as that obtained with the collector at 2,500 volts. Even higher collector efficiency could have been achieved if a collector-diameter to beam-diameter ratio of unity had been employed.

The curve for helix current (no r-f) in Figure 11 shows that in the absence of r-f beam bunching, the minimum collector potential is decreased by approximately 340 volts. The higher collector voltage

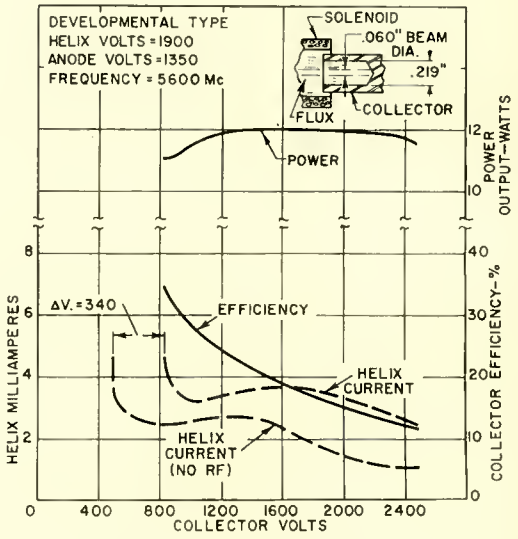


Fig. 11—Collector efficiency of a developmental traveling-wave tube.

required when beam bunching occurs is an approximate measure of the voltage component of the kinetic energy extracted from the electron beam. In this case, the measured r-f energy at saturation was 11 watts, and the apparent additional d-c loss sustained by the collector supply was 12.2 watts. The increase in apparent power imparts to the beam a velocity increment exactly equal to the decrease in velocity which results when r-f energy is extracted.

Conclusions

It has been shown that to achieve higher efficiency in a traveling-wave tube by reduction of collector potential, it is necessary to inhibit

secondary-electron emission. It has also been shown that if a portion of the collector is immersed in a magnetic field, the beam-flow perveance must be adequate to sustain maximum beam current at depressed collector voltage, and that the potential-reduction effects of space charge can be minimized by the use of a fixed-potential region at the center of the coaxial beam-collector system.

Application of these findings has made it possible to increase collector efficiency from the 10 per cent normally tolerated to approximately 35 per cent. This increase was achieved without any attempt to optimize conditions; it should be possible to achieve efficiencies of 40 per cent or more with optimization.

Acknowledgment

The author wishes to express appreciation for the encouragement and guidance given by H. Jenny and M. Nowogrodzki, and the invaluable advice supplied by W. W. Siekanowicz. He also wishes to acknowledge the help of others in the microwave group, especially G. Novak and R. Pekarowitz.

Appendix I

Beam Flow Discontinuity for Brillouin-Focused Beams

In Brillouin flow, an electron beam is constrained so that it has a constant radius (a) determined by the value of an axial magnetic field (B) and the voltage (V_0) at the beam axis.[3] In the case of an axially symmetrical beam completely filling a hollow collector, the Brillouin flow current is given by

$$I = \frac{\pi \epsilon \eta^{3/2} B^2 V_0^{\frac{1}{2}} a^2}{\sqrt{2}}. \tag{6}$$

The voltage at the beam axis (V_0) is related to the voltage at the beam boundary (V_a) by

$$V_0 = V_a - \frac{\eta B^2 a^2}{8}. \tag{7}$$

The geometry upon which Equation (7) is based is shown in Figure 12. When Equations (6) and (7) are combined, the equation for Brillouin flow becomes

$$I = \frac{\pi\epsilon\eta^{3/2}B^2a^2}{\sqrt{2}}\left(V_a - \frac{\eta B^2a^2}{8}\right)^{\frac{1}{2}}. \tag{8}$$

Equation (8) indicates that current flow in a completely filled hollow tube or collector ceases if the beam boundary voltage (V_a) is reduced to a value such that

$$V_a = \frac{\eta B^2a^2}{8}. \tag{9}$$

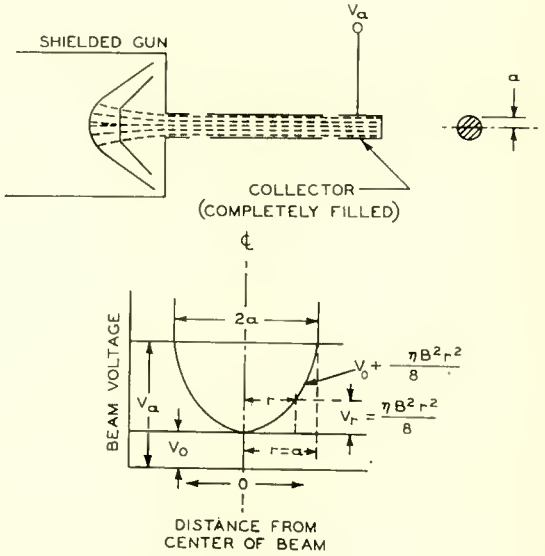


Fig. 12—Variation of potential along the beam radius for Brillouin flow.

Similarly, V_o becomes zero when the equality of Equation (9) is satisfied. The value of V_a which yields the maximum possible beam-current flow when the values for the magnetic field (B) and beam radius, (a) in Equation (8) are maximized, is given by

$$V_a = \frac{3}{16}\eta B^2a^2. \tag{10}$$

The maximum beam current that can be sustained by a given beam-boundary voltage (V_a) in a magnetic field, therefore, is:

$$I_m = 25.4 \times 10^{-6}\, V_a^{3/2}. \tag{11}$$

The potential V_a in Equation (11) corresponds to the voltage of the hollow tube which is completely filled by the electron beam and is employed in the discussion of the general case later in this Appendix. The fact that the constant coefficient of Equation (11) has the same dimensions as that of the equation for a space-charge-limited diode ($K = I/E_b^{3/2}$) suggests that for the case of the completely filled hollow tube, the maximum "beam flow perveance" is 25.4 microamps/volts$^{3/2}$.

In ideal Brillouin flow the electron beam enters the restraining magnetic field without radial velocity components, so that the axial and angular velocities of the beam are constant along the beam radius. In Brillouin flow, therefore, the axial velocity of the entire beam ($\dot{z}$) depends only on the voltage at the center of the beam (V_o) and is given by

$$\dot{z} = \sqrt{2\eta V_o} = \sqrt{2\eta\left(V_a - \frac{\eta B^2 a^2}{8}\right)}, \tag{12}$$

and the tangential electron velocity at the beam periphery (v_θ) is given by:

$$v_\theta = -\sqrt{2}\,\dot{z} = -2\sqrt{\eta V_o}. \tag{13}$$

Equation (13) indicates that when the strength of the magnetic field, the axial velocity of the beam, and the beam radius are held constant, reduction of the potential of the hollow electrode (V_a) below the critical value given by Equation (11) should result in complete collapse of beam flow through that electrode. Any attempt to exceed the 25.4 microamps/volts$^{3/2}$ maximum value of beam-flow perveance given in Equation (11), however, results in an abrupt departure from Brillouin-flow conditions. This departure, which occurs because the increased negative space charge tends to reduce the potential along the beam radius to that of a virtual cathode, manifests itself in reduced total beam current, and appears to produce a partial reversal of beam-current flow. Equation (7) for the voltage appearing at the beam axis (V_o) may be written to accommodate the maximum possible beam current in terms of the minimum beam boundary potential (V_a) given by Equation (10);

$$\begin{aligned} V_o &= V_a - \frac{\eta B^2 a^2}{8}, \\ &= V_a - \frac{2}{3} V_a = \frac{1}{3} V_a. \end{aligned} \tag{14}$$

The uniform axial beam velocity ($\dot{z}$) can, therefore, be stated in terms of the beam boundary voltage:

$$\dot{z} = \sqrt{2\eta V_o} = \sqrt{\frac{2}{3}\eta V_a}. \tag{15}$$

A more general case of beam-flow discontinuity using the same system (magnetically focused beam with shielded electron gun) considers the condition where the beam flow does not completely fill the hollow coaxial tube (see Figure 8). The maximum beam current that can be sustained under these conditions depends on the potential of the confining electrode, the potential gradient of the coaxial system to the beam boundary, and the reduced potential along the radius of the beam due to negative space charge. In the absence of beam flow, there is no space charge, with the result that an equipotential area (neglecting fringe effects) exists across the hollow tube. Beam flow, however, results in the potential gradient and the space-charge effects described above.

The potential of the beam boundary (V_a) which satisfies the relationship expressed by Equation (11) is also valid for the case of partially filled hollow electrodes. The potential of the hollow electrode or outer cylinder of the coaxial system (V_b) is, therefore, the summation of the potential drops, and is given by

$$V_b \equiv V_a + (V_b - V_a), \tag{16}$$

where $(V_b - V_a)$ is the potential difference between the outer cylinder of radius b and inner coaxial beam of radius a. Using Laplace's equations, the potential difference is

$$(V_b - V_a) = \frac{1}{2\pi\epsilon}\frac{I_m}{\dot{z}}\int_a^b \frac{dr}{r} = \frac{1}{2\pi\epsilon}\frac{I_m}{\sqrt{2\eta V_0}}\ln\left(\frac{b}{a}\right) \tag{17}$$

$$= \frac{I_m \ln\left(\frac{b}{a}\right)}{2\pi\epsilon\sqrt{2\eta V_o}} = \frac{I_m \ln\left(\frac{b}{a}\right)}{2\pi\epsilon\sqrt{\frac{2}{3}\eta V_a}}. \tag{18}$$

When Equation (16) for the minimum hollow-tube potential necessary to sustain a current flow I_m of uniform density and the maximum

intensity given by Equation (9) is modified to include Equations (6) and (11),

$$V_b = \left[\frac{I_m \times 10^6}{25.4} \right]^{2/3} + \frac{I_m \ln \left(\frac{b}{a} \right)}{2\pi\epsilon \sqrt{\frac{2\eta}{3}} \left[\frac{I_m \times 10^6}{25.4} \right]^{1/3}} \tag{19}$$

$$= I_m^{2/3} \left[1157 + 1544 \ln \left(\frac{b}{a} \right) \right].$$

Substitution of various values of I_m in Equation (19) yields the voltage of discontinuity of the hollow tube or collector for various values of the ratio b/a. This voltage of discontinuity or critical voltage for nine values of b/a is plotted on log-log coordinates in Figure 13. The voltage for four values of b/a is replotted on cartesian coordinates in Figure 14.

Equation (19) suggests that the beam-flow perveance for Brillouin flow in a partially filled hollow tube can be written as

$$K = \left[\frac{1}{1157 + 1544 \ln (b/a)} \right]^{3/2}. \tag{20}$$

A plot of the beam-flow perveance as a function of b/a is shown in Figure 15. Maximum beam-flow perveance (25.4 microamps/volts$^{3/2}$) is realized when b/a is unity, i.e., when the hollow tube is completely filled. An increase in the ratio b/a decreases the beam-flow perveance and increases the minimum collector voltage required to support a constant beam current in magnetically confined flow.

To determine the validity of Equation (19) for the voltage of beam discontinuity, attempts were made to use this equation to predict the experimental results shown in Figure 7. The points of collector voltage instability for various levels of cathode current shown in Figure 7, which describe the locus of the curve of beam-flow discontinuity, were identified by the onset of oscillations. These points are replotted on log-log coordinates and compared with calculated values in Figure 16. Calculations were based on a beam diameter of 0.060 inch and a maximum collector inside diameter of 0.219 inch, i.e., a ratio b/a of 3.65. There is a deviation of approximately 40 volts between estimated and measured values of the voltage of discontinuity at 40 milliamperes, the rated cathode current of the experimental tube.

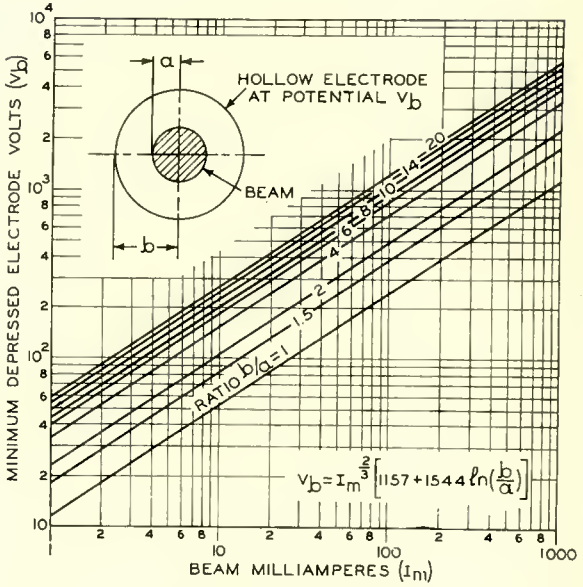


Fig. 13—Maximum possible current sustained in Brillouin flow versus collector potential with variations of b/a.

