General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

NASA TECHNICAL MEMORANDUM

NASA TM-75453

THE PRESENT SITUATION AND FORECASTS OF SEMICONDUCTOR ELEMENTS PERFORMANCE WITHIN THE MICROWAVE RANGE, 1970 - 1985

Bertil Peterson

(NASA-TM-75453) THE PRESENT SITUATION AND FORECASTS OF SEMICONDUCTOR ELEMENTS PERFORMANCE WITHIN THE MICROWAVE RANGE, 1970-1985 (National Aeronautics and Space Administration) 229 p HC A11/MF A01 G3/33

N78-33338

Unclas 33660

Translation of "Nulaege och prognos foer FT elements prestanda inom mikrovagsomradet 1970-1985," Research Institute of National Defence, Stockholm, Sweden, Sept. 1971, 224 pages

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, D.C. 20546 OCTOBER, 1978

STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA TM-75453		

4. Title and Subtitle	5. Report Date
THE PRESENT SITUATION AND FORECASTS OF SEMICONDUCTOR ELEMENTS PERFORMANCE WITHIN THE MICROWAVE RANGE, 1970-1987	October, 1978
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Bertil Peterson Research Institute of National Defense	10. Work Unit No.

9. Performing Organization Name and Address	11. Contract or Grant No.
Leo Kanner Associates Redwood City, California 94063	NASw-3199
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address	Translation
National Aeronautics and Space Administration, Washington, D.C. 20546	14. Sponsoring Agency Code

15. Supplementary Notes

Translation of "Nulaege och prognos foer FT elements prestanda inom mikrovagsomradet 1970 - 1985," Research Institute of National Defence, Stockholm, Sweden, Sept. 1971, 224 pages (N74-13483)

16. Abstract

The present situation and possible developments over the period 1970-1985 for active semiconductor elements in the microwave range are outlined. After a short historical survey of FT techniques, the following are discussed: Generation, power amplification, amplification of small signals, frequency conversion, detection, electronic signal control and integrated microwave circuits.

17. Key Words (Selected by Author(s))	18. Distribution Statement
State of art and future. Semiconductor devices. Microwave. Transistors, diodes, generation, power amplifiers. Low noise amplifiers.	Unclassified-Unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified		

TABLE OF CONTENTS

Introduction 1

Short Summary of the Development of Active Semiconductor Devices in the Microwave Range 1

Generation and Power Amplification 6

The High Frequency Transistor 6

The Development of the Power Transistor 10

Pulse Operation in the Transistor 12

A New Type of Transistor 14

The Frequency Multiplier 16

The Use of Multiple Varactors 22

Bandwidth of the Frequency Multiplier 23

Noise Characteristics of Frequency Multipliers 23

The Avalanche Diode Oscillator 26

Power-Frequency Limits in Avalanche Diodes 27

Noise in the Avalanche Diode Oscillator 35

Electronic Tuning of the Avalanche Oscillator 36

Power Amplification with the Avalanche Diode 36

The Gunn Oscillator 40

Power Limits of the Gunn Diode 42

Efficiency and Forecast of Power Output in the Gunn Diode 50

Noise Properties in the Gunn Diode Oscillator 51

Electric Tuning of the Gunn Diode Oscillator 53

Frequency Locking in the Gunn Diode Oscillator 53

Life Span of the Gunn Diode Oscillator 54

Future Variants of the Gunn Oscillator 54

TABLE OF CONTENTS (Con't)

Power Amplification with the Gunn Diode 55

The Electroacoustic Amplifier 59

Other Semiconductor Oscillators 67

The Tunnel Diode Oscillator 67

The Josephson Oscillator 74

Microwave Generation with Indium Antimonide 77

Comprehensive Forecast for Power Generating Solid-State Devices 78

Small Signal Amplification 81

The Low Noise Transistor 81

Parametric Amplifiers 85

Future Development of the Parametric Amplifier 92

The Tunnel Diode Amplifier 97

Bandwidth 101

Noise factor 102

The Maser Amplifier 104

Future Development of the Maser 106

Other Low Noise Solid-State Amplifiers 108

A Summarizing Forecast for Low Noise Amplifiers 108

Frequency Conversion 111

Up-Converters for the Upper Sideband 111

The Mixer Diode 118

Estimating the Lowest Possible Noise Factor for a Mixer 122

The Broadband Case 122

The Image Frequency Sees an Open Circuit 122

Maximum Power Level for a Mixer Diode 126

TABLE OF CONTENTS (Con't)

The Saturation Level 126

The Mixer Diode's Power Durability 127

Broadband Mixers 128

The Tunnel Diode Mixer 129

The Josephson Junction as a Local Oscillator and Mixer 130

Detection 134

The Detector Diode 134

The Detector's Quadratic Range 141

Less Common Types of Detectors 143

The Josephson Detector 143

The Thermoelectric Detector 145

The Dielectric Detector 145

Combinations of Pre-Amplifiers and Detectors 145

Electronic Switching 149

The Breaker Diode 149

General Information on the High-Frequency Breaker 149

High-Frequency Breakers with PIN diodes 151

General Characteristics 151

Power Durability 153

Attenuators with PIN diodes 162

Phase Shifters with PIN diodes 163

The Phase Shifter's Power Durability 166

PIN diode Limiters 171

Other Signal Manipulation 178

TABLE OF CONTENTS (Con't)

Integrated Microwave Circuits 182

The Integration of Circuits and Active Semiconductor Devices 182

Maximum Frequency for Microstrip Lines and Concentrated Impedance Devices 187

The Power Durability of Microstrip Lines 189

Some General Views on the Future Prospects for Integrated Circuits 189

THE PRESENT SITUATION AND FORECASTS OF SEMICONDUCTOR ELEMENTS PERFORMANCE WITHIN THE MICROWAVE RANGE, 1970 - 1985

Bertil Peterson
Research Institute of National Defense

Introduction

The main purpose of this report is to give an explorative /5* forecast for active semiconductor performance and characteristics within the microwave range around the year 1985. Starting with the available information in the scientific literature, various attempts have been made to find the outermost, physical limits for each component's function. When this has not been possible, more empirical evaluation methods have been used. Starting with recognized trends in development, the forecasted values are shown at a sufficient distance toward their "absolute" maximum values.

Since a technical forecast should be based on both theoretical models and earlier development history, every semiconductor device discussed here is characterized in a brief description of its mode of operation and its development up until 1970. The descriptions are at times more complete than would have been necessary solely for the purposes of forecasting, but this is due to the fact that the report is also hoped to be of use as a reference book covering the many new active semiconductor components which have been introduced during the 1960's. In the reference section after each chapter, the sources of more detailed descriptions of the respective semiconductor components are given. With this addition, the report will be of use to a significantly larger readership than that which is only interested in the forecasting aspects.

Due to the large scope of this report's subject matter, it is impossible for one author to have practical experience with all the discussed semiconductor devices or to have mastered the necessary theoretical background. Therefore, it is hoped that the forecasts presented here will form a basis for an inquiry of a number of specialties on their views on the future of semiconductor devices. If a sufficient number respond, it is planned that a revision and /6 synthesis will later be published as an appendix to this report. It is hoped that this somewhat preliminary effort will develop into a better-organized final product.

Short Summary of the Development of Active Semiconductor Devices in the Microwave Range

Natural polycrystalline semiconductors were used within radio

---

* Numbers in the margin indicate pagination in the foreign text.

technology as detectors at a very early stage. The electron vacuum tube, which at low frequencies yielded better detector and mixer results, was invented around 1925. As a result of this, the development of the crystal diode was temporarily halted. Around 1940, the rapidly growing interest in radar and microwave techniques again put the crystal diode in a position of importance. In particular, the diode was far superior to the converter tube as a mixer in a superheterodyne receiver in the microwave range.

During the subsequent development of semiconductor technology, silicon and germanium crystals were produced with much greater purity, and around 1943, it was discovered that extremely small additions or impurities could increase the qualities of the semiconductor. These technological improvements made possible the extremely significant invention of the transistor in 1948. At the beginning of the 1950's, methods for producing single crystal semiconductors of high purity through a combination of zone refining and the growth junction technique were improved. Diffusion replaced alloyage as a method for introducing impurities and, around 1960, planar and epitaxial processes were discovered. These permitted the inexpensive mass production of semiconductor elements with improved properties and with greater reproductability. Because of the refinement of semiconductor technology, a number of semiconductor diodes with new and interesting characteristics began to appear around 1958.

The first to appear were the tunnel diode and the varactor, both of which could be used in low-noise amplifiers. It appears that the very rapid development of these amplifiers was inspired by the discovery of the maser in 1956. The maser's function was /7 based upon simple quantum mechanical processes and yielded a still unsurpassed low-noise amplification. The major disadvantage was that the semiconductive maser needed to be cooled in liquid helium, and because of the desire to find a mechanism which could give low-noise amplification without cooling, the parametric amplifier, which was based on the non-linear reactance in a ferrite, was developed. Just two years later, the first feasible tunnel diode amplifiers and varactor diode parametric amplifiers were constructed.

The varactor diode made possible the construction of effective frequency multipliers, breakers, reversing switches, limiters, and phase shifters. At the beginning of the 1960's, these circuit functions were greatly improved by two new types of diodes: the pin diode as a high-frequency breaker for high current, and the SR diode (step recovery) for frequency multiplications with a high multiplication factor.

Because of advances in epitaxial and planar processing, two new mixing diodes, both with low flicker effect, were introduced. They were the backward diode, which was developed from the tunnel diode, and the Schottky-barrier diode (Sb-diode) which is a renoded

metal semiconductor overlay.

Around 1963, a new phase of microwave generation began. Theoretical work during the period 1958-1962 had shown the possibilities of direct conversion of direct current to high-frequency power in pn diodes, in certain homogeneous semiconductors and in connection with tunneling through an isolated layer between two superconductive electrodes. Oscillations in homogeneous GaAs led to the development of the Gunn diode, which was named after its inventor. The Gunn diode soon received competition from the avalanche or ATT diode oscillators. (ATT is an abbreviation of "avalanche transit time.") In 1966, a variant of the Gunn diode was invented, called the LSA diode (from "limited space-charge accumulation"), which was characterized by the ability to produce very high peak power. Still another type of avalanche diode oscillator was introduced in 1967. Named the TRAPATT diode oscillator (from "trapped plasma avalanche triggered transit"), it was characterized by very high conversion efficiency and reliability at lower frequencies than the ATT diode. The superconductive oscillator, which is based on the Josephson effect, produces a very low power output and, therefore, has yet to become technically significant. The Josephson effect should, however, be exceptionally valuable in combination with basic electrical measurement techniques and continued research within the millimeter and sub-millimeter range. /8

New construction methods and improved photolithographical techniques led in 1964 to the production of transistors operating within the microwave range and as a natural result, to the development of attendant circuitry for amplifier and oscillator junctions produced in miniaturized form on the semiconductor substratum from which the transistor is produced. This began the development of integrated circuit technology in the microwave range and that development continued very rapidly. These same techniques proved to be useful for connections from electrical to acoustical microwave energy in the form of surface waves, and for conduction of these waves in a crystal. This led to new interest in the already-known interaction between sound elements and electrons in homogeneous crystals. The surface waves created new possibilities for the so-called phononamplifier and the adherent methods for signal manipulation.

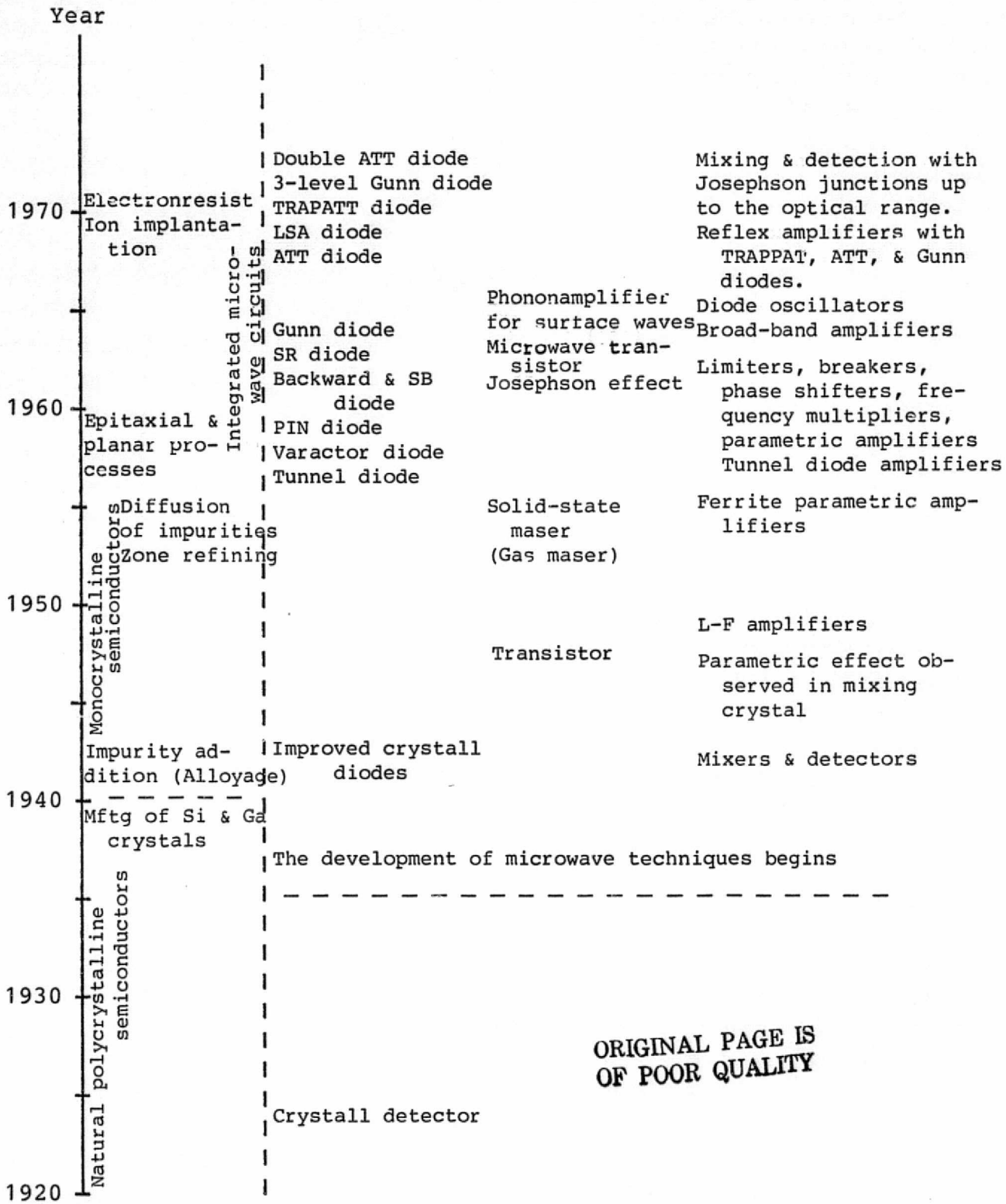
During the 1960's, the ever-increasing need for communications bandwidths has led to a broad research effort, partly within the optical region in connection with the development of the laser, and partly within the millimeter wave length region where all the earlier mentioned negative resistance diodes except for the oscillators have an even greater application in circulation-connected amplifiers. Highly developed techniques are necessary for the manufacture of semiconductor devices in the millimeter wave length range and toward the end of the 1960's, several new methods have been developed, of which ion implantation and electron resistance

are the most important. Ion implantation involves the producing of extremely well-controlled pn-overlays with regard to both the concentration of impurities and their location. The use of electron resist and electron beam exposure processing in connection with the production of various semiconductor and electrode patterns yields /9
much higher results than the normal photoresistance process.

This short history is presented schematically in Figure 1.

The most recent developments within semiconductor technology, as in many other areas, is characterized in general by the even greater utilization of very advanced numerical control methods. Thus, the semiconductor device, upon which modern data processing technology is based, has to a high degree begun to work toward its own development.

FIGURE 1: The development of active semiconductor devices for the microwave range. /10



ORIGINAL PAGE IS
OF POOR QUALITY

Generation and Power Amplification

The High Frequency Transistor

The first transistor, a point contact type, was constructed /11
in 1948 at Bell through the work of Bardeen, Brattain, and Shockley. The enormous significance that this event had for the development of semiconductors was soon demonstrated. In the meantime sixteen years passed before the transistor became usable within the microwave range. Several "milestones" in the development history of the transistor can be worth mentioning: in 1951 the junction transistor was produced; in 1955 the development of the interdigital transistor began; in 1957 the maser transistor was produced; in 1959 planar techniques (with oxide masking) began and in 1960 the production of active junctions through the epitaxial process was started; during 1963 and 1964 the "overlay" transistor [1], which made operations possible within the microwave range, was developed; in 1965 the first transistor capsule with channeled junctions, which produced lower parasitic reactions; in 1967 masking and photoresist processes were refined.

For a low level class A amplifier with a junction transistor (see for example [2,3]):

The power amplification $G \approx \left(\frac{f_{max}}{f}\right)^2$.

where $f_{max} = \sqrt{\frac{\alpha_o f_T}{8\pi r_b C_c}} \approx \sqrt{\frac{f_T}{8\pi r_b C_c}}$ = maximum oscillation frequency

r_b = the base's dispersion resistance.
C_c = pn capacitance.
α_o = power gain at low frequencies.
$f_T = 1/2\pi t_{ec}$ = the frequency at which the power gain = 1 in a continuous p coupling.
t_{ec} = the signal delay time in the emitter-collector.
(f_α is the frequency at which the power gain $\alpha = \frac{\alpha_o}{1 + j\frac{f}{f_\alpha}} = \frac{1}{\sqrt{2}}$

For a silicon transistor f is 1.2 f_T, and for a germanium transistor with bias it is 2 f_T.)

Using the foregoing, we find: /12

$$f_{max} \approx \frac{1}{4}\sqrt{\frac{1}{r_b C_c t_{ec}}}.$$

This formula is also a good approximation for class C power amplifiers.

ORIGINAL PAGE IS
OF POOR QUALITY

In constructing a high-frequency transistor it is necessary to maximize f_{max}. Therefore, $r_b C_c$ should be as small as possible. C_c is minimized by allowing the depletion region to stretch over all of the collector junctions (approximately 1 μ thick), while r_b is minimized by the appropriate formation of the emitter-base electrodes. These commonly take the form of two groups of parallel, adjacent "fingers." It should be observed that the lengthening of the base (and emitter) electrodes increases C_c while at the same time r_b decreases, so that f_{max} remains constant. The lengthening meanwhile brings about improved heat dissipation, increased current and therefore higher output.

The delay time, t_{ec} should also be minimized. This is determined mainly by the electrification time of the emitter capacitance C_e and the transit time through the base and the completion region in the collector. The former is proportionate to C_e but inversely proportionate to the emitter current. Therefore the highest possible emitter current is desired from the smallest emitter surface, or since the current originates mainly from the edge of the emitter, the largest possible length of the emitter's edge to its surface. The transit time through the base is minimized by reducing the thickness of the base as much as possible without increasing r_b, normally down to 0.3 μ. The transit time through the collector's depletion region is not minimized since it is advantageous to minimize C_c by increasing the width of the depletion region, as was previously mentioned. The main improvements during the last years have been brought about by minimizing t_{ec}.

A power transistor must be able to withstand high potentials between the emitter and the collector without having avalanche breakdown occur. (This must also be true of the base and the collector, depending upon which junction is used; junctions with /13
common bases are becoming more common and have several advantages.) High resistance in the collector-epitaxial junction yields a high breakdown potential, but creates disadvantages in the form of increased collector losses and increases the base junction thickness as current increases. This reduces f_T.

Excepting the previously discussed f_T, load resistance r_{Lp} influences the parallel circuit at the point of output and the corresponding input resistance r_es influences the acceptor circuit. This occurs according to the following formula [2]:

$$G = \frac{(f_T/f)^2 R_{Lp}}{4R_{es}} .$$

r_{Lp} is fixed by the desired power output value, ((potential collector amplitude)^2/2r_{Lp}). r_T/f, equal to the signal amplification, should be high. R_{es} becomes small when the transistor's surface is large, i.e., the output power is high. But the emitter inductance L_e increases the input resistance by

a value of $2\pi f_T L_e$. This counteracts the decreasing of R_{es}, and therefore L_e decreases amplification. Naturally, R_{es} can not be made too small, since practical difficulties emerge in transforming it to impedance in the signal source.

Efficiency is also significant in a power transistor. It is proportionate to R_{UTp}/R_{Lp}, where R_{UTp} is the parallel resistance in the transistor's output admittance.

In the case of small signals, $R_{UTp} \approx \frac{1}{\omega_t C_c}$, but with increasing power, C_c increases to a maximum value of $2C_{cb}$, where C_{cb} is the capacitance with the direct potential (overlay potential) V_{cbo}. Therefore, C_{cb} must be minimized for maximum efficiency. C_{cb} is composed of two partial capacitances, that between the collector and the base surface directly above the emitter electrodes (C_{ci}), and that between the collector and the base surface between the emitter electrodes (C_{cy}). In general, and particularly with a junction transistor, C_{cy} is much larger than C_{ci}. Decreasing the base surface between theemitter electrodes would therefore be particularly effective as a means of decreasing C_c.

It has been found empirically that efficiency can be written /14
as $\eta \approx \log f_{max}/f$ [4]. At low frequencies, this has a constant value of approximately 80% (in class C) and within the lower microwave range this approaches 60%.

The most important points in constructing a high-frequency power transistor can be summarized as follows:

1. The formation of the np electrodes is of fundamental importance for obtaining high power at high frequencies. To produce the necessary current, the emitter's edge should be as long as possible in relation to its surface. High current, in its turn, produces high power output and a maximization of the amplification-bandwidth product. The base electrode is designed so that the base current is diffused as evenly as possible. In other words, the diffusion resistance in the base is minimized.

2. The base layer is made as thin as possible without increasing dispersion resistance. In this way, the amplification-bandwidth product is maximized.

3. The collector-base capacitance is minimized as much as possible in order to increase the available amplification and efficiency. (By decreasing C_c, f_T and G are decreased, while V_B and the highest level of power are increased.

4. The nn and np breakdown potential must be large for a high level of high-frequency power. This normally requires a compromise between increased resistance capacitance and decreased efficiency and maximum frequency.

5. The active semiconductor chip is to be mounted so that heat dissipates as effectively as possible and lead-in resistance and inductance are as small as possible. In particular, the emitter inductance must be as small as possible in a power transistor so that amplification is not decreased.

There are two transistor constructions which more or less meet the above requirements. They are the interdigital transistor [5] and the overlay transistor [1]. /15

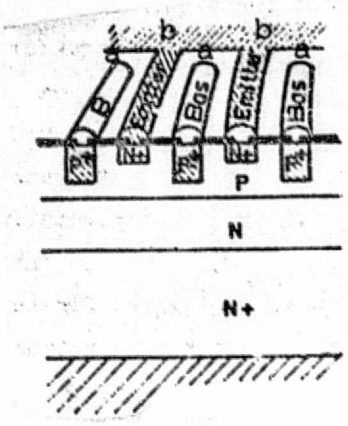


Figure 1. Interdigital structure
a = base
b = emitter

In the first, the emitter takes the form of several strips which are separated by interspersed space contacts (see figure 1). In the overlay transistor, many small emitter regions are diffused into and surrounded by the doped base contact region (see figure 2). The emitters are connected in parallel and with the base region. In this way, the electron configuration becomes very similar to that in the interdigital transistor. The number of small emitter regions is often in the around 100. High-frequency power transistors are always made of silicon by the planar method with npn epitaxial junctions.

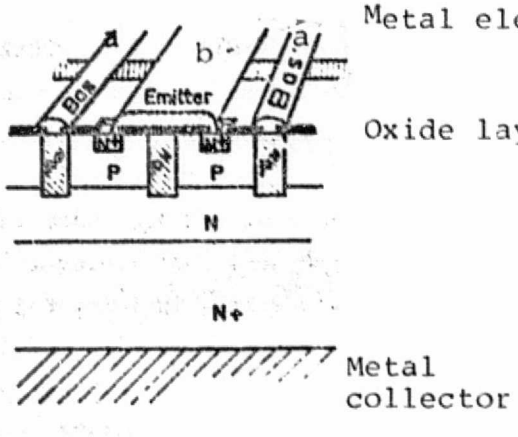


Figure 2. Overlay structure
a = base, b = emitter

ORIGINAL PAGE IS
OF POOR QUALITY

The Development of the Power Transistor

The development of the power transistor can be shown graphically in the form of power frequency curves with the year of invention as the parameter. With the help of [6,7,8 and 9], among others, the improvements in transistor performance which have been reached between the years 1958 and 1969 are shown in this manner in figure 3. (The curves are approximate mid-value curves, representing the best obtained values.) To evaluate the future development of the (bipolar) transistor the effect of new technological methods must be appreciated. The expected transition from conventional photo resist techniques, where the transistor structure is determined through optical techniques, to the much more exact electron beam method, used in combination with electron-sensitive resistance layers, will be extremely significant. Furthermore, the production of very thin base layers by means of ion implantation will become possible. Today's current minimum width of approximately 1μ should through the electron beam method be reduced to approximately 0.1 μ. Further reductions will conflict when the transistor function is dependent on various doped regions, since the distance of the impurity atoms is the order of 0.01 μ. Ion implantation can cause a reduction in the current minimum base layer of approximately 0.3μ by a factor of 0.2, that is to say to approximately 0.06 μ. As a basis to the first major forecast for 1985 it can be said now that the refinement of the transistor structure will yield five times higher current with unchanged capacitance. The power output at relatively low frequencies, for example 1 GHz, should then increase five times from 20 to around 100W. The increased emitter current will also reduce t_{ec}, but since the size of t_{ec} depends mainly on three factors which have previously been discussed, the decrease in t_{ec} will be less than five, probably three times. By the same token, it is seen then that the five-time reduction of the base thickness will reduce t_{ec} by a factor of three. This means that a total increase in f_T by a factor of ten or f_{max} by a factor of three should be possible. Therefore, since f_{max} in 1969 is near 10 GHz, f_{max} should in 1985 be around 30 GHz. Using these values, a forecast curve for 1985 has been drawn in figure 3 in such a way that it has about the same shape as the curve for 1969. It can be interesting to compare this forecast with several other published forecasts [10], which apply to the year 1971, and according to [11], which shows a theoretically calculated maximum limit (see figure 3). /16

In terms of f_{max}, according to those mentioned above the highest obtained values for a bipolar transistor in 1969-1970 is approximately 10 GHz. This can be compared with a field-effect transistor of GaAs, produced using refined optical projection techniques with a width tolerance of 1 μ. This transistor yielded f_{max} 30 GHz, which corresponds to 3 dB of power amplification at 17 GHz [12]. Such a transistor is usually reserved for small signal amplification or very rapid switching functions, while the

ORIGINAL PAGE IS
OF POOR QUALITY

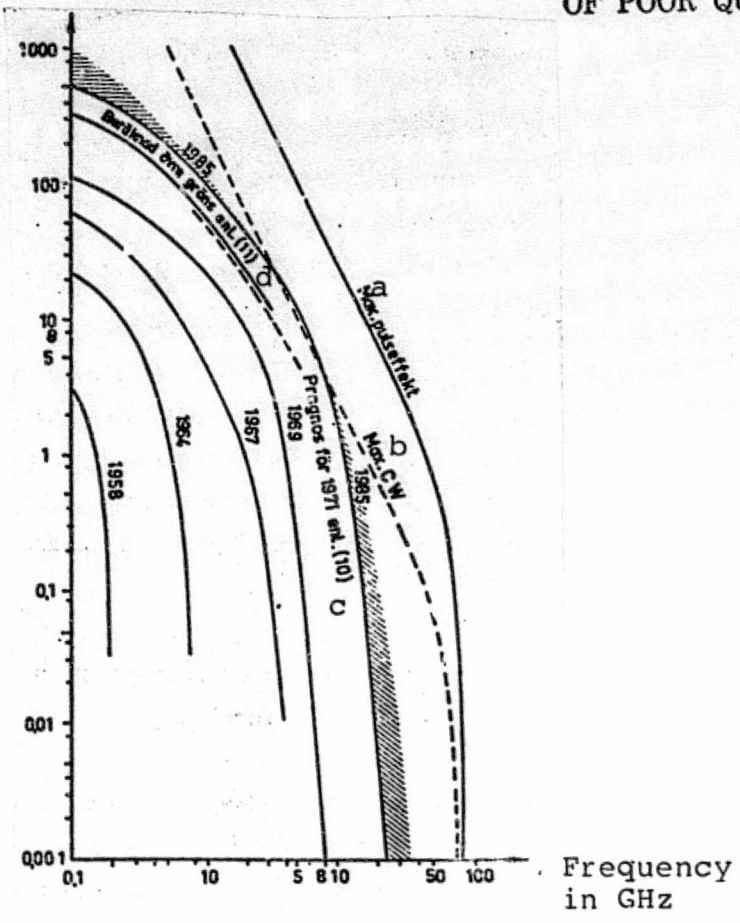


Figure 3.

a = Maximum pulse power

b = Maximum CW power

c = Forecast for 1971 according to [10]

d = Calculated upper limit according to [11]

bipolar transistor will certainly be of use in the generation of higher power. Further information on the field-effect transistor is given in the section on low-noise transistors. /18

Pulse Operation in the Transistor

Measurements and calculations in [4] show that pulse operations in high-frequency power transistors for the present produce peak outputs which lie 2 to 5 times higher than the CW level and that in the future special pulse transistors will be able to produce values 10 times higher.

In [13], Johnson has calculated that the transistor's maximum frequency and power is limited by purely physical principles. If one ignores the heating of the semiconductor and the formation of the transistor structure, an absolute limit can be defined on the basis of how fast energy can be brought to the charge carriers in the semiconductor.

The masimum value for the transistor's cutoff frequency $f_T = 1/2\pi\tau$ will be limited chiefly by the transit time for the charges between the emitter and the collector. This distance can not be made smaller than V_{max}/E_B, where V_{max} is the breakdown potential and E_B is that field strength which produces breakdown. If the saturation speed for the charges is written v_s, we have $\tau_{min} = V_{max}/E_B v_s$ and the maximum value for the cutoff frequency

$$f_T = \frac{E_B v_s}{2\pi V_{max}} .$$

If the material constants E_B and v_s for Si and Ge are inserted, we have $V_{max} f_T = 2 \cdot 10^{11}$ V/s and 10^{11} V/s, respectively.

For a transistor to function, V_{max} must produce in the charges an energy which is greater than that produced thermally. For silicon, the largest f_T with $V_{max} = 1$ V becomes:

$$f_{Tmax} \approx 200 \text{ GHz}.$$

A comparison of the best available high-frequency transistors reveals that the product of their V_{max} and f_T lies about 5 times lower than the theoretical maximum value. This is caused by irregularity in doping and the fact that the electrification time of the capacitance cannot be ignored. With regard to technological limits, it would then be reasonable to set $f_{Tmax} \approx 40$ GHz. /19

Knowing that the current through the transistor can be shown as

$$I_{max} = \frac{Q_{max}}{\tau_{min}} = \frac{C_c V_{max}}{\tau_{min}} ,$$

the following relationship, as in [13], can be established:

$$I_{max} V_{max} X_o f_T^2 = \left(\frac{v_s E_B}{2\pi}\right)^2 ,$$

where I_{max} is equal to the maximum current through the transistor (which does not produce undesirable diffusion in the base junction), and the transistor's output impedance, $X_o = 1/2\pi f_T C_c$.

At low frequencies, the power output can be shown as

$$P_{max_o} \leq \left(\frac{I_{max}}{2\sqrt{2}}\right)\left(\frac{V_{max}}{2\sqrt{2}}\right)$$

$$\therefore P_{max_o} \leq 5/f_T^2 X_o \text{ kW with } f_T \text{ in GHz.}$$

ORIGINAL PAGE IS OF POOR QUALITY

According to [4], the efficiency for a class C transistor amplifier can be written as

$$\eta = 0.7 \log \frac{f_{max}}{f} \text{ when } f > 0.1\ f_{max}.$$

When $f < 0.1\ f_{max}$ the constant $\eta = 80\%$. The output power when $f > 0.1\ f_{max}$ can then be written

$$P_{max} = \frac{5}{f_T^2 X_o}\ 0.7 \log \frac{f_{max}}{f} = \frac{3.5}{f_T^2 X_o} \log \frac{f_{max}}{f}.$$

Given that $f_{max} = 2f_T$, which is normally the case with today's high-frequency transistors [4], and that an X_o value as small as 0.5 Ω can be used, we have /20

$$P_{max_o} = \frac{8}{f_T^2} \text{ kW for } f < 0.2\ f_T$$

and $P_{max} = \frac{7}{f_T^2} \log \frac{2f_T}{f}$ kW for $f > 0.2\ f_T$, with f_T in both cases in GHz.

Calculating P_{max} as a function of f with f_T varying from the aforementioned "absolute" maximum value of 40 GHz down to 1 GHz, produces a family of curves which are approximated by the outermost curve in figure 3, marked "Maximum peak power." Therefore, this curve represents the highest possible values which can be reached in pulse operation with f_T optimized at every signal frequency. One further evaluation of the maximum CW power which a transistor can produce is now possible when it is taken into account, as previously discussed, that CW power comes close to being one-tenth of peak power. The corresponding curve in figure 3 is marked "maximum CW power."

The forecast curve for 1985 which was produced with possible technical improvements in mind, lies somewhat under the "maximum CW" curve. The deviation is largest at low and high frequencies. At low f, it is hardly likely that transistors will be built with $P_{UT} > 1$ kW, even if this were possible. It is more likely that the power output from several transistors in the o.1 - 1.0 kW level will be combined when special high power is needed (lower impedance). The deviation at high frequencies occurs because the "maximum CW power" curve is based upon more than twice as high a high-limit frequency as the "1985" curve. The significant

technological problems with increasing the cutoff frequency entails that the "1985" curve must be seen as more likely within the upper frequency ranges. As a compromise between the two curves, the lined area above the "1985" curve in figure 3 is presented as an area of possible results of the year 1985.