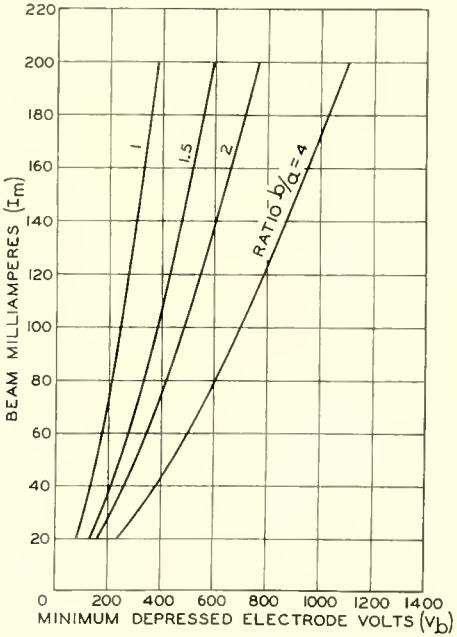


Fig. 14—Maximum beam current versus collector potential for various ratios of b/a for Brillouin flow.

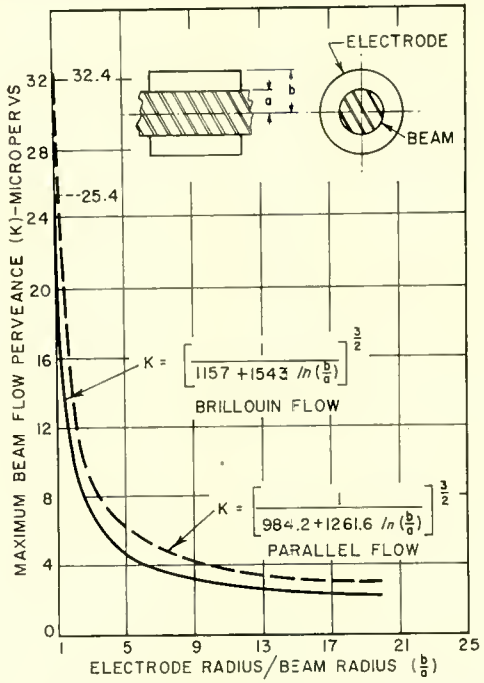


Fig. 15—Maximum beam flow perveance related to b/a.

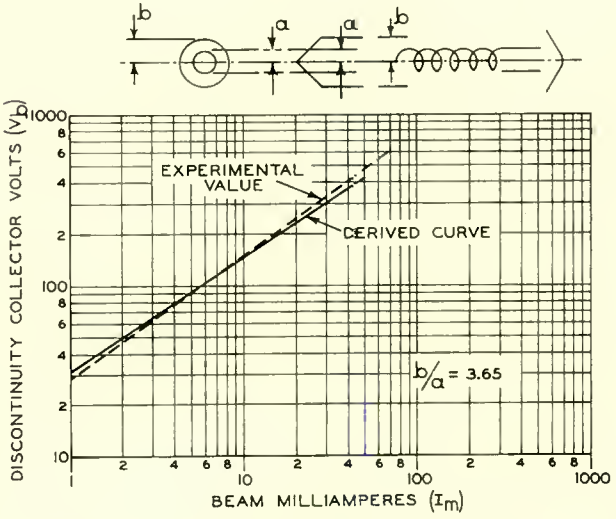


Fig. 16—Comparison of experimental and derived data for potential of beam flow discontinuity.

Figure 17 shows the measured discontinuity characteristic of a developmental tube operated at a hollow-collector potential of 2,000 volts and an initial cathode current of 10 milliamperes. The low value of beam current used permits the collector potential to be depressed to zero without danger that excessive helix dissipation will cause evolution of gas. Depression of the collector potential to approximately 600 volts below the helix voltage results in out-of-phase changes in shield and anode currents. Appreciable anode interception occurs when the collector potential is depressed still further. The negative polarity

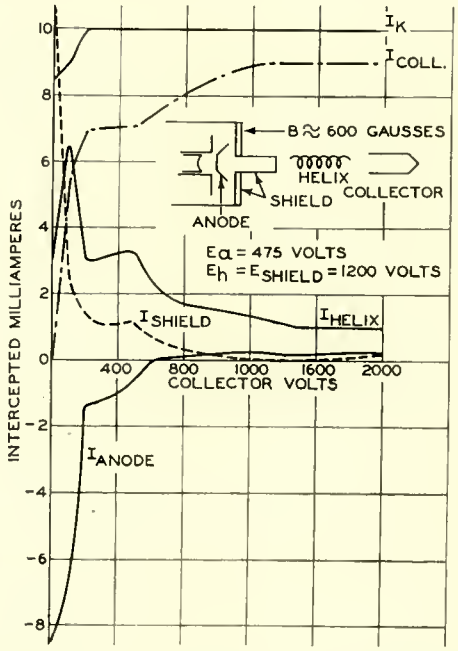


Fig. 17—Intercepted current as a function collector potential.

of the anode current and the out-of-phase relationship between the anode and shield current increments indicate the exchange of secondary electrons between these electrodes. The shield, because of its higher potential, collects the secondary electrons dislodged from the anode. The negative polarity of the anode current is due to the nature of the anode surface and the angle of incidence of the primary beam bombarding that surface. If the secondary-emission ratio of the anode surface is equal to two, the net negative anode current will be equal in magnitude to the beam current. Depression of the collector voltage from 600 volts to cathode potential provides sufficient depression of

the beam space charge to cause a departure from Brillouin-flow conditions. This departure is strongly indicated by the reduction in cathode current which occurs below 200 volts. It is believed to be due to increased space charge in the anode region caused by reflected beam current from the depressed collector region. For the system employed, lowering the collector potential sufficiently, as stated previously, causes a partial reversal in the direction of beam current flow. Although the reversed flow is restrained to a fixed diameter by the axial magnetic field as it travels down the helix, it becomes extremely divergent when it enters the shielded portion of the electron gun. The defocused electrons are intercepted by the anode and initiate the secondary emission to the helix. The voltage of instability for this condition, as calculated from Equation (19) and approximately verified by the experimental data shown in Figure 7, is 200 volts. To determine the effect of focusing on the space-charge conditions at zero collector potential, the strength of the axial magnetic field was reduced to zero. This change returned the cathode current from 8.5 milliamperes to its initial value of 10 milliamperes.

Beam-Flow Discontinuity for Parallel-Flow Beams

For the case of an unshielded gun immersed in a large axial magnetic field, the following relationship based on the limiting current for a hollow tube completely filled by the beam has been derived by Pierce:[6]

$$I_m = 1.963\pi\epsilon\sqrt{2\eta}V_a^{3/2} = 32.4 \times 10^{-6}\, V_a^{3/2}. \tag{21}$$

Pierce also shows that at maximum current the voltage at the beam axis (V_o) is 0.174 of the beam-boundary voltage (V_a) indicated by Equation (21), i.e.,

$$V_o = 0.174V_a. \tag{22}$$

The maximum beam-flow perveance under these conditions, as indicated in Equation (21) is 32.4 microamp/volts$^{3/2}$. The voltage variation as a function of distance along the beam radius is shown schematically in Figure 8.

The sum of the potential drops between the beam boundary and the partially filled hollow cylinder can be estimated in the same manner as for Brillouin flow. However, consideration must be given to the variation in electron velocity across the beam radius;

---

[6] J. R. Pierce, *Theory and Design of Electron Beams*, D. VanNostrand Co., Inc., Princeton, N.J., 1949, pp. 161-164.

$$v_{avg} = \sqrt{2\eta\, V_{avg}} = \sqrt{2\eta}\,\frac{1}{a}\int_o^a V dr$$
$$= \sqrt{2\eta}\ \sqrt{0.528\, V_a}\,, \tag{23}$$

where v_{avg} and V_{avg} are, respectively, the average velocity and average voltage along the beam radius. The approximate potential difference between the beam boundary and the tubular drift electrode is, therefore,

$$(V_b - V_a) = \frac{I_m \ln (b/a)}{2\pi\epsilon\, v_{avg}} = \frac{I_m \ln (b/a)}{2\pi\epsilon \sqrt{2\eta}\ \sqrt{0.528\, V_a}}. \tag{24}$$

Application of the restrictions imposed by Equation (21) for the maximum sustained beam current yields

$$(V_b - V_a) = \frac{I_m \ln (b/a)}{0.2527 \times 10^{-4} \left[\dfrac{I_m \times 10^6}{32.4}\right]^{1/3}}. \tag{25}$$

The voltage of the partially filled tubular hollow electrode is, therefore,

$$V_b \equiv V_a + (V_b - V_a)$$
$$V_b = \left[\frac{I_m \times 10^6}{32.4}\right]^{2/3} + \frac{I_m \ln (b/a)}{.02527 \times 10^{-4} \left[\dfrac{I_m \times 10^6}{32.4}\right]^{1/3}} \tag{26}$$

$$V_b = I_m^{2/3} [984.2 + 1261.6 \ln (b/a)]. \tag{27}$$

The limiting voltage as given by Equation (27) in terms of beam current and various ratios of electrode to beam radius is plotted in Figure 18. Equation (27) suggests that the beam-flow perveance for parallel magnetically restrained beam flow may be defined in the same manner as for Brillouin flow, i.e.,

$$K = \frac{1}{[984.2 + 1261.6 \ln (b/a)]^{3/2}}. \tag{28}$$

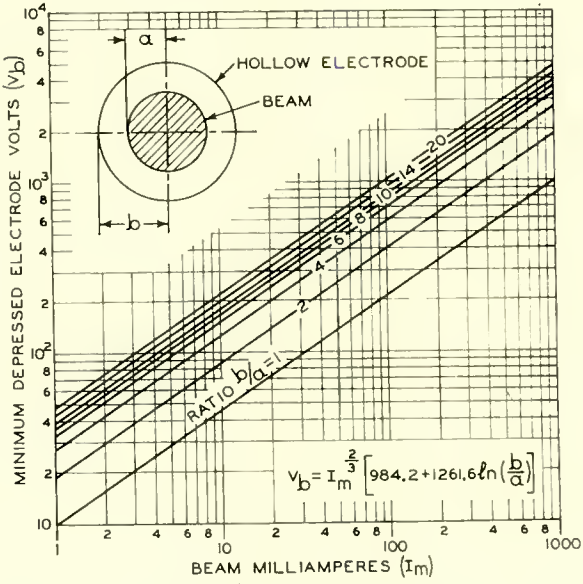


Fig. 18—Maximum beam current versus collector potential for various ratios of b/a for parallel flow.

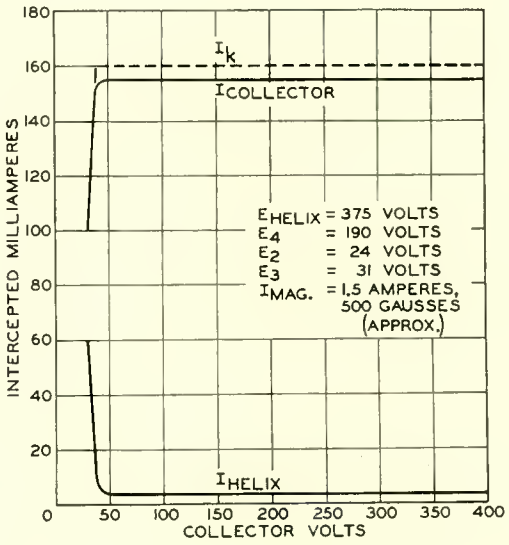


Fig. 19—Focusing characteristics of an experimental low-noise parallel-flow traveling-wave tube.

Plots of K as a function of electrode and beam geometry for both Brillouin flow and parallel-beam flow are shown in Figure 15. The comparison shows that for a given value of b/a, the highest beam current at reduced electrode potentials is obtained in the parallel-flow case. A developmental low-power parallel-flow traveling-wave tube having a beam current of 150 microamperes was tested to determine the degree to which the minimum depressed collector potential correlated with the theory presented above. The magnetically shielded collector employed in this tube adequately inhibited focused flow of secondary electrons from the collector without modifications found necessary for other tube types. As shown in Figure 19, at a solenoid flux density of 525 gausses the collector could be depressed to approximately 33 volts without effect on the collector current. At higher flux densities the collector voltage could be depressed to even lower values without reduction of the collector current.

The minimum voltage required to sustain a beam current of 150 microamperes, as calculated for Equation (27), was approximately eight volts. The calculation was based on an estimated beam diameter of approximately 0.060 inch and a maximum collector inside diameter of 3/16 inch. The 25-volt discrepancy between the measured and calculated values of minimum depressed potential is not great for this low value of beam current.

In this case, the minimum tolerable depressed collector voltage (33 volts) permits operation of the collector at less than 10 per cent of the helix potential, and makes it possible to achieve a collector efficiency of approximately 20 per cent at one milliwatt output.

PROPAGATION CHARACTERISTICS OF SLOW-WAVE STRUCTURES DERIVED FROM COUPLED RESONATORS*

By

Erwin Belohoubek

RCA Electron Tube Division,
Princeton, N. J.

Summary—*A general procedure, based on the perturbation theory for resonated cavities, is given to determine qualitatively the ω-β_0 diagram for slow-wave structures of the coupled-resonator type. After the two basic coupling possibilities (i.e., inductive and capacitive coupling between resonators) are defined, different coupling schemes are considered and their application to slow-wave structures is discussed. By the method outlined, the influence of changes in the geometrical configuration of the structure on the ω-β_0 diagram can be estimated. Finally, the qualitative considerations are supplemented by some measurements on a round waveguide having differently shaped partition walls.*

Introduction

A VERY important characteristic in the design and application of slow-wave structures for traveling-wave tubes is the so-called ω-β_0 diagram. This diagram provides the necessary information about the phase and group velocities of different space harmonics and about the bandwidth of the slow-wave structure. For helix-type structures, reasonably accurate solutions exist and a sizable amount of design information is available. On the other hand, the rigorous field solutions (where they exist) of most of the other slow-wave structures are rarely useful for any numerical evaluation. Consequently, either a fairly rough approximation must be used, or the ω-β_0 diagram must be derived from an assumed equivalent circuit of the structure. For some structures neither method is very satisfactory.

When an attempt is made to establish a common procedure for the investigation of different slow-wave structures, it is found that a very large group of structures (nearly all of the high-power structures fall into this category) can be derived from coupled resonators. These slow-wave structures consist of resonators of identical shape, coupled together and so arranged geometrically that a beam shot along the array can interact with a strong axial electrical field. Sometimes the coupling hole may serve also as the hole through which the electron beam

* Manuscript received April 25, 1958.

passes; sometimes coupling hole and beam hole are separated. The loaded-transmission-line type of structure can also be fitted into this picture if the case of diminishing loading is considered as the extreme case of full coupling between resonators.

The basic procedure for finding the principal shape of the ω-β_0 diagram for structures of this type is as follows: the particular structure under consideration is assumed to consist originally of uncoupled resonators, and the E- and H-field configurations for the two cutoff points of the passband ($\beta_0 L = 0$ and $\beta_0 L = \pi$) are drawn. Then a slight amount of coupling is introduced and, on the basis of a small perturbation of the original field configuration, the effect on the frequencies of the two cutoff points is investigated. For some structures it may be more advantageous to start with a homogeneous transmission line, introduce a slight amount of loading, and then determine how the frequencies of the points $\beta_0 L = 0$ and $\beta_0 L = \pi$ have changed.

As long as the field lines are not appreciably changed by the introduction of coupling, the frequency of the particular cutoff point remains constant. However, if, as an example the E lines are strongly perturbed by cutting a slot into the partition between resonators, the resonant frequency will increase. On the other hand, if a slot is cut which affects mainly the H-field configuration, the frequency of the cutoff point under consideration will decrease. The direction of the frequency change (Δf) caused by the perturbation can be determined from the relationship $\Delta f/f = \Delta W/W$, where ΔW represents the change in stored energy introduced by the perturbation and W represents the total stored energy of the structure.[1,2] Although this formula is actually valid only for a structure at resonance, it can be used in this case because the group velocity at the cutoff points of the passband is zero and, therefore, the resulting standing wave corresponds to the field configuration in a resonated cavity. The quantitative change of Δf for a given coupling hole cannot be determined in this way because the relationship $\Delta f/f = \Delta W/W$ is exactly valid only for the limit $\Delta W \to 0$. However, the relationship can be used to obtain a qualitative picture of the passband behavior for different coupling arrangements.

In the following, the two basic possibilities for coupling between resonators, pure inductive coupling and pure capacitive coupling, are discussed and then somewhat more complicated coupling types and their practical application to slow-wave structures are investigated.

---

[1] J. Müller, "Untersuchung über elektromagnetische Hohlräume," *Hochfrequenztechnik und Elektroakustik*, Vol. 54, p. 157, 1939.

[2] J. C. Slater, *Microwave Electronics*, D. Van Nostrand Co., Princeton, N. J., 1950, p. 80.

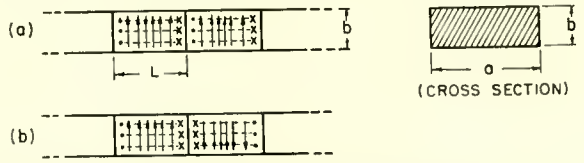


Fig. 1—Rectangular waveguide with periodic partitions (no coupling). (a) Field pattern for $\beta_0 L = 0$; (b) Field pattern for $\beta_0 L = \pi$. (Solid lines = E field; Dashed lines = H field).

The ω-β_0 diagrams shown throughout the paper refer to the fundamental space harmonic of the slow-wave structure, which shall be defined as the space harmonic having the largest amplitude. This definition is in agreement with the ω-β_0 curves measured by the resonance method.[3]

Pure Inductive Coupling of Resonators

Pure inductive coupling can be defined as a coupling scheme where only H lines link the cavities; the E lines do not cross the common partition wall between cavities. An example of this type of coupling can be realized as follows: consider a rectangular waveguide propagating in the TE_{10} mode which has partitions periodically placed across the guide. Figure 1 shows such a structure having the field lines drawn in for the two cutoff points, $\beta_0 L = 0$ and $\beta_0 L = \pi$. The resonance frequency for both points is the same, and, corresponding to the relationship $2L = \lambda_g$ (λ_g is the guide wavelength for the TE_{10} mode), lies at $f = c\sqrt{1 + (L/a)^2}/2L$. When a small vertical slot is cut in the partition, as shown in Figure 2, a certain amount of inductive coupling is introduced. The field configuration for $\beta_0 L = \pi$ shown in Figure 2 (b) is not changed, and, as long as the partition wall is very thin, the resonant frequency does not change either. For $\beta_0 L = 0$, however, the H lines link the cavities, as shown in Figure 2 (a), indicating that

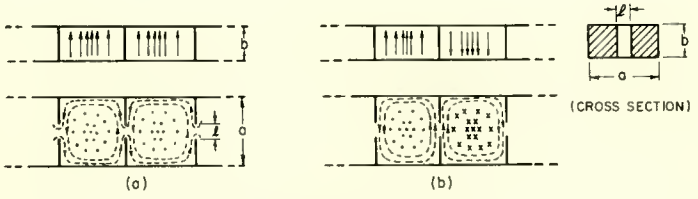


Fig. 2—Pure inductive coupling. (a) Field pattern for $\beta_0 L = 0$; (b) Field pattern for $\beta_0 L = \pi$.

[3] E. J. Nalos, "Measurement of Circuit Impedance of Periodically Loaded Structures by Frequency Perturbations," *Proc. I.R.E.*, Vol. 42, p. 1508, October, 1954.

some of the H lines have been strongly perturbed from their original paths with a corresponding decrease in frequency. The ω-β_0 diagram for various degrees of coupling is shown in Figure 3. When l becomes equal to a, the propagation characteristic approaches that of the plain waveguide corresponding to full inductive coupling. Thus, the inductive coupling lowers the passband below the resonance frequency of the resonators when no coupling is present. In this special case, the point $\beta_0 L = 0$ moves down in frequency while $\beta_0 L = \pi$ remains constant.

It may be worthwhile to mention that the structure for this simple case can be represented quite well by an equivalent circuit which consists of a transmission line[4] periodically loaded by a normalized shunt susceptance B/Y_0. The relation between propagation constant and frequency is then given by

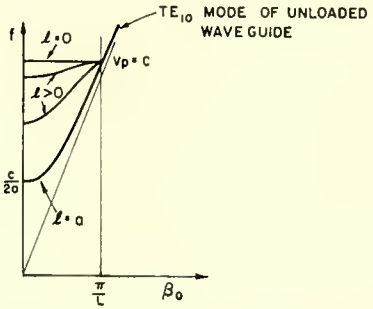


Fig. 3—Pure inductive coupling of various degrees for TE_{10}-mode propagation in a rectangular waveguide. (v_p = phase velocity; c = velocity of light).

$$\cos \beta_0 L = \cos \frac{2\pi}{\lambda_g} L - \frac{B}{2Y_0} \sin \frac{2\pi}{\lambda_g} L. \qquad (1)$$

For $B = 0$, the equation yields the propagation constant of the plain waveguide. This approximation by an equivalent circuit does not take into account any higher-order-mode coupling between the irises which occurs when the higher waveguide modes, excited at the irises, are not fully attenuated in the waveguide sections between the partitions. Good quantitative results can only be obtained from this analysis, therefore, if the cell length is large compared to the size of the introduced discontinuities.

This coupled-cavity structure is, of course, not useful for a direct application to traveling-wave tubes because no axial E field is present

[4] M. H. Steward, "Traveling-Wave Tubes Using Folded Slow-Wave Structures," Technical Report No. 16, Stanford University, California, p. 49, 1952.

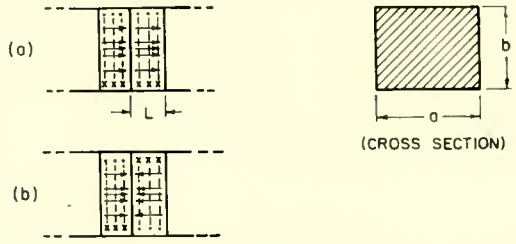


Fig. 4—Rectangular TE_{101} resonators placed with the top surfaces against each other. (The period L corresponds to the former height b of the waveguide of Figure 1). (a) $\beta_0 L = 0$; (b) $\beta_0 L = \pi$.

for interaction with the electrons, but it represents a simple example of pure inductive coupling.

Pure Capacitive Coupling of Resonators

Pure capacitive coupling can be defined as a coupling arrangement in which only E lines link the resonators. As an example of this type of coupling, rectangular waveguide sections resonated in the TE_{10} mode are used, and are connected by placing the top surfaces together, as shown in Figure 4. If a small hole is cut in the center of the partition, the E lines penetrate into the next cavity leading to pure capacitive coupling. The E lines are not affected for the point $\beta_0 L = 0$ (a in Figure 5), but are strongly perturbed for $\beta_0 L = \pi$ (b in Figure 5). The resulting ω-β_0 diagram, therefore, shows a constant lower cutoff frequency* ($\beta_0 L = 0$), while the other edge of the passband moves up in frequency as the size of the hole increases. When all partitions are removed, the propagation characteristic becomes that of the TM_{11} mode, as shown in Figure 6. The basic characteristic of capacitive

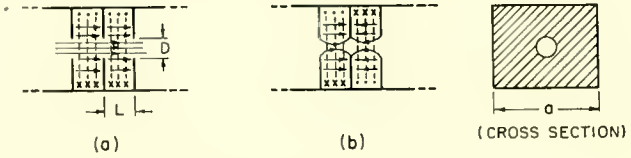


Fig. 5—Pure capacitive coupling. (a) Field pattern for $\beta_0 L = 0$; (b) Field pattern for $\beta_0 L = \pi$.

* The lower cutoff frequency is exactly constant only as long as the wall thickness is infinitely small. However, when the wall thickness is appreciable, perturbation considerations show that the point $\beta_0 L = 0$ first increases in frequency and, with increasing coupling, goes back to the TM_{11} cutoff frequency of the plain waveguide. (A similar consideration applies to the cutoff frequency $\beta_0 L = \pi$ for the pure inductively coupled structure.)

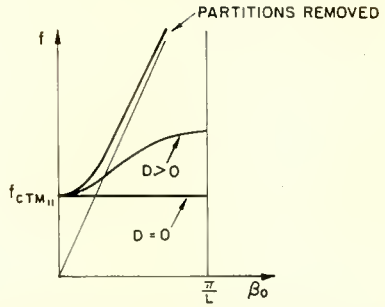


Fig. 6—Pure capacitive coupling of various degrees for TM_{11}-mode propagation in a rectangular waveguide.

coupling, therefore, is that it raises the passband above the resonance frequency of the resonators when no coupling is present. As in the case of pure inductive coupling, this type of structure can be approximated by an equivalent circuit which consists of a waveguide operated in the TM_{11} mode and periodically loaded with capacitive irises. The same precautions about a numerical evaluation in the case of closely spaced irises also apply here.

A well-known practical example of this type of coupling is the disk-loaded round waveguide, which is used primarily in linear accelerators.

Coupling Through a Resonant Slot

Figure 7 shows an example of resonators coupled by means of resonant slots. If the distance between the partitions in the rectangular waveguide structure is large compared to the height of the waveguide, higher-order-mode coupling between the slots can be neglected and the coupling is purely inductive. The resonance frequency of an iris in a rectangular waveguide can be determined[5] graphically as indicated in Figure 8. The general behavior of the iris, however, does not change if the slot is moved in a vertical direction; consequently, the results can also be applied to the slot in Figure 7. A given resonance frequency

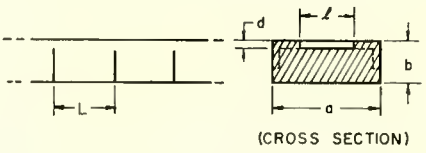


Fig. 7—Coupling through a resonant slot.

[5] L. N. Ridenour, *Principles of Microwave Circuits*, McGraw-Hill Book Co., New York, N. Y., 1948, p. 170.

can be achieved, therefore, either with a narrow and short slot, or with a somewhat broader and longer slot.

For slots having different values of l and with $d = b$, as in the case of pure inductive coupling, the slot does not become resonant except in the extreme case where $l = a$, in which case the slot resonance coincides with the cutoff frequency of the unloaded waveguide. Now consider the case when $d < b$ and the slot length l increases slowly from zero to a. For small values of l the slot is resonant at a very high frequency, far outside the range of the lowest passband. The ω-β_0 diagram, therefore, is similar to that shown in Figure 3. As l increases, the resonance frequency of the slot decreases until finally it crosses the upper cutoff frequency ($f = c\sqrt{1 + (L/a)^2}/2L$) of the passband. If losses are neglected, a partition with a slot at resonance represents an

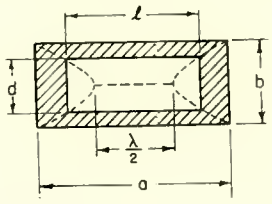


Fig. 8—Approximate determination of the resonance frequency of a slot. (TE_{10}-mode wave propagation assumed). All rectangular slots whose edges lie on the drawn hyperbola have the same resonance frequency (the distance of the vertices of the hyperbola corresponds to $\lambda/2$).

infinite shunt impedance, and, therefore, the ω-β_0 curve must intersect the propagation curve of the unloaded waveguide at this frequency.