A New Type of Transistor /21

The so-called metal-based transistor could become very significant for expanding the transistor's frequency range upward. As the name implies, the base electrode is formed by a metal plate which must be very thin, around 100 A in order to pass through emitter electrons. Furthermore, it must be of absolutely constant thickness. This construction produces in principle very low base resistance (approximately 80 times lower than in a semiconductor) and short transit time through the base. However, the technological problems with its development have so far been difficult to overcome. The outlook for this type of transistor is therefore very uncertain, but it is still worthy of mention here.

REFERENCES

1. Carley, et al, "The Overlay Transistor, Part I," Electronics, (August 1965). /22

2. Lee, H.C., "Microwave Power Transistors," Microwave Journal, p. 51, (February 1969).

3. Cuccia, C.L., "Low Noise Microwave Transistors," Microwave Journal, p. 77, (February 1968).

4. Meyer, Auth, "Overlay Transistors as CW and Pulsed Oscillators," Microwave Journal, p. 59, (August 1969).

5. Tatum, J.G., "Fingers in the Die," Electronics, p. 98, (February 1968).

6. Webster, R., "Microwave Transistors," Microwave Journal, p. 99, (July 1966).

7. Carley, D., "A Worthy Challenger for R-F Power Honors," Electronics, p. 98, (February 1968).

8. Gundlach, R., "R_x for R-F Power Transistors," Electronics, p. 84, (May 1969).

9. Small, T., "UHF and Microwave Devices, Present and Future," Texas Instruments Report, (1965).

10. Cooke, H., "Microwave Transistors, Part I: Power Devices," Microwaves, p. 46, (December 1969).

11. Magill, L.M., "Watt-Megahertz Ratings Run Second to High Reliability in Foreign R-F Power Transistors," Electronics, p. 88, (April 1970).

12. Drangeid, et al, "High-Speed GaAs SB Field-effect Transistors," Electronics Letters, p. 228, (April 1970).

13. Johnson, E.O., "Physical Limitations and Power Parameters of Transistors," RCA Revue, p. 163, (June 1965).

The Frequency Multiplier

The nonlinear resistance in point-contact diodes was used /23
at an early stage for the generation of harmonics. However, the very high diode losses entailed low efficiency ($\leq 1/n^2$, where n stands for the multiplication factor).

Harmonics can also be generated by the nonlinear reactance which the voltage-produced variable capacitance in a varactor yields. Since the varactor is produced so that high-frequency losses are small (high cutoff frequency), efficiency in the generation of harmonics can be very high.

Since the aim here is to evaluate maximum obtainable power levels, only frequency doublers will be discussed. These produce highest efficiency and power output. When a higher multiplication factor is needed, a cascade rectifier circuit with an appropriate number of doublers will yield the best results. Naturally, very high n values can be used in connection with low output power and the demand that the size of the multiplier be as small as possible. n values of 3 and 4 are common with varactors, while values between 10 and 20 are common when using so-called SR (step recovery) diodes. In the case of the latter, for instance, an output frequency in the X-band is produced with only several hundred mW output power with a maximum efficiency of 5%.

An attempt to evaluate how thefunction of a frequency doubler is influenced by the varactor parameters will be made here. This is based of [1, figure 8.7]. See figure 4.

If a common value for P_{diss}, V_B, R_S, and f_c can be found, the broken lines in figure 4 give the maximum value for f_{IN} (allowable dissipation) and the corresponding output power P_{UT}.

Instead of using the curve, the maximum f_{IN} for any P_{diss} can be calculated from:

$$f_{IN\ max} \approx \frac{f_c}{V_B}\sqrt{2R_S P_{diss}} \quad \left(= \frac{1}{V_B}\sqrt{\frac{f_c P_{diss}}{\pi C_j}}\right). \qquad /24$$

Efficiency, which is a function of f_{IN}/f_c, is determined most simply then from figure 5. (According to [1, figure 8.7] or [2, figure 12].) Therefore, the power output is calculated as:

$$P_{UT} = \frac{\eta}{1-\eta} P_{diss}.$$

The desired common value for P_{diss}, V_B, R_S, and f_c can be determined empirically by studying the varactor data collected in figure 6.

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

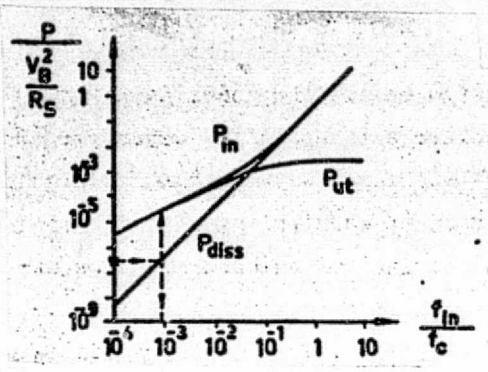


Figure 4.

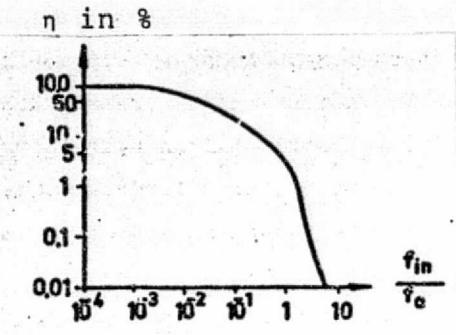


Figure 5.

The desired relationship between P_{diss}, V_B, R_S and f_c can be determined empirically from the data in figure 6.

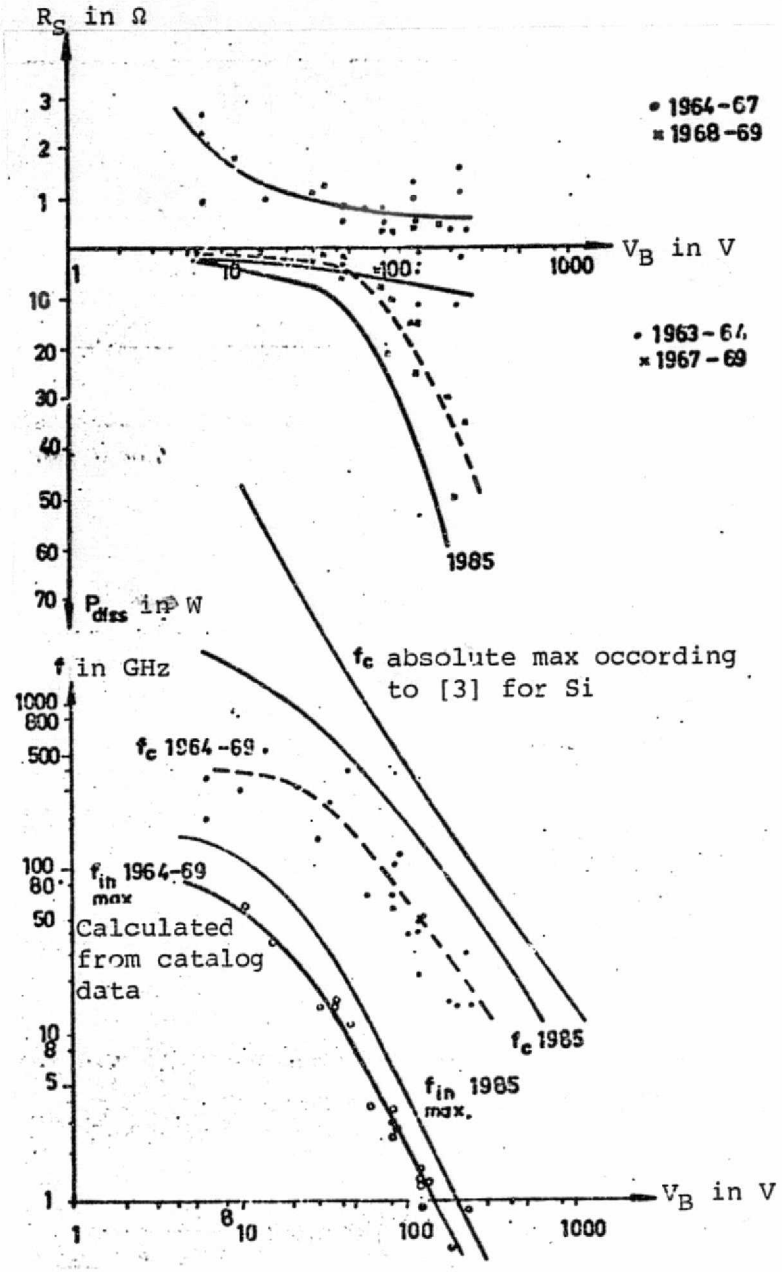
Because of improved heat dissipation, it can be assumed that by 1985 P_{diss} will have increased by a factor of 2 for all V_B values (see "1985" curve in figure 6a). Also, it is assumed that f_c will more than double for all V_B values (see "f_c 1985" curve in figure 6b), mainly by reducing the series resistance in half. (C_j remains approximately unchanged so that the input resistance and the load resistance, both proportional to $1/\omega_{IN}C_j$, also are unchanged. According to the shape of f_{INmax}, the result will be a doubling in 1985 ("f_{INmax}" curve in figure 6b). With the aid of the "1985" curve in figure 6, P_{UT} can now be calculated as a function of f_{UT} for the year 1985. See table 1.

Table 1. Calculated maximum power output for frequency doublers. /26

f_{IN} GHz	f_{UT} GHz	f_c GHz	η	P_{diss} W	$P_{UT} = \frac{\eta}{1-\eta} P_{diss}$ W	Necessary P_{IN} W
1	2	75	0,75	60	180	240
5	10	200	0,63	30	51	81
10	220	350	0,60	19	28	47
50	100	900	0,40	7	4,6	11,6
100	200	1500	0,30	4	1,7	5,7
200	400	1500	0,10	1	0,1	1

P_{diss} (or P_{UT}) is calculated according to [1] based on the following assumptions:

$$\left(\frac{\frac{1}{C_{min}} - \frac{1}{C_o}}{\frac{1}{C_{min}} + \frac{1}{C_o}}\right)^2 = 1$$



Figures 6A and 6B.

ORIGINAL PAGE IS
OF POOR QUALITY

and that

$$f_c = \frac{\frac{1}{C_{min}} - \frac{1}{C}}{2\pi R_S} \approx \frac{1}{2\pi C_{min} R_S}$$

which can be restated as $\frac{1}{C_o} << \frac{1}{C_{min}}$. But $C(V) = \frac{C_o}{\sqrt{1 - V/0.5}}$ for an abrupt junction.

With $V_B = 6V$, we obtain $C(-6) = 0.28\ C_o$.

According to equation 8.42 in [1], P_{UT} at $V_B = 6V$, which is to say that $f_{UT} = 100$ GHz, is multiplied by a factor of

$$k = \left(\frac{\frac{C_o}{C_{min}} - 1}{\frac{C_o}{C_{min}} + 1}\right) \cdot \left(1 - \frac{C_{min}}{C_o}\right) \approx 0{,}5 \text{ for } \frac{C_{min}}{C_o} = 0{,}28.$$

At $f_{UT} = 2$ GHz, this factor $k \approx 0.85$.

Furthermore, variations in circuit loss with f can be introduced according to the following: /27

Figure of merit $Q \sim \frac{\lambda}{\delta} \sim \sqrt{\lambda} \sim \sqrt{1/f}$ (δ = depth of penetration) and

$$Q_o = 2\pi = \frac{\text{stored energy}}{\text{losed energy per period}}$$

This yields the result that the power loss is proportionate to $\sqrt{f}$.

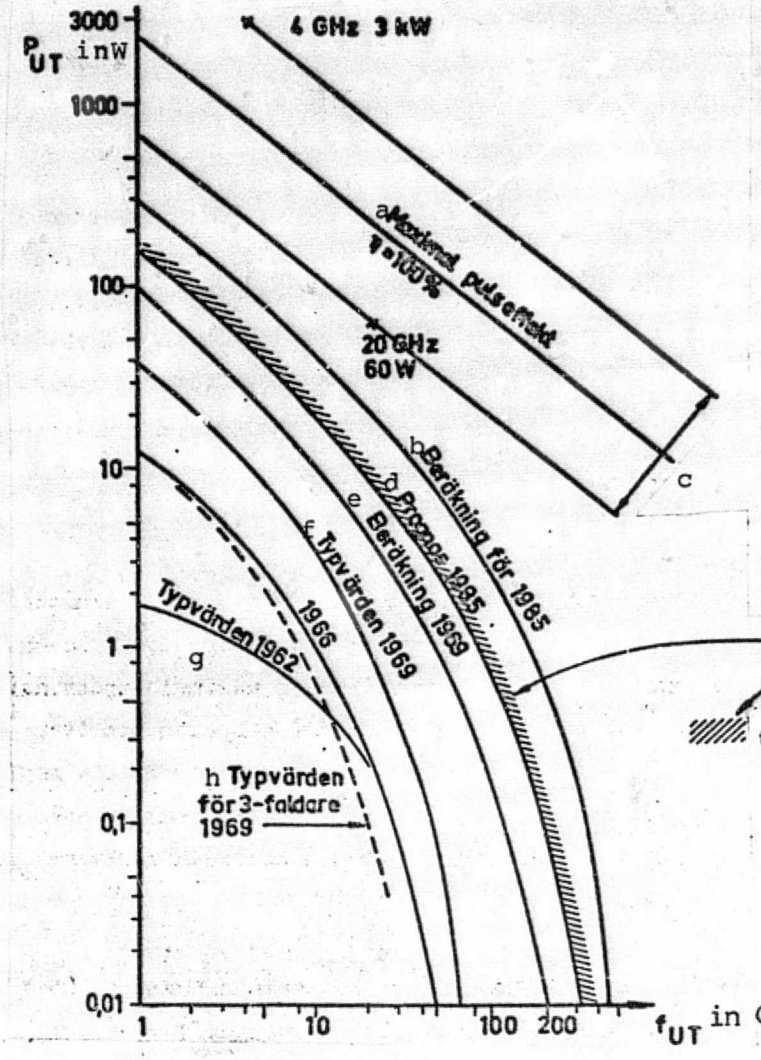
With $f_{UT} \approx 1$ GHz, normally $\eta_C \approx 97\%$ (see [4])	The losses =	3%.
With $f_{UT} = 10$ GHz,	the losses =	9%.
" 20 "	"	13%.
" 100 "	"	30%.
" 200 "	"	42%.
" 400 "	"	60%.

		Correction k
With $f_{UT} = 1$ GHz, the correction factor $\eta_C =$	0.97	k = 0.9
" 10 "	0.91	0.8
" 20 "	0.87	0.8
" 100 "	0.7	0.5
" 200 "	0.6	0.5
" 400 "	0.4	0.4

f_{UT} GHz	Approximate total correction factor $\eta_c k$	Uncorrected P_{UT} W	Corrected P_{UT} W
2	0.9	180	160
10	0.7	51	36
20	0.7	28	20
100	0.35	5	2
200	0.3	2	0.6
400	0.2	0.1	0.02

These corrected power values have been incorporated in figure 7 and are labeled "Calculations for 1985." As was shown above, this calculation is based on evaluated varactor data for the year 1985. /28

It can be interesting to test the same method of calculation on the basis of known varactor data for the year 1969. (Since both f_{IN} and f_c according to figure 6B were doubled from 1969 to 1985, the efficiency remains constant according to figure 5 despite the doubled frequency. At the same time, according to figure 6A, P_{diss} will double, which in turn will double P_{UT}. The desired curve for 1969 can be calculated easily by taking half of both P_{UT} and f_{UT} in the calculations for 1985.) The resulting curve is shown in figure 7, marked "Calculation for 1969." The typical powers reached in the years 1962, 1967 and 1969 ("state of art") are also shown in figure 7. It is of particular interest now to determine that the calculations for 1969 yield significantly higher output power than what is in reality the case. For this reason the calculated powers for 1985 should also be too large. A reduction of the values calculated for the 1985 curve in proportion to the difference between the calculated and the actual power values for 1969 should therefore yield a likely forecast curve, labeled "Forecast 1985" in figure 7. It has been shown in [2] that the maximum power output in a varactor with an abrupt junction can be increased by a factor of 4 in relation to the maximum value according to [1], if the varactor is overdriven in the forward bias region. When taking into account the highest allowable power losses in the varactor, the increase in power which can be utilized becomes insignificant. The efficiency without overdriving lies between 10 and 70% (as in table 1), and overdriving will yield 13 to 73%, in other words an increase in power output of 34 to 16%. In a graduated junction with an n value somewhat under 0.1 (the n value is the exponent which determines the current-dependent junction capacitance), efficiency can be further increased approximately 5% in comparison to an abrupt junction. This corresponds to a 20% increase in power output. Therefore, overdriving can produce a 50% increase in the power values of the forecast curve (shown by the lined area in figure 7). If thermal problems are ignored, the maximum peak power according to [1] can be calculated from: /30



a) Maximum pulse power η=100%

b) Calculation for 1985
c) Variation with the impedance level
d) Forecast 1985
e) Calculation for 1969
f) Typical value1969

Area of possible power increases through overdriving.

g) Typical value 1962
h) Typical value 3-times 1969

Figure 7.

$$P_{UT} = 0.002 \frac{V_B^2}{R_S} \quad (\text{for } f_{IN}/f_c \geq 0.1).$$

Judging from the previously shown varactor data curves and the assumptions for 1985, varactors with $V_B \approx 300$ V and $R_S \approx 0.2$ ohm can be expected. These yield $P_{UT\ max} \approx 1$ kW for frequencies greater than 5 GHz. By overdriving the varactor, the absolute limit becomes approximately 4 kW. The efficiency is given by $\eta = 0.39 (f_c/f_{IN})^2$%. (For example, in the X-band η=5% with f_c= 35 GHz for V_B = 300 V.) Difficult transient phenomena resulting from thermal and other causes can in the meantime be expected to arise with pulse operation of the frequency multiplier.

The maximum power level in a varactor, without regard to

thermal problems, can also be evaluated according to a method given by De Loach [5]. The calculation is based in terms of the maximum voltage V_m which can be applied to the varactor's depletion region L without causing breakdown, and on the shortest time which is necessary to eliminate the charges at saturation speed v_s and to create the depletion region. When $\tau = L/v_s$, one half the period of the maximum frequency f_τ, $f_\tau = v_s/2L$.

The calculation yields the following expression for the maximum power which (at a 100% efficiency level) can be converted to the frequency f_τ:

$$P_m = \frac{1}{X_\tau f_\tau^2} \frac{E_C^2 v_s^2}{4\pi} (m+1)^{\frac{1}{m}}$$

where $$E_C = (2f_\tau/v_s\alpha_o)^{\frac{1}{m}} = E_B(m+1)^{-\frac{1}{m}}$$

and
E_B = the field strength which produces breakdown
α_o = an ionization constant
m = constant

$$X_\tau = \frac{1}{2\pi f_\tau C_{min}}$$

For silicon, $E_B = 3 \cdot 10^5$ V/cm, $m = 6$ and $v_s = 10^7$ cm/s. /31

At a constant impedance level, P_m is proportional to $f_\tau^{-1.67}$ and if one assumes that the impedance level must increase in proportion to $f_\tau^{-\frac{1}{2}}$, the variation of P_m can be described as $f_\tau^{-2.17}$. (This is approximated as $P_m \sim f_\tau^{-2}$.) The above values yield:

$$P_m = \frac{1}{X_\tau} \left(\frac{800}{f_\tau\,[GHz]}\right) \text{ [W]}.$$

According to the catalog data for varactors, $C_j = 3$ pF is a typical value with an output frequency of 4 GHz and 0.3 pF is typical at 20 GHz. If one assumes that C_j and with it the impedance level in the future must remain unchanged, two values for P_m can be calculated from these values: 3 kW at 4 GHz and 60 W at 20 GHz. These points are shown in figure 7 and through them is drawn the $P_{UT} \sim f^{-2}$ line. The relatively high level of impedance at 20 GHz have evidently reduced P_m unnecessarily. Therefore, a line between these two, labeled "maximum pulse power," has been chosen as an "absolute" value for P_{UT}.

The Use of Multiple Varactors

The maxiumum values derived from single varactors, which have previously been discussed, can naturally be increased significantly by the summation of the output powers from several varactors in separate circuits. However, the power level can also be increased buy using a multiple varactor. In this case, several semiconductor chips are connected either serially or in parallel, or both in series and in parallel, in order to obtain the appropriate level of impedance. The power levels obtained through such a combination technique is very hard to predict. It should, however, be possible

with a multiple varactor consisting of 3 x 3 or 4 x 4 diodes in a series-parallel combination. According to the experiments performed in [6], the power output increases approximately quadratically with the number of serially connected varactors, which means that the hypothetical multiple varactor should be able to produce at least 100 times more power output that a single varactor, assuming that heat dissipation problems do not set a lower limit. Preliminary attempts to develop a special capsule for a multiple varactor have been reported [7]. This capsule could conduct away heat at least 20 times more effecticely that the normal power varactor case. Future multiple varactors can therefore be expected to produce 10 to 100 times higher power levels than what was presented in the previous forecast curves. /32

Bandwidth of the Frequency Multiplier

In commercially available multipliers, the bandwidth rarely exceeds 5%, and only in exceptional cases almost exceeds 10%.

Calculations by a number of authors (see the summary in [2]), have shown that a 30% bandwidth should be possible with simple tuning of the input and output circuits in a doubler, using a varactor with low parasitic reactance. With multiple input and output circuits, 66 to 80% bandwidth should be obtained, in other words, the octave level or higher. The former level has been produced experimentally [8] in a coupling consisting of a pair of varactors adapted with 5 circuits and which produced 2 to 4 GHz (66%) with an efficiency level of 50%. Bandwidths much over an octave can hardly be expected in the future, but octave bandwidth should be obtained up to the X-band with 40% efficiency. Naturally, efficiency declines with increased bandwidth, so if, for example, the source of the X-band is made narrow (10%), the efficiency should increase to around 70%, according to [2, p. 222, figure 41].

Noise Characteristics of Frequency Multipliers

According to calculations, the noise factor in the frequency multiplying function of the varactor is highly insignificant except for those unavoidable increases produced through frequency multiplication. The total frequency noise in the multiplier's output is therefore nearly equal to n times the noise in the input, where n is the multiplication factor[9]. In this manner, the contribution of FM noise is calculated as $\Delta f_{rms} < 0.1$ Hz at $f_m = 10$kHz and $\Delta f_{rms} < 1$ Hz at $f_m = 100$ kHz within 1 kHz bandwidth. Multipliers are normally used in crystal stabilized, cavity stabilized, or varactor-tuned transistor-oscillators and followed by a transistor-amplifier. Because of this, it can be appropriate to discuss the noise additions from various active components and the size of the noise with various stabilization from the ground oscillator. /33

A varactor-tuned transistor-oscillator, followed by a

multiplier, typically yields in the X-band $\Delta f_{rms} \approx$ 100 Hz for f_m = 1 - 10 kHz (within i Khz bandwidth). The corresponding value for a crystal-tuned oscillator with an amplifier before the multiplier lies around 1 Hz, but when f_m > 10 kHz, the frequency noise begins to rise and when $f_m \approx$ 200 kHz it has increased by a factor of 10 and at 1 MHz by a factor of 100. This increase in noise stems from the phase noise in the power amplifier. In a cavity-tuned ascillator, $\Delta f_{rms} \approx$ 10 Hz when f_m = 1 kHz, but as f_m increases, the noise decreases and at f_m = 100 kHz is less than 1 Hz. This shows that a combination of crystal and cavity switching can produce interesting results[10].

Future improvements of an oscillator-multiplier's noise characteristics can be expected to occur along the following lines:

1. New types of crystals and improved transistors will increase the ground oscillator's frequency significantly and consequently decrease the required multiplication factor.

2. The noise factor in a power amplifier will be reduced.

3. The use of various feedback-reducing circuits will increase.

It can be guessed that these methods will be able to reduce frequency noise by a factor on the order of 10 and consequently produce nearly as good noise data as the best klystron oscillators.

Amplitude noise, which was not mentioned above, is much lower than frequency noise and normally lies about 120 dB under the signal power.

REFERENCES

1. Penfield, Rafuse, "Varactor Applications," MIT Press, (1962). /34

2. Scanlan, J.O., "Analysis of Varactor Harmonic Generators," Advances in Microwaves, Academic Press, pp. 165 - 236, (1967).

3. Penfield, P., "Maximum cut-off frequency of Varactor Diodes," Proc. IEEE, pp. 422 - 423, (April 1965).

4. Watson, H.A., Microwave Semiconductor Devices and their Circuit Applications, McGraw-Hill, p. 243, (1969).

5. De Loach, B., "Recent advances in solid-state microwave generators," Advances in Microwaves, Academic Press, pp.43 - 88, (1967).

6. Irwin, Shaw, "A composite varactor for simultaneous high power and high efficiency harmonic generation," Trans. IEEE, ED, pp. 466 - 471, (May 1969).

7. Tang, D., "High power package for multichip application," Trans. IEEE, pp. 799 - 800, (May 1969).

8. Kotzebue, Matthaei, "The design of broad-band frequency doublers using charge-storage diodes," Trans. IEEE, MTT, p. 1077, (December 1969).

9. Frilley, Grandchamp, "Design of low-noise solid-state microwave sources," Proc. of Microwave Applications of Semiconductors, London, (July 1965).

10. Back, I., "Reducering av FM-brus hos en FT-generator pa X-bandet," KTH report, (December 1967).

The Avalanche Diode Oscillator

In 1958, D. Read [1] showed how a semiconductor diode of /35
special construction would be able to function as a microwave oscillator. The avalanche breakdown in an abrupt pn junction was shown to produce a current which lay 90° after the applied voltage, and if the charges were allowed to pass through a drift field of sufficient length, consisting of an intrinsic semiconductor in which the charges move at the saturation speed, an additional 90° current delay was produced. Such a Read diode (pnin or npip) should then yield a negative resistance atound the transit-time frequency and as a microwave oscillator produce an efficiency of 30%. However, this type of diode was difficult to produce and not until 1966 was a Read diode able to oscillate within the microwave range. One year earlier [2], X-band oscillations were generated in pulse operations with a normal pn junction, in which the breakdown and drift-space regions were combined. This was name the ATT oscillator, after "Avalanche Transit Time." After optimizing the doping profile and improving the heat insulation, both pn and pin diodes could generate CW power between 5 and 50 GHz with approximately 10% efficiency. Bell later coined the name IMPATT for this oscillator (from "IMPact Avalanch Transit Time").

A new phase in the development of the avalanche diode oscillator was begun in 1967. Several researchers at RCA [3] observed high power output and very high efficiency with the generation of a harmonic to the transit-time frequency. An ATT diode with a transit-time frequency within the X-band when coupled with a double resonance circuit produced a pulse operation of several hundred watts around 1 GHz with up to 60% efficiency. This diode, like several other ATT diodes, was made of silicon. Later, germanium diodes also produced CW oscillations in the highly effective "anormal" mode (5 W, 0.5 GHz, 43%) [4]. Several partially conflicting theoretical explanations for the function of the anormal mode have been offered in 1969 and this has resulted in nearly as many new names for the oscillator. Bell has, for example, proposed the "TRAPATT" mode (from "TRApped Plasma Avalanch Triggered Transit") /36
[5,6], and Cornell University has suggested "ARP mode" (from Avalanche Resonance Pumped") [8,9]. The common ground between the various theories is still that a normal transit-time oscillation is presumed to be necessary in order to start the low-frequency switching between high current-low voltage and high voltage-low current, which characterizes the TRAPATT mode and produces the high level of efficiency. Experiments to date show that the efficiency increases with the ratio of the transit-time frequency to the output frequency, and that a ratio of 2 produces 10% efficiency. The negative resistance is both the ATT and the TRAPATT mode has also been used in the construction of circulator-coupled microwave amplifiers with high output power.

Power-Frequency Limits in Avalanche Diodes

According to De Loach [10], the absolute power-frequency limit can be most easily evaluated in a Read-type avalanche diode. The condition that the avalanche breakdown shall not occur in the drift region of length L yields the maximum voltage $V_m = E_B L$. The maximum value of the charge which ban be injected in the drift region is determined by the field strength which on the one edge is E_B and on the other edge must not be less than that field E_s which gives the charges the saturation speed v_s. The maximum current $I_m = dQ_m/dt = C(\delta v/\delta x)(\delta x/\delta t)$ the becomes (using maximum E and minimum v) approximately $I_m = CE_B v_s$, where C is the capacitance in the drift region. If it is assumed as earlier that the maximum frequency $f_m = v_s/2L$ (which allows the whole charge to emerge during one-half of a period), then

$$P_m X_c f_m^2 = \frac{E_B^2 v_s^2}{4\pi}, \text{ where } P_m = I_m V_m, \text{ the maximum power which can}$$

be fed into the diode.

$$X_c = \frac{1}{2\pi f_m C} \quad \text{with } C = \epsilon \frac{A}{L}.$$

Considering that V_B varies somewhat with L, and that the impedance level, as earlier presented, should increase proportionally to $\sqrt{f_m}$, it can be shown that P_m is proportional to $f_m^{-2.17} \approx f_m^{-2}$. In [10], it is predicted for avalanche diodes with common breakdown and transit-time regions that V_m and also I_m (since the current will be produced by both the holes and the electrodes) will double. This yields /37

$$P_m X_c f_m^2 = \frac{E_B^2 v_s}{\pi}.$$

With $E_B \approx 300$ kV/cm and $v_s = 10^7$ cm/s (the value for silicon), the absolute limit (at 100% efficiency) for power as a function of frequency should result from the expression

$$P_m = \frac{3 \cdot 10^6}{f_m^2[\text{GHz}] X_c} \quad [\text{W}].$$

In order to approximate a practical maximum power limit with pulse operations, one can use a 10% efficiency level and a diode capacitance large enough for X_c to be near 1. In this case

$$P_{peak} = \frac{3 \cdot 10^5}{f_m^2[\text{GHz}]} \quad [\text{W}].$$

A curve constructed from this formula is drawn in figure 9 and labeled "Maximum peak power. η = 10%, [10]."

In [8,9], Hofflinger has given a method for calculating the product of the power output and the load resistance for both the normal ATT (or transit-time) mode and the ARP mode. His method

for calculating the transit-time mode is briefly recapitulated here. The product of the output power and the load resistance can be written

$$P_{UT}R_L = V_B I \eta R_L$$

$$\text{With } R_L = r\,\frac{L^2}{\varepsilon v_S A}$$

and $V_B = E_B L$,

$I = J_O A$ /38

$$J_O = \frac{\varepsilon v_S}{\lambda L} j$$

$$\theta = 2\pi f \frac{L}{v_S}$$

we obtain

$$P_{UT}R_L = \eta r j \theta^2 \times \left(\frac{v_S}{2\pi}\right)^2 \frac{E_B}{\lambda} \times \frac{1}{f^2}$$

scaling factor | material factor | frequency factor

In the above expressions, the following are:

- ε = the semiconductor's dielectric constant (10^{-12}As/Vcm for Si)
- v_S = the saturation speed of the charges (10^7 cm/s)
- L = the thickness of the active semiconductor junction
- A = the cross-sectional surface of the active semiconductor junction
- λ = an ionization parameter defined from the expression for the ionization speed $\alpha = \alpha_o e^{\lambda E}$ and therefore $= (d\alpha/dE)/\alpha$ cm/v
- E_B = the average field strength at breakdown
- J_O = the current density
- r = the normalized load impedance
- j = the normalized current density
- θ = the transit-time angle.

Both calculations and measurements show that a relatively high negative resistance combined with maximum efficiency is obtained when the transit-time angle is 2 radians. The avalanche resonance frequency w_A indicates that the avalanche area's equivalent parallel reactance is infinitely large and that the negative resistance is maximized, should also lie near the transit-time frequency. With the following expressions for w_A:

$$w_A^2 = \frac{2.7 v_S}{\varepsilon L} \lambda J_O = w^2$$

and

$$\theta = w\frac{L}{v_s} = 2$$

we obtain $J_o = 1.5\,\frac{\varepsilon v_s}{\lambda L}$. In other words, j is a constant equal to 1.5. /39

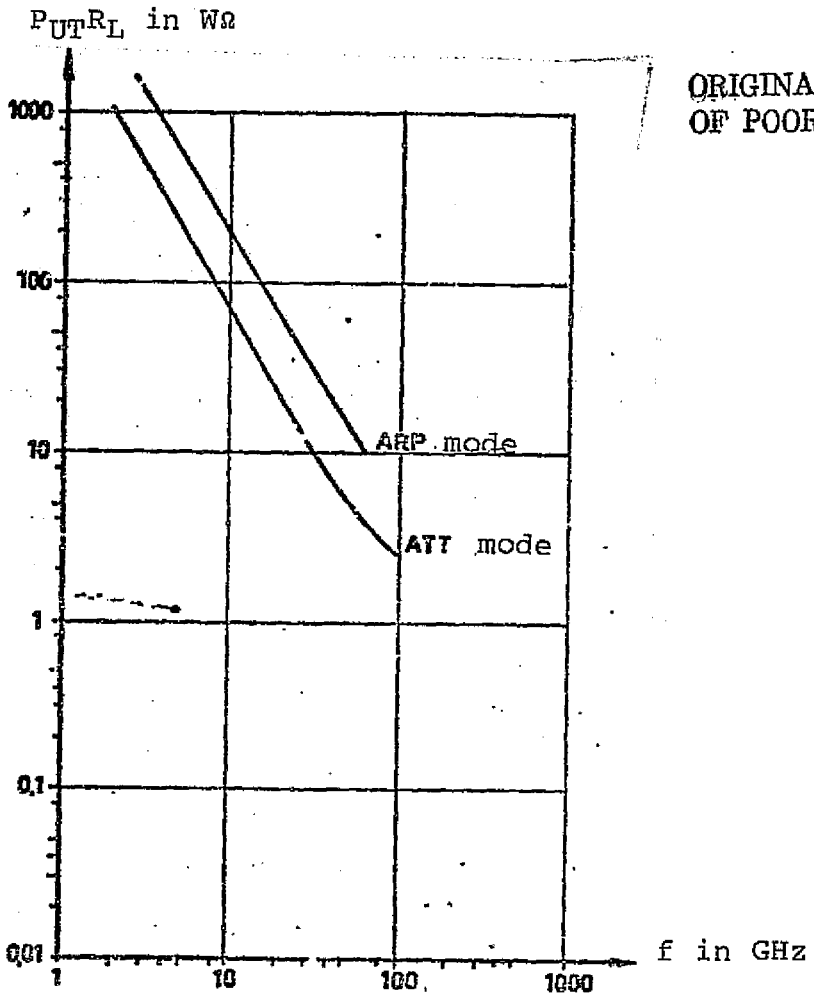
Hofflinger shows that $E_B = f(L)$ and $\lambda = f(E_B)$. For a particular transit-time angle, f will determine L, L will determine E_B, and E_B will determine λ. The ratio E_B/λ will, therefore, represent a certain frequency dependence. All the factors in the product $P_{UT}R_L$ are determined without r and η. r can, for instance, be determined by measuring the impedance in a diode oscillator (this will yield the large signal value for r; the small signal value can be otherwise calculated), while η can be set approximately equal to 0.1 ([8] indicates that η varies between θ and j).

The end result of calculating $P_{UT}R_L$ as a function of frequency is shown in figure 8 and is proportionate to $f^{-1.6}$. The deviation from f^{-2} occurs because of the frequency dependency with E_B/λ. The curve is valid for Si (or GaAs) diodes. The values produced by Ge are 4 times less.

Hofflinger also reports in [9] that the total thermal resistance R_T for a circular Si disc as a function of its surface, A, with ideal mounting on a diamond heat sink (see [23] for a detailed discussion of heat problems in ATT diodes). For any particular frequency, the amount of power per unit of surface, P_{IN}, which is required to make the diode oscillate can be calculated. The total power $P_{IN} = p_{IN}A$ yields the temperature rise in the diode $\Delta T = P_{IN}A_T$, which, therefore, is a function of the diode surface and, through P_{IN}, dependent upon frequency. (With earlier descriptions, the calculation of P_{IN} can be explained in the following manner: at any particular frequency there are determined values for L, E_D and λ which determine the current density J_o, the diode voltage $V_D = E_B L$ and consequently $p_{IN} = J_o V_D$). For example, then, at the frequencies 3, 10, and 50 GHz, the current density will equal 300, 1300 and 12,000 A/cm^2, and the diode voltage will be approximately equal to 250, 90 and 24 V. In other words, $p_{IN} = 7 \cdot 10^4$, $1.2 \cdot 10^5$ and $2.9 \cdot 10^5$ W/cm^2. With an allowable temperature rise of $\Delta T = 180^oC$, the maximum diode surfaces are approximately $2 \cdot 10^{-3}$cm^2 (3 GHz), 10^{-3}cm^2 (10 GHz) and 10^{-4} cm (50 GHz). The corresponding maximum input power becomes 140 W, 120 W and 29 W. The $P_{UT}R_L$ curve in figure 8 gives the corresponding load resistance of 36, 6 and 2 ohm.

The highest conceivable peak power is yielded from the $P_{UT}R_L$ curve in figure 8 under the assumption that the lowest feasible load resistance is 1 ohm. This the produces 500 W at 3 GHz, 75 W at 10 GHz, 6 W at 50 GHz and 2 W at 100 GHz. The aforementioned maximum CW and peak power as a function of frequency are shown in figure 9, where the best obtained laboratory results for the years /40

1965, 1967 and 1969 are also shown.



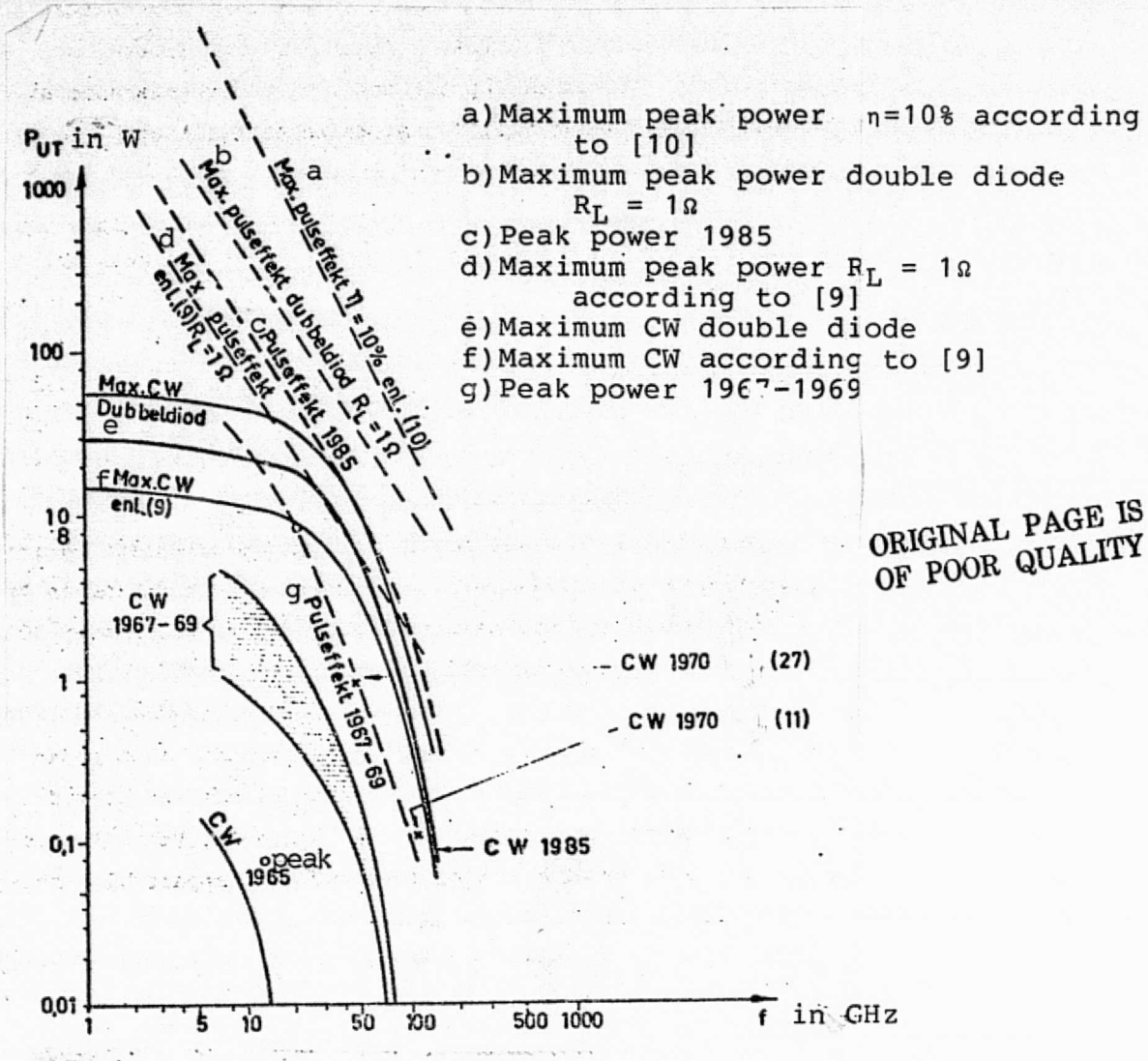
ORIGINAL PAGE IS
OF POOR QUALITY

Figure 8. The product of the power output times the load resistance for ATT and ARP oscillators according to [8,9].

For the anormal mode, Hofflinger [9] has used several simplifying assumptions to estimate the product /41

$$P_{UT}R_L = \frac{V_B^2}{16},$$

in which a 50% efficiency level is given. If this high efficiency level could be reached with one-half the transit-time frequency, the output frequency would correspond to $\Theta = 1$ radian. The transit-time frequency 20 GHz, which requireds that $E_B = 320$ kV/cm and $L = 1.8 \cdot 10^{-4}$cm, so that $V_B = 58$ V, would then produce $P_{UT}R_L$=200 Wω at 10 GHz. In fugure 8, the curve labeled "ARP mode" was calculated in this manner.



ORIGINAL PAGE IS OF POOR QUALITY

Figure 9. Power output for ATT diodes as a function of frequency.

According to [9], the feeding power of a diode in the ARP mode /42 is 2 to 3 times larger compared to a ATT diode with the same dimensions. (The necessary current density is approximately 5 times higher, but the working voltage is only half as large due to current rectification, which therefore produces a power factor of 2.5.) The higher efficiency in the ARP mode means that the loss effect is approximately the same as in the ATT mode, and also in this way the temperature rise is the same. Assuming that the feeding power is 2 times higher and the efficiency is 5 times higher in the ARP mode, the output power will be 10 times higher than that in an equally large ATT diode. The power, however, is produced at a frequency which yields a harmonic to the transit- time frequency and in

the best case is equal to one-half the transit-time frequency. The maximum CW power in this case is shown in figure 10. Also shown is the maximum peak power which can be reached given the assumption that the load can be transformed tc 1 ohm. The best results obtained to date with the anormal mode are also shown.

ORIGINAL PAGE IS
OF POOR QUALITY

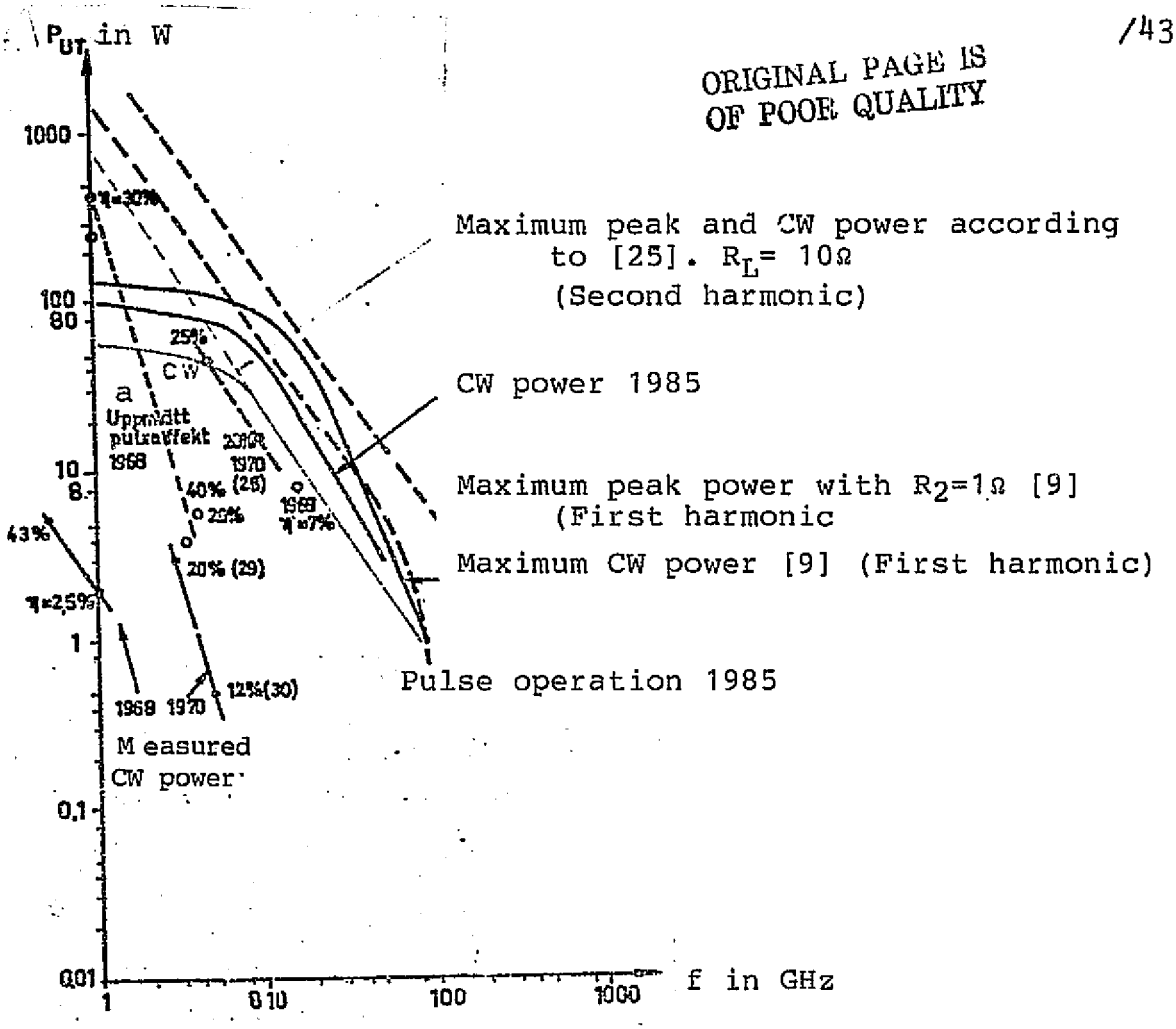


Figure 10. Power output in the ARP or TRAPATT mode.

Later work by Evans and Scharfetter [24,25] has produced good correlation between experiments and calculations and as a result is in strong support of the TRAPATT mode theory. In [25], CW and peak power is calculated for a TRAPATT diode working at one-third of the transit-time frequency, with R_L = 10 ohm and a level of efficiency between 40 and 50%. The results of the calculations are presented in figure 10. Given that $P_{UT}R_L$ is a constant, the estimation of 1 ohm is in good agreement with the maximum peak power obtained by [9]. The CW power is then somewhat smaller than the curve presented

by [9] at low frequencies and somewhat larger at frequencies over 10 GHz. (Information on the heat sink is missing.)

Continuing experimental research has shown that when ground oscillations in the TRAPATT mode are not charged, power can be extracted from harmonics with fairly good efficiency. However, data published to date indicates that improvements will hardly be possible in conjunction with the above forecast. From its discovery in 1965 until 1967, the ATT diode was developed very rapidly, mostly because the necessary silicon technology already existed. The high CW and peak power which was obtain in research laboratories in 1967 have still, with some exceptions, not been surpassed. This indicates that continuing improvements in performance should occur at a slow pace. Present efficiency levels between 5 and 10% will be raised to 15 to 20%, according to newly present large signal calculations [12,13] in Read diodes. With improvements in efficiency of this magnitude up until 1985, power output can be expected to be approximately twice as large as calculated in [9] for both CW and pulse operation, or approximately 20 W CW in the X-band. The newly begun development of a double diode [26] could bring /44
about further improvements. Such a diode consists of a p^+pnn^+ formation, differing from the usual, which is either p^+nn^+ or n^+pp^+ formation. The maximum field used to produce avalanche breakdown is located in the middle of the diode and on the respective sides there is one drift region for electrons and one for holes. This means that the output power per unit of surface area is approximately doubled and that the impedance per surface unit is also doubled. The product of the power output and the impedance will then be 4 times larger than in a normal ATT diode with a drift region. At the same time, improved efficiency is obtained. Diodes manufactured through ion implantation have already produced 1 W CW with 50 GHz with an efficiency of 13.3% [27], compared to the maximum of 0.45 W from a simple diode.

With an assumed efficiency of 15%, the maximum power limit with pulsed operation will total approximately 6 times higher than given in [9]. However, difficulties in connection with high CW power will arise with a two-sided heat sink, and will lead to a significantly lower power increase. The likely increase factor will lie around 4. This power limit for double diodes is shown in figure 9. Because the current-displacement phenomena in the substrate at very high frequencies leads to an uneven current density distribution in circular diodes, the efficiency level will decrease significantly [14]. If one assumes a value of 15 to 20% with the X-band, the value at 100 GHz will have fallen to at least 10%. This estimate is based upon the fact that to date the best obtained efficiency levels within the X-band are around 10% and at 100 GHz only 3%. The tendency towards reduced efficiency at high frequency can in some ways be counteracted by the construction of diodes with very thin stubstrates or with ring formations.

Two forecast curves based upon the above reasoning, labeled "CW 1985" and "Peak power 1985," are also shown in figure 9. They are shown up to a level of 150 GHz, where the drift region's thickness is around 0.2 μ. It is preferred that an abrupt pn junction shall be obtained with a much thinner layer, which is still expected to be possible through multiple ion implantation (several successive implantations with various acceleration voltages can eventually produce very large variations in concentration within 0.2 μ.) This will likely bring about the construction of ATT oscillators at 300 GHz with power output as high as 1mW. /45

It is hard to judge which power levels will be reached by 1985 with the ARP or the TRAPATT mode. The estimated maximum power values shown in figure 10, which assume that the power output will will be obtained at one-half the transit-time frequency, are probably much too optimistic. With regard to the published data, it seems more likely that a 50% efficiency will be reached at one-third to one-fourth of the transit-time frequency. Therefore, the given maximum power values are used at the given frequency to produce the forecast curves in figure 10. This also produces a good compromise with the calculations in [25]. With a maximum transit-time frequency of approximately 300 GHz, the maximum ARP frequency the becomes 75 to 100 GHz.

The previously mentioned p^+pnn^+ double diode [26] is expected to be of special significance for the further development of the TRAPATT oscillator. In [30], the major advantages in being able to optimize the one diode section for normal ATT oscillator operations and the other for TRAPATT operations were shown. A normal ATT oscillation is of course necessary to start the TRAPATT oscillator. The optimization is aimed at creating the highest possible negative conductance in the ATT diode with a certain current and at the same time keeping the necessary (transient) start-up current low for the TRAPATT diode. This can not be done simultaneously in one diode. Preliminary measurements indicate 0.5 W CW power at 4.8 GHz with 12% efficiency.

According to this forecast, the ARP mode will allow higher CW output power than the transit-time mode at frequencies under 15 GHz. Between 1 and 10 GHz, the power is two to three times higher and efficiency is at all times approximately three times higher.

In terms of the maximum peak power, the ARP and the ATT modes will yield approximately the same power level: several kW at 1 GHz and approximately 100 W in the X-band. /46

It should be observed that a large number of avalanche diodes can easily be produced on the same substrate [15]. Because of this, combined parallel and series couplings should be able to produce at least ten times higher power levels than discussed above for a diode with an unchanged level of impedance. Instead of circular diode structures, very small, oblong pn junctions can be of great significance because of their lower thermal resistance.

Noise in the Avalanche Diode Oscillator

AM noise in an X-band diode is saturated within the 1 kHz bandwidth, at a distance of 10 kHz from the carrier wave, and typically around 100 to 110 dB under the signal power, in the best case, according to [18], at a value of 125 dB. The effective value Δf_{rms}, used to characterize FM noise, has been shown (see for example [16]) to depend on the mid-frequency f_o, the outer Q value, Q_{ex}, the bandwidth B and the output power P_{UT}, according to the following formula:

$$\Delta f_{rms} = \frac{f_o}{Q_{ex}} (kTB/P_{UT})^{\frac{1}{2}}, \text{ (where kTB indicates thermal noise).}$$

The lowest value for Δf_{rms} reached within the X-band with an output power of about 100 mW and $Q_{ex} \approx 150$ is on the order of 25 Hz, in which B = 1 kHz (standard value) [18]. One characteristic of the avalanche diode oscillator is that both the AM and the FM noise is constant with varying distance (f_m) from the carrier frequency. In this way, the flicker noise in an oscillator of this type should be insignificant. Certain new measurements of commercially available oscillators indicate, however, in contradiction to this, some increase in FM noise and a large increase in AM noise near the carrier wave. For example, for 100 kHz to 10 kHz distance from the carrier wave, FM noise increases 10 Hz and AM noise 10 dB, according to [17]. On the other hand, new measurements by [18] dispute the results in [16] for f_m values less than 100 Hz. The divergent results must be due to differences in the diode's construction. /47

Despite the fact that the avalanche process will in principle generate noise, theoretical research shows that both AM and FM noise in an ATT oscillator should be able to be reduced to a level lower than what has been obtained to date. At the MOGA conference in 1970, it was reported [22] that an X-band ATT oscillator with a Schottky-barrier contact (and high Q-cavity?) produced $\Delta f_{rms} \approx$ 2.5 Hz/kHz at a distance of two to 10 kHz from the mid-frequency. This represents a 10-time improvement over the best normal ATT diodes. Noise characteristics nearly as good as those in the Gunn diode are therefore expected in the future.

An increase in P_{UT} to 1 W and in Q_{ex} to 1000 should, therefore, lead to $\Delta f_{rms} \approx 3$ Hz, a value which is as low as the best Gunn diode oscillators. For obtaining the highest P_{UT} and with that, the lowest FM noise, special series-couplings of several diodes will be significant. Improvements in efficiency will also be important in this regard and in this way the TRAPATT mode can eventually lead to a lower noise level. Certain experimental results indicate, however, that noise is not invariably reduced with higher power output. In this manner, the lowest noise is not produced when the load is optimized for maximum power. However, the power reduction is not so large (< 3 dB) when the load is optimized for minimum noise [18].

Electronic Tuning of the Avalanche Oscillator

Since the equivalent inductance in the avalanche region varies with direct current through the diode, the oscillation frequency can in principle be changed with the current. Frequency changes are normally on the order of 0.5 to 1 MHz/mA. In the X-band, frequency changes of approximately 10 MHz can be easily obtained, but at the same time the power is greatly changed because the diode's negative resistance also varies with the current.

A more linear frequency tuning with nearly constant power output is obtained by coupling a varactor diode to the oscillator circuit. Coupling a more powerful varactor produces a larger electrical tuning area, but at the same time increases losses in the varactor which in turn reduces power output. Varactor-tuning of avalanche diode oscillators has so far been used very little, but an electronic tuning region of 10% should be reached without difficulty. The combined effect of several varactor diodes should in the future produce a tuning region of up to one octave (66%). /48

Power Amplification with the Avalanche Diode

The avalanche diode can be of great significance for amplification in frequency ranges above those in which the transistor can be used. A normal two-gate amplifier is obtained when the avalanche diode's negative resistance, which yields reflection amplification, is connected to a circulator. Both the usable frequency range and the power output level are assumed to be very close to earlier-presented conditions in the avalanche diode oscillator. According to actual measurements, the noise factor remains relatively high, around 30 to 40 dB in a silicon diode and around 10 dB lower in a germanium diode. (However, the noise factor is less significant in connection with large signal amplification.) An X-band amplifier with a diode and 15 dB amplification typically produces a bandwidth of several percent and a power output between 0.5 and 1 W.

Since the negative resistance in the avalanche diode is large within a relatively wide region around the avalanche resonance, (around 4 GHz in the X-band), broadband methods will surely be developed which will lead to amplification with a 20% bandwidth or greater. Amplifiers with several diodes will produce octave bandwidth. Attempts have already been made to combine the amplification from several diodes with circuits which are detuned in relation to each other. For example, three diodes, each with its own minature circulator, have produced 40 dB amplification, 200 MHz bandwidth and 30 mW power output in the X-band [19]. Three direct-coupled diodes in a periodic waveguide structure have produced 120 MHz 1 dB-bandwidth with approximately 15 dB amplification and 11 GHz [20].

Amplification with avalanche diodes which function in the ARP /49
or the TRAPATT mode has also been reported [21]. A diode which oscillated at 10.4 GHz produced a saturation amplification of 12 dB and a peak power output of 6.5 W at 3.1 GHz with an efficiency of 25%. The bandwidth was 3%. The power output from such an amplifier is assumed to be close to the forecast curve in figure 10. It will be of special interest between 1 and 15 GHz, where it will probably produce more power output than the transistor and the ATT diode amplifier.

A special for of saturated amplification which is of great interest in connection with a number of diode oscillators is the frequency-locking of the oscillator which can be produced from a much weaker control signal. The necessary locking power is produced when the control frequency recedes from the oscillator frequency according to the following formula:

$$\frac{f_{osc} - f_{lock}}{f_{osc}} = \frac{1}{2Q_{ext}} \left(\frac{P_{lock}}{P_{osc}}\right)^{1/2}.$$

The locking power P_{lock} typically lies 15 to 25 dB under the oscillator power P_{osc}, when $f_{osc} - f_{lock} \leq 1$ MHz. Of relevant interest is the fact that if the locking oscillator has lower noise than the power oscillator, the noise level in the power oscillator will drop. A reduction in FM noise near the carrier frequency by a factor of 10 to 30 can easily be obtain with a phase-locking signal at a relatively high power-level. Frequency and phase-locking is discussed further in connection with the Gunn Diode.

REFERENCE

1. Read, W.T., "A proposed high-frequency, negative resistance diode," Bell Systems Techn. Journal, (March 1958). /50

2. Johnston, R.L., et al, "A silicon diode microwave oscillator," ibid., (February 1965).

3. Prager, Chang, Weisbrod, "High power, high efficiency silicon avalanche diodes at ultra-high frequencies," Proc. IEEE, p. 586, (April 1967).

4. Iglesias, Evans, "High-efficiency CW IMPATT operation," Proc. IEEE, p. 1610, (September 1968).

5. Johnston, R.L., et al, "High-efficiency oscillations in Ge avalanche diodes below the transit-time frequency," Proc. IEEE, p. 1611, (September 1968).

6. Evans, W.J., "Circuits for high-efficiency avalanche-diode oscillators," Trans. IEEE, MTT, p. 1060, (December 1969).

7.

8. Hofflinger, B., "Fundamental aspects of ATT diodes with distributed multiplication," Proc. high frequency generation and amplification, Cornell University, (1967).

9. Hofflinger, B., "Recent developments on avalanche diode oscillators," Microwave Journal, (March 1969).

10. De Loach, B.C., "Recent advances in solid state microwave generators," Microwave Journal, (March 1969).

11. See Electronics, p. 96, (April 1969).

12. Scharfetter, Gummel, "Large-signal analysis of a silicon Read diode oscillator," Trans. IEEE, ED, (January 1969).

13. Mouthaan, K., "Nonlinear analysis of the avalanche transit-time oscillator," Trans.IEEE, ED, (November 1969).

14. De Loach, B.C., "Thin skin IMPATTs," Trans. IEEE, MTT, p.72-74, (January 1970).

15. Zettler, Cowley, "Batch fabrication of integral-heat-sink IMPATT diodes," Electronic Letters, (December 1969).

16. Josenhans, J., "Noise spectra of Read diode and Gunn oscillators," /51 Proc. IEEE, p. 1478, (October 1966).

17. Varian-catalogue, High-Q IMPATT oscillators, (April 1969).

18. Cowley, A.M., "Design and application of silicon IMPATT diodes," Hewlett-Packard Journal, (May 1970), and (July 1970) p. 9, correction.

19. Schere, Barett, "A broadband multistage avalanche amplifier at X-band," ISSCC Digest, (1969).

20. Ivanek, Reddi, "X-band oscillator and amplifier experiments using avalanche diode periodic structures," ISSCC Digest, (1969).

21. Hofflinger, Snapp, Stark, "High-efficiency avalanche resonance pumped amplification," ISSCC Digest, (1969).

22. Goldwasser, Berenz, Lee, Dalman, "IMPATT and Gunn oscillator noise," MOGA 70, speech 12.5.

23. Clorfeine, A.S., "Power dissipation limits in solid-state oscillator diodes," Microwave Journal, p. 93, (March 1968).

24. Evans, Scharfetter, "Characterization ofavalanche diode TRAPATT oscillators," Trans IEEE,ED, p. 397, (May 1970).

25. Scharfetter, D.L., "Power-frequency characteristics of the TRAPATT diode mode of high efficiency power generation in germanium and silicon avalanche diodes," Bell System Techn. Journal, p. 799, (May-June 1970).

26. Scharfetter, Evans, Johnston, "Double-drift-region (p^+pnn^+) avalanche diode oscillators," Proc. IEEE, p. 1131-33, (July 1970).

27. Seidel, Scharfetter, "High-power millimeter wave IMPATT oscillators with both hole and electron drift spaces made by ion implantation," Proc. IEEE, p. 1135-36, (July 1970).

28. Ying, Kramer, "X-band silicon TRAPATT diodes," Proc. IEEE, p. 1285-86, (August 1970).

29. Evans, Johnson, "Improved performance of CW silicon TRAPATT oscillators," Proc. IEEE, p. 845, (May 1970).

30. Evans, Seidel, Scharfetter, "A novel TRAPATT oscillator design," Proc. IEEE, p. 1294-95, (August 1970).

The Gunn Oscillator

Instabilities in GaAs in connection with high electrical fields were first observed by J. B. Gunn in 1963 [1]. The possible existence of such instabilities, however, had been shown theoretically by Ridley, Watkins [2] and Hilsum [3] several years earlier. /52

The energy levels in GaAs are constructed in such a way that if an electron at room temperature is accelerated in a field larger than 3.5 kV/cm, it will rise to a higher energy level in which its effective mass is 15 times greater than at the lower level. As the field strength increases up to approximately 10 kV/cm, more electrons will rise to the higher level and become less mobile. At even higher field strength, the electron distribution approaches a state of equilibrium. The resultof this process is that the average speed of the electrons, and consequently the current through the semiconductor, is proportionate to the energy field from 0 to 3.5 kV/cm, and remains relatively constant above that level. The diminishing current above the threshold field corresponds to a negative differential resistance which can give way to oscillations. Several oscillation and amplification processes can arise depending upon the thickness L of the semiconductor layer, the electron concentration n_o, the applied voltage and the characteristics of the outer circuit. Four main groups can be identified:

1. If $10^{12} < n_oL < 10^{13}cm^{-2}$, a powerful excess of electrons is usually built up in a thin semiconductor junction, in which $n >> n_o$ in the collection region and $n < n_o$ in the adjacent depletion region. This rearrangement of the charges increases rapidly up to an equilibrium state and causes a powerful, localized increase in the energy field (called domain). If the outer circuit has a low Q-value, the domain built-up around the cathode will travel throughout the semiconductor at a speed of 10^7 cm/s. When the domain is taken up by the anode, another is built-up at the cathode. This periodic flow, the frequency of which is determined by the thickness of the semiconductor, is called transit-time mode.

2. If $n_oL \approx 10^{12}$ cm^{-2}, and the variations in doping in the semiconductor are very small, only one group of excess electrons is created within a very thin layer. This concentration of charges is called the accumulation domain, as opposed to that in section 1, which is called the dipolar domain. As in 1, a periodic transfer of current through the semiconductor determined by the transit-time is created. Disturbances in the doping level can easily transform an accumulation domain into a dipolar domain. /53

3. If $n_oL < 10^{12}cm^{-2}$, no domain is created. However, the negative dynamic resistance can be utilized for small signal amplification in a circulator-connected reflection amplifier. This is the so-called stable amplification mode.

4. Even if $n_oL \geq 10^{12}$ cm^{-2}, domain building can be suppressed by various methods, and one can then speak of a pure negative

resistance mode.

The transit-time mode mentioned in paragraph 1 can with the appropriate formation of resonance circuits be made to function both over and under the natural transit-time frequency. The following sub-groups can be delineated:

1a. If the resonance circuit is tuned to a frequency above the transit-time frequency and has a sufficiently high Q-value, the wandering domain will be discharged by the high-frequency field before it reaches the anode. With increased frequency, a larger portion of the semiconductor will be inactive, which leads to decreased frequency. However, up to a level of 1.4 times the transit-time frequency the decrease is insignificant. This method of operation is termed the abbreviated transit-time mode.

1b. When the resonance circuit is tuned to a frequency under the transit-time frequency, to as low as one-half its value, the domain will travel throughout the semiconductor junction. However, the initiation of the next domain will be delayed because the negative portion of the high-frequency period has lowered the diode voltage under a critical level. Not until the high-frequency voltage increases again can a new domain be created. This so-called delayed transit-time mode usually works with good efficiency down to 0.6 times the transit-time frequency.

1c. A Gunn diode with a built-up dipolar domain can experience, particularly at lower frequencies, negative resistance, which can easily cause undesirable oscillations in the voltage-feeding circuits. If these are not suppressed with appropriate measures, they can unfavorably affect the desired high-frequency. These vibrations at low frequencies are called the relaxation mode. /54

1d. The negative dynamic resistance in the dipolar domain can also be utilized in a circulator-coupled amplifier at frequencies high above the transit-time frequency. As opposed to amplifiers with sub-critically doped GaAs (paragraph 3), high values are achieved in the saturation level and the power output. This function can be called the wandering-domain amplifier mode.

The mode described in paragraph 4 which produces a negative resistance without the creation of either an accumulation domain or a wandering domain, can be divided into the following variations:

4a. The LSA mode (from limited space-charged accumulation) [4] can be created with a sinusoidal signal if the ratio of the charge concentration to the frequency n_o/f lies between 10^4 and $2 \cdot 10^5$ $cm^{-3}s$ (with a square-wave signal, the upper limit is 10^6 $cm^{-3}s$). The resonance circuit in this case is tuned to a frequency high enough so that a domain does not have time to be

built during the positive portion of the high-frequency period. However, the collection area which is building up at the cathode must have enough time to discharge completely through the dielectric relaxation process during the negative portion of the high-frequency period. The result of this is that the entire semiconductor functions as a negative resistor and can produce an oscillation frequency which is completely independent of the thickness of the semiconductor junction.

When n_oL approaches $10^{13}cm^{-2}$ and n/f approaches $10^6cm^{-3}s$, the accumulation does not have time to completely discharge and dipolar domains can occur, even though they do not become fully developed. This is called the hybrid mode and is used often to extract high power with a harmonic to the transit-time frequency.

4b. The creation of dipolar domains can be prevented through frequency-selective loading of the high-frequency circuit. At the other frequencies, a negative resistance remains which can be utilized, for example, in a reflection amplifier. This is an amplifier mode with a circuit-suppressed domain. /55

4c. Under certain conditions (L small and n_oL relatively large), only a static domain is created in the region of the anode. This brings about a stable negative resistance, usable for amplification within a wide range of frequencies. This amplifier mode with static domain is probably closely related to that in paragraph 4b and can eventually become identical to it.

4d. In another type of amplifier or oscillator, the creation of the domain is suppressed by using the altered field distribution of a dielectric junction in contact with a thin, active GaAs junction. This is a mode with dielectrically-suppressed domain.

Power Limits of the Gunn Diode

The Transit-time Mode

A. Sasaki [5] has shown that the power output from a Gunn diode with wandering domain can be written, ignoring any thermal limitations, as:

$$P = \left[\frac{(E_B - E_S)L}{\sqrt{2}}\right]^2 \frac{1}{R_L} = \frac{(E_B - E_S)^2}{2R_L} \frac{v_d^2}{f^2},$$

where

E_B = the electric field from the applied direct voltage
E_S = the lowest electric field in which a domain can exist
v_d = the domain's drift speed
f = transit-time frequency
R_L = the load resistance.