As l increases, this intersection point moves down in frequency until, when $l = a$, it finally reaches the cutoff frequency of the plain waveguide as shown in Figure 9. If l is increased further, as shown by the dashed lines in Figures 7 and 9, the lower cutoff point is lowered even more to approximately $f = c/2l$ (d assumed to be small). The coupling is purely inductive throughout the whole range; i.e., the passband is perturbed below the resonance frequency when no coupling is present. For small coupling holes, one cutoff frequency remains constant while the other moves down in frequency; the passing through a slot resonance, however, causes both cutoff points to move down in frequency.

The extreme case ($l = a$) may also be treated advantageously by another approach. In this case a plain waveguide is used instead of

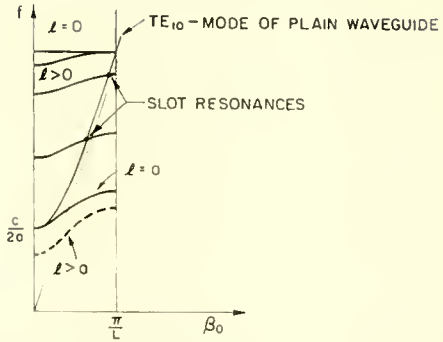


Fig. 9—Coupling through a resonant slot ($L > b$, d = constant, l varied).

uncoupled resonators, and a study is made of the way in which the propagation constant is affected by the introduction of periodically spaced partitions with height s. The first case considered is the field configuration of the cutoff point $\beta_0 L = 0$, as shown in Figure 10(a). As can be seen both the H and E fields are perturbed by the introduction of the partitions. For a rectangular waveguide operated in the TE_{10} mode at cutoff, the E- and H-field configurations are given by

$$E_y = \frac{C\pi}{a} \sin \frac{\pi x}{a},$$
$$H_z = j \frac{C\pi^2}{\beta_0 Z_0 a^2} \cos \frac{\pi x}{a}. \tag{2}$$

When a partition with height s and width t is introduced, the frequency deviation Δf can be derived as follows:

$$\frac{\Delta f}{f} = \frac{\Delta W}{W}, \tag{3}$$

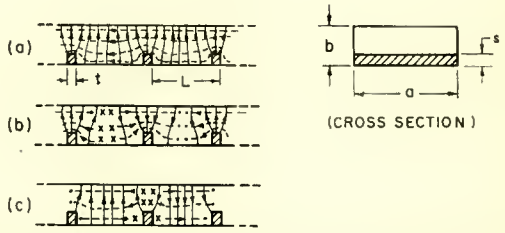


Fig. 10—Field configurations in a periodically loaded rectangular waveguide. (a) $\beta_0 L = 0$; (b) $\beta_0 L = \pi$; (c) $\beta_0 L = \pi$ (next higher pass band).

$$\Delta W = \Delta W_{el} + \Delta W_{magn} . \tag{4}$$

The change in the average stored electrical energy caused by the introduction of a small partition is calculated to be

$$\Delta W_{el} = -\frac{1}{4}\int_{\Delta V} \epsilon_0 EE^* \, dV = -\frac{st\epsilon_0}{4}\int_0^a \frac{C^2\pi^2}{a^2}\sin^2\frac{\pi x}{a}\, dx = -\frac{C^2 st\epsilon_0\pi^2}{8a} . \tag{5}$$

Analogously, the change in average stored magnetic energy is

$$\Delta W_{magn} = \frac{1}{4}\int_{\Delta V} \mu_0 HH^* \, dV = \frac{st\mu_0}{4}\int_0^a \frac{C^2\pi^4}{\beta_0^2 Z_0^2 a^4}\cos^2\frac{\pi x}{a}\, dx = \frac{st\mu_0}{8}\frac{C^2\pi^4}{\beta_0^2 Z_0^2 a^3} . \tag{6}$$

After substituting for

$$\beta_0 Z_0 = \frac{2\pi}{\lambda}\sqrt{1-\left(\frac{\lambda}{2a}\right)^2}\sqrt{\frac{\mu_0}{\epsilon_0}}\frac{1}{\sqrt{1-\left(\frac{\lambda}{2a}\right)^2}} = \frac{2\pi}{\lambda}\sqrt{\frac{\mu_0}{\epsilon_0}} , \tag{7}$$

and with

$$\lambda = 2a$$

the change in average magnetic stored energy is

$$\Delta W_{magn} = \frac{st\epsilon_0}{8}\frac{C^2\pi^2}{a} . \tag{8}$$

The changes in magnetic and electrical stored energy, therefore, just cancel each other, and the total change in frequency, Δf, is zero. This conclusion is, of course, strictly correct only for the limit where the height of the perturbing partition goes to zero. However, the result also applies for high and very thick partitions because, in the extreme case where the partition thickness becomes equal to the period L, a homogeneous waveguide again results, although with a different height b, and the cutoff frequency remains unchanged.* The cutoff frequency

* See also Marcuvitz, *Waveguide Handbook*, McGraw Hill Book Co., p. 252; 1951. Note that the susceptance of a thick capacitive obstacle goes to zero for $\lambda_g \to \infty$, which corresponds to the case $\beta_0 L = 0$.

$\beta_0 L = 0$, therefore, is independent of the height and the width of the partitions.

Similar considerations applied to the other cutoff point $\beta_0 L = \pi$, show that, although both the E and H lines are perturbed, the change in the average stored magnetic energy ($\Delta W_{magn.}$) now is smaller than the change in the average stored electrical energy (ΔW_{el}), and, therefore, the upper cutoff frequency is lowered, as shown in Figure 11.

Still another field configuration is possible for $\beta_0 L = \pi$, as shown in Figure 10(c). In this case, the introduction of a very thin partition perturbs neither the electrical nor the magnetic field. The resonance frequency, therefore, lies on the propagation curve of the plain waveguide. It represents the lower cutoff frequency of the next higher

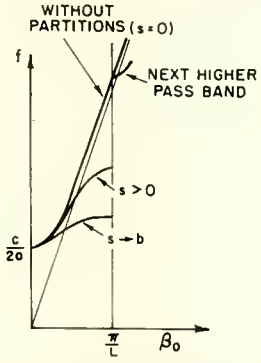


Fig. 11—ω-β_0 diagram for a periodically loaded rectangular waveguide. (Height, s, of the partitions varied).

passband, but, unlike the lower cutoff frequency of the first passband, is sensitive to the width of the partition, t.

The case $l = a$ can also be treated quantitatively by an equivalent circuit approach, as shown by Lines, Nicoll, and Woodward.[6]

Mixed Coupling

"Mixed" coupling is characterized by the presence of both E_z and H_z lines in the coupling aperture. A simple example of mixed coupling in which capacitive and inductive coupling are combined by the use of separate holes will be discussed first.

For this example, the structure discussed in the case of pure capacitive coupling (Figure 5) is used as a starting point. If, in

[6] A. W. Lines, G. R. Nicoll, and A. M. Woodward, "Some Properties of Waveguides with Periodic Structure," *Proc. I.E.E.*, Vol. 97, Part III, p. 263, July, 1950.

addition to the capacitive coupling hole, a slot is cut on the periphery of the partition wall as shown in Figure 12, a combination of the two basic coupling types, capacitive and inductive coupling, is achieved. An investigation of the field perturbations caused by the introduction of the coupling holes shows that the point $\beta_0 L = 0$ remains constant, independent of the degree of coupling for both coupling types (this statement is true only for small apertures, as will be seen later), while the point $\beta_0 L = \pi$ is raised in frequency for the capacitive coupling and lowered for the inductive coupling. Thus, the addition of an inductive slot will decrease the originally capacitive passband more and more until finally, despite two coupling holes in the partition walls, the passband diminishes entirely which means the net coupling is zero.

However, mixed coupling can also occur when only one coupling slot along the edge of the partition is present, provided the slot opening is very long and the partitions are closely spaced. The structure shown

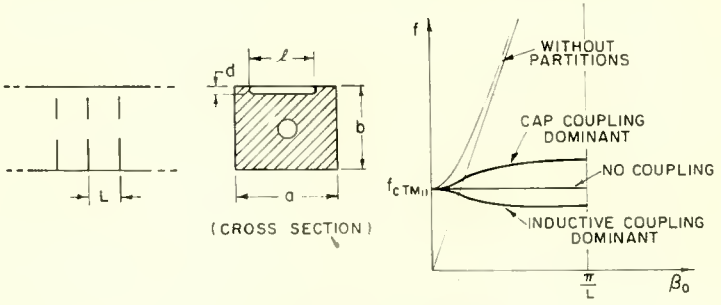


Fig. 12—Mixed coupling.

in Figure 12 again is used for this case, but without the capacitive coupling hole in the center. First, the two extreme cases, l very small, and $l = a$, are considered, and then the shape of the passband for the intermediate slot lengths is investigated.

For a short slot, the coupling is purely inductive and a passband such as that shown in Figure 14(a) is obtained. On the other hand, when $l = a$, it can be shown that as in the case of inductive coupling through a resonant slot, the lower cutoff point of the passband coincides with the cutoff frequency of the TE_{10} mode of the plain waveguide. For $\beta_0 L > 0$, however, longitudinal E fields are present throughout the cell and in the slot opening due to the proximity of the next partition, leading to an additional capacitive coupling which raises the upper cutoff frequency close to the point where $(b—d)$ becomes $\lambda_g/4$. The ω-β_0 diagrams drawn for two extreme values of the period L, shown in Figure 13, illustrate the influence of the capacitive coupling caused by

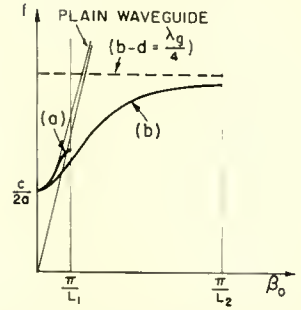


Fig. 13—Influence of the spacing of the partitions on the pass band for coupling through a long slot ($l = a$). (a) Widely spaced partitions; (b) Closely spaced partitions.

the close spacing of the partition walls. It must be realized that this result is derived from a perturbation of the TE_{10} mode, while the uncoupled resonators originally had a field configuration corresponding to a TM_{11} mode.

Thus far, the passbands for the two extreme cases have been determined (Figure 14(a) for l very small, and Figure 14(d) for $l = a$). For all intermediate slot lengths, therefore, ω-β_0 curves such as those shown in Figure 14 must be expected. As the slot opening becomes larger, E lines start to link adjacent resonators due to the close spacing of the partitions, and mixed coupling occurs. The inductive part of the coupling is responsible for the lowering of the total passband, while the additional capacitive coupling causes the cutoff point $\beta_0 L = \pi$ to

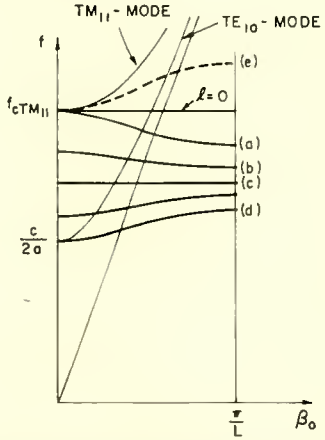


Fig. 14—Coupling through long slots with closely spaced partitions. Curves (a) to (d) represent the shapes of the lowest pass band for various slot lengths. Curve (e) is the next higher passband which arrives at $f_{cTM_{11}}$ for greater slot lengths.

rise in frequency. The over-all effect, therefore, is a reduced lowering of this point compared to the other cutoff point, $\beta_0 L = 0$, which finally leads to a reversal in the slope of the ω-β_0 diagram (Figure 14(c)).* The transition from uncoupled resonators to an unloaded waveguide by first making a slot in the partition, and increasing the slot length a and then reducing the height, s, of the partition to zero, is accompanied by a mode change. The field configuration in the structure at the beginning is that of a pure TM mode at cutoff which finally becomes the TE mode of a plain waveguide. Between these two extremes, the presence of mixed coupling leads to a diminishing passband for a certain slot length due to the additional capacitive coupling.

It should be noted that the frequency $f_{cTM_{11}}$ remains a cutoff frequency at all times because, when the point $\beta_0 L = 0$ of the earlier discussed passband decreases in frequency, another higher passband arrives with its lower cutoff point at $f_{cTM_{11}}$, as shown in Figure 14(e). This higher passband is actually a perturbed TE_{10} mode which comes down in frequency as the slot length increases. When it reaches $f_{cTM_{11}}$, its lower cutoff point remains constant, while the lower passband now starts to decrease in frequency.

Many slow-wave structures are included in the category of mixed coupling. The planar magnetron structure, which consists of a ridged waveguide with transverse slots cut into the ridge, corresponds geometrically to the structure shown in Figure 7, but with very long slots and closely spaced partitions to provide the necessary longitudinal E field for interaction with the electron beam. Contrary to the previous case, the lower cutoff frequency of this structure, depends on the thickness of the partitions and lies between $c/2a$, the cutoff frequency of the plain waveguide, and f_c, the cutoff frequency of a ridged waveguide having a ridge of the same cross-section as that of the partitions (Figure 15).