With E_B = 25 kV/cm (approximately equal to the practical maximum value, which, however, does not produce maximum efficiency), then E_S = 1.5 kV/cm, $v_d = 10^7$ cm/s and f in GHz, we obtain:

$$P_{max}R_L \approx \frac{28{,}000}{f^2_{GHz}} \text{ W}\Omega. \qquad /56$$

If it is further assumed that R_L = 1 ohm is the lowest useful load, $P_{max} = 28{,}000/f_{GHz}$ is obtained.

Using more precise calculations in which efficiency is maximized, Copeland [6] has calculated the somewhat lower value

$$PR_L = \frac{12{,}000}{f^2_{GHz}} \text{ W}\Omega, \text{ which applies for } \eta = 7\%; \ (PR_L \sim \eta).$$

For a Gunn diode functioning in the transit-time mode:

$$n_o L \approx 10^{12} \text{ cm}^{-2},$$

and $$f = \frac{v_d}{L} = \frac{10^7}{L_{cm}} \text{ Hz}.$$

The specific resistance can be written as:

$$\rho = \frac{1}{en_o\mu} = \frac{10^{-5}}{e\mu f_{Hz}} \Omega\text{cm} = \frac{10^{10}}{f_{Hz}} \Omega\text{cm},$$

where μ = the mobility, (6000 to 8000 cm^2/Vs.

If one assumes that the cross-sectional surface area in the Gunn diode is limited only because the dimensions must be small in comparison to the wavelength, the maximum surface can be written as:

$$A_{max} = \left(\frac{\lambda_g}{4}\right)^2 = \left(\frac{\lambda_o}{4\sqrt{\varepsilon}}\right)^2$$

where λ is the wavelength at which ε = 1.

With ε = 10.8 for GaAs, we have

$$A_{max} = \frac{5.2 \cdot 10^8}{f^2_{Hz}} \text{ cm}^2.$$

In this way, the low field-resistance becomes /57

$$R_o = \rho \frac{L}{A_{max}} = \frac{1}{52} \Omega, \text{ independent of f.}$$

According to Copeland, maximum efficiency is achieved when the load $R_L \approx 50R_o \approx 1\ \Omega$, with $E_B \approx 10$ kV/cm. For higher E_B values, R_L must be increased, presumably to $100R_o$, with the earlier assumption of E_B = 25 kV/cm, in which $R_o \approx 2\ \Omega$, independent of f. In this way, a value midway between the calculations of Sasaki and Copeland

ORIGINAL PAGE IS
OF POOR QUALITY

can feasibly be set at

$$P_{max} = \frac{14,000}{f^2_{GHz}} \text{ W.}$$

This should apply to efficiencies somewhat under 10% and approximates the maximum power output in pulse operation. This relationship is shown in figure 11.

By using the conclusion of Knight [7], the maximum CW power from a Gunn diode can be calculated. For an epitaxial $n^{++}nn^{+}$ junction diode, with the n^{++} junction (of thickness δ) connected to an infinitely large heat sink, and with an active junction of thickness L and radian r, the maximum temperature rise in the active junction will be

$$\Delta T = (T_{vs} + \frac{QLr}{K_{st}}) e^{QL(\frac{L}{2K_1} + \frac{\delta}{K_2})} - T_{vs},$$

where T_{VS} = the heat-sink temperature ≈ 300°K,

Q = the applied power per unit of volume. This is nearly constant in the diode and given as JE = the current density times the field strength.

K_{st} = the thermal conductivity in the heat sink

$\frac{K_1}{T}$ = the thermal conductivity in the n junction

$\frac{K_2}{T}$ = the thermal conductivity in the n^{++} junction.

This can be written as /58

$$Q = JE = (v_s n_o e)E_B = \mu E_s n_o e E_B = \frac{E_s E_B}{\rho} = f_{Hz} E_s E_B \cdot 10^{-10} = 0,1\, f_{GHz} E_s E_B.$$

With E_B chosen to maximize the efficiency, according to Copeland, (or, E_B ≈ 8000 V/cm and E_S = 2000 V/cm), then $Q = 1.6 \cdot 10^6 f_{GHz}$.

Since the efficiency in a Gunn diode decreases rapidly at temperatures over 200°C, the highest allowable $\Delta T \approx 200^{\circ}$. To find a dimension variable to meet these requirements, we set $r = \alpha r_{max}$, where $\alpha \leq 1$. Since the power output is proportionate to r^2, the calculated power output becomes $\alpha^2 P_{max}$.

A calculation using the following data yields the results shown in table 2: K_{st} = 3.88 W/cm°K for copper, 16 W/cm°K for diamond; K_1 = 150 W/cm; K_2 = 120 W/cm; $\delta = 4 \cdot 10^{-4}$cm = 4 μ; and $r_{max} = \lambda g/8 = 1.15/f_{GHz}$.

In the previous calculations, consideration has not been given /60 to the fact that efficiency (or negative resistance) over 100 GHz

Table 2. Calculated maximum power output for a Gunn diode.

f GHz	α_{cu}	$\alpha_{diamond}$	$P_{UT} = \alpha^2 P_{max}$ $P_{UT_{cu}}$ W	$P_{UT_{diamond}}$ W	
1	$1.68 \cdot 10^{-3}$	$6.72 \cdot 10^{-3}$	0.04	0.64	$\Delta T = 250^{o}$
5	0.132	0.528	10	160	
10	0.316	1	14	140	
30	1	1	15.5	15.5	
100	1	1	1.4	1.4	
300	1	1	0.15	0.15	

These results are also shown in figure 11, where the values for 1 GHz have been modified somewhat. The difficulty in producing CW operation at low frequencies can be lessened somewhat by operating in the delayed transit-time mode, in which the lowest frequency is approximately 0.6 times the transit-time frequency. In this manner, the power values in the table's first column should be valid down to 0.6 GHz.

can be expected to diminish through the effect of the finite relaxation time during the transfer of electrons between the two energy states. According to certain calculations, efficiency should be zero at 300 GHz, but this is still not regarded as completely proven.

It should be pointed out that if the diode's efficiency is calculated from table 2, so that

$$\eta = \frac{P_{UT}}{\text{applied power}} = \frac{P_{UT}}{QL\pi r^2},$$

a value somewhat over 20% results. This is still approximately twice as large as that used in the beginning assumptions in the expression for P_{max}. This is due to different assumptions in the two calculations, differences in the geometry of the diodes, and the assumed field strength in the determination of Q. The CW power curves in figure 11, calculated from table 2, must therefore be assumed to show maximum powers which are too high. A calculation of output power as 10% of the feeding power produces a value approximately one-half as large.

The LSA Mode

Next to the transit-time mode, the LSA mode is the most promising, especially for the generation of high peak power. Since the thickness of the active junction can be increased independently of the frequency, both the input and the output power can be raised significantly. The thermal limitations in CW operation at

ORIGINAL PAGE IS
OF POOR QUALITY

/59

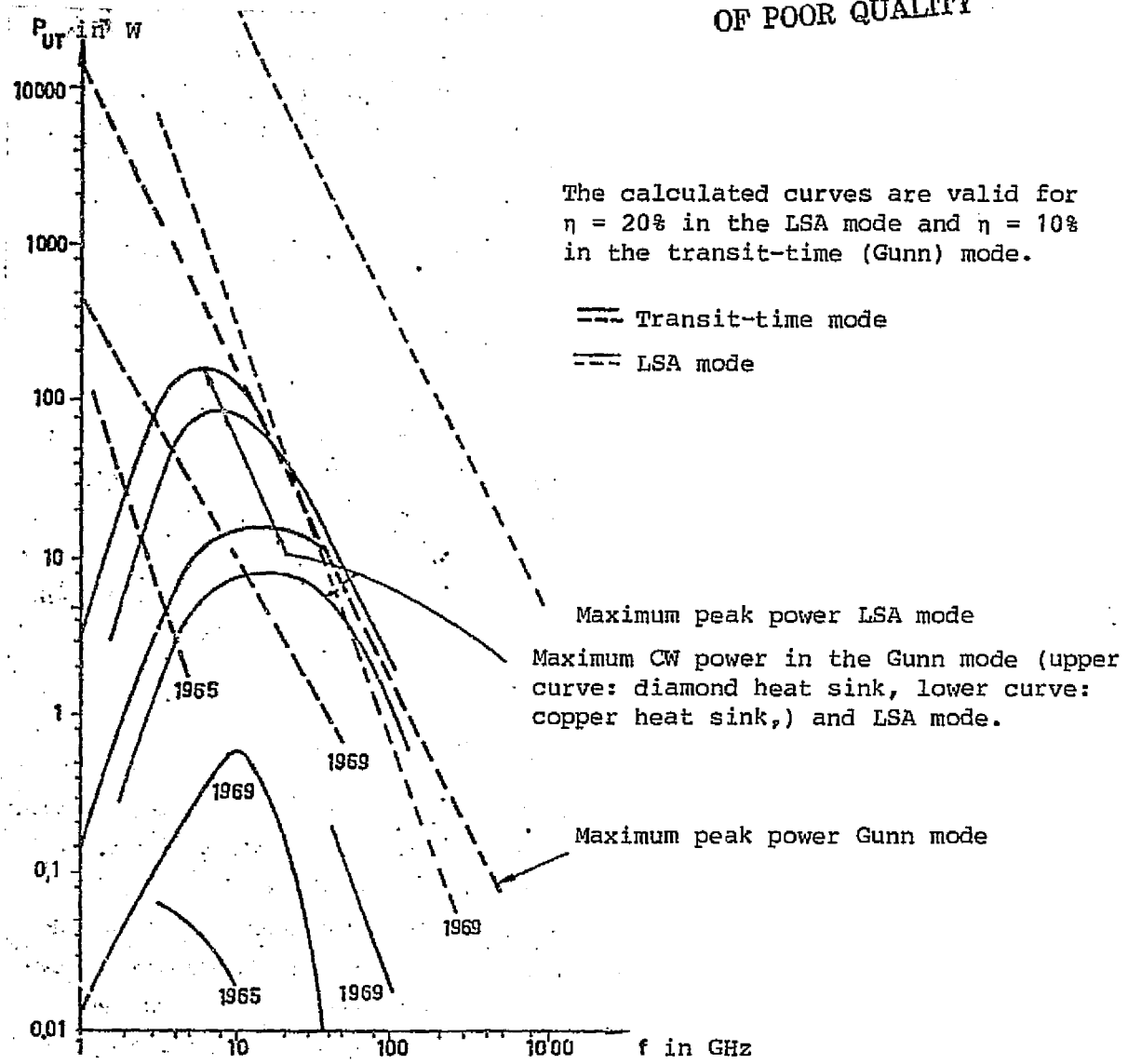


Figure 11. The continuous curves indicate CW operation. The broken curves indicate pulse operation. Curves labeled by year represent the best obtained results.

the centimeter wavelength entail that the power output can not be expected to be much larger than that from a Gunn diode in the transit-time mode. On the other hand, the LSA mode becomes very interesting for CW operation within the millimeter wavelength because of the difficulties in producing the thin junction which a Gunn oscillator must have in this range.

The product of the power output and the load resistance for an

ORIGINAL PAGE IS
OF POOR QUALITY

LSA didde can be written

$$PR_L = \frac{V_{pow}^2}{R_L} R_L = \left(\frac{E_B - E_p}{\sqrt{2}} \frac{\lambda}{4}\right)^2,$$

where E_p stands for the field strength at which the negative differential mobility stops, in other words, the field strength for the maximum drift speed of the charges equal to the threshold field. Naturally, the high-frequency amplitude must be large enough for the total field strength during any particular time to be somewhat under E_p, so that the accumulated charges are spread. However, for this estimation the lowest energy field is $E_p = 3.5$ kV/cm. In this expression, it is also assumed that the active junction's thickness is small in comparison to the wavelength, so that /61

$$L_{max} = \frac{\lambda}{4} = \frac{\lambda_o}{4\sqrt{10,8}} = \frac{2,28}{f_{GHz}} \text{ cm.}$$

According to Copeland [8], the maximum efficiency is 18% when $E_B \approx 25$ kV/cm and $R_L = 50R_o$, where

$R_o = \rho\frac{L}{A}$ = the low-field resistance.

Therefore: $P = \frac{(E_B - E_p)^2}{100} \frac{L_{max}A}{\rho}$.

Therefore, the diode's cross-sectional surface must be maximized for maximum power. A dimension can be made equal to $\lambda/4$, but according to Copeland [9], the other dimension must not exceed f_{GHz}^{-1} cm, if the variation in the high-frequency energy field is to be less than 10%, the level necessary for maximum efficiency. In this way,

$$A_{max} = \frac{\lambda}{4} \frac{1}{f_{GHz}} = \frac{2,28}{f_{GHz}^2} \text{ cm}^2$$

and $$P_{max} = \frac{(E_B - E_p)^2}{100} \frac{2,28^2}{\rho f_{GHz}^3} \text{ W,}$$

where $$\rho = \frac{1}{e\mu n_o}$$

and $$10^4 < \frac{n_o}{f_{Hz}} < 2 \cdot 10^5 \text{ s/cm}^3.$$

Assume that $n_o/f_{Hz} \approx 10^5$ a/cm^3, and $\mu = 8000$ cm^2/Vs. With the values given above, we then have /62

$$P_{max} \approx \frac{3 \cdot 10^6}{f_{GHz}^2} \text{ W,}$$

which, therefore, indicates maximum peak power at approximately

18% efficiency. In comparison to pulse operation in the transit-time mode, the power output is 200 times larger. This relationship is shown in figure 11.

To estimate the maximum CW power in the LSA mode, Knight's conclusion is used again. To obtain sufficiently low Q in this case, which is to say, to avoid excessive temperature rises, $n_o/f_{Hz} = 0.5\ 10^5 s/cm^3$ is chosen. According to Copeland, the values E_b = 25 kV/cm and $R_L = 50R_o$ remain unchanged and still close to maximum efficiency. Observe that the power output in this case is reduced to half compared with the above. To maintain $\Delta T \approx 200^oC$, the LSA diode values $L = \alpha L_{max}$ and the equivalent radius $r = \beta r_{max}$ are chosen. In the following calculations, $2r_{max}$ is assumed to equal the mid-value of the diode's length and width, so that $r_{max} = 0.3/f_{GHz}$ cm. In this manner, the maximum CW power can be expressed:

$$P_{max} \approx 0{,}5\alpha\beta^2 \frac{3\cdot 10^6}{f_{GHz}^2} W = \alpha\beta^2 \frac{1{,}5\cdot 10^6}{f_{GHz}^2} W.$$

ORIGINAL PAGE IS
OF POOR QUALITY

The results of the calculations for a copper or diamond heat sink are summarized in table 3. The results are also shown in figure 11. Here again, the effects of the finite relaxation time on power output have been ignored. The maximum powers shown in the table have not been optimized with any great precision. The diode's thickness, the equivalent radius and n_o/f can all vary. The latter value has been kept constant (equal to 10^5 s/cm^3) and the diode thickness L has be held approximately as thin as in the transit-time mode so that the exponents in the expression for ΔT will not be too large. The equivalent diode radius has then been made as large as possible. The largest diameter is obtained at 5 to 10 GHz with a diamond heat sink, and is approximately 1.7 mm. /63

As figure 11 shows, the maximum power limits for CW operation in the LSA mode lie somewhat lower than those in the transit-time mode. This is due, among other things, to the extra limitation of the diode dimension which was introduced according to Copeland [9]. One advantage of the LSA mode whould be that the predicted efficiency levels are more than twice as large. However, the values of 8 and 18% involve pulse operation and in the experiments done to date, the levels of CW operation have been much lower. According to [19], the efficiency level is reduced rapidly with an increasing temperature gradient in the active junction and approaches zero at a temperature differential of approximately 70°. The assumption of a maximum temperature differential of around 200° used in the above calculations would then be altogether too large. However, it is likely that the efficiency decreases along the temperature gradient can be counteracted more or less completely through an internal doping gradient. The active junction would then be "electrically homogeneous" with the temperature distribution obtained in CW operations. One condition which supports this speculation is that /64

f GHz	α_{cu}	$\alpha_{diamond}$	β_{cu}	$\beta_{diamond}$	Maximum Power Output P_{cu} W	$P_{diamond}$ W
2	0.0044	0.0044	0.015	0.06	0.4	6
5	0.0044	~0.005	0.14	0.57	5	95
10	0.0044	~0.005	0.33	1	7	75
30	0.005	0.01	0.95	1	7.5	15
100	~0.005	~0.01	1	1	1	2

Table 3. Calculated maximum power output for an LSA diode.

when the active junction is evenly heated, efficiency starts to approach zero around 350°C. A double-sided heat sink could also contribute to a lowering of the temperature differences in the active junction. Since the diode's n^+ junction must be relatively thick against the upper heat sink, the upper will not be as effective as the lower. Therefore, any large increases in allowable power loss will not result from a double-sided heat sink, but the temperature profile in the diode will become more even. The calculations of the maximum CW power assuming temperature increases of around 200°C within the diode must still be judged as very optimistic.

It should also be emphasized that the power limits apply to an infinitely large heat sink, so that the "practical" maximum power will be significantly lower.

Also shown in figure 11 are the results obtain in the Gunn and LSA modes using years as the parameter.

The high peak power obtained to date in the LSA mode has invariably been generated with so-called compensated GaAs, which, from a thermal standpoint, is very unstable. This has meant that only very short pulses have been able to be generated at low frequencies of repetition. It is expected by 1985 that the thermal material problems in a pulsed LSA diode will have been solved through the production of thick, thermally stable epitaxial junctions. However, the heat dissipation in diodes dimensioned for maximum peak power, in other words, with very thick active junctions, is so poor that particularly high mean power is still not found. In a diode constructed for very high peak power, a utilization factor higher than 10^{-4} cannot be expected. According to the calculated curves, this value produces a mean power which is approximately 50 times less than the power from an optimized CW diode. A compromise between high peak power and mean power should, in the X-band, lead to several kilowatts of peak power and several watts of mean power. /65

Efficiency and Forecast of Power Output in the Gunn Diode

The maximum power levels calculated above (according to Copeland) involve 10% efficiency in the domain mode and approximately 20% in the LSA mode. These values are obtained with sine wave high-frequency voltage and the relation $r \approx 2$ between the maximum and minimum electron speeds. In a good material, larger values for r can occur, and then the efficiency increases in proportion to $(r - 1)/(r+1)$. Higher efficiency is also achieved by the appropriate tuning of the harmonics, so that the form of the high-frequency voltage approaches a square wave.

A number of calculations indicate in this manner that efficiency levels of approximately 30% should be obtainable in both the transit-time and the LSA modes. However, individual experimental results contradict these figures. The highest reported value with pulse operation in the L-band with a utilization factor of 10^{-4} is 32% [10]. With increased utilization factors and higher temperatures, however, efficiency declines rapidly (see the foregoing discussion in connection with [19]). In CW operation, efficiency over 5% is seldo reached. If it is assumed that the efficiency in pulse operation is around 30% for both the transit-time and the LSA modes, the earlier calculated values for P_{max} will increase by a factor of three in the transit-time mode and 1.5 in the LSA mode. If the efficiency level in CW operation is assumed to be 10%, P_{maxCW} in figure 11 will, according to earlier discussions, be reduced by a factor of two in the transit-time mode and by at least as large a factor in the LSA mode. This yields approximately the same power level for both modes. These efficiency-modified curves are shown in figure 12, where the likely forecast curves for 1985 are also presented. In the latter, an excessively fast power reduction ($P \sim f^{-3}$) has been indicated at frequencies over 100 GHz, but the size of this reduction is very uncertain. This also means that the maximum frequency which can be generated with an LSA diode at present cannot be predicted. It can be speculated that it will lie somewhat over 200 GHz. Diodes functioning in the transit-time mode /66
can hardly be expected to produce at frequencies higher than 150 to 200 GHz, since the active junction would then have to be thinner than 1 μ and, therefore, the substrate upon which the epitaxial junction is built would have to have a smoothness of better than 1000 A.

The forecast curves show the superiority of the LSA mode when compared to the transit-time mode in pulse operation. If GaAs of sufficient purity and homogeneity can be produced in the necessary dimensions, it can be assumed that LSA modes alone will be used for the generation of high peak power. The LSA mode will also dominate the area of CW power at frequencies over 50 GHz and, due to its presumably better noise characteristics, also at other frequencies. Observe that the curve indicating peak power applies at low utilization factors ($\leq 10^{-4}$). If this factor is increased, thermal factos will cause the maximum peak power to decrease.

ORIGINAL PAGE IS
OF POOR QUALITY

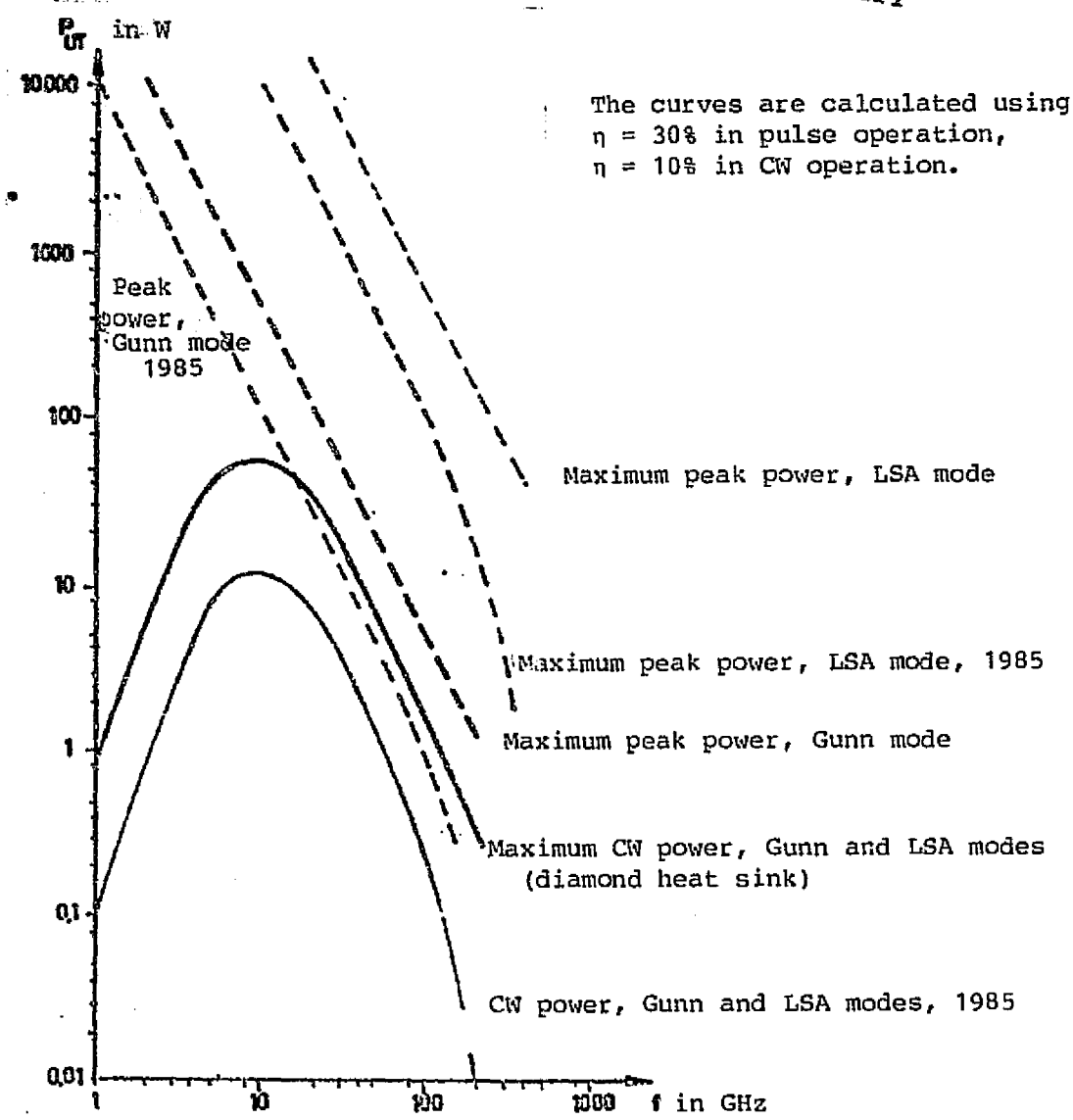


FIGURE 12.

Noise Properties in the Gunn Diode Oscillator

The Gunn diode is characterized by very good noise properties, but as in the ATT diode, both AM and FM noise depend upon the outer Q-value (Q_{ex}) of the resonance circuit. As opposed to the ATT diode, both AM and FM noise decrease with increased frequency distance (f_m) from the carrier wave. Noise reduction is usually on the order of 3 to 6 dB per octave with regard to f_m. Among the best noise data observed to date with X-band Gunn diodes functioning in the transit-time mode (with Q_{ex} = 300) are: AM noise 110 dB under the oscillating power at f_m = 1 kHz, and 125 dB at f_m = 10 kHz with a test bandwidth of 1 kHz. FM noise, expressed in the effective value of the frequency deviation, is Δf_{rms} = 10 Hz at 1 kHz and Δf_{rms}= 4 Hz

at 10 kHz, still with a test bandwidth of 1 kHz. (If one wishes to express the FM noise as a power relation between the noise and the oscillator power, this can be calculated for noise in double sidebands according to $N/P = (\Delta f_{rms})^2/f_m^2$.) For comparison, noise data for a good reflex klystron is given here at f_m = 10 kHz: AM noise, -125 dB; Δf_{rms} = 5. In other words, the noise characteristics of the Gunn oscillator are definitely in the same class as good reflex klystrons.

If the Gunn oscillator is connected with an out stabilizing /68
cavity with extremely high Q, for example, 35,000, the FM noise can be further reduced. Measured within 1 kHz, Δf_{rms} = 1.5 Hz at f_m = 1 kHz, and Δf_{rms} = 0.4 Hz at f_m = 10 kHz. The AM noise lies at -155 dB. An invar stabilizing cavity yields further advantages, reducing the temperature-induced frequency changes in the Gunn oscillator from the normal value of -0.5 MHz/°C to less than -10kHz/°C. One disadvantage is naturally the unavoidable coupling-losses through the cavity, which normally reduce the power output 6 to 10 dB. But, this reduction in power can be compensated by allowing the high-stabilized signal to lock a high-power Gunn oscillator into a low Q-cavity, so that the noise characteristics are worsened only insignificantly. The use of a non-transmitting cavity behind the diode can yield very low stabilizing losses.

Except for the resonance circuit, noise in the Gunn oscillator is naturally dependent on the characteristics of the semiconductor and the electronic process through it. A reduction in the diode's "intrinsic" noise is therefore very significant. The noise theories concerning various types of Gunn diodes are still far from complete, but it is believed that diodes with developed dipolar domains are the worst in terms of noise characteristics. The following mechanisms can be noise-generating:

1. Unevenness and irregularity in connection with domain building and growth.

2. Speed variation in the domain caused by variations in the load concentration in the semiconductor.

3. The occurence of electron deposits.

Through improved material technology, future improvements should be expected in all of these points, but a transition to diodes without dipolar domains will probably produce the best solution. Individual measurements indicate, for example, that a LSA oscillator has 10 dB lower AM noise than a Gunn oscillator with dipolar domains [11]. In terms of noise characteristics, then, future Gunn oscillators should be as good and eventually better than the best contemporary 2-cavity klystrons.

Electric Tuning of the Gunn Diode Oscillator

The frequency of the Gunn oscillator can be varied within a small range of less than 1% by altering the voltage, which alters the diode's domain-capacitance and negative resistance. The size of the frequency change depends much upon the outer circuit, but normally has a value of 1 to 10 MHz/V. With larger changes, the output power will vary significantly, and then it is better to tune with the help of a varactor diode coupled to the resonance circuit. The tuning range with a normally-coupled varactor approaches 10%. Attempts to utilize several coupled varactors indicate the possibility of achieving tuning in the octave range. However, tuning in the octave bandwidth usually occurs by means of a coupled YIG ball, in which the resonance frequency is varied in a magnetic field. The scanning frequency with YIG tuning is limited to several hundred Hz, but frequencies on the order of 10 MHz can be realized with varactor tuning. Since the varactor diode can also withstand higher power than the YIG ball, development will concentrate on the varactor-tuned oscillator. /69

Frequency Locking in the Gunn Diode Oscillator

As with the avalanche diode oscillator, the conditions for frequency-locking can be described by the following equality:

$$\frac{|f_{osc} - f_{lock}|_{max}}{f_{osc}} = \frac{1}{2Q_{ext}} \sqrt{\frac{P_{lock}}{P_{osc}}}$$

$|f_{osc} - f_{lock}|_{max}$ is the maximum frequency range within which the oscillator remains locked with a certain locking-power. This can easily be determined experimentally and is inversely proportional to Q_{ext}.

The phase angle between the oscillator and the locking-oscillator can be written as

$$\Phi = \Phi_o + \arcsin \frac{f_{osc} - f_{lock}}{|f_{osc} - f_{lock}|_{max}}$$

ORIGINAL PAGE IS
OF POOR QUALITY

or

$$\Phi = \Phi_o + \arcsin \frac{f_{osc} - f_{lock}}{f_{osc}} 2Q_{ext} \sqrt{\frac{P_{osc}}{P_{lock}}}$$ /70

or

$$\Phi + \Delta\Phi = \Phi_o + \arcsin \left[2Q_{ext} \left(1 - \frac{f_{lock}}{f_{osc} + \Delta f_{osc}}\right) \sqrt{\frac{P_{osc} + \Delta P_{osc}}{P_{lock}}} \right] .$$

These expression show that all the changes of f_{osc} and P_{osc} will produce a phase shift. Both f_{osc} and P_{osc} in a Gunn diode oscillator are now highly dependent upon the diode voltage V_B and the temperature T. For an X-band oscillator with a low Q-cavity, $\Delta f_{osc}/\Delta T = -0.5$ MHz/°C and $\Delta P_{osc}/\Delta T = -0.05$ mW/°C (with $P_{osc} \approx 15$ mW). If V_B is held constant while the surrounding temperature varies, large phase shifts can result from temperature variations of only several degrees if P_{lock} is small. According to the expression for $\Phi + \Delta\Phi$, a high Q-value can appear to be very undesirable. But a cavity with high Q, which experiences insignificant length extension, has the advantage that $\Delta f_{osc}/\Delta T$ becomes much smaller than in a low Q-cavity. In this manner, changes in Q cannot be expected to have a large effect on phase conditions, while the effect of temperature on the semiconductive material's characteristics will be the most deciding factor. Measurements and calculations of an X-band oscillator by H. R. Holliday [15], show, for instance, that when $P_{osc}/P_{lock} = 20$ dB, $\Delta\Phi = \pm 90^{\circ}$ at $\Delta T = \pm 5^{\circ}$C or at $\Delta V_B \approx \pm 0.4$ V. With an increase in P_{osc}/P_{lock} to 10 dB, $\Delta\Phi$ diminishes to approximately $\pm 10^{\circ}$ with the same ΔT and to $\pm 20^{\circ}$ with the same ΔV_B.

These figures point to difficult practical problems in using several phase-controlled Gunn diode oscillators in, for example, a series-parallel antenna. However, it can be assumed that continued research into materials will produce a more thermally-stable GaAs or a completely new semiconductor with more appropriate characteristics. (The above-discussed problems in phase shifting are generally applicable to negative resistance oscillators, but they /71 have been fully treated here in connection with the Gunn diode because of its high sensitivity to temperature.)

Life Span of the Gunn Diode Oscillator

Previously completed life span tests have yielded 10^5 hours as an estimation for MTBF ("Minimum time before failure"). At the 90% level of confidence, MTBF is greater than $3 \cdot 10^4$ hours (the measurements are by Varian). Hopefully, the life span in the future will be increased at least by a factor of 100, so that MTBF will be on the order of 10^7 hours.

Future Variants of the Gunn Oscillator

The future prospects for the Gunn diode oscillator have been evaluated here using the GaAs semiconductor. This has to date produced better results than several other known mixed semiconductors which display the Gunn effect. Naturally, it is not precluded that new semiconductor combinations can yield even better results, foremost in the area in increased efficiency and increased thermal stability. In this context, modification of the basic principle of the Gunn oscillator can be significant. For instance, Hilsum and Rees [16] have pointed out that a semiconductor

with three distinct electron energy levels, instead of the normal two, could produce improved efficiency together with better operations in the LSA mode.

Power Amplification with the Gunn Diode

As was already mentioned, a Gunn oscillator with wandering dipolar domains possesses a negative resistance at low frequencies and up to ten times the transit-time frequency [12]. An amplifier with high saturation power is then produced in a circulator coupling. Completed experiments show that this maximum output power lies somewhat over half of what the diode produces as an oscillator in the transit-time mode. The following is given as an example of the type of data which to date has been obtained from such an amplifier: f_o = 6 GHz, Δf = 2%, G = 10 dB, P_{max} = 60 mW, and the noise factor F = 20 dB. Very large improvements in terms of output power and bandwidth must be seen as possible within the near future.

A combination of several circulators with separate diodes /72 should at least produce octave bandwidth, and amplification of very high frequencies can be possible. If, for example, the thickness of the active layer is reduced to 2μ, the corresponding transit-time frequency to 50 GHz, and a charge density n_o is chosen around $5 \cdot 10^{16}$, so that the condition that $n_oL >> 10^{12}$ cm^{-2} is met, usable amplification should be obtained up toward 200 GHz. The noise factor will probably not be low in this amplifier.

The most attractive alternative is an amplifier with high n_o, which will make possible high amplification and output power, but with suppression of domain-building, so that the noise remains low. This could be realized in several ways:

(1) A frequency-selective load circuit is built to suppress domain-building. A stationary domain is possibly built at the anode [17].

(2) The active material is made at a right angle to the domain's direction of travel, and it is made so thin that the domain cannot grow. Without making the semiconductor extremely thin, domain-building can also be suppressed if a dielectric disc with high ϵ is laid on the GaAs layer [13].

(3) By forcing a signal of sufficiently large amplitude and high frequency upon the active layer, the same negative resistance principle which is utilized in the LSA oscillator should be able to be used in a circulator-coupled power amplifier [14].

Amplifiers designed according to the principle in (1) should

be relatively difficult to realize, but newly published information, which is probably related to an amplifier of this type, shows that very good results can be obtained: a two-stage amplifier (two circulators and two diodes) yielded 10 dB amplification within the 4.5 to 8 GHz band with an output power of 0.2 W and a noise factor of 15 dB [18]. Several promising experiments have been done according to the principles in (2), while (3) has yet to be tested. According to the calculations, however, a 1 mm long diode being fed with 50 W should be able to produce 177 W at 10 GHz in pulse operation. A special advantage with (2) is that its planar structure can give /73 good possibilities to the effective cooling of the active semiconductor.

Irregardless of which principle of amplification is shown to be best, the above discussion shows that Gunn diodes will be significant for broadband power amplification at high frequencies in addition to their use as oscillators. The output power levels should be on the same order of size as given for Gunn oscillators, and multiple-octave bandwidths should be possible. This type of amplifier could therefore replace the travelling-wave tube at low power levels.

REFERENCES

1. Gunn, J.B., "Microwave oscillations of current in III-V semiconductors," Solid State Com. vol. 1, (September 1963).

2. Ridley, Watkins, "The possibility of negative resistance effects in semiconductors," Proc. Phys. Soc., (August 1961).

3. Hilsum, C., "Transferred electron amplifiers and oscillators," Proc. IEEE, (Fenruary 1962).

4. Copeland, J.A., "A new mode of operation of bulk negative resistance diodes," Proc. IEEE, (October 1966).

5. Sasaki, A., "Frequency dependencies of power and efficiency of transit-time oscillations in two-valley semiconductors," Proc. IEEE, p. 1757, (October 1968).

6. Copeland, J., "Theoretical study of a Gunn diode in a resonant circuit," Trans. IEEE, p. 55, (February 1967).

7. Knight, S., "Heat flow in n^{++}-n-n^{+} epitaxial GaAs bulk effect devices," Proc. IEEE, p. 112, (January 1967). /74

8. Copeland, J., "LSA-oscillator-diode theory," JAP, p. 3096, (July 1967).

9. Copeland, J., "Doping uniformity and geometry of LSA oscillator diodes," Trans. IEEE, ED, p. 497, (September 1967).

10. Berson, Enstrom, Reynolds, "High-power L- and S-band transferred-electron oscillators, RCA Revue, p. 20, (March 1970).

11. Copeland, J., "The LSA oscillator," Proc. of high frequency generation and amplification, Cornell University, (1967).

12. Thim, H., "Linear microwave amplification with Gunn oscillators," Trans. IEEE, ED, p. 517, (September 1967).

13. Kataoka, et. al., "Suppression of travelling high-field-domain mode oscillations in GaAs by dielectric surface loading," Electr. Lett., p. 48, (February 1969).

14. Hashizume, Kataoka, "Transferred-electron negative-resistance amplifier," Electr. Lett., p. 34, (January 1970).

15. Holliday, H.R., "The effect of operating parameters and oscillor characteristics upon the phase angle between a locked X-band Gunn oscillator and its locking signal," Proc. IEEE, (July, 1970).

16. Hilsum, Rees, "Three-level escillator: A new form of transferred-electron device," Electr. Lett., p. 277, (April 1970).

17. Mircea, Magarshack, "Simulation sur ordinateur de la diode Gunn et du circuit associe," Colloque International sur la Microelectronique avancee, Paris, (April 1970).

18. Hardeman, L., "Progress in transferred-electron device circuits"(reference from the 1970 IEEE Convention), Microwaves, p. 16, (May 1970).

19. Christenson, Woodward, Eastman, "High peak-power LDS operation from epitaxial GaAs," Trans. IEEE, p. 732-38, (September 1970).

The Electroacoustic Amplifier

In the electroacoustic or phonon amplifier, amplification is achieved in principle in the same way as in a travelling-wave tube. If the atoms in the crystal grid of a piezoelectric semiconductor are made to vibrate in an appropriate manner, an acoustic wave will spread causing a wandering disturbance in the crystal's inner electric field. When the electrons in the semiconductor are accelerated by an external field to a speed higher than the above-mentioned speed of the disturbance, they will transfer part of their kinetic energy to the acoustic wave's energy field and, in that manner, amplify the acoustic vibrations. (The energy of the vibrations in a crystal can be described in the form of quanta, which are called phonons. In this way, the phonon amplifier gets its name.) The speed at which the acoustic waves spread is approximately 10^5 times less than the speed of light, which means in part that the electrons do not need to be accelerated with so high a voltage to produce interaction, and in part that the wave length for microwave frequencies will be extremely small (for example, 1 μ at 3 GHz). This property means that the microwave-acoustic functions lend themselves well to great miniaturization and particularly to the time delay of electric signals. /75

A number of different types of waves can be utilized within microwave acoustics. In the earliest experiments, volume waves were used, but in the last years the use of surface waves has become more common. In both cases, the waves can be both longitudinal and transversal. The construction of an effective audio-acoustic converter for exciting acoustic waves forms one of the major problems in this context. In the beginning, the concentrated, high-frequency energy field in the cavity was allowed to directly affect the (cooled) piezoelectric crystal, in which the acoustic waves were to be produced. Later, separate converters were produced through evaporation techniques on the surface of a side of the crystal. A $\lambda/2$ thick layer of CdS placed between two metal films produced a fairly effective but narrow band converter at room temperature. Through the production of multiple layers (several layers $\lambda/2$ thick consisting of various materials), the adaption of the acoustic wave to the main crystal was improved and, as a consequence, the conversion efficiency increased. Conversion losses as low as 4.5 dB at 800 MHz and 5 dB at 1,600 MHz have been obtained with a band width around 6%. Since 1965, when White [1] constructed an effective, so-called Rayleigh electro-acoustic converter for surface waves, most of the research in microwave acoustics has been aimed at the utilization of surface waves. The surface-wave connector consists of an interdigital structure with a distance of one-half wavelength between the centers of two adjacent fingers. The connection to the crystal increases with the number of fingers, but at the same time the bandwidth decreases. Naturally, the conversion efficiency is also dependent upon the constant of the piezoelectric coupling, and lithium /76

niobate is particularly good in this respect. A surface-wave interdigital converter has the disadvantage of exciting the elastic waves in two directions at right angles to the length of the fingers. The consequent loss of 3 dB can be eliminated if two interdigital structures, whose centers are located an odd number of square wavelengths from each other, are fed with two signals, one of which is 180° behind the other. Because of interference, the acoustic wave is excited in only one direction.

The width of and distance between the fingers of the interdigital pattern are produced with dimensions as small as 1 μ using the conventional photoresist process. This corresponds to a maximum frequency of approximately 800 MHz. With an electron beam functioning in the electronresist process, structural widths of 0.3 μ, corresponding to a frequency of 2.5 GHz, have been produced [2]. This should be the best obtained result to date. The coupling is made of lithium niobate ($LiNbO_3$), and with an 8% bandwidth, the conversion losses in the converter are 10 dB (inclusive of the 3 dB directional loss).

The most important result of calculations of the function of both volume and surface-wave amplifiers is given in [3]. For volume wave amplification, the following is applicable:

The maximum amplification $\hat{G} = \frac{0.5G_o}{f/f_D + f_c/f}$ dB /77

yields $\frac{v_o}{v_a} = \frac{f}{f_D} + \frac{f_c}{f} + 1$,

where $G_o = 8.6\pi K^2 f_c \frac{L}{v_a}$ dB;

K = the electromechancial coupling constant;

the dielectric relaxation frequency $f_c = \frac{1}{2\pi\rho\varepsilon\varepsilon_o} = \frac{1800}{\rho\varepsilon}$ GHz, in which ε is the relative dielectric constant and ρ is the resistivity in ohm cm;

the charges' diffusion frequency $f_D = 63 \cdot 10^{-10} \frac{v_a^2}{\mu}$ GHz, in which μ is the mobility in cm^2/Vs and v_a is the acoustic wave's speed in cm/s;

v_o = the electron speed = μE_o;

E_o = the field strength from the applied direct voltage;

and L = the length of the crystal.

If the above expression for the amplification is maximized with regard to frequency, the maximum value becomes:

$$\hat{G}_m = \frac{G_o}{4}\sqrt{\frac{f_D}{f_c}} \text{ dB}$$

with $f_m = (f_c f_D)^{\frac{1}{2}}$ and the corresponding $\frac{v_o}{v_a} = 1 + \frac{2f_m}{f_D}$.

ORIGINAL PAGE IS
OF POOR QUALITY

According to [3], an investigation of what bandwidth the amplification mechanism in a volume-wave electroacoustic amplifier can produce shows that octave bandwidth or greater can be easily reached. In practice, this means that the electroacoustic converter will set the limit for the bandwidth; in other words, at present between 20 and 40%.

The previous descriptions show that the power loss per unit /78
of volume from the applied direct current can be written as

$$P_v = \frac{1}{\rho} E_o^2$$

which with maximum amplification becomes

$$P_v = 2\pi f_c \varepsilon\varepsilon_o E_o^2 = 2\pi \frac{f_m^2}{f_D} \varepsilon\varepsilon_o E^2$$

With some typical values inserted, E_o = 2 kB/cm, ε = 12, this becomes

$$P_v \simeq 27\ f_c\ \text{W/mm}^3.$$

For CdS, for example, the power loss is around 10 W/mm^3, which produces difficulties in CW operations. This thermal problem can be solved by placing the crystal in a transverse magnetic field. With this the effective high-frequency mobility will diminish (while the corresponding value for direct current is unchanged), which leads to the changing of f_c and f_D to the effective values f_c' and f_D' so that $f_c' f_D' = f_c f_D$, in other words, unchanged frequency for maximum amplification, but

$$f_c' = \frac{f_c}{1 + (\mu B)^2}.$$

In a semiconductor with high mobility, a powerful magnetic field can decrease f_c and consequently P_b. (With $\mu = 10^4$ cm^2/Vs and $B = 10^4$ Gauss, P_b is only cut in half.)

The maximum output power with maximized amplification is

$$P_m \simeq \frac{4 v_s^3 \varepsilon\varepsilon_o (f_m/f_D)^2}{\mu^2 K^2} \text{ Watt/ unit surface area.}$$

Efficiency then becomes

$$\eta = \frac{P_m}{LP_v} = \frac{4{,}34}{\hat{G}_m} \frac{f_m/f_D}{(1 + 2(f_m/f}$$

If the highest possible efficiency is desired, the amplification $\hat{G}_m$ must be held low and a crystal with a low electromechanical coupling constant, K, becomes feasible. /79

To illustrate the above relationship, the data obtained with a CdS volume amplifier for transverse waves with ρ = 550 Ω cm will be used.

For CdS, $K^2 = 0.04$, $\varepsilon = 9$, $v_a = 1.8 \cdot 10^5$ cm/s, $\mu = 300$ cm^2/Vs and $f_D = 0.68$ GHz. This yields $f_c = 0.36$ GHz and $f_m = 0.5$ GHz. If the amplifier's length $L = 0.1$ cm, $\hat{G}_m = 76$ dB. The ratio v_o/v_a becomes $4.5 \cdot 10^5$, which yields $E_o = 1.5$ kV/cm, $V_o = 148$ V, $P_v = 4$ W/mm^3 and $\eta = 0.7\%$. If the crystal has a cross-sectional surface area of 1 mm^2, $P_m = 24$ mW. Therefore, very high amplification is obtained with a short crystal. Amplification levels over 80 dB cannot be utilized, however, since the amplifier's internal noise reaches such a high lever at the output that saturation occurs.

No complete theory for calculating the function of a surface-wave amplifier has yet been published. According to [3], for a double-layer amplifier, in which an acoustic wave in $LiNbO_3$ is amplified by the electrons in a very close silicon layer (on a sapphire plate), the following amplification equasion applies:

$$G = \frac{G_o\left(\frac{v_o}{v_a} - 1\right)}{\left(\frac{v_o}{v_a} - 1\right)^2 + C_s^2\left(1 + \frac{f}{f_a}\right)^2} \text{ dB},$$

ORIGINAL PAGE IS
OF POOR QUALITY

where

$$G_o = 8{,}68 \frac{d}{\rho v_a^2 \varepsilon_p} \frac{\Delta v_a}{v_a} 2\pi f L \text{ dB};$$

and $C_s = \frac{d}{\rho v_a \varepsilon_p}$ (the space charge constant),

$f_a = \frac{v_a}{2\pi h \varepsilon'_p}$ ($\varepsilon'_p = 50.2$ for $LiNbO_3$),

d = the semiconductor junction's thickness, /80

h = the distance between the $LiNbO_3$ and Si layers,

$\frac{\Delta v_a}{v_a}$ = an interaction parameter, dependent upon K^2 (= o.o246).

This formula, which applies when $f << f_D$ and d and h are small in comparison to the Rayleigh wave length, shows that the amplification will diminish rapidly if $f > f_a$ (and C_s^2 is large). Since h can hardly be made smaller than 200 A, with $f_{amax} \approx$ 0.6 GHz, $b_a = 3.5 \cdot 10^5$ cm/s. This entails difficulties in reaching high amplification at frequencies over 1 GHz. The amplification around 0.3 GHz is approximately 10 dB/cm at h = 1,580A and 60 dB/cm at h = 560A . The surface-wave amplifier's band width,

saturation power and efficiency are about as large as with a volume amplifier when the converters are not taken into account. Again, without considering the effect of the interdigital converter, the surface-wave amplifier's band width is comparable to the volume amplifier's. As already mentioned, the converter's band width depends upon the conversion losses, so that an increased number of fingers diminishes both the losses and the band width. Judging from results obtained to date, a surface-wave amplifier obtains at most twice as large a band width as a volume amplifier at the same frequency and with the same conversion losses. The biggest advantage with the surface-wave amplifier is that power loss can easily be avoided and, in this manner, it yields itself to CW operations. Efficiency levels over 10% have been obtained experimentally with low surface-wave amplification of approximately 10 dB.

An evaluation of the electroacoustic amplifiers usefulness and especially its future significance is relatively complicated. Since the electroacoustic amplifier covers about the same frequency range as transistor amplifiers, it must be mainly compared with them. The most characteristic properties of the electroacoustic amplifier are its simple construction and small dimensions. It functions approximately the same as a multi-stage transistor amplifier containing a minimum of 50 circuit elements [4]. This means that the electroacoustic amplifier, especially /81
those of the surface-wave type, could at low frequencies compete very strongly as an alternative to the transistor amplifier, despite the fact that through mass-production and the use of concentrated integrated circuits it can be both small and inexpensive. However, with the need for high CW power and efficiency, coupled with low-level amplification within a very wide band, the transistor amplifier should remain dominant even at very high frequencies (larger than 3 GHz).

There is, however, a special area in which the electroacoustic amplifier is and will continue to be very significant. This is in connection with acoustical delay lines for high-frequency signals. The previously discussed characteristics of a piezoelectric material are such that a 1 cm long crystal delays a high-frequency signal as much as a 100 m long coaxial cable; in other words, approximately 1 μs. At 1 GHz, the damping in such a cable is about 20 dB, while the damping in the crystal is only 1 dB, to which the conversion loss of 10 to 20 dB must then be added. In such a delay line, consisting of a piezoelectric crystal and an electroacoustic converter, there are already several of the main elements of an electroacoustic amplifier. In this way, they lend themselves to easy integration in the structure. The amplifier easily compensates both damping in the converter and in the crystal. Especially in connection with large delay (high crystal damping), the internal amplifier can be of great value since it reduces the dynamic range of the delay lines insignificantly, while compen-

sation by an external amplifier diminishes the dynamic range by the value of the crystal damping. Furthermore, a very important function of the internal amplifier is to reduce the influence of reflection in the converter. It has been shown in a passive delay line to be difficult to maintain the signal, which is first reflected at the output point and then at the input point, 20 dB lower than the primary output signal. With acoustic amplification, the electron speed can be reduced so that the reverse damping becomes much larger than the amplification, so that the twice-reflected signal is at least 40 dB lower than the primary signal.

It is believed that the surface-wave type of electroacoustic /82
amplifier will be most significant in the future because of its large flexibility in terms of input and output coupling of signals and because of suitable manufacturing methods, which create good possibilities for its adaption to other integrated circuits produced with planar processing. The highest possible frequency limit for this type of amplifier depends partly on the size of the converter's interdigital structure, and partly on the possibility of creating an effective coupling between the (probably necessary) separate layers for electroacoustic and electron conductants. The best conceivable method for producing very fine structures, pattern-definition using an electron beam in the electronresist process, has already been mentioned [2]. The method is described more completely in [5], where it is pointed out that the electron beam should have a diameter which is 4 to 5 times less than the smallest dimension in the structure being produced. Since an electron beam can be focused to a diameter approaching 25A , this should mean that the production of fingers with a width of 125A will be possible. However, other problems, such as resistance inhomogeneity, making the utilization of this solution impossible. Therefore, if one assumes that the width (and distance) of the smallest possible finger will be 500A , it will be possible to excite a Rayleigh wave-length of 0.2μ , which corresponds to a frequency of 17 GHz. According to earlier discussions, the surface-wave amplifier with separate crystals for electroacoustic and electron conductants can hardly function over 1 GHz due to poor interaction through the necessary air space between the crystals. A way to overcome this frequency limitation is suggested in [6]. The appropriate conductivity is given to a thin surface layer on the piezoelectric crystal using ion implantation. Maximum effectiveness is then achieved in the coupling without damping the acoustic wave by any large bodies in the conducting crystal (or its substrate). If ion implantation is carried out in several adjacent square surfaces instead of one continuous rectangular surface, the electrons can be accelerated to the appropriate speed with significantly less voltage. Voltage is fed to the square areas in parallel and the acoustic wave can be said to be amplified in several stages. An added advantage is that power loss would be reduced.

Through the development of the technological methods outlined here, it seems feasible that the frequency range of the surface-wave type of electroacoustic amplifier will be expanded further upward to the X-band around the year 1985. The amplifiers will be mainly used in connection with acoustic delay lines and have up to 40% bandwidth and 70 dB amplification. For separate amplification functions, except for extreme miniaturization at relatively low frequencies, the transistor will maintain its advantage over the electroacoustic amplifier. The relatively high conversion losses and low efficiency in the electroacoustic amplifier mean that the transistor can produce both a lower noise factor and higher output power. /83

REFERENCES

1. White, Voltmer, "Direct piezoelectric coupling to surface elastic waves," Appl. Phys. Lett., p. 314-16, (December 1965).

2. Lean, Broers, "S-band surface acoustic delay lines," ISSCC, (1970).

3. Kino, Reeder, "Microwave acoustic amplifiers," Microwave Journal, p. 79-85, (March 1970).

4. Dobriner, R., "Electroacoustic amplifiers about to leave lab?" Electr. Design, p. 17-18, (April 1966).

5. Lean, Broers, "Microwave surface acoustic delay lines," Microwave Journal, p. 97-101, (March 1970).

6. Yoder, M. M., "Systems exploitation of the multifunctional properties of praetersonics technology," Proc. 1970, Electronics Components Conference, p. 455-469, (1970).

Other Semiconductor Oscillators

The Tunnel Diode Oscillator

The interesting possibility of electronically utilizing the /84
quantum-mechanical tunnel effect was discovered in 1958 by Leo Esaki [1]. His experiment showed that a semiconductor diode with very high doping (concentration of impurities) in both the p and n layer, displayed a negative dynamic resistance within a small area of forward bias. The high level of doping in such a diode (10^{19} to 10^{20} doped atoms per cm^3, as opposed to 10^{16} to 10^{17} in a normal diode), means that the overlay is extremely thin, on the order 100 A. Since it is possible for the electrons to penetrate (tunnel) through such a thin potential barrier without having sufficient energy to climb over it, the appearance of the tunnel diodes I-V characteristic can be easily explained. Figure 13 shows schematically the energy levels and the size of the currents in a tunnel diode with varied bias. In Figure 13-A, there is zero bias, and since the electron density is as large in the valence bands on the p-side as in the conduction bands on the corresponding level of the n-side, two small, equally large tunnelling currents flow in opposite directions across the abrupt junction. The net current is therefore zero. With reverse bias (Figure 13-B), the tunnelling current from the n-side remains unchanged, while the current from the p-side increases greatly, since the number of empty energy levels which can receive electrons from the p-side increases rapidly with negative voltage. The result is then that reverse current increases rapidly with reverse bias. With increasing forward bias, an increasing number of electrons will tunnel from the n-side to the empty levels in the valence bands on the p-side, until a maximum current is obtained at the energy level shown in Figure 13-C. If the bias is increased still further, the tunnelling current, lacking empty energy levels on the p-side, will decrease to nearly zero, as shown in Figure 13-D. At the same time, however, normal diffusion current will begin to flow, since the electrons on the n-side have sufficient thermal energy to climb over the potential barrier. Figure 13-E shows the I-V characteristic which is obtained through the process discussed above, and the letters a-d correspond to the /85
current configurations in Figures 13 A-D. Figure 13-E also shows some of the I-V curve's most characteristic points: peak tunnelling current I_p, valley current I_v, and the corresponding voltages V_p and V_v. The negative dynamic resistance which the tunnel diode exhibits between V_p and V_v can naturally be utilized for constructing oscillators and reflection amplifiers, and since the negative resistance is practically independent of the transit-time phenomenon, one should expect it to be used at extremely high frequencies. The upper frequency boundary will, however, be dependent upon the junction capacitance C_j, which bypasses the negative resistance R_j (absolute value), and the unavoidable /86
series resistance R_s in the semiconductor. The total series

resistance is then zero at the resistive cutoff frequency f_r, which is easily calculated as

$$f_r = \frac{(R_j/R_S - 1)^{\frac{1}{2}}}{2\pi R_j C_j} .$$

Above f_r, the diodes total resistance is positive and no oscillations can occur.

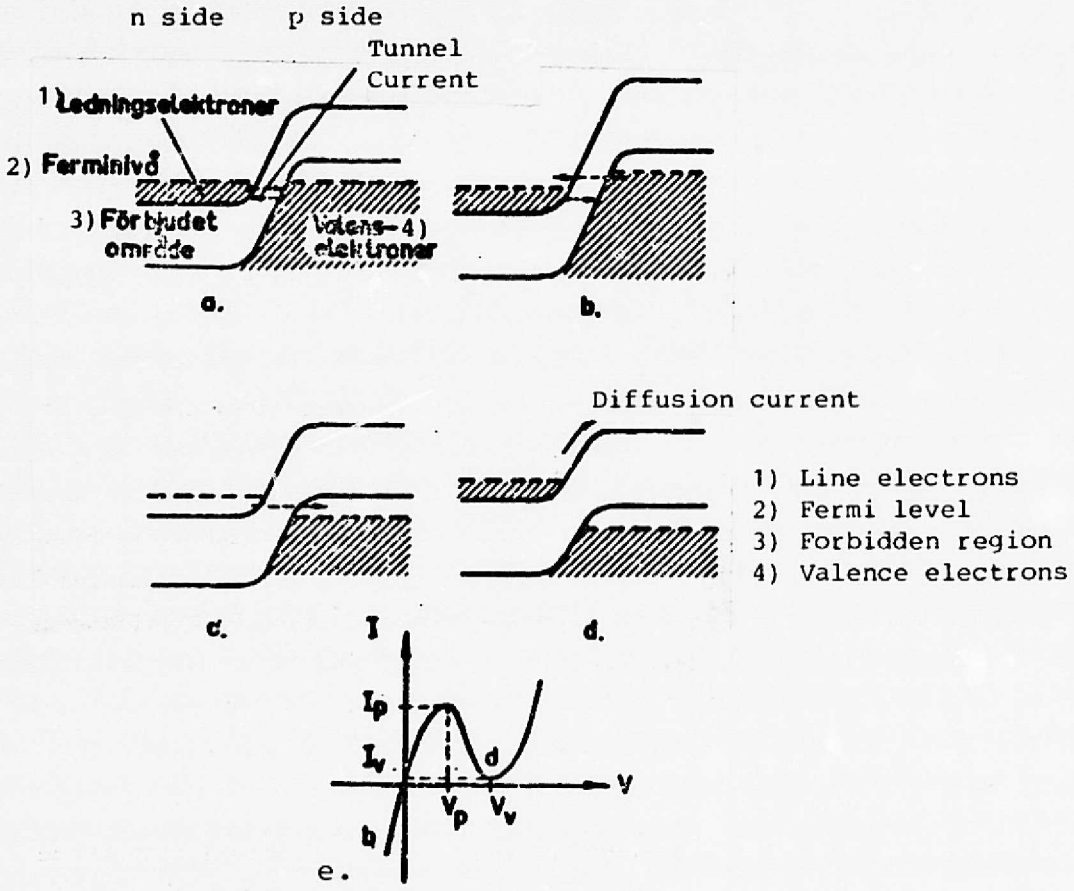


FIGURE 13A-E

An enclosed diode is also characterized by its self-resonant frequency f_s, at which the lead resistance L_s resonates with the capacitive junction. The equivalent circuit for the tunnel diode is shown in Figure 14, in which the diffusion capacitance C_s is drawn with a broken line since it is usually practical to include it in the external circuit. When C_s is ignored, the series reconant frequency becomes

$$f_s = \frac{\sqrt{(R_j^2 C_j/L_s) - 1}}{2\pi R_j C_j} .$$

ORIGINAL PAGE IS
OF POOR QUALITY

The maximum output power from a tunnel diode oscillator can, according to [2], be estimated from the expression

$$P_{max} = \frac{3}{16}(I_p - I_v)(V_v - V_p) \approx 0{,}2\, I_p(V_v - V_p),$$

which is a good approximation for frequencies up to 0.3 f_r; in other words, up to frequencies where the diode's parasitic elements C_j and R_s have little effect. To obtain the highest possible output power, I_p and $V_v - V_p$ must be increased. However, with reference to the discussion in connection with Figure 13, the conclusion can be drawn that $V_v - V_p$ must be significantly less than the semiconductor's bandgap, which for Ge, GaSb, Si and GaAs is respectively 0.67, 0.68, 1.1 and 1.4 V. The only way to produce a larger increase in power is then to increase I_p.

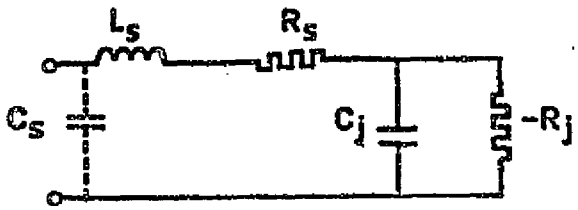


FIGURE 14

According to [2], the current density with peak current is /87

$$J_p \sim \frac{\sqrt{n}}{e^{\frac{\text{const.}}{\sqrt{n}}}},$$

where n = the number of doping atoms per cm^3.

If the junction's cross-sectional surface is A, then

$$I_p = J_p A \sim \frac{A\sqrt{n}}{e^{\frac{\text{const}}{\sqrt{n}}}}.$$

In this manner, I_p can be increased with A and n, but the way in which this will effect the diode's other properties must be investigated. Changes in R_j and C_j, which can influence f_r, are of special interest. It can be written that

$$R_j \sim \frac{1}{I_p} \sim \frac{e^{\frac{\text{const}}{\sqrt{n}}}}{A\sqrt{n}};$$

$$C_j = A\sqrt{\frac{een}{2(\phi_c - V_j)}},$$

ORIGINAL PAGE IS
OF POOR QUALITY

where ϕ_c = the contact potential,
and V_j = the forward bias.

In this way, $R_jC_j \sim e^{\frac{\text{Const}}{\sqrt{n}}}$ and is independent of A. R_j/R_s, which is also found in the expression for the resistive cutoff frequency, will decrease slowly as A increases (both R_j and R_s decrease). f_r decreases slowly with increasing A, but increases with increasing n (when $R_j >> R_s$). Therefore, if high power at high frequency

is desired, A can be increased somewhat, but the doping should be increased first. However, increases in n (or A) cannot be too large since both R_j, R_S, and their difference will be very small and then the external load resistance must also be made extremely small, and this is particularly difficult at high frequencies. According to [3], the highest obtainable load is proportionate to $\sqrt{f}$. /88

It is pointed out in [2] that it is very difficult to build a high-frequency oscillator for frequencies above the diode's self-resonant frequency. This means that the way in which the changes in A and n discussed above can influence f_s should be investigated. It is then found that f_s decreases as A and n increase, and that this can only be counteracted by reducing L_s.

P_{max}, given earlier as the maximum output power from a tunnel diode oscillator, applies as indicated only at frequencies under f_r, the frequency at which oscillations cease and therefore output power approaches zero. The way in which output power decreases with frequency with regard to the effects of C_j and R_s can be calculated easily. If the power given by the load resistance R_L is expressed as P_L, then

$$\frac{P_L}{P_{max}} = \frac{R_L}{R_L + R_s}.$$

According to [2] and [4], the effective negative resistance which produces P_{max} is

$$R_D = 2R_j.$$

Since R_D is bypassed by C_j, the real negative series resistance will be

$$\frac{R_D}{1 + w^2C_j^2R_D^2}$$

and with stable oscillations, it must have the same size as the loss resistance; in other words, /89

$$R_L + R_S = R_D/(1 + w^2C_j^2R_D^2).$$

For osciallation at frequency w_r, it is necessary that

$$R_S = \frac{R_D}{1 + w_r^2C_j^2R_D^2}.$$

These expressions for R_L and R_S yield

$$\frac{P_L}{P_{max}} = 1 - \frac{1 + w^2C_j^2R_D^2}{1 + w_r^2C_j^2R_D^2} = 1 - \frac{1 + 4w^2C_j^2R_j^2}{1 + 4w_r^2C_j^2R_j^2}$$

From this it is concluded that the output power is very close to the maximum level at frequencies far below f_r, but decreases rapidly near f_r.

According to the calculations in [5], the efficiency for a tunnel diode oscillator with an optimized working point and load is close to 30% (excluding the losses in the stabilizing resistance).

To summarize, the tunnel diode oscillator's output power increases with the diode surface A, and the doping n. Increasing the surface, however, results in a lower cutoff frequency and increasing the doping quickly leads to the need for a very low load resistance. At the same time, the inductance in the diode's leads must be reduced for practical reasons when A and n increase. When the oscillation frequency nears the cutoff frequency f_r, output power decreases rapidly.

The best results obtained to date, according to [3], are shown in Table 4.

Table 4. Results reached with TD oscillators.

Frequency GHz	Power Output mW	Efficiency %	Material
5	9	-	p-GaAs
10	2	2	p-GaAs
50	0.2	-	n-GaAs

ORIGINAL PAGE IS OF POOR QUALITY

The results given here are from 1964, and since then no improvements have been observed, probably because the research is completely directed toward other types of diode oscillators. Continued development of the tunnel diode will probably produce certain improvements using very highly-doped GaAs. The quantitative results of such development is hard to forecast, but 10 to 20 mW in the X-band and several tenths mW at 100 GHz could possibly be generated. However, the most important result of this appraisal is that the tunnel diode, like the power oscillator, will not be in a position to compete with other diode oscillators. Using series combinations of several highly-doped tunnel diodes, the impedance level can be raised, and this would, of course, increase the output power at high frequencies, as long as the increased series inductance does not create any problems. Both series and series-parallel combinations of tunnel diodes have already been tested [6]. /90

It should be mentioned that among the tunnel diode oscillator's other characteristics is thermal instability. This is due to the fact that the junction capacitance varies with temperature. Aside from direct temperature stabilization, this can be counteracted by weakly coupling the diode to a frequency-determined high-Q circuit. This, however, partially decouples the available negative resistance, which reduces its utility at very high frequencies. In this way, a stability of - 1 MHz/°C is normally

obtained in a 1 mW, S-band oscillator. Information on the noise characteristics of the tunnel diode oscillator does not exist at present, but the FM noise for a X-band oscillator with output power of 0.1 mW is given in [7] as 3 Hz measured within 1 kHz and 70 kHz distance from the carrier wave. The FM noise is, therefore, somewhat lower than in a good Gunn oscillator.

REFERENCES

1. Esaki, L., "A new phenomenon in narrow germanium p-n junctions," Phys. Rev. vol 109, p. 603, (1958). /91

2. Sterzer, F., "Tunnel diode devices," Advances in Microwaves, Academic Press, p. 1-41, (1967).

3. De Loach, B., "Recent advances in solid state microwave generators," Advances in Microwaves, Academic Press, p. 67, (1967).

4. Chow, W. F., Principles of tunnel diode circuits, John Wiley and Sons, Inc, p. 170, (1964).

5. Degen, W., et. al., "Berechnung der Schwingungsamplitude und der Leistung eines Tunneldiodenoszillators," AEU, p. 569-575, (September 1964).

6. Thomson, G.W., "Microwave oscillators using series and series-parallel combinations of tunnel-diodes," Proc. IEEE, p. 535, (May 1965).

7. International Microwave Corp., "Diode oscillators feature low noise," Electronics, p. 180, (November 1966).

The Josephson Oscillator

In 1962, B. D. Josephson did a theoretical examination of /92
the tunnelling effect in a superconductor [1] and showed the possibility of developing several new technically useful phenomenon. The quantum mechanical background to these phenomena is very complicated, but they result in the so-called Josephson junction producing in part a finite direct current without the application of direct-current voltage and in part in an alternating current with the application of direct-current voltage, V_o, the frequency of which is determined by the equality

$$f = \frac{2eV_o}{h} = 483{,}6 \text{ MHz}/\mu\text{V},$$

where e = the electron charge
and h = Planck's constant.