The round waveguide, which operates in the TM_{01} mode with inductive coupling between cells, has been extensively used in high-power traveling-wave amplifiers.[7] The ω-β_0 diagram of this structure (provided the coupling is mainly inductive) has a shape such that group and phase velocity have opposite signs for the fundamental space harmonic, which allows fairly dispersion-free operation in the first forward space harmonic. The structure shown in Figure 16 employs two coupling slots in order to obtain a greater bandwidth. As will be shown

---

* The diminishing passband in this case, however, does not represent a case of zero net coupling, because the passband is raised in frequency to $f_{cTM_{11}}$ when the slot opening is closed entirely.

[7] M. Chodorow and E. J. Nalos, "The Design of High-Power Traveling-Wave Tubes," *Proc. I.R.E.*, Vol. 44, p. 649, May, 1956.

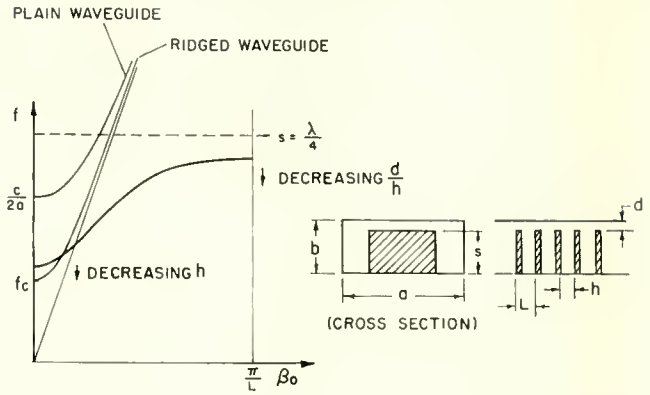


Fig. 15—Planar magnetron structure (f_c = cutoff frequency of equivalent ridged waveguide). The arrows in the ω-β_0 diagram indicate the direction in which the cutoff points are moved for changing certain dimensions.

later, a structure of this type can also be explained from a different point of view which gives a better understanding of the possible field configurations in the structure and provides additional design information.

The clover-leaf structure shown in Figure 17, is an example of an advantageous application of mixed coupling.[8] In the ordinary case (a round waveguide operating in the TM_{01} mode and having partitions with a circular hole in the center and a slot along the periphery), the influence of the additional inductive coupling reduces the total bandwidth and, therefore, is undesirable. In the clover-leaf structure, however, fingers penetrate radially inward and modify the H field in such a way as to reverse the direction of the H lines in alternating cavities along the coupling slots for both cutoff points. If the fingers were

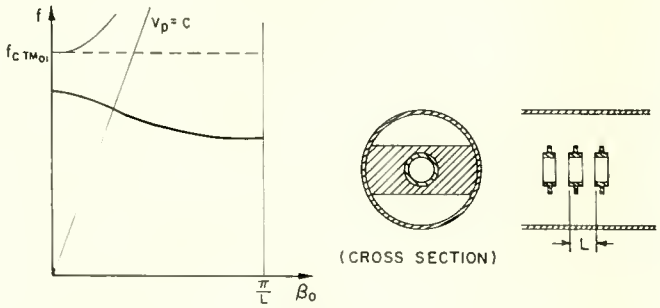


Fig. 16—Inductively coupled cavities with two coupling slots.

[8] M. Chodorow and R. A. Craig, "Some New Circuits for High-Power Traveling-Wave-Tubes," *Proc. I.R.E.*, Vol. 45, p. 1106, August, 1957.

axially aligned, the structure would employ ordinary inductive coupling, which would result in a very narrow bandwidth with the lower cutoff frequency increased considerably due to the perturbation of the H field by the fingers (dotted line in Figure 17). If, however, the fingers are alternately rotated by 90 degrees in every second cell, the influence of the inductive coupling on both cutoff points is reversed. That means, while for normal inductive coupling the point $\beta_0 L = 0$ remained constant and $\beta_0 L = \pi$ decreased in frequency with increased coupling, in this case the effect is just the opposite. The point $\beta_0 L = 0$ decreases

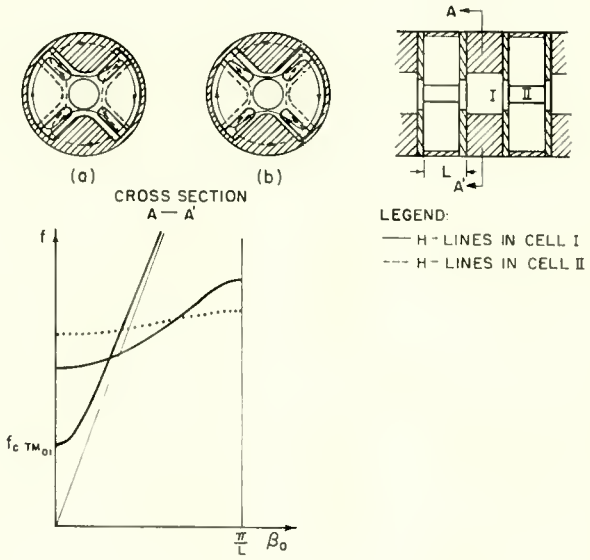


Fig. 17—Clover-leaf structure. (The actual structure has 8 slots instead of 4). (a) H lines drawn for $\beta_0 L = 0$ *before* the 4 slots are cut out; (b) H lines drawn for $\beta_0 L = \pi$. The dotted line in the ω-β_0 diagram represents the propagation characteristic for the case, all perturbing fingers aligned (ordinary inductive coupling).

in frequency and $\beta_0 L = \pi$ remains constant. This so-called "negative mutual-inductive coupling," therefore, increases the original bandwidth produced by the capacitive coupling and allows wide-band operation of the structure in the fundamental space harmonic.

Coupling Through Loops

Another important possibility for coupling between adjacent resonators is the use of coupling loops.[8] This type of coupling affects primarily the H lines and, therefore, the passband is lowered below the line of no coupling present (Figure 18). As in the previous case,

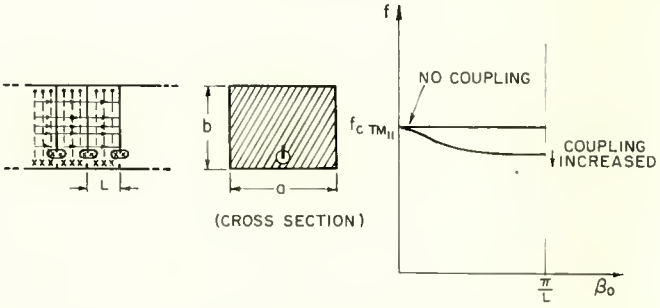


Fig. 18—Inductive coupling through loops.

rectangular cavities resonated in the TE_{101} mode of the same arrangement as described for pure capacitive coupling are used. The field configuration shown in Figure 18 corresponds to $\beta_0 L = 0$. For this case, the presence of the coupling loops does not perturb the H lines, and, therefore, the cutoff frequency $\beta_0 L = 0$ remains constant for different degrees of coupling. On the other hand, the field configuration for the cutoff point, $\beta_0 L = \pi$, is strongly affected, and, because the coupling has inductive character, the point $\beta_0 L = \pi$ decreases in frequency as the coupling increases (larger loop area).

Instead of coupling loops of the above type, S-shaped coupling loops such as those shown in Figure 19 can also be used. An investigation of the field pattern shows that this case is analogous to the "negative mutual-inductive coupling" through apertures, and, as before, the behavior of the two cutoff frequencies is reversed. While for the upper cutoff point, $\beta_0 L = \pi$, the current introduced in the loop on one side is continued in the corresponding current on the other side of the partition and the frequency is unperturbed, for the point $\beta_0 L = 0$ the currents are in opposite direction and the frequency is lowered with increasing coupling.

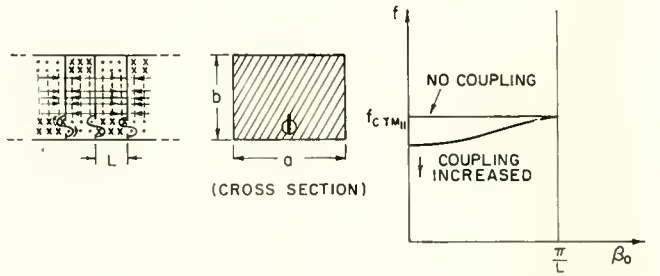


Fig. 19—Negative mutual inductive coupling through loops.

Coupling Between λ/2-Wires

Figure 20 shows an array of parallel wires extending across a waveguide. A structure of this type has two possible field configurations. In the symmetrical case all E lines either start or end on the wires. This field configuration is the important one because it provides a longitudinal E field for interaction with the electrons. The asymmetric field configuration, shown in Figure 20(c), corresponds to the TE_{10} mode of the waveguide and is only slightly perturbed by the presence of the wires. However, this asymmetrical mode, which has an ω-β_0 diagram corresponding to a slightly capacitively loaded waveguide, has no appreciable E_z component, and, therefore, will not be considered further.

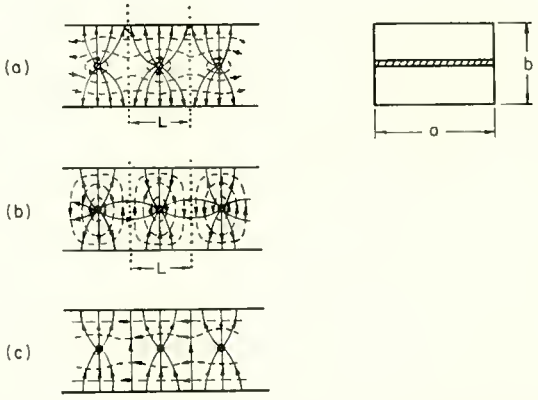


Fig. 20—Uncoupled λ/2 wires in a rectangular waveguide.
(a) Field distribution for $\beta_0 L = 0$ } Symmetric mode
(b) Field distribution for $\beta_0 L = \pi$ }
(c) Asymmetric mode ($\beta_0 L = 0$).

If the structure is separated along planes halfway between the wires, shown by the dotted lines in Figure 20(a), the field lines for the cutoff point $\beta_0 L = 0$ are essentially unperturbed, and the sections can be considered as coaxial resonators each having two wide, longitudinal slots. The resonant frequency of such a resonator is determined solely by its length, a, and is equal to $c/2a$. The other cutoff field distribution, $\beta_0 L = \pi$, is shown in Figure 20(b). In this case, the placing of shorting planes along the dotted lines does not perturb the field lines, and the resonators again have a resonance frequency at $c/2a$. The structure has, therefore, the same frequency for both cutoff field configurations; both inductive and capacitive coupling are present but the amounts are just equal and cancel each other, as in the case of mixed coupling described earlier.

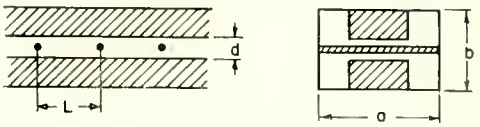


Fig. 21—Inductively coupled $\lambda/2$ wires.

The same result can be obtained also in a somewhat different way. The above structure is again considered, but with the addition of partitions halfway between the wires. If a slot is cut in the center of the partition to allow pure capacitive coupling, the point $\beta_0 L = 0$ increases in frequency while the point $\beta_0 L = \pi$ remains constant. On the other hand, if a slot is cut at the edge of the partition wall to provide pure inductive coupling, the frequency of the point $\beta_0 L = 0$ decreases. Therefore, when the partitions are entirely removed, both inductive and capacitive coupling are introduced and, consequently, the cutoff point $\beta_0 L = 0$ remains constant at $f = c/2a$.

Thus far, the structure consists of effectively uncoupled resonators and has no passband. In order to obtain a passband with finite bandwidth, the inductive coupling must be unbalanced with respect to the capacitive coupling. For example, a ridge along the structure can be added, as shown in Figure 21. Obviously, both points $\beta_0 L = 0$ and $\beta_0 L = \pi$ will be affected by the introduction of the ridge. However, for $\beta_0 L = \pi$ only some of the E lines (determined by the ratio b/L) are perturbed while in the case $\beta_0 L = 0$ all E lines are perturbed. When $b/L \gg 1$, for example, there is a strong E_z field between wires and relatively little E_t field for $\beta_0 L = \pi$. Therefore, the influence of the ridge is very strong for $\beta_0 L = 0$, but very slight for $\beta_0 L = \pi$. On the other hand, as b/L approaches 1, an appreciable E_t field also is present for the case $\beta_0 L = \pi$, and the difference in the change of frequency for $\beta_0 L = 0$ and $\beta_0 L = \pi$ is much less pronounced (Figure 22). Actually the intro-

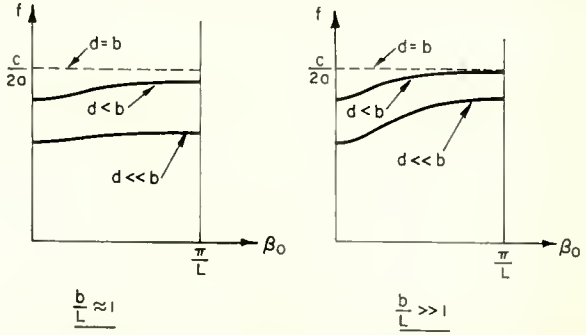


Fig. 22—ω-β_0 diagrams of the inductively coupled $\lambda/2$ wire structure for two different values of b/L. (For infinitely thin wires, the point $\beta_0 L = 0$ approaches the cutoff frequency of the corresponding ridged waveguide).

duction of the ridge does not affect the inductive coupling between wires, but reduces the number of E_z lines linking one wire with the other and therefore, reduces the capacitive coupling. The net effect is a dominant inductive coupling which results in the passband being lowered below the frequency with no coupling present.