In principle, the Josephson junction consists of two very thin superconductive metal surfaces, separated by an extremely thin isolating layer (approximately 10A), which often takes the form of an oxide on one of the metal surfaces. The electrodes can be made, for example, from the metals Pb, Sn, Nb, or an alloy such as Pb-Bi. The pair of electrons, so-called Cooper pairs, which produce the superconductivity, will break apart at a certain voltage, whose value depends upon the characteristics of the metals and temperature. This voltage determines, according to [2], the upper cutoff frequency for the Josephson oscillator, and yields as a maximum value for Sn, Pb and Pb-Bi 350, 650 and 1,000 GHz, respectively. An experimental investigation [3] with Nb shows, however, that a voltage ten times greater than that which will separate the Cooper pair will still yield oscillations at a frequency which is, therefore, more than ten time greater than the maximum frequencies given in [2], in this case around 8,000 GHz. Aside from the fundamental tone, the Josephson junction is also an effective harmonic generator due to its non-linear I-V characteristic. Despite the fact that the oscillation frequency is in principle continually variable with voltage, output power is usually observed at certain discrete voltage (and frequency) levels, which is connection with the junction itself creating a resonance circuit. As a result of the low voltage values (20 μV for the X-band and approximately 2 mV for 100 GHz), the output power from the Josephson oscillator must be very low. The feeding current, /93
which normally lies around 10 or more mA, should certainly be able to be increased in order to increase the power, but the adaption to the external load would then be even more difficult to realize. It is estimated, for example, in [2] that the output power, because of poor adaption, will decrease by a factor of 10^{-4} in a "normal" Josephson junction, which is 250 μ wide and 10 A thick. (The junction then is regarded as a band lead which radiates out in the free space.) The power radiated at 10 GHz is estimated at

ORIGINAL PAGE IS
OF POOR QUALITY

$5 \cdot 10^{-12}$W. By reducing the junction's width to 2.5 μ, the impedance in the band lead would increase 100 times to approximately 1 ohm. If it is assumed that this impedance could be transformed in a loss-free manner to antenna impedance, and that the current could remain constant by, for example, increasing the length of the junction, the output could then be increased to $5 \cdot 10^{-12}$W, or negative 63 dBm. These approximations show that the output power from a Josephson junction is very low. This type of oscillator is still of very great interest since it is the only solid-state device which can generate sub-millimeter waves. The low power means that its most interesting application will be as a local oscillator for mm and sub-mm waves (up to 10,000 GHz), at which point the Josephson junction would function simultaneously as a mixer.

REFERENCE

1. Josephson, B. D. "Possible new effects in superconductive tunneling," Phys. Letters., p. 251-253, (July 1962).

2. Langenberg, Scalapino, Taylor, "Josephson-type superconducting tunnel-junctions as generators of microwave and submillimeter wave radiation," Proc. IEEE, p. 560,(April 1966).

3. McDonald, D. G., et. al., "Harmonic generation and submillimeter mixing with the Josephson effect," Appl. Phys. Lett., p. 121, (August 1969).

Microwave Generation with Indium Antimonide

Since it was discovered around the year 1964 that indium /94
antimonide under certain conditions could emit microwave radiation, very extensive work, on both the experimental and the theoretical plane, has gone on without producing any results of practical value. An explanation to the phenomenon (or phenomena) is naturally of great interest to the field of physics, but since a definite explanation is still lacking, it can be judged as very unlikely that microwave generation using InSb will produce any advantages in comparison to other generation methods and semiconductor devices. In the interest of completeness, however, a very short summary of the experiments done and the results obtained to date is given, based on [1], which itself includes 93 references.

When an electric field is applied to n or p-type cooled NiSb, low-power electromagnetic radiation is emitted. Most of the experiments have been done at 77°K, but investigations at temperatures from 2.4 to 250° K have been recorded. Radiation occurs with electric fields as low as several volts/cm, at which point the output power is on the order of 10^{-9} W. With an electric field of 100 V/cm or greater, the output power increases to around 10^{-6}W.

The measured frequencies range from several MHz, at which point they appear as an oscillation in the voltage-feeding circuit, and within the microwave range up to approximately 100 GHz, appearing in the form of directly emitted microwave energy. The radiation is emitted in the form of a wide spectrum, usually extending across all of the frequency range mentioned, with an energy distribution which is approximately inversely proportionate to the frequency. Some form of resonance which would lead to a relatively narrow-band spectrum is thought to have been detected in only a few experiments. It should be noted in this context that the methods used to date for the output-coupling of microwave energy have probably been highly ineffective. It is possible that a better understanding of the generation mechanism can lead to the construction of much better output-couplings, but in spite of this, the level of output power will probably remain low in comparison to other solid-state generators. In a number of experiments, the /95
indium antimonide is also subjected to an external magnetic field, at right angles or parallel to the electric field. The direction and size of the magnetic field changes the level of output power and the threshold level (at which point oscillations begin) of the electric field. However, this will probably not be of fundamental significance to the generation mechanism.

Among the theories which have been proposed to explain the emission of electromagnetic radiation from InSb, the following are thought at present to be the most likely: interaction between the charges and acoustic waves (phonons) can produce amplification of thermal noise. The mechanism which then out-couples the high-

frequency energy is at present unclear. In connection with high (local) electric fields leading to the creation of electron holes, initiated instabilities can be amplified by collisions of -charge waves in the electron-hole plasma. Continued research will probably reveal that several more processes are of significance in explaining all of the experimental results within this complex area.

Comprehensive Forecast for Power Generating Solid-State Devices

The earlier deduced forecast curves for the various solid-state devices, in which output power is given as a function of frequency, have been combined in Figure 15. In terms of CW power, the transistor will dominate up to several GHz. In ranges from 2 to 20 GHz, the TRAPATT-mode avalanche diode oscillator will be the best alternative when both high power and efficiency is needed. The Gunn, or LSA, and the ATT diode oscillator will be very significant within ranges 20 to 200 GHz, even if the frequency multiplier yields about the same power level. The total efficiency of a power oscillator followed by a multiplier is still much lower than that of a diode oscillator. Furthermore, when the multiplier's greater complexity and cost is considered, it becomes clear that it cannot compete with the diode oscillator. In connection with frequency and phase locking of diode oscillators, frequency multipliers for low output power and a high multiplication factor at every varactor stage can be of continuing significance. It is conceivable that the tunnel diode oscillator will be usable /96 in special circumstances demanding low-power consummation. Lastly, the Josephson oscillator creates the possibility to cover the submillimeter range, but only with very low output power (approximately 10^{-10} W).

Frequency GHz	Maximum CW power W	Efficiency %
0.1	1000	50-70
1	100	50-70
10	30	50
100	1	10
200	0.1	1
300	0.01	0.1
1000	10^{-10}	-
10000		

Figure 5. CW powers from solid state devices 1985.

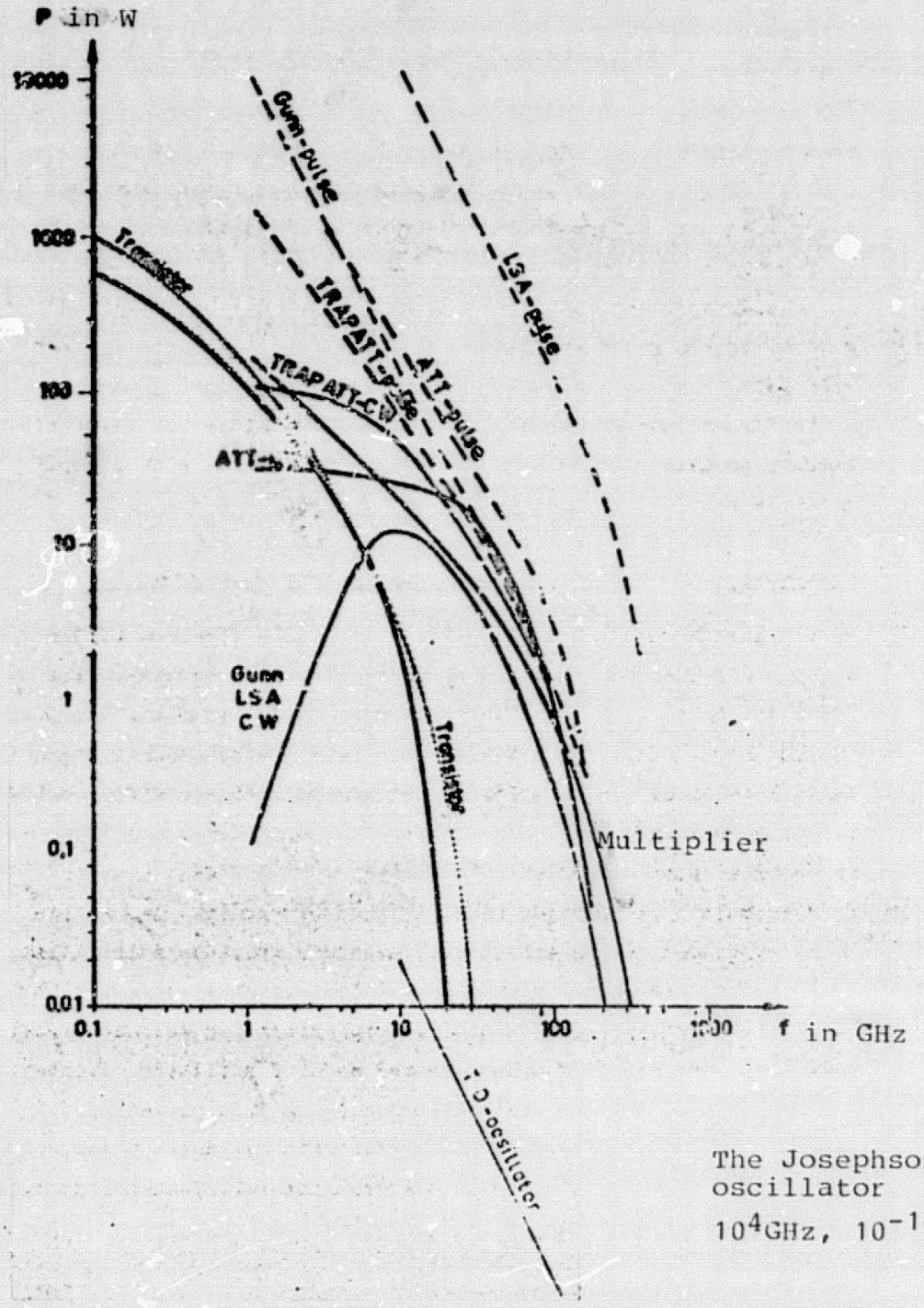
In terms of pulse operation, approximately the same power level for both the Gunn, ATT, and TRAPATT diodes can be expected, with the latter having the advantage with the highest efficiency. The highest level of output power is obtained from the LSA diode, but only in connection with very short pulses and a low utilization factor. /97

The maximum CW powers to be expected in 1985 are summarized in Table 5, in which approximations of the

ORIGINAL PAGE IS
OF POOR QUALITY

appropriate efficiency levels are also shown.

FIGURE 15



The Josephson oscillator

10^4GHz, 10^{-10}W

ORIGINAL PAGE IS
OF POOR QUALITY

REFERENCES

1. Glicksman, M., "Summary of microwave emission from InSb: Gross features and possible explanations," IBM Journal Res. Develop., p. 626, (September, 1969).

Small Signal Amplification

The Low Noise Transistor

The highest possible power amplification and lowest noise factor are sought in a small signal amplifier. The methods to maximize the amplification have already been discussed in connection with the power transistor. Ignoring the parasitic reactances, the noise factor for a high-frequency transistor's upper frequency range can be written according to Fukui [1] as: /98

$$F = 1 + \frac{r_b}{Rg} + \frac{kT}{2eI_cRg} + \frac{eI_c(r_b + Rg)^2}{2kTRg}\left(\frac{f}{f_T}\right)^2,$$

where r_b = the base's diffusion resistance.

R_g = the source resistance,

$\frac{kT}{eI_c} = \frac{\text{the differential emitter resistance}}{\alpha_o}$,

α_o = the low frequency value of the current amplification with a common-base coupling,

k = Boltzmann's constant,

T = the absolute temperature,

e = the electron charge,

and I_c = the collector direct current.

In the second term, r_b corresponds to the addition of purely thermal noise from the base resistance. The third term represents the fluctuation noise in the current at the emitter-base and collector-base junctions (equivalent to a thermal noise, according to Van der Ziel). The last term represents a current-distribution noise with pure distribution-noise characteristics. It is normally attributed to the collector [2].

When $f > 0.5\ f_T$, the minimum value of the noise factor, still according to Fukui [1], can be approximated as:

$$F_{min} \approx 1 + h\left(1 + \sqrt{1 + \frac{2}{h}}\right), \text{ där } h = \frac{eI_cr_b}{kT}\left(\frac{f}{f_T}\right)^2.$$

For an experimental-type microwave transistor, L 78 B, F = 6 dB at 3 GHz, when f_T = 2.5 GHz. This yields $h = 0.8\left(\frac{f}{f_T}\right)^2$. If it is assumed for estimation that r_b can be reduced so that the factor of 0.8 is lowered to 0.5 by 1985, and the maximum value of f_T is then 15 GHz, the lowest possible F-value can be calculated as a function of the frequency. (For other low-noise transistors in 1969, the factor is often between 1 and 2.) In particular, the transition from Si to other semiconductors with higher mobility will decrease r_b and increase f_T. The improvements between /99

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

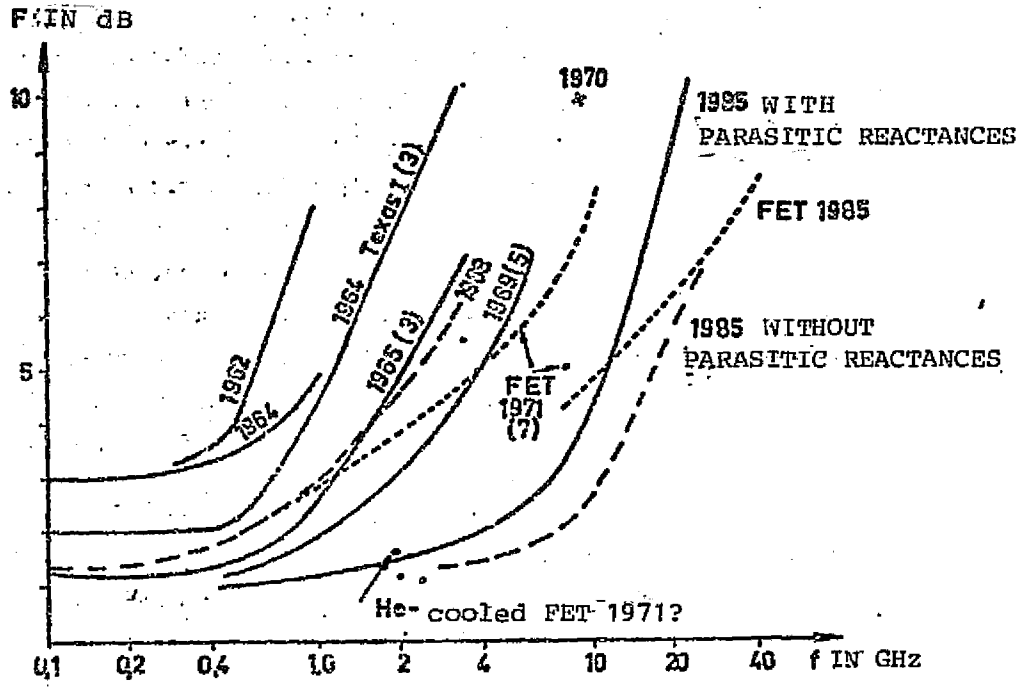


FIGURE 16

1962 and 1969 in the transistor's noise factor is shown in Figure 16 as a function of frequency. The lowest possible noise factor for 1985, calculated according to the previous assumptions, is drawn as a broken curve marked "1985 without parasitic reactances."

It is clear from [1] that parasitic reactance at present will be significant in the S-band, where it will increase F approxi- /100
mately 1 dB (several dB in the upper S-band). With future improvements in mounting (combined with integrated circuits) it can be assumed that the parasitic reactance will increase F by approximately 0.5 dB in the S-band and approximately 2 dB in the X-band. These corrections are taken into account in the forecast curve marked "1985 with parasitic reactances."

With $f_T \approx 15$ GHz, as assumed above, the low-noise transistor amplifier with regard to the feasible amplification per stage can be expected to be useful up to frequencies between 15 and 20 GHz. This evaluation is most applicable to normal bipolar transistors. However, it can be of interest to note how refined manufacturing techniques have recently led to the development of field-effect transistors with Schottky-contact modulation at very high f_{max}. One such transistor, made from GaAs, is reported in [6] to have $f_{max} = 30$ GHz and 3 dB amplification at 17 GHz.

In [7], f_{max} values of 40 to 60 GHz are reported. This should make it possible to built an amplifier up to 30 GHz. The noise factor over 4 GHz is at present somewhat lower in a field-effect transistor than in a bipolar transistor. (According to undocumented information, in 1971, a helium-cooled field-effect transistor exhibited a noise factor 1.5 dB at 2 GHz.) The noise factor results given in [7] are shown in Figure 16. The curve is for a so-called MOSFET of silicon, while the best single value, 5 dB at 8 GHz, was probably obtained with GaAs. If the field-effect transistor is improved by 1985 about as much as the bi-polar transistor, the dotted curve shown in Figure 15 will result. This shows that the field-effect transistor can be very significant at frequencies over 10 GHz.

In a transistor amplifier, the output level at 1 dB gain compression is typically around +10 dBm, and the corresponding dynamic range (calculated within 1 MHz) is typically 90 dB.

Integrated-types of low-noise amplifiers already exist which produce up to 50 dB amplification within a range of 0.1 to 2 GHz with a maximum output power of several mW and a noise factor around 6.5 dB. An amplifier for the 1 to 4 GHz range with 30 dB amplification and a noise factor 8 to 10 dB is also commercially available. The outer dimensions of these amplifiers are less than 10 x 5 x 5 cm. The broad-band techniques utilized here probably cannot be improved too much more, but will be utilized more often at ever-higher frequencies as the transistors are improved. This /101
will make possible around 1985 the building of a compact unit consisting of 4 or 5 integrated amplifiers, which will cover the frequency range from 0.1 to 20 GHz and have significantly less volume than a normal travelling-wave tube.

REFERENCES

1. Fukui, H., "The noise performance of microwave transistors," TIE^3 ED, p. 329. (March 1966).

2. Nielsen, E. G., "Behavior of noise figure in junction transistors," PIRE, p. 957, (July 1957).

3. Small, T., "UHF and microwave devices, present-future," TI report, (1965).

4. Kotyczka, Strutt, "Noise measurements of Si planar microwave transistors in the frequency range 4 - 8 GHz," Electronics Letters, p. 478-79, (July 1970).

5. "Fundamental transistor oscillator invades X-band," Microwaves, p. 74, (December 1969).

6. Drangeid, et. al., "High-speed GaAs, SB field-effect transistors," Electronic Letters, p. 228, (April 1970).

7. Davis, R. T., "Front end designs, assaulting old noise barriers," Microwaves, p. 32, (April 1971).

Parametric Amplifiers

The principles of the so-called parametric amplifier, in- /102
stabilities in non-linear energy-inhibiting systems, have been known for about 100 years. In 1936, Hartley described a device with a mechanically variable capacitance which was very close to a parametric amplifier. The first experiments with parametric-type oscillators and amplifiers were done at the end of the 1940's. In these, the non-linear element was a coil with a ferrous center. In 1948, Van der Ziel [1] described the mixing qualities of non-linear capacitance and also pointed out the possibility of low-noise amplification. The first parametric microwave amplifier, constructed in 1957, utilizied a ferrite as the non-linear reactance [1, 3], but a year later, the first experiments were done with a semiconductor diode as the active element [4]. The subsequent development concentrated completed on utilizing semi-conductor diodes in which the capacitance varies with the applied voltage, so-called varactor diodes.

The function of the parametric amplifier depends upon bringing a high-frequency, so-called pumping power to the varactor diode at a pumping frequency f_3, which is larger than the signal frequency f_1. The varactor's non-linear reactance then produces the mixing frequencies $f_3 \pm f_1$. If the varactor is coupled to two resonance circuits, one for f_1 and one for the so-called no-load frequency $f_2 = f_3 - f_1$, high-frequency energy will be transferred from the pumping source and result in a negative resistance in both the signal and the no-load circuits. In a normal parametric amplifier, both the signal source and the load are coupled over a circulator to the signal circuit while the no-load circuit usually has no load. The amplified signal can also be coupled out to a load connected to the no-load circuit. In this case, the amplifier is called a lower side-band tranducer. In both these cases, the amplification with varying load can be increased to very high values and convert to oscillation. In a third type of amplifier, the signal is coupled out through an upper side-band circuit, $f_3 + f_1$. This is called an upper side-band converter and can have maximum amplification of $(f_3 + f_1)/f_1$. Despite is unquestionable stability and low internal noise, it has not been utilized as a /103
low-noise amplifier for two reasons: the amplification is usually too low, and the following mixer exhibits worsened properties because it must work at a higher frequency than the signal. The meaning of high amplification, G_1, and the effect of the noise factors of the low-noise amplifier, F_1, and its following stages, F_2, are given in the following expression for the total noise factor:

$$F_T = F_1 + \frac{F_2 - 1}{G_1}.$$

The latest development of the parametric amplifier has been completely dominated by the circulator-coupled type, in which the amplified signal has the same frequency as the feeding signal. An amplification G_1 = 20 dB is easily obtained while maintaining good stability. It is possible to use this amplifier without a circulator. In this case, separate input and output couplings to the signal circuit are used, but because very large amplification variations result from very small changes in the signal source's impedance, the stability becomes very bad. For that reason, this report will concentrate completely on the circulator-coupled type.

A simplified equivalent circuit for this is shown in Figure 17-A, in which BPF stands for band-pass filter for the given frequency. The varactor is shown without parasitic elements (these can be thought to be included in the outer circuits) and is characterized by the average capacitance $C_D = 1/S_D$ produced with applied pumping power, and a change in capacitance of the amplitude $2C_1 = 1/2S_1$ with the pumping frequency. The varactor's series resistance is R_S and its dynamic characteristics are usually characterized by the cut-off frequency $f_c = 1/2\pi R_S C_o$ (where Q = 1) and $2S_1/S_o$ (or, alternatively, $2C_1/C_o$), or by the product $Q_n = (S_1/2S_o)(f_c/f_n) = m(f_c/f_n) = m/\omega_n C R_S$, in which m is, therefore, the modulation ratio.

A calculation of the varactor's equivalent impedance at frequency f_1 shows that it corresponds to the series resistance R_S and an impedance of

$$-\frac{(mf_c R_s)^2}{f_1 f_2 (R_s + R_{L2} + R_2 - jX_2)} \qquad /104$$

which with the resonance $X_2 = 0$, becomes a purely negative resistance. The signal circuit's equivalent circuit is shown in Figure 17-B in which all thermal-noise generators are also drawn. (However, the circuit losses, represented by R_1 and R_2, are assumed to be so small that their noise factor can be ignored.) In Figure 17-B, v_v represents the noise which is transformed from the no-load circuit by via the varactor into the signal circuit. T_o is the standard temperature in the generator-resistance (290°K), T_d is the varactor temperature and T_2 is the load temperature in the no-load circuit.

With the help of figure 17-B, the reflection amplication can be written

$$G = \left| \frac{R_G - \left(R_s + R_1 + jX_1 - \dfrac{m^2}{\omega_1 \omega_2 C_o^2 (R_{T_2} - jX_2)} \right)}{R_G + \left(R_s + R_1 + jX_1 - \dfrac{m^2}{\omega_1 \omega_2 C_o^2 (R_{T_2} - jX_2)} \right)} \right|^2$$

ORIGINAL PAGE IS
OF POOR QUALITY

With resonarce and high amplification, $X_1 = X_2 = 0$, and:

$$R_G + R_s \simeq \frac{m^2}{\omega_1 \omega_2 C_o^2 (R_{T_2} - jX_2)}$$

$$\therefore G \simeq \frac{4R_G^2}{\left(R_{T_1} - \frac{m^2}{\omega_1 \omega_2 C_o^2 R_{T_2}}\right)^2}$$

The noise factor $F = 1 + \frac{T_e}{T_o}$ is calculated from:

$$F = \frac{\text{Total noise in the output}}{G(\text{Thermal noise in the input})} = \frac{N_{UT}}{GkT_oB}$$

From Figure 17, with high amplification and when $T_2 = T_d$: /105

$$F \simeq 1 + \frac{T_d}{T_o}\left(\frac{R_s}{R_G} + \frac{\omega_1}{\omega_2}\frac{R_{T_1}}{R_G}\right).$$

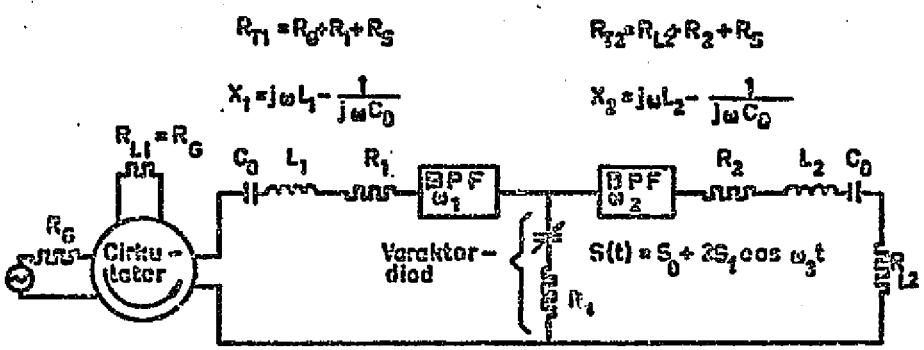


FIGURE 17A

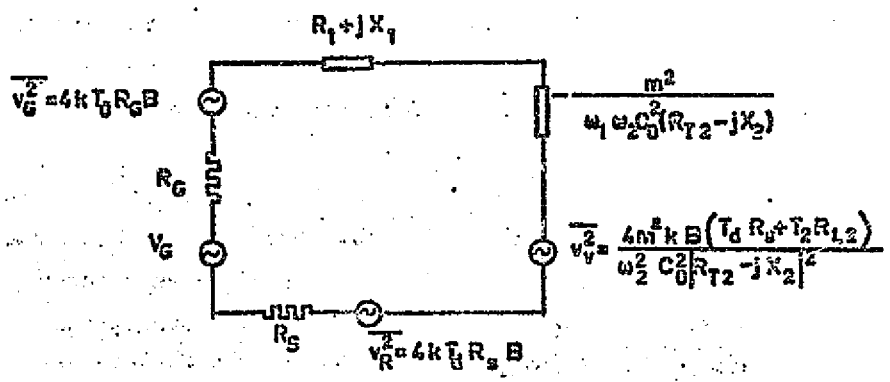


FIGURE 17B

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

In [5, 6], a noise figure M is defined as

$$M = \frac{F - 1}{1 - \frac{1}{G}},$$

which is shown to be independent of the amplification:

$$M = \frac{\left(m\frac{\omega_c}{\omega_2}\right)^2 + 1 + \frac{R_{L2}}{R_s}}{m^2\frac{\omega_c^2}{\omega_1\omega_2} - 1 - \frac{R_{L2}}{R_s}} \frac{T_d}{T_o}$$

The lowest noise figure then yields /106

$$\frac{\omega_2}{\omega_1} = \sqrt{1 + \left(\frac{m\omega_c}{\omega_1}\right)^2 \frac{1}{1 + \frac{R_{L2}}{R_s}}} - 1;$$

so that

$$M_{min} = \frac{2}{\sqrt{1 + \left(\frac{m\omega_c}{\omega_1}\right)^2} - 1} \frac{T_d}{T_o}.$$

This expression shows that the lowest noise figure is obtained in two ways: by choosing a cutoff frequency ω_c much larger than the signal frequency ω_1, and by cooling the varactor and the possible load in the no-load circuit. Also, the noise figure's minimum value is obtained by choosing $R_{L2} = 0$.

It is naturally very advantageous to obtain low noise without cooling, and in this way it is meaningful to utilize varactors with high cutoff frequencies. In [7], the absolute upper limit for f_c is estimated at 10^4 to 10^5 GHz, and [8] reports that GaAs varactors with $f_c > 2000$ GHz have been developed. (It should be observed here that f_c is calculated from an impedance measurement at relatively low frequency, normally in the X-band. Surface effects, however, cause the varactor losses to increase significantly at very high frequencies and lead to the corresponding reduction in f_c. This problem can possibly be solved by coating the varactor surface with gold up to the active pn junction.) If one assumes that $f_c = 3000$ GHz is an obtainable value and if one sets $m = 0.33$, the following results are obtained for an X-band amplifier: $f_2/f_1 = 100$ and $MM_{min} = 0.02$, or, that $F = 1.02$ 0.1 dB or $T_e = 6^oK$. However, this would require a pumping frequency of 101 $f_1 = 1010$ GHz, and the no-load circuit at 1000 GHz can hardly be separated from the pumping circuit. These figures show that the minimum noise factor cannot possibly be reached in practice. With $f_c = 300$ GHz, we obtain $f_2/f_1 = 9$ and $M_{min} = 0.2$; in other words, $F = 1.2 = 0.8$ dB or $T_e = 60^oK$. These requirements are fully realizable. These breakdown calculations show that a /107 noise factor of 1 dB should be obtained with an X-band amplifier at room temperature. At 1 GHz, the minimum noise factor is several tenths of a decibel. In this case, circulation damping is assumed to zero, whereas it will normally increase the noise factor by only about 0.2 dB.

To obtain extremely low internal noise in a parametric amplifier it is usually most appropriate to cool the varactor diode. The expression for the noise figure M shows that it is proportionate to the diode temperature T_d, naturally under the condition that m and f_c do not change with the temperature. This last requirement is hardly fulfilled in all varactors, but highly doped diodes, especially of GaAs, have been developed which function without deterioration at temperatures as low as 4°K. At the low noise temperatures which are obtained in this manner, several phenomena are now exhibited which in the uncooled amplifier were without meaning. In the cooled state, however, they are very significant. If, for example, the varactor is pumped so that a direct current I_o is produced, [6] shows that the noise temperature will increase a minimum of

$$\Delta T_{I_o} = \left(1 + 2\frac{\omega_1}{\omega_2}\right)\frac{eI_o}{2(\omega_1 C_c)^2 R_G^2}.$$

(when f_1 = 4 GHz, 3°K/μA is a typical value for GaAs and 25°K/μA is typical for Si).

The heating of the varactor's active junction due to absorbed pumping power is also very significant for the noise temperature. This is discussed at length in [6] and the most important consequences are that:

1. The optimal frequency ration f_2/f_1 for minimum noise will be much lower than in an uncooled amplifier. The value is dependent on the varactor's heating, f_c and m, but will lie between 2 and 5.

2. An effective heat sink in the varactor diode will be very significant in obtaining minimum noise.

3. An optimized, cooled load in the no-load circuit can reduce the noise up to 20% in comparison to an unloaded circuit. /108

The low noise in the cooled parametric amplifier also causes damping in the circulator and the leads to have a much greater effect on the total noise, especially if the damped parts are at room temperature. If the damping factor is L_1, and the temperature of the leads (including the circulator) is t_1, the damping factor at the output is L_2, and the noise temperature in the following stage is T_{e2}, then the system noise temperature will be

$$T_e = (L_1 - 1)T_1 + L_1 T_{e1} + \frac{L_1(L_2 - 1)T_1 + L_1 L_2 T_{e2}}{G_1},$$

where T_{e1} and G_1 are the noise temperature and the amplification

ORIGINAL PAGE IS
OF POOR QUALITY

in the parametric amplifier. In connection with a helium-cooled amplifier, the circulator should definitely be cooled as well. If, for example, $T_{e2} = 40°K$, $L_1 = L_2 \approx 0.3$ dB, and $G_1 = 20$ dB, the result is that $T_e \approx 4°K$. If the circulator were maintained at room temperature, however, T_e would rise to approximately 25°K. In connection with a liquid nitrogen cooled amplifier, the worsening is not so large if the circulator is not cooled. This is due to the fact that the losses in a cooled circulator are normally larger than in one at room temperature. If $L_1 = 0.3$ dB in a cooled circulator, $L_1 = 0.1$ dB can be obtained in a circulator at room temperature. This results in about the same T_e.

In addition to the noise factor, the bandwidth of the parametric amplifier is also very significant in its utilization. A calculation of the amplification-bandwidth product, $\sqrt{G}\,B$, using the above expression for the amplification, gives the bandwidth for a single-tuned circuit with a varactor without parasitic elements. When $m = 0.33$ and $\omega_c/\omega_1 = 30$, such a calculation yields a maximum bandwidth of 10% at G = 20 dB, and nearly 30% at G = 10 dB. If m is increased to 0.5 and ω_c/ω_1 to 60, the maximum bandwidth will be approximately 14% at G = 20 dB and approximately 40% at G = 10 dB. This requires that the ratio ω_c/ω_1 is 0.7 and 0.5, respectively, times the value which produces the lowest noise. In other words, a certain deterioration in the noise factor must /109 be accepted in order to obtain the largest bandwidth. These calculations indicate that bandwidths of approximately 30% should be obtained, in every case through the cascade-coupling of several stages with relatively low amplification at each stage.

A more complete calculation of the bandwidth, taking into account the varactor's parasitic elements and the necessary band-pass filters, has been done in [9] and several typical numerical results are shown in [6]. From these it is taken that the maximum bandwidth will diminish rapidly above the varactor's series-resonance frequency, which at present in a diode capsule is around 10 GHz. With a single-tuned signal circuit and G = 20 dB, the bandwidth is approximately 1% in the X-band and 3% in the S-band. Multiple tuning of the signal circuit can at most increase these values to 15% and 25%, respectively. However, by constructing a balanced varactor-pair with multiple-tuning, maximum values of 20% in the X-band and 40% in the S-band can be reached. (The two adjacent varactors in this type of amplifier create a series-resonance circuit for the no-load frequency and work in parallel at the signal frequency. The no-load circuit will in any case be very broadbanded and the no-load frequency is chosen to be equal to the varactors' series-resonance frequency. If the amplification is reduced and several amplifier stages are used to produce the desired total amplification of approximately 20 dB, bandwidths of 50% in the L- and S-bands and approximately 30% in the X-band should be possible. Several examples of broad-

band parametric amplifiers are shown in Table 6.

Table 6. Performance of parametric amplifiers.

f_1 GHz	G dB	Δf %	Comments	Reference
1	19	20	1 varactor, 2 stages, double-tuning.	32 in [6]
2	20	40	Varactor-pair, 4 stages, double-tuning.	36 "
4	16	15	1 varactor, 1 stage, double-tuning.	33 "
9	19	5	1 varactor, 1 stage, double-tuning.	35 "
18	19	1	1 varactor, 1 stage, double-tuning.	[10]

(Between 1958 and 1962 parametric amplifiers of the transient-wave type, with several varactor stages, were both theoretically and practically investigated. Despite the fact that large bandwidth could be obtained in principle in this manner, the construction of the amplifiers was so complicated that they never became, and probably never will become of practical significance.) /110

As the expression for the amplification shows, its constant will mainly depend upon holding the pumping power, in other words m, constant. A parameter for the amplification stability can be defined in the following manner:

$$GS = \frac{dG/G}{d\alpha/\alpha} \text{ med } \alpha = \frac{m^2 S_o^2}{\omega_1 \omega_2 (R_G + R_s)(R_{L2} + R_s)}.$$

It is shown in [5] that GS is minimized when $R_{12} = 0$ and is approximately proportionate to $\sqrt{G}$. The change in amplification is typically 0.1 dB when G = 20 dB with a 0.01 dB change in the pumping power. The best way to improve the amplification stability is to reduce G and to use several stages to obtain the desired total amplification.

Relatively wide banded pumping and no-load circuits are required so that small changes in the pumping frequency shall not cahnge G. This requirement is usually fulfilled. All small changes in the circuit dimensions resulting from temperature changes or vibrations can also change G. Therefore, the circuits are designed as small and stable as possible, and by cooling the circulator and other circuits to a certain constant temperature, improved amplification stability is obtained, since all expansion with temperature variation is eliminated.

The dynamic range of an uncooled parametric amplifier, calculated from the noise level within 1 MHz bandwidth and at 1 dB amplification compression, is typically around 80 dB. The

ORIGINAL PAGE IS
OF POOR QUALITY

C-2

corresponding saturation level at the output is -20 to -10 dBm.

Future Development of the Parametric Amplifier /111

It has already been shown that there are many problems in maintaining extremely low noise in uncooled amplifiers:

1. The requirement that varactors have very low losses at millimeter-wave frequencies is difficult to fulfill.

2. The high necessary pumping frequency is a drawback.

3. Filtration problems in separating close pumping and no-load frequencies.

The rapid development of cooling techniques, which has already resulted in compact, reliable cooling apparati for temperatures down to 20°K, gives strong support to the increased utilization of cooled parametric amplifiers. The noise temperature will not be any lower than what has already been obtained, but the technical performance will be improved through more compact cooling and amplification units, producing low noise and large bandwidth, and through the utilization of solid-state oscillators as pump sources. A parametric amplifier cooled to 20°K, followed by another amplifier at room temperature, produces a total noise temperature of approximately 25°K. If the first amplifier is cooled to 4.2°K and a second to 20°K, the total noise temperature will be 4°K. The varactor diode's characteristics must be improved in order to achieve close to maximum bandwidth simultaneously with low internal noise. Large bandwidth requires high pumping (large m) but at the same time the pump-heating of the varactor must be avoided by improving the heat sink and by reducing the loss resistance at low temperatures.

It is assumed to be likely that in any case, parametric amplifiers cooled to 20°K, in connection with a decrease in the varactor's parasitic reactances and the utilization of balanced varactor pairs in several stages, will display the following bandwidth in 1985:

Frequency	Bandwidth
1 - 4 GHz	40%
8 - 12 GHz	20%
15 - 20 GHz	10%
30 - 40 GHz	3%

The possibilities of constructing parametric amplifiers for very high frequencies will be limited by the availability of pump oscillators. If one assumes that these will be developed up to 300 GHz, and use a no-load-signal ratio of 3 (for a cooled /112

amplifier), the maximum signal frequency will be approximately 75 GHz. Such an amplifier will, however, only have good noise characteristics if the signals are fed without phase information into both the partially overlapping signal and no-load channels. For example, this is the case in radio astronomical measurements, in which the signal consists of broadband noise. In other cases, the noise factor is increased by 3 dB.

Parametric amplification at even higher frequencies can possibly be realized by using a Josephson junction operating simultaneously as a pump oscillator instead of a varactor [11]. The Josephson current is not expected to contribute any noise, and low noise operation at approximately 300 GHz is theoretically possible. The Josephson junction has been shown to function as an oscillator at 8,000 GHz (see the section on the Josephson oscillator), but at levels over 1,000 GHz a normal, noise-generating tunnel current begins to flow. Therefore, low noise amplification can be expected at pumping frequencies up to that level.

Since the circulator-coupled parametric amplifier has been the main topic here, it is of interest to know if the circulator causes any frequency limitations. In the lower frequency range, there already exists circulators which function well down to 0.1 GHz. With regard to very high frequencies, [12] mentions that circulators for 94 GHz and 220 GHz have been produced. However, the losses can be assumed to be larger than desirable at these high frequencies. Octave bandwidths are already common for frequencies up to the X-band and it will probably be relatively easy to maintain at all frequencies circulator bandwidths larger than the parametric amplifier's bandwidth. This shows that the circulator will not limit the performance of the parametric amplifier, with the exception of a noise factor which can be significant at extremely high frequencies.

A two-port parametric amplifier without circulator and with /113
non-reciprocal amplification characteristics has already been built [13, 14]. Two stages are normally used, one for the up-conversion and one for the down-conversion of the frequency, together with a feed-back coupling from the second to the first stage. However, all such couplings have a tendency to be narrow-banded and therefore are not expected to be very significant, especially since circulators can be manufactured in very miniaturized form and directly integrated in, for instance, stripline circuits.

In certain applications, the use of a narrow band, but electronically-tunable amplifier can be significant. If a certain deterioration of the noise factor is allowable and the need for rapid tuning is not too large, this operation can be easily

obtained by coupling a tunable YIG filter before a broadband parametric amplifier. It should be pointed out here that narrowband, electronically-tunable, frequency-conversion low noise amplifiers have been constructed, and that they could be of greater significance in the future. Upper and lower sideband frequency converters with broadband pumping circuits, but narrow band sideband circuits, are treated in [15]. By tuning the pumping frequency, the signal frequency can be tuned up to an octave. Even larger tuning ranges, possibly several octaves at low frequency, might be possible. Another method is given in [16]. A broadband, low amplification is achieved through frequency conversion in the upper sideband and a narrowband, high additional amplification is achieved regeneratively through tuning a circuit at the lower sideband (in other words, the no-load circuit). A varactor-tuned no-load circuit could in this way produce an electronically-tunable signal frequency.

REFERENCES

1. Van der Ziel, "On the mixing properties of nonlinear capacitances," J. Appl. Phys. Vol. 19, p. 999, (1948).

2. Suhl, H., "Proposal for a ferromagnetic amplifier in the microwave range," Phys. Rev. p. 384, (April 1957).

3. Weiss, M., "A solid-state microwave amplifier and oscillator using ferrites," Phys. Rev., p. 317, (July 1957). /114

4. Heffner, Kotzebue, "Experimental characteristics of a microwave parametric amplifier using a semiconductor diode," Proc. IRE, p. 1301, (June 1958).

5. Khan, P. J., "Optimum design of varactor diode parametric amplifiers," Proc. High frequency generation and amplification, Cornell University, (1967).

6. Uenohara, M., "Cooled varactor parametric amplifiers," Advances in microwaves, Academic Press, (1967).

7. Penfield, P. Jr., "Maximum cut off frequency of varactor diodes," Proc. IEEE, p. 422, (April 1965).

8. Burrus, C. A., "Planar diffused GaAs millimeter-wave varactor diodes," Proc. IEEE, p. 1104, (June 1964).

9. De Jager, J. T., "Maximum bandwidth of nondegenerate parametric amplifier," Trans IEEE, MTT, p. 459, (1964).

10. Kinoshita, Maeda, "An 18 GHz single-tuned parametric amplifier with large gain-bandwidth product," Trans. IEEE, MTT, p. 409, (July 1970).

11. Russer, P., "Parametric amplification with Josephson junction," AEU, p. 417, (August 1969).

12. Feldman, N. E., "Syllabus on low-noise microwave devices," Microwave Journal, p. 59, (July 1969).

13. Thompson, G. H. B., "Unidirectional lower sideband parametric amplifier without circulator," Proc. IRE, p. 1685, (November 1961).

14. Maurer, Locherer, "Theorie nichtresiproker Schaltungen mit gleicher Eingangs- und Ausgangsfrequenz unter Verwendung nichtlinearer Halbleiterelements," AEU, vol. 15, p. 71.

15. Matthaei, G. L., "Design theory of up-converters for use as electronically-tunable filters," Trans. IRE, MTT, (September 1961).

REFERENCES

16. Luksch, Stachejko, "A new approach to simplified parametric amplifier tuning," Proc. IRE, p. 1540, (June 1962).

The Tunnel Diode Amplifier

The development of the tunnel diode and the physical phenomena which explain its function have already been treated in connection with the tunnel diode oscillator. The negative resistance produced when an appropriate direct voltage is applied to the tunnel diode can naturally be utilized in a reflection amplifier. A circulator-coupling is almost always used to separate the feeding and the amplified signal, and to give good stability against impedance variations in the source impedance. /115

To calculate the characteristics of the tunnel diode amplifier the equivalent circuit in figure 18 is used.

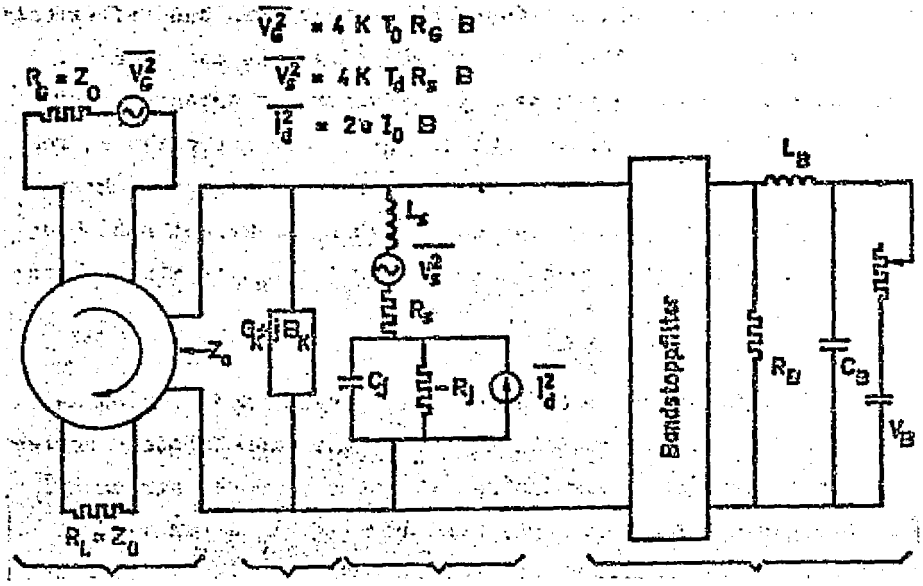


FIGURE 18.

The voltage supply and stabilizing circuits on the right side of the figure have no effect on the high-frequency circuit within the amplifier's band because of the construction of the band stop filter. R_B is coupled in to the high-frequency circuit outside of the amplifier band and loads the diode so that no oscillations can occur.

If the diode's parasitic elements, R_S and L_S are ignored, the reflection amplification can be written /116

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

$$G = \frac{\left(\frac{1}{Z_o} + \frac{1}{R_j}\right)^2 + \left(\omega C_j - \frac{1}{\omega L_k}\right)^2}{\left(\frac{1}{Z_o} - \frac{1}{R_j}\right)^2 + \left(\omega C_j - \frac{1}{\omega L_k}\right)^2}$$

At the resonance frequency $\omega_o = \frac{1}{\sqrt{L_k C_j}}$ and with high amplification ($Z_o \approx R_j$)

$$G_o = \frac{4R_j^2}{(R_j - Z_o)^2}.$$

A calculation of the bandwidth yields $2\Delta\omega = (R_j - Z_o)/R_j^2 C_j$. Therefore, the amplification-bandwidth product $\sqrt{GB} = 1/\pi R_j C_j$.

Through the utilization of special band-widening circuits, built as bandpass filters and placed between the diode and the load, the reflection coefficient and the amplification can be held nearly constant within a much wider band. Calculations in [1] show that the utilization of 2, 8, and an infinite number of impedance devices produce respectively approximately 2, 7 and 10 times greater $\sqrt{GB}$ than a single-tuned circuit. These are the cut-off values which apply when R_s and L_s are very small. In particular, the effect of an appreciable L_s will be to reduce the bandwidth.

The tunnel diode amplifier's noise factor can be calculated with the help of the noise sources identified in figure 18. In part, the tunnel diode series resistance produces a thermal noise voltage, $\sqrt{v_s^2}$, and the current I_e produces through the junction layer a fluctuation noise current $(I_d)^{\frac{1}{2}}$. I_e is composed of two opposing currents, each of which produces an equally large noise factor. Within the diode's negative resistance region, the current is completely dominant in one direction and one can set $I_e = I_o$ = the direct current through the diode. The noise calculations are most easily done in a series circuit [2] (in which the parallel combination of $-R_j$ and C_j is included in the equivalent series resistance $-R_e$, with the noise voltage $(i_n R_e R_j)^{\frac{1}{2}}$ and the sum of all the reactances set at 0). The noise factor definition /117

$$F = \frac{\text{total noise in } Z_o}{GkT_oB}$$

results in $F = \dfrac{1 + \frac{eI_o}{2kT_o} R_j}{\left(1 - \frac{R_s}{R_j}\right)\left(1 - \left(\frac{f}{f_r}\right)^2\right)}$,

where the diodes resistive cutoff frequency

$$f_r = \frac{\sqrt{\frac{R_j}{R_s} - 1}}{2\pi R_j C_j}$$

under the condition that the amplification is high and the diode has a standard temperature $T_o = 300^oK$. The lowest noise factor is obtained by minimizing the noise constant $eI_oR_j/2kT_o = 20I_oR_j$ and the ratios R_s/R_j and f/f_r. R_s/R_j is typically around 0.1 and f/f_r is rarely chosen at less than 0.25, since the difficulties in stabilizing the amplifier against undesirable oscillations would then be too large. Thus, the denominator adds approximately 1 dB to the noise factor. The value of the noise constant depends upon which semiconductor is used, but also to a certain degree and in an unknown manner, upon the construction of the diode itself. The lowest values obtained were 0.7 to 0.86 for GaSb, 1.2 for Ge and 1.5 for GaAs, which should lead to a noise factor on the order of 3.3 dB, 4.3 dB and 4.9 dB, respectively. InAs and InSb semiconductors have several properties which are favorable for low noise factors [3]. Among other things, the high mobility yields a small R_s and the small band gap contributes to a low R_j. Unfortunately, the band gap is so small that the diffusion current makes operating at room temperature impossible. However, cooled tunneled diodes made from these materials with very low noise factors are possible. Isolated tests with cooled tunneled diode amplifiers have already been reported [4]. Ge tunnel diodes were successfully cooled down from 300 to 80^oK, and in so doing the noise factor (at 1.3 GHz) was reduced from 4.9 to 3.1 dB. This relatively large reduction is difficult to explain. It can be shown that if the tunnel diode's temperature T_d deviates from T_o, the numerator in the expression for the noise factor will have the extra term

$$-\frac{R_s}{R_j}\left(1 + (\omega C_j R_j)^2\right)\left(1 - \frac{T_d}{T_o}\right), \qquad /118$$

ORIGINAL PAGE IS
OF POOR QUALITY

which can be written

$$-\left[\frac{R_s}{R_j} + \left(1 - \frac{R_s}{R_j}\right)\left(\frac{f}{f_r}\right)^2\right]\left(1 - \frac{T_d}{T_o}\right).$$

With the normal values $R_s/R_j \approx 0.1$ and $f/f_r = 0.25$, and with $T_d \approx 0$, this term will yield a maximum reduction in the thermal noise on the order of -0.16. This should reduce the noise factor in Ge by 0.3 dB, while a reduction of 1.8 dB was measured. The

ORIGINAL PAGE IS
OF POOR QUALITY

cooling changes R_j insignificantly and the product $R_j I_o$ was nearly constant. These figures show that the fluctuation noise must for some reason diminish with the temperature. The results of any investigations showing how the fluctuation noise changes with temperature have yet to be published. This means that it is difficult at present to evaluate the improvements in the noise factor which could be obtained by cooling the tunnel diode. By using a semiconductor with a small band gap, R_j should be reduced to one-third of its value in GaSb. If one assumes at least as large a reduction of R_s, a noise factor of around 1.8 dB should be obtained with normal fluctuation noise, and a possible reduction in fluctuation noise resulting from the low temperature should then lead to a value of around 1 dB.

The use of the methods discussed above for reducing the noise factor necessarily entail a drawback in the form of reduced output power from the amplifier. As in the previously discussed tunnel diode oscillator, the maximum output power is proportionate to the diode's peak current and the difference between the valley voltage and the peak voltage. But a low value for the noise constant requires both low battery current, and therefore low peak current, and a small band gap, corresponding to a small difference between the valley voltage and the peak voltage. If the attempt is made, without regard to the noise factor, to maximize the output power in connection with a certain semiconductor, the peak current must be increased as much as possible. The upper limit is then set by the stability requirements for the /119 tunnel diode, which according to [5] can be written as

$$\frac{R_s}{R_j} < 1;$$

$$\frac{L_s}{R_j^2 C_j} < 1 \quad \text{(when } \frac{R_s}{R_j} \text{ is close to one).}$$

If the expression for the amplification-bandwidth product is introduced and $I_{p\ max}$ is set approximately equal $1/(R_{j\ min})$, the stability requirement can be written as [6]:

$$I_{p\ max} < \text{constant} < \frac{1}{\pi\sqrt{GB}L_s}.$$

Thus, small values for $\sqrt{GB}$ and L_s at the same time produce higher peak current. The relationship between the constants for GaAs, Ge and GaSb is 1:0.5:0.27. If the maximum voltage modulation is taken into account, the relationship between the respective semiconductor's maximum output power will be 1:0.2:0.08.

To obtain a tunnel diode amplifier with both a low noise

factor and a high output power level (large dynamic region), one can build the first amplifier stage with a Ge or GaSb diode and the second stage with a GaAs diode. In this manner a dynamic region of approximately 84 dB is created (as normally calculated from the noise effect within 1 MHz and up to the 1 dB compression level). The maximum value of 90 dB given in [7] should be reached with a final stage containing two push-pull diodes.

Tunnel diode amplifiers have so far been built for frequencies up to 20 GHz. The maximum frequency must naturally lie under the diode's resistive cut-off frequency f_r and to obtain an acceptable noise factor it should be at least three times lower. To obtain high f_r, the product R_jC_j must be minimized, and it has been shown to decrease exponentially with increased doping. Since oscillations in the diode must be avoided, it is necessary /120 at the same time to reduce the series inductance about as much as R_jC_j. From the invention of the tunnel diode in 1958 until 1964, f_r reached the level of 4 GHz and between 1964 and 1968 it increased to 50 GHz. Diodes with f_r up to 30 GHz could according to the catalog data have L_s equal to 0.3 nH, but at f_r equal to 40 to 50 GHz, L_s decreases to 0.1 nH. It is mentioned in [6] that series inductances down to 0.02 nH have been realized, which with a rough approximation would make possible stable diodes with f_r values up to 150 GHz. As was mentioned earlier, tunnel diode oscillators already operate around 100 GHz. With the above line of reasoning it is likely that tunnel diode amplifiers will be made useful up to 30 to 50 GHz.

The results of the possible further development of the tunnel diode and the adherent amplifier can be summarized as follows:

Bandwidth. With $\sqrt{GB} \approx 1/R_jC_j \approx f_r$, and for practical reasons the relation $f \approx 0.3\ f_r$, one obtains $B \sim f$, in other words that the percentage bandwidth is constant, independent of the frequencies. However, the diode's parasitic reactances will influence this relationship and reduce the bandwidth further as frequency increases. It follows from the stability criteria that, for example, the series inductance must be reduced about the same factor as f_r (and f) isincreased by. One can therefore assume that a future amplifier for the 50 GHz level will have about the same bandwidth as an X-band amplifier today. The decreased parasitic reactances will of course also increase the bandwidth at lower frequencies. Considering that today's single-tuned tunnel diode amplifiers have approximately 10% bandwidth between 1 and 10 GHz with moderate amplification (10 to 15 dB), the following forecast can be made:

> The use of a moderate number of broadband circuits in combination with reduced parasitic reactances will lead to at least 60% (octave) bandwidth in the L and S-bands, 40% in the X-band and 10% to 15% within the range of 30 to 50 GHz.

Noise factor. Any great improvements of the noise factor in uncooled tunnel amplifiers are not to be expected. In the expression for the noise factor, the ratio f/f_r will remain about the same, R_s/R_j can possibly be reduced somewhat, but will increase with increasing frequency, and the noise constant can possibly be made somewhat smaller than at the present. This can lead to a reduction of about 1 dB in the noise factor for 1969, which according to [7] at 1, 10 and 18 GHz is 3, 5 and 7 dB. Thus, by 1985 the following noise factor values could be reached: /121

f in GHz	F in dB
1	2
10	3
20	5
40	7

A further reduction in the noise factor in liquid nitrogen cooled tunnel diode amplifiers is obtained in two ways: through the use of semiconductors with small band gaps, and with an experimentally-confirmed extra reduction of the noise constant. The total reduction in noise as a result of this is hard to estimate, but can be assumed to lead to a noise factor which is at least 1 dB lower than in the table above. Another advantage with cooling is that the tunnel diode's constant temperature results in a more stable amplification function.

It is also possible that a superconducting tunnel diode will be fully developed by 1985. This would consist of one tin and one aluminum electrode separated by a thin oxide layer (but not so thin that Josephson-tunneling occurs) and is calculated to produce a noise temperature of approximately $4^{\circ}K$ [8] in a helium-cooled amplifier. However, the level of output power would be very low, on the order of -60 dBm and such a diode is not expected to be able to operate at such a high frequency as a normal tunnel diode, but could still be useful within the lower microwave range.

REFERENCES

1. Scanlan, Lim, "A design theory for optimum broadband reflection amplifiers," Trans. IEEE, MTT, vol. 12, p. 504, (1964). /122

2. Chow, W. F., Principles of tunnel diode circuits, John Wiley and Sons, (1964).

3. Armstrong, L.D., "Tunnel diodes for low noise microwave amplification," Microwave Journal, (August 1962).

4. Stachejko, "Investigation of a L-band tunnel diode amplifier at low temperatures," Proc. IEEE, p. 1368, (November 1964).

5. Smilen, Youla, "Stability criteria for tunnel diodes," Proc. IRE, p. 1206, (July 1961).

6. Sterzer, F., "Tunnel diode devices," Advances in microwaves, Academic Press, (1967).

7. Feldman, N. E., "Syllabus on low-noise microwave devices," Microwave Journal, p. 59, (July 1969).

8. Penney, J. D., "Potential of superconducting-tunnel diode amplifiers as high-frequency amplifiers," Proc. IEEE, p. 1628, (October 1969).

The Maser Amplifier

/123

The possibility of producing high-frequency energy amplification by stimulated emission in an excited atom or molecular system was first discussed in 1953 by Weber [1], and later in 1955 by Basov and Prokhorov [2]. The first so-called maser amplifier, which was based on the sorting out of excited atomis in ammonia gas, was built in 1955 [3], and the name "maser" was suggested as an abbrevation of "Microwave Amplification by Stimulated Emission of Radiation." Theories for a continually-functioning three layer maser were published in 1955 [2] for a gas, and in 1956 [4] for a crystalline, paramagnetic material. In contrast to the gas maser, the solid-state maser can operate only if the crystal is cooled to a very low temperature (with liquid helium). This will bring about the added adcantage of the amplifier having an extremely low internal noise. The first experimental results with a paramagnetic maser operating in the X-band were published in the beginning of 1957 [5]. In this case, a good connection between the high-frequency field and the paramagnetic atoms was obtained by placing the crystal in a microwave cavity. However, this effective connection resulted in a very small bandwidth (< 0.1%). Consequently, the next important stage in the development of the maser was the construction of the traveling-wave maser [6], in which a "slow" electromagnetic wave in a delay line produces a broadband coupling to the active crystal. A bandwidth of approximately 1% was then obtained, and by varying the pump frequency and the magnetic field, the signal frequency was tuned up to 20%. Furthermore, the traveling-wave maser produces non-reciprocal amplification and the feed-back damping can be easily increased by placing a number of ferrite devices in the delay line.

The amplification mechanism in a maser has a quatum mechanical basis. The atoms in a paramagnetic crystal possess a quantified magnetic moment which causes them to exist in a number of discrete energy states or energy levels. The difference in energy between these levels varies with the outer magnetic field, and a diagram of the energy levels appears in figure 19.

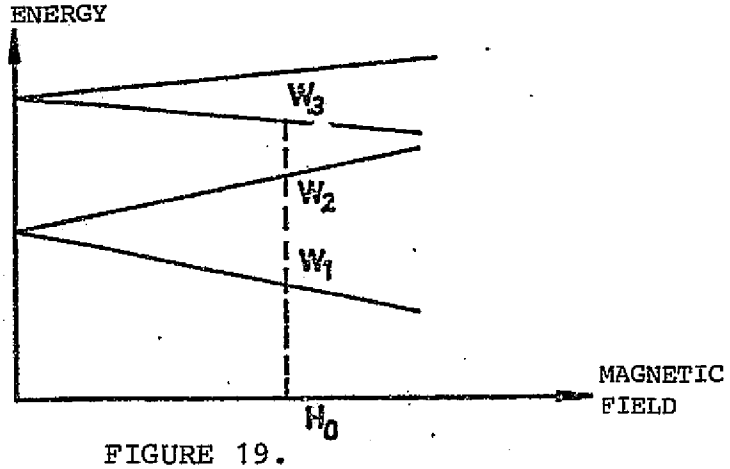


FIGURE 19.

124

If thermal equilibrium exists at the absolute temperature T, the ratio between the number of atoms, indicated by N_1, at energy level W_1, and the additional number of atoms, N_2, at the nigher energy level, W_2, can be written as:

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

$$N_2/N_1 = e^{-\frac{W_2 - W_1}{kT}}$$

where k equals Bolzmann's constant.

Therefore, $N_2 < N_1$ in a state of equilibrium. If such a crystal is radiated with a frequency f, possessing the photon energy

$$fh = W_2 - W_1$$

in which h equals Planck's constant, a resonance phenomena arises. Either a photon is absorbed and an atom makes the transition from energy level W_1 to W_2, or else an atom at level W_2 is stimulated, reducing its energy to W_1 and emitting a photon in phase with the impinging photon. There is an equal probability that either one or the other transition shall occur if one assumes an equal distribution of atoms in each energy state. In a crystal at thermal equilibrium, in which most of the atoms are in the lower energy state, the number of atoms making the /125
transition to the higher energy level will dominate, and the crystal will absorb radiation. Due to so-called relaxation processes, an excited atom will have a tendency to return to the lower level after a certain time. If the relaxation process between two levels is weak and the impinging radiation is strong, there will be about the same number of atoms at each level. If the powerful radiation, or pump frequency, corresponds to the energy difference $W_3 - W_1$ in figure 19, it can be easily observed that the number of N_3 atoms at level W_3 will be larger than the number of N_2 atoms at the W_2 level. (The relaxation from W_3 to W_2 should then be as low as possible.) A signal frequency $f_s = (W_3 - W_2)/h$ will then mainly result in transition from W_3 down to W_2, so that a net amplification is obtained through stimulated emission.

If the amplification is high and the damping in the delay structure is negligible, the maser's noise factor can, according to [7], be written

$$F = 1 + \frac{hf}{kT_o(1 - \frac{N_2}{N_3})}.$$

According to the above expression, N_2 will diminish as the crystal temperature diminishes and when $N_2 << N_3$, a X-band maser has the minimum value $F_{min} \approx 1.002$, or $T_{e\,min} \approx 0.5^\circ K$. The lowest noise temperatures obtained in practice lie around 2 to $10^\circ K$ [8].

The power output from a maser at 1 dB amplification compression is approximately -25 dBm [8].

The frequency range within which the maser is usable is nearly unlimited. Within the microwave range, maser amplifiers have been built between 1 and 70 GHz [8,9]. As predicted by Schawlow and Townes [10] in 1958, the principle of the maser is applicable within both the visible and the infrared portion of the spectrum. During the last decade, very extensive research has been done using these ideas as a basis. The main attempt has been to porduce coherent oscillations. Such oscillators are /126
called lasers, with the L standing for light (as opposed to the "M" in microwave). The active media in a laser can be either a gas or a liquid, to which pump power is brought in the form of incoherent light. However, semiconductor crystals which are pumped with a direct current have also been utilized. Laser techniques are not within the scope of this report, and are mentioned here only as indicative of the possibility of increasing the maser's frequency range.

Future Development of the Maser

The maser's most important characteristic, its extremely low internal noise, is already so close to an absolute minimum that further improvements are not possible in the amplification mechanism itself. One major disadvantage is that most of the useful masers to date require cooling with liquid helium in order to be functional. Since cooling techniques appear to be developing rapidly and will possibly soon result in compact, reliable, but not inexpensive cooling apparatus for temperatures as low as 4°K, this drawback will probably become less significant. Given the state of current cooling techniques, 20°K can be set as the lowest temperature obtainable using relatively inexpensive, small and reliable cooling systems. This and other reasons has led to the development of the maser amplifier, which operates at relatively high temperatures (up to 78°K). If, however, a microwave oscillator is used as the pump source, the maser's noise temperature will rise to the same level as the crystal temperature and furthermore, bandwidth will be reduced. According to [11], it is still possible to obtain very low internal noise at increased crystal temperatures by using so-called optical pumping via a laser. This is to be seen as a very likely line of development since optical pumping must also be used to develop maser amplifiers for frequencies up to the sub-millimeter range.

The maser's major disadvantage, its small bandwidth, is directly dependent upon and limited by the width of the levels describing the energy state of the system of atoms. Despite an intensive investigation of various crystal material, nothing has /127
been found in over ten years to expand the amplifier bandwidth by more than a few percentage points at most. It seems therefore very likely that the maser will continue to be a narrowband amplifier. Even with possible improvements in the crystal structure, a 10% bandwidth is assumed to be the upper limit.

REFERENCES

1. Weber, J., "Trans.," IRE, ED, p. 1, (June 1953).

2. Basov, Prokhorov, "Possible methods of obtaining active molecules for a molecular oscillator," Journal of Experimental and Theoretical Physics, p. 184, USSR, (July 1955).

3. Gordon, Zeiger, Townes, "The maser-- new type of microwave amplifier, frequency standard and spectrometer," Physics Review, p. 1256, (August 1955).

4. Bloembergen, N., "Proposal for a new type solid state maser," Physics Review, p. 324, (October 1956).

5. Scovil, Feher, Seidel, "Operation of a solid state maser," Physics Review, p. 762, (January 1957).

6. De Grasse, Schulz-du Bois, Scovil, "Three-level solid-state traveling-wave maser," Bell Systems Technical Journal, p. 305, (March 1957).

7. Witthe, F. P., "Molecular amplification and generation of microwaves," Proc. IRE, p. 291, (March 1957).

8. Feldman, N. E., "Syllabus on low-noise microwave devices," Microwave Journal, p. 59, (July 1969).

9. Hughes, Kremenek, "70-Gc Maser," Proc. IEEE, (May 1963).

10. Schawlow, Townes, "Infared and optical Masers," Physics Review, p. 1940, (December 1958).

11. Hsu, Tittel, "Optical pumping of microwave masers," Proc. IEEE, p. 185, (January 1963).

Other Low Noise Solid-State Amplifiers /128

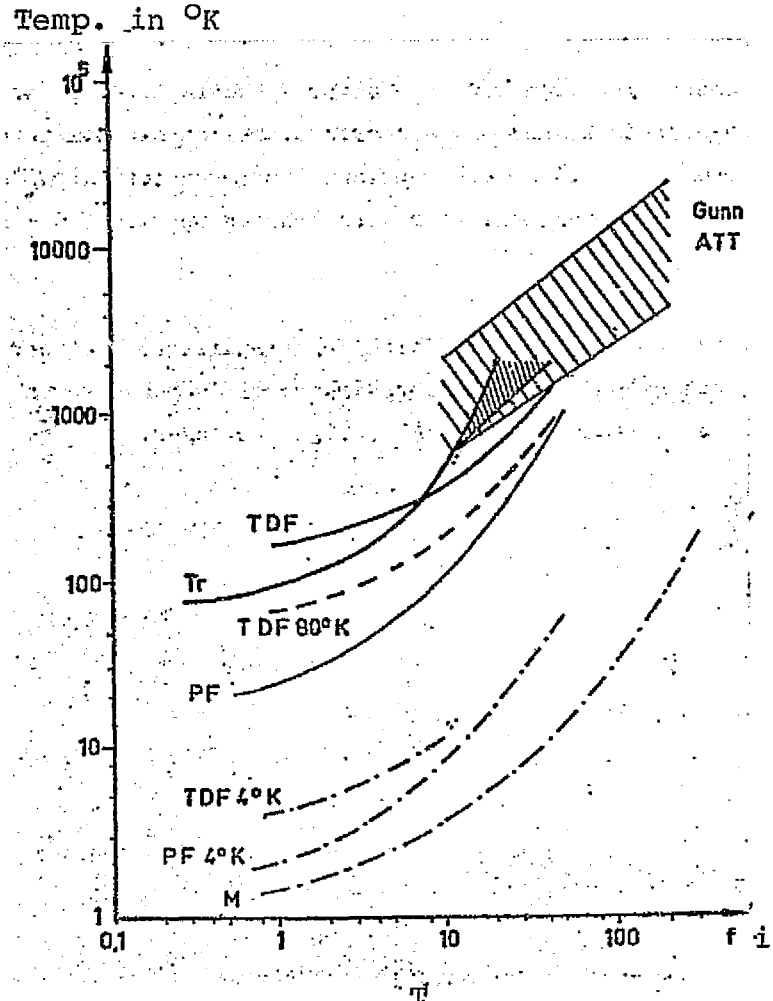
The function of circulator-coupled avalanche and Gunn diodes as broadband amplifiers was discussed in the chapter on oscillators. These amplifiers so far have high noise factors in the X-band of approximately 30 dB for a GaAs avalanche diode and approximately 15 dB for a Gunn diode with domain suppression. Since an avalanche diode with a Schottky barrier contact was reported at the 1970 MOGA conference to yield 10 dB lower FM noise than what is normal for an oscillator, it is assumed that significantly lowered factors will be obtained in an amplifier coupling. Also, a Gunn diode with domain suppression should yield a lower noise factor through increased homogeneity and improved contacts. Future diodes, especially developed for low noise amplification, will probably produce amplifiers with a noise factor of 5 to 10 dB. Particularly low internal noise can therefore not be expected from these types of amplifiers, but they are still of great interest since they can be used at very high frequencies and furthermore, are of the broadband type. Both Gunn and avalanche diode amplifiers should have upper cutoff frequencies of approximately 150 to 200 GHz. It is estimated that bandwidths of 30 to 67% will be reached, and possibly more when several diodes are included in the amplifier.