A similar analysis can be made of the capacitively coupled $\lambda/2$-wire structure. In this case, the ridge affects mainly the inductive coupling,

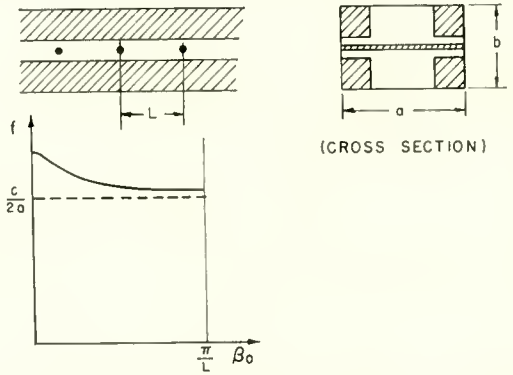


Fig. 23—Capacitively coupled $\lambda/2$ wire structure with corresponding ω-β_0 diagram.

and makes the capacitive coupling dominant, as shown in Figure 23. Here, the passband is raised above the frequency when no coupling is present. The same considerations with respect to the ratio b/L apply.

A practical example of these structures is the inductively coupled $\lambda/2$-wire structure used by Karp[9] for the generation of millimeter waves. The structure is operated in the first backward space harmonic as a backward-wave oscillator.

Coupling Through Alternately Displaced Coupling Holes

For reasons of simplicity, a rectangular waveguide with closely spaced partitions is again used. However, unlike the structures discussed thus far, where the coupling slots were axially aligned, every second slot is displaced vertically, as shown in Figure 24(b). A structure of this type belongs to the group of "detour" structures, and has a somewhat different behaviour. From a standpoint of strict field theory, the structure can still be considered as consisting of coupled cavities, but when the partitions are closely spaced the higher-order-

[9] A. Karp, "Backward-Wave Oscillator Experiments at 100 to 200 KMc," *Proc. I.R.E.*, Vol. 45, p. 496, April, 1957.

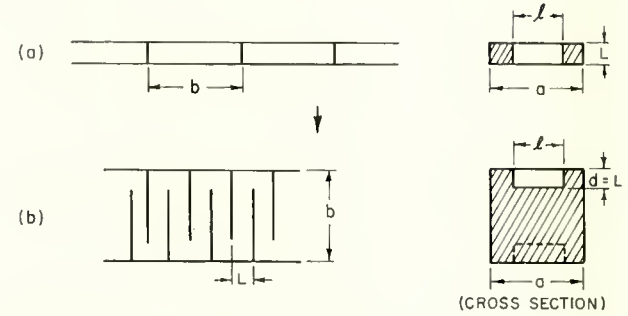


Fig. 24—Coupled cavities structure with alternately displaced coupling slots (a) Inductively loaded rectangular waveguide; (b) Structure with displaced coupling holes, obtained by folding of structure (a).

mode coupling between the displaced slots is different from that between aligned slots. When the partitions are very widely spaced the influence of higher-order-mode coupling is negligible, and there is no appreciable difference between the ω-β_0 diagrams for structures having aligned or displaced slots.

The structure having closely spaced partitions can be considered as an inductively loaded rectangular waveguide of height L, as shown in Figure 24 (a), which has been folded into the form shown in Figure 24 (b). The straight rectangular waveguide with periodic partitions, shown in Figure 24 (a), was discussed under pure inductive coupling. Because the folding process leads to an additional phase shift of π radians per cell, the ω-β_0 diagram of the straight loaded waveguide shown in Figure 25 (a) has to be shifted by $-\pi/b$ radians, and β_0 must be multiplied by b/L due to the difference in cell length. The final ω-β_0 diagram of the folded structure shown in Figure 25 (b) is actually drawn for positive phase velocity rather than positive group velocity, so that a direct comparison can be made with the ω-β_0 diagram

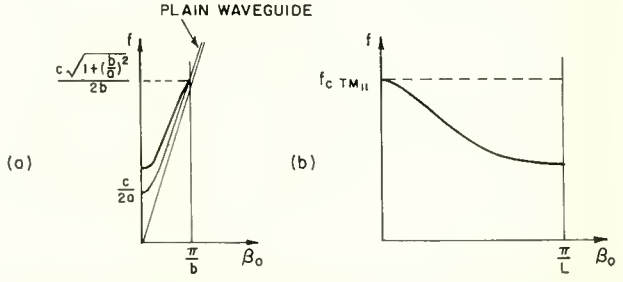


Fig. 25—(a) ω-β_0 diagram of inductively loaded waveguide shown in Figure 24a; (b) ω-β_0 diagram of structure with displaced coupling slots, obtained by folding the inductively loaded waveguide.

for the structure having aligned slots.*

For very small coupling holes, there is only slight coupling between slots and, consequently, the influence of the higher-order-mode coupling is not appreciable. Therefore, the ω-β_0 diagrams for the structure with displaced slots and for the similar structure with aligned slots are essentially the same. For larger holes, however, it has already been shown that in the structure with aligned slots, the upper cutoff point of the passband will start to decrease in frequency, leading to a reduction in bandwidth with a further increase in coupling. For the same hole size, the equivalent slot in the structure of Figure 24 (a) is far from being resonant in the frequency range considered, and, therefore, the upper cutoff point of the structure with the displaced holes will still remain constant. The structure with aligned holes experiences a mode change from a TM_{11} to a TE_{10} mode when the slot length is increased, while the structure in Figure 24 (a) remains constantly in the TE_{10} mode. For rather narrow coupling slots (for example when d is markedly smaller than L) the equivalent aperture of the structure in Figure 24 (a) can also become resonant and, therefore, cause a lowering of the upper cutoff frequency. The main difference between the two structures, however, is that the structure with aligned holes passes through a point where the passband diminishes entirely, while the passband of the structure with displaced coupling holes increases steadily with the elongation of the slots.

Measurements of ω-β_0 Diagrams for Various Types of Coupling

In order to verify some of the qualitative considerations presented in this paper, the ω-β_0 diagram was measured for various coupling types. The measurements were performed on a round waveguide structure consisting of rings and disks alternately stacked on top of each other and then securely clamped together (Figure 26). A round waveguide structure was chosen mainly to facilitate the machining process. Although most of the considerations discussed thus far are based on different arrangements of rectangular resonators, the results can also be applied, with some caution, to structures consisting of round waveguide sections. The TE_{10} and TM_{11} modes in a rectangular waveguide are closely similar to the TE_{11} and TM_{01} modes, respectively, in a round waveguide. However, due to the somewhat different field configuration, the coupling slots must be modified slightly to obtain the same passband behavior as in the case of rectangular resonators.

*As is well known, the ω-β_0 diagram is mirror-symmetrical around the line of zero phase shift and can be drawn for either a positive or negative phase velocity of the fundamental space harmonic, depending on the orientation of the chosen coordinate system.

The measurements were taken on a structure with four cells employing the resonance method.[3] In this method, the slow-wave structure is resonated by placing two shorting planes at suitable points along the structure, and then the corresponding β_0 of a particular resonance frequency is identified through observation of the field pattern of the resulting standing wave. Coupling holes of various shapes and sizes were cut out of the cell partitions, and the influence on the ω-β_0 diagram was determined.

First, a series of tests were conducted with slots of different length. A structure in which the slots are aligned and the cell length is small compared to the diameter of the waveguide corresponds very closely to the structure used in the discussion of mixed coupling. The influence of the slot length (the slot width is kept constant at 0.300 inch) on the ω-β_0 diagram is shown in Figure 27. The measured curves are in good agreement with the curves shown in Figure 14. The higher passband, a perturbed TE_{11} mode, decreases in frequency as the slot length

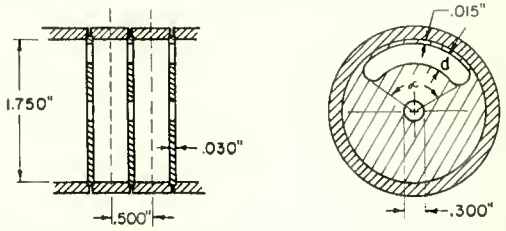


Fig. 26—Round waveguide with periodic partitions.

increases and undergoes a mode change to a perturbed TM_{01} mode. The lower passband, originally a perturbed TM_{01} mode, becomes for very large slots a perturbed TE_{11} mode. For a structure with very closely spaced partitions, similar mode changes also occur at higher frequencies whenever the TE_{11} mode, decreasing in frequency from $f = c\ [1 + (2L/\lambda_{cTE_{11}})^2]^{1/2}/2L$, approaches a passband of one of the higher waveguide modes. The cutoff frequency of the TE_{11} mode has a field configuration which prevents its excitation in a short-circuited structure of finite length. This agrees well with the measured results, where the dashed lines in Figure 27 indicate that no resonance $\beta_0 L = 0$ could be detected.

However, one discrepancy shows up for very long slots. The reversal in the slope of the passband occurs below the cutoff frequency of the TE_{11} mode, while in the rectangular structure it is expected to occur above the cutoff frequency of the corresponding TE_{10} mode, as shown in Figure 14. This shift below $f_{cTE_{11}}$ is due to the different position and shape of the slot with respect to the field configuration of the TE_{11}

mode. For example, in the rectangular structure the condition $l = a$ implies that the lower cutoff frequency of the passband is fixed at $f = c/2a$ and only the upper cutoff frequency is influenced by the slot width. In the round structure, due to the different field configuration, no such simple relation exists. There must be a whole set of passbands for which the lower cutoff frequency is fixed at $f_{cTE_{11}}$ and only the upper cutoff point is moving, but the slots for these passbands will not be of equal length. In the round structure both length and width have to be changed to keep the lower cutoff frequency constant at $f_{cTE_{11}}$.

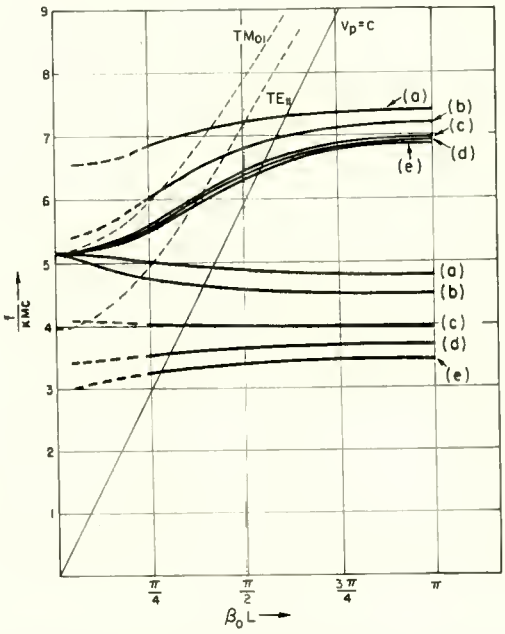


Fig. 27—ω-β_0 diagrams for various slot lengths ($d = .300$ inch = constant). (a) $\alpha = 70°$, (b) $\alpha = 90°$, (c) $\alpha = 120°$, (d) $\alpha = 140°$, (e) $\alpha = 155°$. Only the two lowest pass bands are shown.

For example, a slot having an opening which corresponds to an angle of 140° (d in Figure 27) has a resonance that is obviously below $f_{cTE_{11}}$. Now, in order to obtain a passband behavior similar to the one in the rectangular structure, the whole passband must be raised in frequency. In the discussion of coupling through a resonant slot, it was shown that the passband is reduced as the slot is shortened, while an increase in the slot width widens the passband and also raises the whole passband in frequency. The slot in the round structure, therefore, must be widened to avoid a slope reversal below the cutoff fre-

quency of the TE_{11} mode. Measurements with a slot 0.500 inch in width confirmed this expectation, as shown in Figure 28. Thus, the behavior of both structures is alike if the correct slot dimensions are used.

Another series of measurements was made to find the effect of vertical displacement of every second slot. The same disks were used as before, but every second one was rotated 180°. The results, shown in Figure 29, are in good agreement with those obtained earlier for the rectangular guide. The slot dimensions for the curves (b) and (c) in Figure 29 are such that the slots are resonant within the passband,

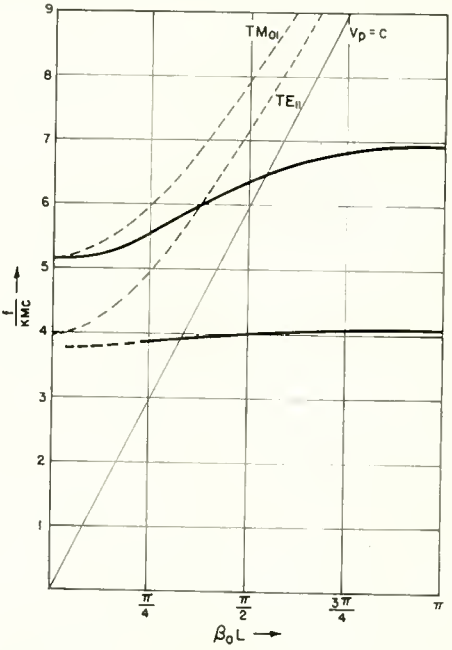


Fig. 28—ω-β_0 diagram for slot with increased width ($d = .500$ inch, $\alpha = 140°$).

and the upper cutoff frequency is lowered below $f_{cTM_{01}}$, but contrary to the case where the slots are aligned, the passband does not show a reversal in the slope.

Finally, some measurements were taken with partitions having two inductive coupling slots cut out, as shown in Figure 30. This structure, originally suggested by Nalos,[7] requires special attention because it can be explained from two different viewpoints. In one approach, the structure is considered as round waveguide sections having mainly inductive coupling between sections. This procedure follows closely that outlined in the discussion of mixed coupling. For the dimensions used

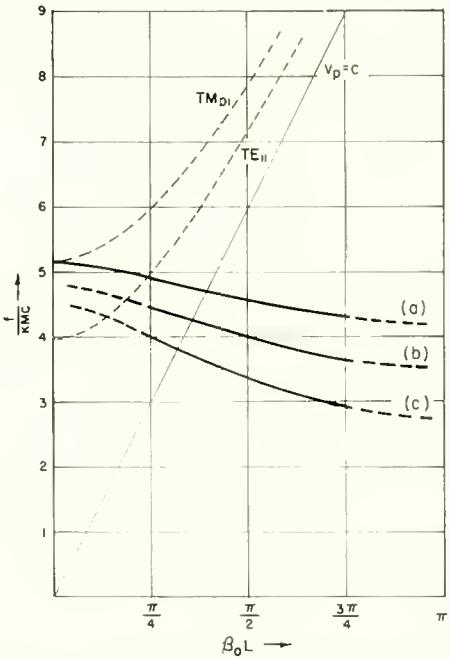


Fig. 29—ω-β_0 diagrams for a structure with displaced slots (d = .300 inch = constant). (a) $\alpha = 90°$, (b) $\alpha = 120°$, (c) $\alpha = 155°$. Only the lowest pass band is shown.

in this particular case, the slots are very long, and, therefore, the upper cutoff frequency of the passband is shifted somewhat below $f_{c_{TM_{01}}}$. However, little additional design information can be derived from this approach.

The alternate approach to an analysis of this structure provides a much better understanding of its behavior. If the structure is compared with the capacitively coupled wire structure shown in Figure 23, a very close geometrical similarity becomes evident. The bars replace

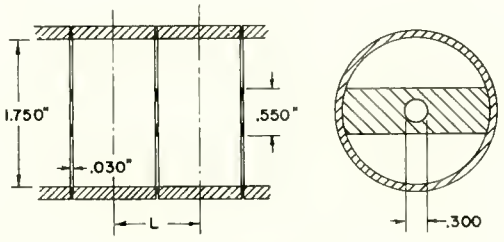


Fig. 30—Slow-wave structure having partitions with two inductive coupling slots.

the round thin wires, and the curvature of the round waveguide produces the same enhancement of the capacitive coupling as the inductive ridge does in the coupled wire structure.

There is no discrepancy in the definition of the coupling. If the structure is considered as being derived from originally uncoupled round waveguide sections, the partition walls, which separate adjacent cells, coincide with the planes of the bars. In this case the lines crossing through the partition planes are mainly H lines, and, therefore, the coupling is inductive (passband perturbed downwards in frequency

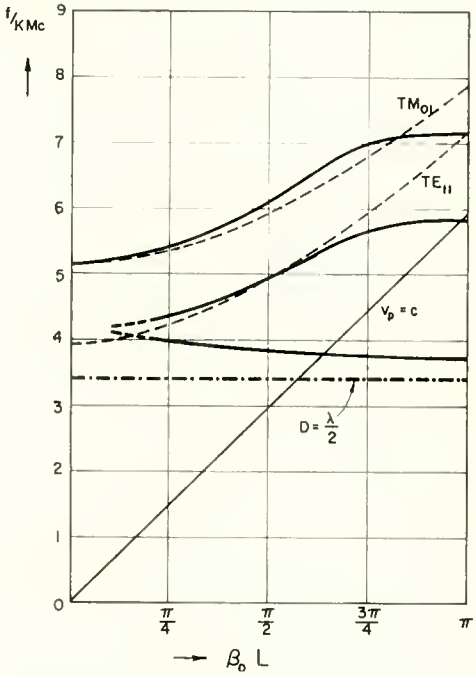


Fig. 31—ω-β_0 diagram for the structure shown in Figure 30 (cell length $L = 1$ inch).

from the line with no coupling present). On the other hand, the coupled wire structure has been derived from uncoupled coaxial resonators. Here the partitions (now halfway between the bars) could be removed without causing any net coupling, while the introduction of the ridge, or in this case the curvature of the waveguide, produces a dominant capacitive coupling. This coupling is caused by E lines crossing the partition planes between the bars. Because of the different ways of defining the partition planes, therefore, the same structure is said to employ in one case inductive coupling and in the other case capacitive coupling. The two approaches do not contradict each other.

Measurements performed on two structures with different cell lengths, L, shown in Figures 31 and 32, indicated good agreement with the theoretical considerations. The lower cutoff frequency of the structure lies somewhere above $f = c/2D$ depending on the ratio of D/L and the cross sections of the bars.* If the period, L, is short compared to the diameter, D, the influence of the inductive ridge on the field configuration of the cutoff point $\beta_0 L = \pi$ is reduced,

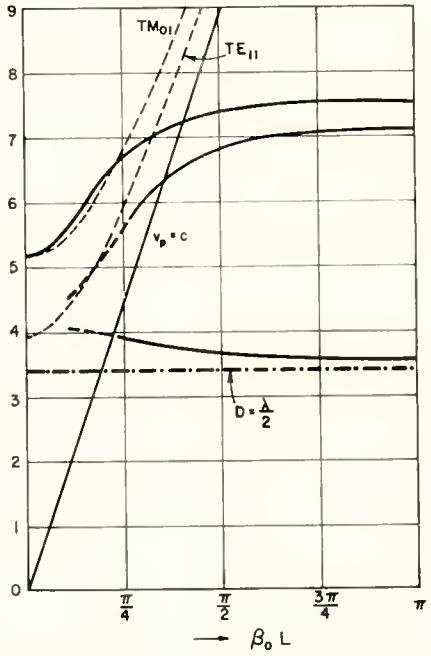


Fig. 32—ω-β_0 diagram for the structure shown in Figure 30 (cell length L = .330 inch).

but no appreciable change results for the field configuration of $\beta_0 L = 0$. Consequently, a reduction in L brings the cutoff point $\beta_0 L = \pi$ closer to $f = c/2D$ and therefore increases the bandwidth.

The presence of the next higher passband is also easily understood, if the structure is considered as a variation of the coupled-wire structure. This passband, with its lower cutoff frequency approaching the cutoff point $\beta_0 L = 0$ of the first passband, can be identified as the asymmetrical mode, shown in Figure 20 (c) which has no appreciable E_z field. Perturbation measurements performed by moving a bead along

* The use of drift tubes which cause additional capacitive loading may lower the point $\beta_0 L = \pi$ even somewhat below $f = c/2D$.

the axis of the structure confirmed this expectation. The third passband, starting from the cutoff frequency of the TM_{01} mode, originates from a mainly capacitively loaded TM mode.

Conclusion

In this paper, a general procedure has been outlined by which a large number of structures can be evaluated from one common point of view. This approach gives a good qualitative picture of the behavior of different types of slow-wave structures and clarifies the close relationships and analogies that exist between certain of these structures.

The basic rules about the two major coupling types are: inductive coupling, caused by H lines linking the cavities, tends to lower the passband; capacitive coupling, caused by E lines linking the cavities, tends to raise the passband. Which cutoff point remains constant and which one moves (sometimes, both points move) cannot be decided offhand and will differ from structure to structure. If the two basic coupling types are extended to the limit of full coupling, it is found that pure capacitive coupling leads to a TM mode of propagation, while pure inductive coupling leads to a TE mode of propagation.

The two possible approaches that can be taken to explain the ω-β_0 diagram of different structures are then clearly indicated. In one case the investigation starts with uncoupled resonators and the effect on the two cutoff points is determined when coupling is introduced. In the other case the investigation starts with the extreme case of full coupling, which corresponds to a plain waveguide, and the influence on the propagation characteristic is determined when periodically spaced loading plates are introduced. Some structures (for example, the disk-loaded round waveguide) can be explained equally well by either approach.

Finally, it should be pointed out that the structures discussed in this paper do not represent all the possible slow-wave structures that can be explained on this basis, but merely are basic examples for certain coupling types. The method outlined in the paper is, in principle, applicable to all slow-wave structures which can be derived from one of the two extreme cases of no coupling or full coupling between cells.

RCA TECHNICAL PAPERS†

First Quarter, 1958

Any request for copies of papers listed herein should be addressed to the publication to which credited.

"Activation of Silver-Magnesium and Copper-Beryliium Dynodes," A. H. Sommer, *Jour. Appl. Phys.* (March) (Letter to the Editor) 1958

"Analysis of the Effect of Nuclear Radiation on Transistors," J. J. Loferski, *Jour. Appl. Phys.* (January) 1958

"Automatic Cuing of Television Film Projectors," B. F. Melchionni, *Jour. S.M.P.T.E.* (February) 1958

"Beam-Landing Errors and Signal-Output Uniformity of Vidicons," R. G. Neuhauser and L. D. Miller, *Jour. S.M.P.T.E.* (March) 1958

"On Closed-Form Expressions for Mean Squares in Discrete Continuous Systems," J. Sklansky, *Trans. I.R.E. PGAC* (March) 1958

"Color Processing in RCA Video Tape Recorder," A. H. Lind, *Broadcast News* (February) 1958

"Considerations Affecting the Rise and Decay of Cathode Currents in Receiving Tubes," E. R. Schrader, *RCA Review* (March) .. 1958

"Controlled Thermonuclear Fusion — Promise of the Future," G. Warfield, *I.R.E. Student Quarterly* (February) 1958

"Creative Engineering Methods," A. N. Goldsmith, *Trans. I.R.E. PGEM* (March) 1958

"Crystallographic and Magnetic Studies of the System $(NiFe_2O_4)_{1-x} + (NiMn_2O_4)_x$," P. K. Baltzer and J. G. White, *Jour. Appl. Phys.* (March) 1958

"Decomposition Method for Producing p-n Junctions in InP," K. Weiser, *Jour. Appl. Phys.* (February) (Letter to the Editor) 1958

"Differential Method of Lag Compensation in Photoconductive Devices," H. Borkan and P. K. Weimer, *RCA Review* (March) 1958

"Discussion of 'Statistical Treatment of Sampled-Data Control Systems for Actual Random Inputs,' by M. Mori of Japan," J. Sklansky, *Trans. ASME* (February) 1958

"Electronic Extension of Optical Observation Techniques," V. K. Zworykin, *Science* (March 28) (Letter to the Editor) 1958

"An Electrostatically Focused Traveling-Wave-Tube Amplifier," K. K. N. Chang, *RCA Review* (March) 1958

"Elimination of Cold-Junction Error in Thermocouple Measurements in Electron Tubes," I. S. Solet, *Rev. Sci. Instr.* (Notes) (January) 1958

"Emission Mechanism of Cold-Cathode Arcs," K. G. Hernqvist, *Phys. Rev.* (February 1) 1958

"Engineers Should Write!," W. O. Hadlock, *Electronic Industries* (January) 1958

"Etched I-F Amplifier Pares Color TV Cost," L. Ruth, *Electronics* (March 14) 1958

"Formula for Platform Poise," R. J. Norko, *Trans. I.R.E. PGEWS* (March) 1958

"Ham Shack Trouble-Shooter," Part II, R. Samuel, *RCA Ham Tips* (February) 1958

"High-Fidelity Audio Amplifiers Using RCA-6973 Beam Power Tubes," *RCA Application Note AN-173*, Electron Tube Division, Radio Corporation of America, Harrison, N. J. (February) 1958

† Report all corrections or additions to *RCA Review*, RCA Laboratories, Princeton, N. J.

"High-Voltage Photovoltaic Effect," L. Pensak, *Phys. Rev.* (January 15) (Letter to the Editor) 1958
"Hollow-Cathode Glow Discharge in Mercury Vapor," K. G. Hernqvist, *RCA Review* (March) 1958
"How to Use High-Speed Motion-Picture Cameras," D. Colasanto, *Machine Design* (January 23) 1958
"A Hysteresis Effect in Cadmium Selenide and Its Use in a Solid-State Image Storage Device," F. H. Nicoll, *RCA Review* (March) 1958
"Improved Developmental One-Inch Vidicon for Television Cameras," L. D. Miller and B. H. Vine, *Jour. S.M.P.T.E.* (March) 1958
"Improved Power Fault Protection," J. T. Mark, *Electronic Design* (March 5) 1958
"Intrinsic Optical Absorption in Germanium-Silicon Alloys," R. Braunstein, A. R. Moore, and F. Herman, *Phys. Rev.* (February 1) 1958
"KGHL-TV Gets First Traveling Wave Antenna," *Broadcast News* (February) 1958
"Large-Area Germanium Power Transistors," B. N. Slade and Jane Printon, *RCA Review* (March) 1958
"Medium Screen Color Television Projection," S. L. Bendell and W. J. Neely, *Jour. S.M.P.T.E.* (March) 1958
"Methods for Determining Amplitude-Modulation Rejection Performance of Frequency-Modulation Detectors," R. J. Schultz, *Trans. I.R.E. PGBTR* (February) 1958
"Microwave Ferrite Frequency Separator," H. Rapaport, *Trans. I.R.E. PGMTT* (January) 1958
"More Senses Make More Sense," C. N. Hoyler, *Trans. I.R.E. PGEWS* (March) 1958
"Mountains Increase Radio Reception," M. P. Bachynski, *Canadian Electronics Engineering* (January) 1958
"A New Aural and Visual TV Exciter," H. E. Small, *Broadcast News* (February) 1958
"Noise Shield for Microphones Used in Noisy Locations," M. E. Hawley, *Jour. Acous. Soc. Amer.* (March) 1958
"Operating Range of a Memory Using Two Ferrite Plate Apertures Per Bit," M. M. Kaufman and V. L. Newhouse, *Jour. Appl. Phys.* (March) 1958
"Parametric Amplification Using Low-Frequency Pumping," S. Bloom and K. K. N. Chang, *Jour. Appl. Phys.* (March) (Letter to the Editor) 1958
"Photographing C-R Tube Images," R. Samuel, *Radio-Electronics* (February) 1958
"Photovoltages Larger than the Band Gap in Zinc Sulfide Crystals," S. G. Ellis, F. Herman, E. E. Loebner, W. J. Merz, C. W. Struck, and J. G. White, *Phys. Rev.* (Letter to the Editor) (March 1) 1958
"A Plea for Maximum Utility in Government Contract Reports Covering Research and Development," E. W. Herold, *Proc. I.R.E.* (January) (Letter to the Editor) 1958
"Progress Report on Improved B-W Picture Performance Available on Color-TV Sets," E. R. Klingeman, *Service* (January) 1958
"Properties of Photovoltaic Films of CdTe," B. Goldstein, *Phys. Rev.* (January 15) (Letter to the Editor) 1958
"Proposed Negative-Mass Microwave Amplifier," H. Kroemer, *Phys. Rev.* (Letter to the Editor) (March 1) 1958
"A Pulse Emission Test for Field Testing Hot-Cathode Gas Tubes," *RCA Application Note AN-172*, Electron Tube Division, Radio Corporation of America, Harrison, N. J. (February) 1958
"Radiation Induced Noise in p-n Junctions," W. H. Fonger, J. J. Loferski, and P. Rappaport, *Jour. Appl. Phys.* (March) 1958
"RCA's Contribution to Defense," A. L. Malcarney, *Signal* (January) 1958
"Simple Shop Practices for Short Runs," H. J. Ackerman, L. J. Caprarola, and G. C. Walther, *American Machinist* (January 13) 1958

"A Solid-State Amplifying Fluoroscope Screen," B. Kazan, *RCA Review* (March) 1958
"Some New Structure-Type Targets for the Vidicon—An Analysis of Their Operation," S. A. Ochs and P. K. Weimer, *RCA Review* (March) 1958
"Some Properties of Gallium Arsenide-Germanium Mixtures," D. A. Jenny and R. Braunstein, *Jour. Appl. Phys.* (March) (Letter to the Editor) 1958
"Sputtering of Surfaces by Positive Ion Beams of Low Energy," R. E. Honig, *Jour. Appl. Phys.* (March) 1958
"A Stabilized Monitor for Color Television Picture Quality Control," E. E. Gloystein and N. P. Kellaway, *Jour. S.M.P.T.E.* (March) 1958
"Synchronous and Exalted-Carrier Detection in Television Receivers," J. Avins, T. Brady and F. Smith, *Trans. I.R.E. PGBTR* (February) 1958
"Talos Defense Unit," H. W. Phillips, *Signal* (March) 1958
"Ten Recommendations to Meet a Challenge," D. Sarnoff, *Signal* (February) 1958
"A Ten-Watt High-Quality Transistorized Audio Power Amplifier," R. Minton and C. F. Wheatley, *I.R.E. Student Quarterly* (February) 1958
"Theoretical Investigation of the Electronic Energy Band Structure of Solids," F. Herman, *Rev. Mod. Phys.* (January) 1958
"Theory of Interstitial Impurity States in Semiconductors," P. E. Kaus, *Phys. Rev.* (March 15) 1958
"The Thyristor," L. E Barton, *Electronic Design* (March 19) 1958
"The 'Thyristor'—A New High Speed Switching Transistor," C. W. Mueller and J. Hilibrand, *Trans. I.R.E. PGED* (January) .. 1958
"Transistorized Memory Monitors Earth Satellite," C. S. Warren, W. G. Rumble, and W. A. Helbig, *Electronics* (January 17) .. 1958
"Transistorized Turntable Preamplifier," G. C. Weilenmann, *Broadcast News* (February) 1958
"Tropospheric Scatter Propagation—A Summary of Recent Progress," H. Staras, *RCA Review* (March) 1958
"Variation of Transistor Parameters with Temperature," C. R. Eshelman, *Semiconductor Products* (January/February) 1958

Correction:

The following corrections apply to the paper entitled "Hollow-Cathode Glow Discharge in Mercury Vapor," by K. G. Hernqvist, which appeared on pages 35-48 of the March 1958 issue:

On page 43, in the last two lines, the phrase which reads ". . . a microwave signal whose frequency, f, is *larger* than . . ." should, instead, read ". . . a microwave signal whose frequency, f, is *lower* than . . ."

On page 44, in the first three lines of text below the caption for Figure 4, the symbols for signal frequency (f) and plasma frequency (f_e) have been transposed. The lines should read:

(1) For $f_e \geqq f$, the plasma is almost opaque;

(2) For $f_e \ll f$, the plasma is transparent;

(3) For f_e nearly equal to but less than f, the . . .

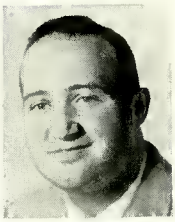
AUTHORS

ERWIN BELOHOUBEK received the degree of Diplom-Ingenieur in 1953 and the Ph.D. in Electrical Engineering in 1955 from the Technical University in Vienna, Austria. From 1953 to 1955 he worked as Research Assistant at the Institute for High-Frequency Techniques of the same university. He joined the RCA Tube Division in Harrison, N. J. in 1956 and transferred to Princeton in 1957, where he is engaged in work on traveling-wave tubes and magnetrons.

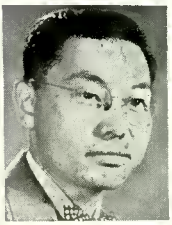
MURLAN S. CORRINGTON received the B.S. degree in Electrical Engineering in 1934 from South Dakota School of Mines and Technology, and the M.Sc. degree in 1936 from Ohio State University. From 1935 to 1937 he was a graduate assistant in the Physics Department of Ohio State University. In 1937 he joined the Rochester Institute of Technology, where he taught mathematics, mechanics, and related subjects. Since 1942 he has been engaged in Mathematical Engineering in the Advanced Development Section of the RCA Victor Television Division, Radio Corporation of America, at Camden, New Jersey. He is Manager of Audio, Acoustics, and Antennas for the Section. He is a member of Sigma Pi Sigma, the Society for Industrial and Applied Mathematics, a Fellow of the Acoustical Society of America, and a Fellow of the Institute of Radio Engineers.

K. G. HERNQVIST—(see *RCA Review*, Vol. XIX, No. 1, March 1958, p. 35.)

EDWARD W. HEROLD was born in 1907 in New York City. He received the B.S. degree in physics from the University of Virginia in 1930, and the M.S., also in physics, from the Polytechnic Institute of Brooklyn, 1942. From 1924 to 1926 he was associated with the Bell Telephone Laboratories, and from 1927 to 1929 with E. T. Cunningham, Inc. Since 1930 he has been with the Radio Corporation of America, first at Harrison, New Jersey, and, since 1942, at RCA Laboratories, Princeton, New Jersey, where he is Director, Electronic Research Laboratory. Temporarily, Mr. Herold is Associate Project Manager and the senior RCA representative in "C" Stellarator Associates, a group employed on a thermonuclear fusion research project at Princeton University. Mr. Herold has specialized in research and development on electron tubes and semiconductor devices. He is a member of Phi Beta Kappa, Sigma Xi, and a Fellow and member of the Board of Directors of the Institute of Radio Engineers.



MORTON KANEFSKY received the B.S. degree in Electrical Engineering from the University of Pennsylvania in 1957. Upon graduation he joined RCA Laboratories at Princeton, N. J. He is presently attending Princeton University as a member of the RCA Graduate Study Program. Mr. Kanefsky is a member of Eta Kappa Nu, Tau Beta Pi, Sigma Tau, Pi Mu Epsilon, and the Institute of Radio Engineers.



T. MURAKAMI received the B.S. degree in E.E. from Swarthmore College in 1944, and the M.S. degree from the Moore School of Electrical Engineering, University of Pennsylvania in 1947. From 1944 to 1946 he was an assistant and research associate in the Department of Electrical Engineering at Swarthmore College. Since 1946 he has been with the Advanced Development Section of the RCA Victor Television Division, Camden, N. J., working on radio frequency circuit development. Mr. Murakami is a Senior Member of the Institute of Radio Engineers and a member of Sigma Xi.

FRANK H. NORMAN attended the University of Connecticut and Rutgers University. From 1926 to 1927 he worked on problems of train control and electric locomotive power with the New Haven Railroad. From 1927 to 1935 he was engaged by the New York Edison Company to work in their meter laboratory. In 1935 he engaged in the design and construction of custom amateur equipment and continued in this field until 1941 when he joined the staff of WTIC as studio and master control engineer. In 1942 he joined the Remington-Rand Electronic Division where he worked on television pickup tube production and development. Mr. Norman joined RCA Laboratories in 1947 where he has been engaged in work on microwave relays, high-power transmitting tube development, and gaseous electronic problems.

DONALD W. PETERSON received the B.S. degree in E.E. from the University of Wisconsin in 1936. In that year, he joined the Service Department of RCA Manufacturing Company, Camden, N. J., shifting to the research division of that company in 1939. In 1942 he was transferred to RCA Laboratories Division in Princeton, N. J., where he is currently engaged in work on antennas, transmission lines and propagation. Mr. Peterson is a Member of the Institute of Radio Engineers.

JAMES W. SCHWARTZ received the B.S. and M.S degrees in Engineering Physics from Cornell University in 1951 and 1952 respectively. He has been employed at the Corning Glass Works and at Oak Ridge National Laboratories. In December, 1952, he joined the Technical Staff of RCA Laboratories where he became engaged in electron optics and picture reproducer research. In June, 1958, he joined the Kaiser Aircraft and Electronics Company, Oakland, Calif. Mr. Schwartz is a member of the Institute of Radio Engineers and Sigma Xi.

AUBREY W. VOSE graduated from the general course in Communications Engineering at RCA Institutes, New York, in 1939. He joined Westinghouse Electric in 1940 as Test Equipment Designer, transferring to the Engineering Department in 1942 where he spent the next three years as Design Engineer on the SCR-584, AN/APS-6 and other ground and airborne radar systems. In 1946 he entered into partnership in a consulting firm. In 1948 he joined the Houston Corporation where he worked on military airborne radar. In 1950 he joined RCA. He received the RCA Award of Merit in 1956 for his work on weather radar. Mr. Vose is a Member of the Institute of Radio Engineers and is a Registered Professional Engineer in the State of California.

GEORGE WARFIELD received the B.S. degree in Physics from Franklin and Marshall College in 1940. From 1941-1944 he was a teaching assistant in Physics at Cornell University. From 1944-1946 he worked on the proximity fuse project at McQuay-Norris Mfg. Co. He returned to Cornell after the war and received the Ph.D. in Physics in 1949. Since that time he has been on the faculty of the Department of Electrical Engineering at Princeton University with the present rank of Associate Professor. In addition he has, since 1951, been serving as a member of the RCA Laboratories Technical Staff on a part-time basis. Dr. Warfield is a member of the American Physical Society, the American Society for Enginering Education, Sigma Xi, and Phi Beta Kappa.

FRED V. WILSON received the A.B. degree in physics from the University of California at Berkley in 1949. At North American Aviation, Los Angeles, he was employed as an aerodynamicist doing aircraft stability and fire control evaluation, and at West Coast Research Company, Los Angeles, on missile flight test programs. Since 1954 Mr. Wilson has been with RCA Defense Electronic Products at Los Angeles where he has been engaged in systems engineering work in connection with radar systems optimization and performance, weather radar detection and penetration studies, proximity warning and doppler navigator systems analysis, and radar counter-countermeasures. Mr. Wilson is a Member of the American Physical Society and an Associate Member of the Institute of Radio Engineers.

HERBERT J. WOLKSTEIN received the B.S. degree in Electrical Engineering from Newark College of Engineering and is presently working toward an MS degree in the same field. Before joining RCA, he was employed at the Research Laboratories of National Union Electric Corporation as Project Engineer on the design of special-purpose beam-deflection and computer tubes. He joined the Microwave Tube Engineering activity of the RCA Electron Tube Division in 1955 and has since been engaged in the development of traveling-wave tubes. He is the author of a number of articles pertaining to this work. Mr. Wolkstein is a member of the Institute of Radio Engineers.