A Summarizing Forecast for Low Noise Amplifiers

Based on the previous discussions, the noise temperatures for various low noise amplifiers in 1985 are shown as a function of frequency in figure 20. Figure 20 shows that the maser has the lowest noise temperature, but that the helium-cooled parametric amplifier is only insignificantly worse in this respect and has significantly harger bandwidth up to approximately 20 GHz. At higher frequencies, the maser produces larger bandwidth and between 50 and 300 GHz, it might be the only available low noise amplifier. Helium-cooled amplifiers should remain very expensive. However, it is possible that optical pumping will make the maser less expersive by allowing temperatures around 20°K. The parametric amplifier can be constructed for any temperature between room temperature and 4°K, and one can envision a helium-cooled amplifier continuing to operate with a non-functioning /129
cooling system with only a somewhat higher noise factor as a result. By using a diode oscillator, the pumping source necessary for a parametric amplifier will become very inexpensive and compact.

Both regular and cooled tunnel diode amplifiers will have /130
somewhat higher noise than parametric amplifiers. In spite of this, they could be of use when extreme miniaturization and low driving power is needed, in which case the smaller dynamic range would not be significant.

This demands for only moderately low noise factors in combination with the desire for large bandwidth, the transistor amplifier will dominate up to approximately 20 GHz, at which point the Gunn and the avalanche diode will take over and possibly produce amplification up to approximately 150 to 200 GHz.



ORIGINAL PAGE IS
OF POOR QUALITY

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

	Approximate percentage bandwidth
ATT = avalanche diode amplifier	30 - 67%
Gunn = Gunn " "	30 - 67%
TDF = tunnel " "	10 - 67%
Tr = transistor amplifier	40 - 120%
PF = parametric "	1 - 40%
M = maser amplifier	5 - 10%

Figure 20. Noise temperature as a function of frequency for various types of low-noise amplifiers in 1985.

Completely new experiments with helium-cooled field-effect transistors indicate that in 1935 the cooled transistor amplifier could possibly produce somewhat lower noise than indicated by the PF curve in figure 20.

Frequency Conversion

Up-Converters for the Upper Sideband

Up-converters for high output power have become somewhat significant in connection with the building of microwave-range link couplings solely with semiconductor components. In this case, the signal coming into a repeater is mixed down to an intermediate frequency, with which a transistor amplifier can easily produce the desired power amplification. Afterwards, frequency conversion occurs using the mixing function of a varactor diode whose local oscillator can be either a Gunn or avalanche type. /131

If the intermediate frequency is written f_1, the oscillator or pumping frequency as f_2 and the output frequency as f_3, then for an upper sideband up-converter $f_1 + f_2 = f_3$. The conversion also produces a power amplification close the value of f_3/f_1 for a loss-free varactor. An up-converter for the lower sideband is characterized by the condition $f_2 - f_1 = f_3$ and by an amplification which can be increased arbitrarily up to the oscillation limit. However, the former converter is preferred because of its higher stability, and since f_1 is normally much smaller than f_3, sufficient amplification is still obtained.

Since the intermediate frequency f_1 consists of a modulated signal, it naturally has a certain bandwidth and can be described by a spectral function $f_1(f)$ centered around f_1. In absolutely distortion-free conversion, a frequency displacement occurs in the input spectrum so that the output spectrum can be written

$$F_3(f + f_2) = kF_1(f),$$

in which k equals the conversion amplification.

K is a constant in an ideal converter. However, in practice k will vary for several reasons and, as a result, cause a certain frequency distortion. Among other things, the low-pass filter for f_1 and the band-pass filter for f_3 will produce a certain change in the amplitude of both the input and the output signals within the medium-frequency bandwidth. Since the amplification is proportionate to f_3/f_1, the amplification must vary somewhat within the band, especially with large medium-frequency bandwidths. Furthermore, the signal reflected by the band-pass filter at the lower sideband (and at other frequencies which might possibly arise) will have various reactions on the varactor as the input frequency varies, which will also produce variations in amplification. Finally, a varying power level in the medium-frequency signal can change the varactor's medium-frequency impedance and, therefore, amplification. Naturally, this effect will only be exhibited if the medium-frequency power is so large that it is not negligible /132

in comparison to the pumping power.

The last influence will not be significant in an up-converter which must fulfill a very high linear requirement. The medium-frequency power must then, in any case, be kept very low so that undesirable intermodulation products are not created by the numerous input signals which, in such cases, are usually present. Up-converters of this type, made with extremely good linearity for use in communication satellites are described in [1]. Both output power and conversion efficiency will then be low, 1 mW and 10%, respectively. Within a ±40 MHz band centered around 4165 MHz, the variation in amplification was only ±0.1 dB. The M-F frequency was centered around 80 MHz. Such a low-level converter is always used in combination with a post-connected travelling-wave tube.

If the requirements for low distortion and a large bandwidth are not too great, an up-converter for the upper sideband can produce enough output power to make a post-amplifier unnecessary. The dimension curves for such an up-converter, optimized for maximum output power, are shown in [2]. The most important curves have been reproduced in Figures 21 and 22. Figure 21 shows the normalized maximum output power

$$p = \frac{P_{UT3} S_{max}}{2\pi f_3 (V_B - \Phi)^2}$$

as a function of f_3/f_c, where

P_{UT3} = the output power at f_3. /134

S_{max} = the varactor's maximum elasticity. If the varactor's lowest junction capacitance is written as $C_{j\ min}$, then

$$S_{max} = \frac{1}{C_{j\ min}}$$

V_b = the varactor's breakdown voltage.

= the contact potential equal 0.5 V for silicon.

f_c = $1/(2\ R_s C_{j\ min})$ = the varactor's cutoff frequency with the forward bias $-V_b$, where R_s = the varactor's series resistance.

It can then be written that

$$P_{UT3} = p \frac{f_3 V_B^2}{f_c R_s} \quad \text{when } V_B >> \Phi.$$

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

p = normalized
power output in W

/133

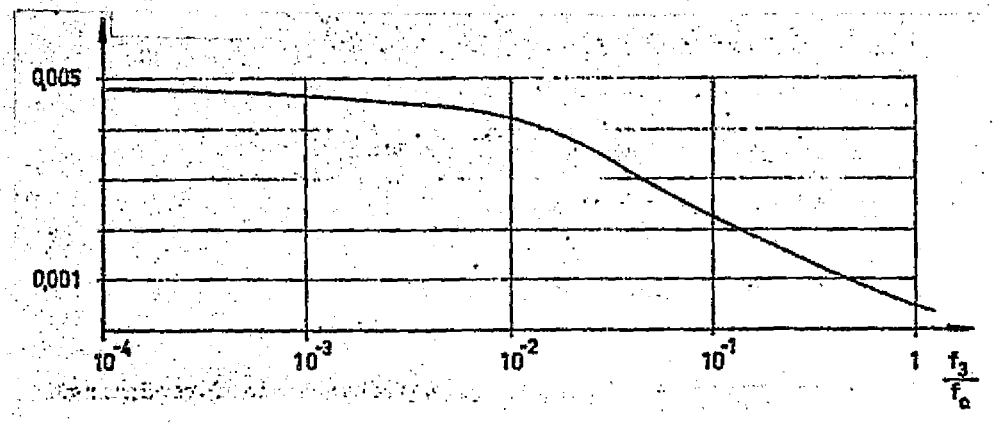


FIGURE 21.

The conversion efficiency from f_2 to f_3 is obtained from figure 21, in which it is also given as a function of f_3/f_c, under the condition that f_1 is much smaller than f_2, so that $f_2/f_3 \approx 1$.

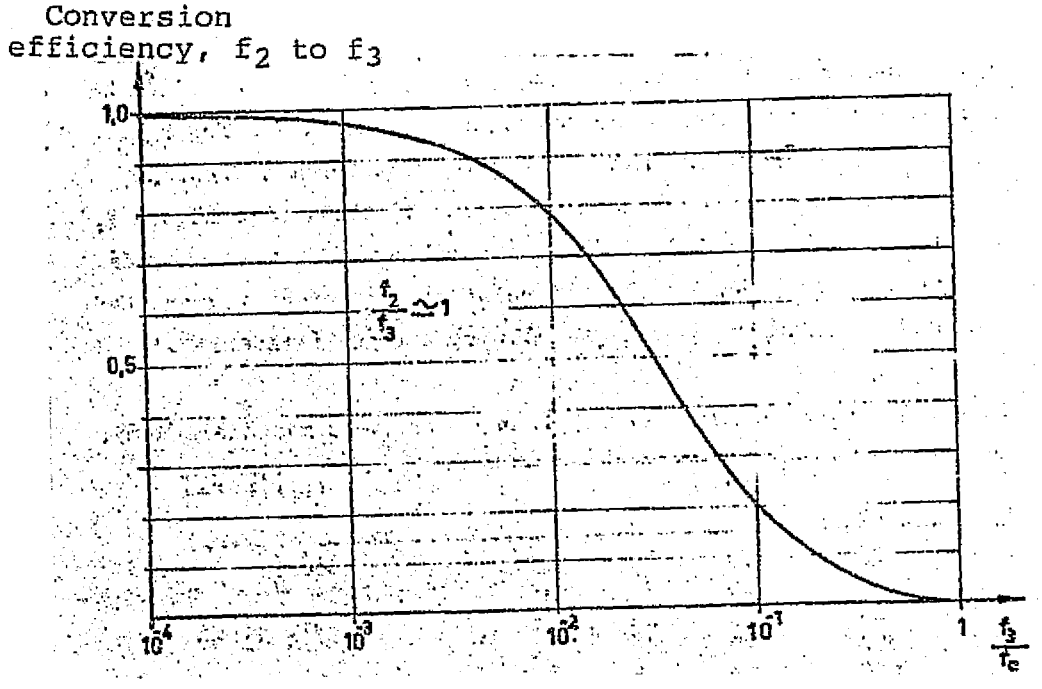


FIGURE 22.

By using the empirically determined relationship between V_B and r_s, the value for f_c and the maximum power loss, which is given in Figure 6-A and B in the chapter headed "The Frequency Multiplier," the forecasted values for P_{UT3} can be calculated directly, if one proceeds from the forecast curves in the figure. The calculation is done using the lowest possible f_c (in other words, highest V_B) which gives a permissible loss level in the varactor. It is shown then that maximum output power is obtained with relatively low efficiency. The results of such a breakdown calculation are presented in Table 7.

Table 7. Calculated maximum power output from a frequency converter, 1985.

f_3 GHz	f_c GHz	P_{IN2} W	η %	P_{UT3} W	Loss corrected P_{UT2} in W
1	20	170	45	75	68
3	50	100	35	35	30
10	120	65	23	15	11
50	350	20	10	2	1
150	650	10	5	0.5	0.1

As with the frequency multiplier, if one introduces a circuit efficiency, decreasing from 0.97 at 1 GHz to 0.65 at 150 GHz, and correct for the losses in the band-pass filter at the output with a factor of 0.94 at 1 GHz down to 0.34 at 150 GHz, the result is the loss-corrected values for output power furthest to the right in the table. (The filter losses are estimated based on a damping value 0.45 dB in the S-band.) The corresponding forecast curve is shown in Figure 23. It applies to a fully-tuned varactor with an abrupt junction. /135

By overdriving the varactor, the output power should be several times larger. The overdriven up-converter for the upper sideband is treated in [3, 4]. According to [3], the maximum output power can be written /136

$$P_{UT3} \approx \beta \frac{f_3}{f_c} \frac{V_B^2}{R_s}, \text{ when } V_B >> \Phi \text{ and } f_3 << f_c$$

in which β equals the normalized power output, corresponding to p in [2].

The conversion efficiency can be written

$$\eta = \frac{P_{UT3}}{P_{IN2}} \approx e^{-\alpha(f_3/f_c)} \quad \text{if } f_3 << f_c.$$

ORIGINAL PAGE IS OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

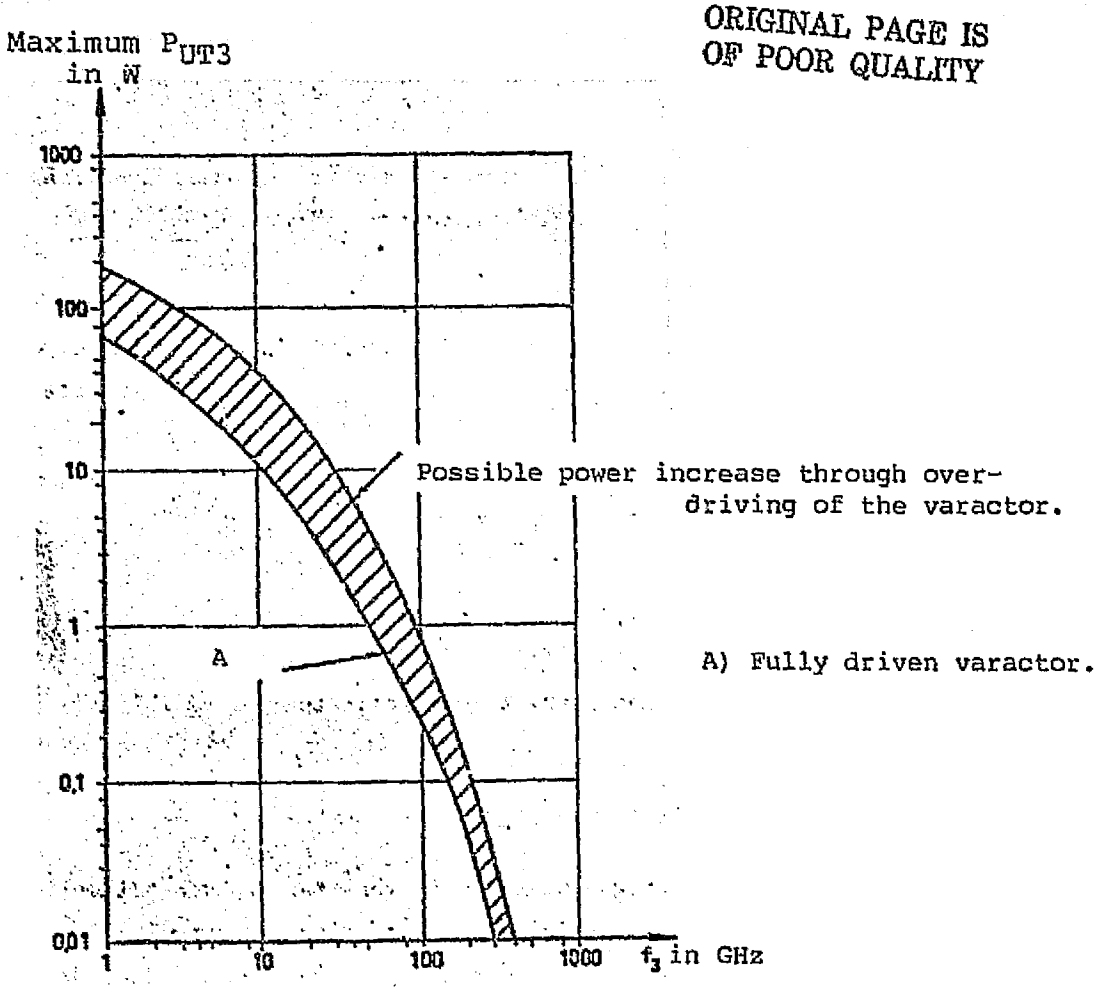


FIGURE 23.

If the ratio f_3/f_c is large enough to make η small, the obtained power output will be corrected by a factor of (1 + 2)/3.

Using the assumption that $f_3 >> f_1$, the results of the calculations in [3] show that a varactor with an abrupt junction produces maximum power $\beta = 28 \cdot 10^{-3}$ W with a $\alpha = 20$. If $P_{IN1} << P_{IN2}$, the power output can be written

$$P_{UT3} = \frac{P_{diss}}{\frac{1}{\eta} - 1} = \frac{P_{diss}}{e^{\alpha(f_3/f_c)} - 1}$$

in which P_{diss} equals the allowable losses.

Both of the expression above for the output power yield the following equality

$$28 \cdot 10^{-3} \frac{f_3}{f_c} \frac{V_B^2}{R_s} = \frac{P_{diss}}{e^{20(f_3/f_c)} - 1}$$

With the help of Figures 6-A and B in the chapter titled "The Frequency Multiplier," the V_B values which satisfy this equasion can be calculated. R_S, f_C and P_{diss} are all functions of V_B according to this statement. After the maximum powers obtained through these calculations are multiplied by the correction factor (1 + 2)/3, we obtain:

f_{UT3} GHz	f_C GHz	P_{IN2} W	P_{UT3} W	Loss corrected P_{UT3} in W
1	45	220	110	100
10	250	60	20	15
150	1400	6	0.3	0.1

/137

This breakdown calculation shows that overdriving the reactor will produce a maximum of 50% higher power output than at normal tuning. The power increase is indicated by the lined area in Figure 23. Overdriving does not increase the power output more because of the loss limitations in the varactor. It is shown in [4] that without such limitations the power output can be increased by at least five times through overdriving. An estimation using limited losses based upon [4, Figure 7 and 10] and Figure 6-A and B produced the following maximum powers:

f_{UT3} GHz	f_C GHz	P_{IN2} W	P_{UT3} W	Loss corrected P_{UT3} in W
1	45	260	180	160
10	200	70	70	50
150	800	8.5	1.5	0.3

These values are assumed to be the upper limits for power output with overdriving and they form the boundary of the lined overdriving region in Figure 23.

ORIGINAL PAGE IS
OF POOR QUALITY

REFERENCES

1. Broderick, Cuilwik, Hutchison, "Varactor up converters for satellite use," Microwave Journal, p. 57, (June 1967).

2. Watson, H. A., Microwave semiconductor devices and their circuit applications, Mc Graw-Hill, chapter 8, p. 262,(1969).

3. Grayzel, A., "The overdriven varactor upper side-band up converter," Trans. IEEE, MTT, p. 561, (October 1967).

4. Gewartowski, Minetti, "Large-signal calculations for the overdriven varactor upper-sideband up converter operating at maximum power output," Bell Systems Technical Journal, p. 1223, (July-August 1967).

The Mixer Diode

Historically, the point-contact diode was the first semiconductor device to become significant within microwave technology. It was first used in fairly primitive application as a detector, but at the beginning of the 1940's, it developed rapidly in connection with the large interest in radar techniques, in which its most important function was as a mixer diode with low internal noise. From the end of the 1940's until 1960, the mixer diode was improved only insignificantly, but around 1961, two new types of diodes appeared: the backward diode and the Schottky-barrier or Sb diode. Both were characterized by having very low flicker noise, which made them particularly in doppler radar. In the last years, mainly technological improvements have resulted in Sb diodes for increasingly high frequencies (up to 35 GHz) and point-contact diodes with increased durability for stray pulses. /138

If a diode with the non-linear I-V characteristic

$$I = I_S(e^{\alpha V} - 1)$$

is subject to a strong local oscillation signal of amplitude V_o and frequency ω_o and a much weaker signal of amplitude V_S and frequency ω_S, one can easily see that a summation and difference frequency will be created with an amplitude proportionate to V_S and, in addition, dependent upon the diode constants I_S and . Furthermore, a direct current component proportionate to V_o^2 and harmonics to ω_o will be created. In calculating the characteristics of the mixer, attention is only given to the impedance conditions in the signal frequency $_S$, in the difference or intermediate frequency $\omega_o - \omega_S$, and the summation or image frequency $\omega_o + \omega_S$. In the original work by [1], and in the summary by [2], the mixer diode is treated first as a purely non-linear conductance, and the input conductance is calculated at signal frequency, the output conductance at the intermediate frequency, and the conversion losses with the high-frequency signal's conversion to a medium-frequency signal. These quantities are shown, first of all, to be dependent upon the diode characteristic constants I_S and α, and also upon the local oscillator amplitude V_o and the possible forward voltage V_B. /139

In addition, they are influenced by the load conductance which the image frequency has. One usually identifies three cases:

1. The image frequency has the same source impedance as the signal, and this is called the broadband case.

ORIGINAL PAGE IS
OF POOR QUALITY

2. The image frequency has no load.

3. The image frequency is short-circuited.

It is of special interest to consitute that the conversion losses which are normally measured according to (1) will decreased in case (2) but increase in case (3). In case (2), therefore, a part of the signal power converted to the image frequency will be reflected and transferred to the intermediate frequency. Measurements of point-contact diodes in [2] show that the conversion losses can in this way be decreased from 5.4 dB to approximately 4.4 dB. (It must be observed that the mixer's medium-frequency resistance will increase at the same time from, typically, 360 to approximately 930 Ω, so that the adaption to the medium-frequency amplifier must be corrected.)

The idealized diode model discussed so far is only valid at low frequencies. Within the microwave range, the junction capacitance C_j and the diffusion resistance R_s become very significant, as do the diode capsule's stray capacitance C_p and the series inductance L_s. The corresponding equivalent circuit is shown in Figure 24. In addition to a change in the high-frequency impedance, an increase in the conversion losses also occurs, which will be more fully discussed later.

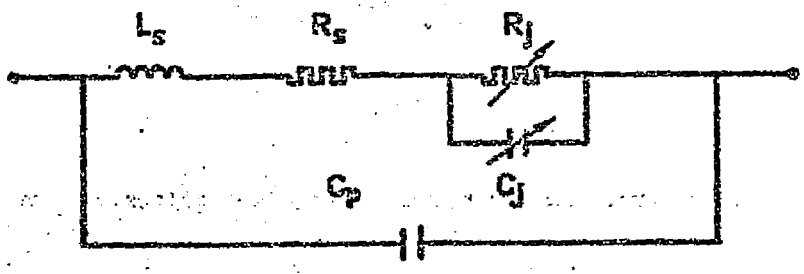


FIGURE 24.

The mixer diode's most important property, its low noise function, is characterized by a noise factor which can be written /140

$$F = L(F_{mf} + t - 1)$$

or $$F = 10 \log [L(F_{mf} + t - 1)],$$

where L = the total conversion loss,
t = the total normalized noise temperature,
and F_{mf} = the mid-frequency amplifier's noise factor (usually having a standard value of 1.41 = 1.5 dB).

ORIGINAL PAGE IS
OF POOR QUALITY

The mixer's noise factor always refers to its single-channel noise factor, unless otherwise specified. In other words, this signifies that value which is obtained by adding 3 dB to the measured value in a mixer which has equally-loaded signal and image frequency channels. Naturally, measuring the noise in a mixer with an image-frequency filter produces the single-channel value.

L is composed of several partial losses, of which the first described the reflection losses in the high-frequency and medium-frequency sides. If the standing-wave ratio there is S_1 and S_2, respectively, we obtain

$$L_1 = 10 \log \frac{(S_1 + 1)^2}{4S_1} + 10 \log \frac{(S_2 + 1)^2}{4S_2}.$$

The earlier mentioned loss resulting from the diode's parasitic elements can be written [2, 3]

$$L_2 = 10 \log \frac{P_{IN}}{P_{R_j}} = 10 \log \left(1 + \frac{R_s}{R_j} + \omega^2 C_j^2 R_s R_j\right).$$

R_j is an average value dependent upon the local oscillator power. If $R_j = 1/C_j$, the minimum value is obtained:

$$L_{2min} = 10 \log \left(1 + 2\omega C_j R_s\right).$$

The actual conversion losses in the non-linear conductance L_3, are shown in [3] to be wholly determined by $I' = d(\log I)/d(\log V)$, i.e., the derivitive of the I-V chacteristic drawn on a log-log graph. In the so-called broadband case, a point-contact diode (1N 21G) with $I' = 4.3$ produces $L_3 = 4.3$ dB, while a Sb diode with $I' = 8.8$ produces $L_3 = 3.6$ dB. The theoretical minimum limit for L_3 is 3 dB in this case. /141

A calculation in [4] shows that an Sb diode with an unloaded image frequency should produce $L_3 = 2$ dB. This should, therefore, reduce L_3 by 1.6 dB. Reductions measured so far have been about half of this, 0.8 dB.

Since the size of the total conversion loss $L = L_1 + L_2 + L_3$ dB is now being discussed, the value of the normalized noise temperature t must be examined before the noise factor can be determined. [3] indicates that

$$t = \frac{2}{L}\left[t_d\left(\frac{L}{2} - 1\right) + 1\right]$$ in the broadband case,

and $$t = \frac{1}{L}\left[t_d(L - 1) + 1\right]$$ in the case of open or short-circuited image frequencies;

where

$$t_d = t_{Lo} + t_w + \frac{K_n I_d}{f},$$

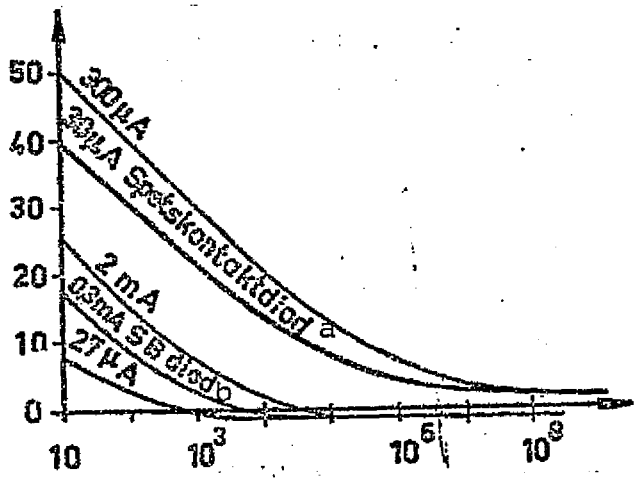
t_{Lo} = the noise added by the local oscillator. This can be ignored when using a balanced mixer.

t_w = the additional thermal and fluctuation noise. A typical value for a point-contact diode is 1.2 and for an Sb diode is 0.3.

$\frac{K_n I_d}{f}$ indicates the flicker or 1/f-noise, where I_d is the diode current in μA and f is the frequency in Hz. The constant K_n is equal to 1.8 Hz/μA in an Sb diode and approximately 1000 times larger in a point-contact diode, as shown in figure 25.

/142

The diode's normalized noise temperature t_d in dB (with $t_{Lo} = 0$)



a) Point-contact diode
b) Sb diode

Figure 25.

ORIGINAL PAGE IS
OF POOR QUALITY

Estimating the Lowest Possible Noise Factor for a Mixer

The broadband case: For a typical 2 GHz Sb diode, R_s = 11 ohm and C_j = 0.8 pF. This yields L_2 = 0.9 dB. L_1 = 0.1 dB with SVF = 1.2. Typically, L_3 = 3.6 dB. This makes L = 4.6 dB = 2.89. If the flicker noise and the local oscillator noise are ignored, we obtain t_d = 0.8. From this, we can calculate t = 0.94, and with F_{mf} = 0.5 dB = 1.12, we obtain the noise factor F = 3.07 = 4.9 dB. It should be observed that F_{mf} is assumed to be 1 dB lower than the standard value 1.5 dB, but mid-frequency amplifiers with a noise factor of 0.5 dB can already be built now.

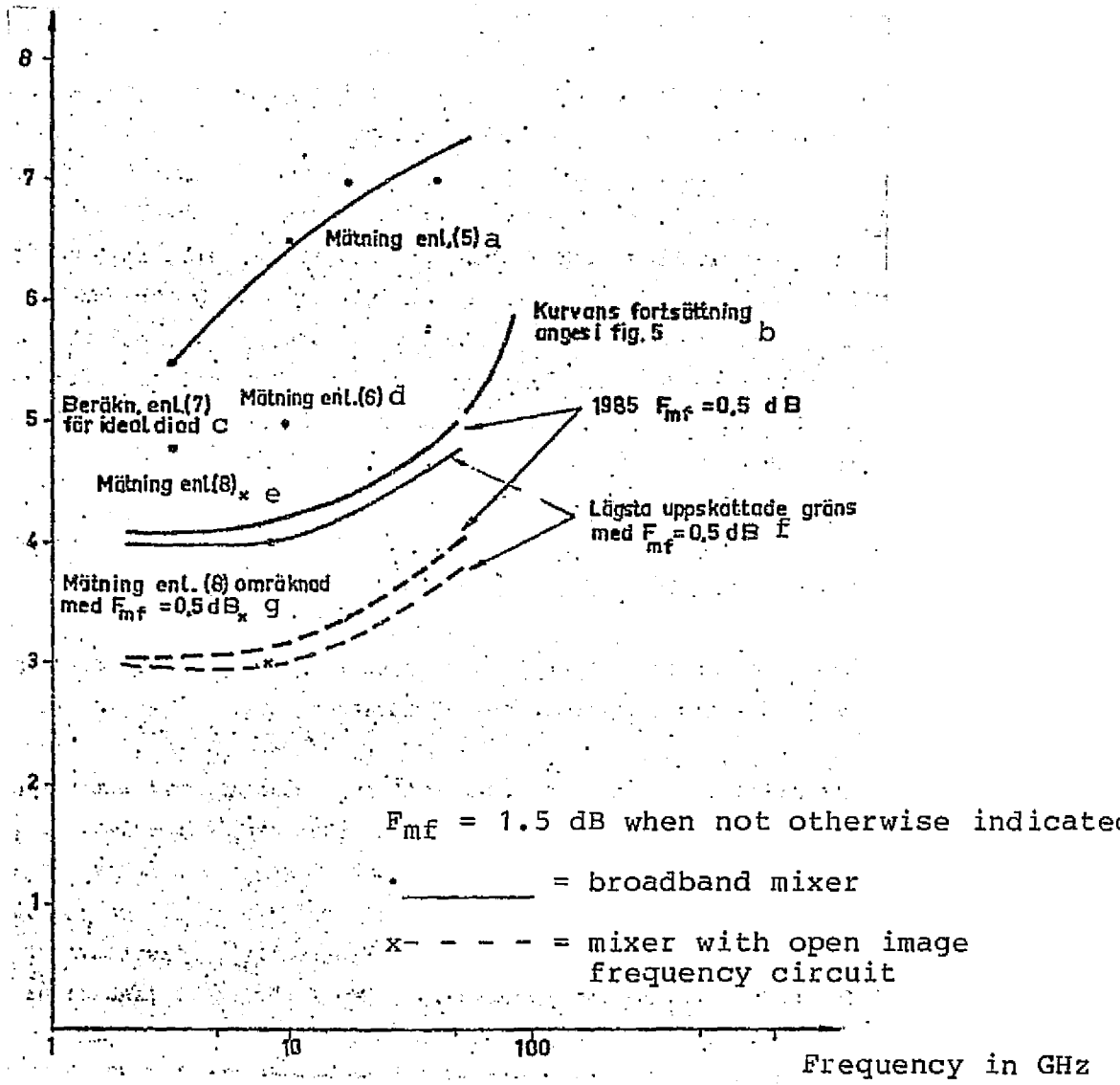
If one assumes that the product R_sC_j can be reduced to one-half, then L_2 = 0.5 dB. By further improving the I-V characteristic, L_3 can probably be reduced down to 3.2 dB, so that L = 3.8 dB = 2.4. This reduced R_s can probably produce t_d = 0.7. This would make t = 0.95 and F = 2.57 = 4.1 dB. A noise factor around 4 dB should, therefore, be assumed to be the lowest obtainable value. According to the catalog data, it is thought to be possible to hold R_sC_j and, as a result, also the noise factor close to a constant value up to the X-band. With F_{mf} = 1.5 dB, [5] indicates the following noise factors: 5.5 dB at 3 GHz, 6.5 dB at 9.4 GHz, 7 dB at 16 GHz and 7 dB at 35 GHz. [6] measured F = 5 dB at 10 GHz. These numbers are shown in Figure 26 and a curve has been drawn the estimated F_{min} values parallel to the measured values. It is assumed that the forecast values for 1985 will lie very close to the latter curve. /143

In the most recent contribution to the broadband mixer theory [7], consideration has been given to the effect of the junction capacitance varying with the LO-signal. It was found that this caused a reduction in the conversion losses of approximately 0.5 dB, but that the noise factor for the most part was unchanged (approximately 5.5 dB in the S-band). A total conversion loss of 3.3 dB was calculated, and 3.5 dB was measured. A totally idealized diode with $R_s = C_j = 0$ is assumed to produce a noise factor of 4.8 dB at 3 GHz (with F_{mf} = 1.5 dB, as usual). /144

The image frequency sees an open circuit: According to the calculations given earlier in [4], L_3 in the very best case can decrease to 2 dB. For a 2 GHz diode, L_2, as in the broadband case, will be at least 0.5 dB and L = 0.1 dB. This yields L = 2.6 dB = 1.82. With t_d = 0.7, we have t = 0.87. If F_{mf} = 0.5 dB = 1.12, the lowest possible noise factor will be F = 1.80 = 2.6 dB. The signal frequency bandpass filter which is used to reactively terminate the image frequency, produces, however, an unavoidable damping of several tenths of a decibel, which, therefore, must be added to the F-value. In this manner, the minimum value for the noise factor is approximately 3 dB.

ORIGINAL PAGE IS
OF POOR QUALITY

Single channel
noise factor in dB



a) Measurement from [5]
b) The curve's coninuation given in fig. 5.
c) Calculated according to [7] for an ideal diode.
d) Measurement from [6]
e) Measurement from [8]
f) Lowest limited estimated with F_{mf} = 0.5 dB.
g) Measurement from [8] using F_{mf} = 0.5 dB.

Figure 26.

In currently available diodes, F_3 has been measured between 2.8 and 3 dB. Using the same calculation as above, this produces F = 3.3 dB, to which the filter loss must be added, producing a total of approximately 3.7 dB.

These calculations show that an open image frequency circuit can produce a noise factor approximately 1 dB lower than in the broadband case.

There are only a few published results concerning measurements of the noise factor in mixers with open image frequency circuits. The minimum noise factor measured in [8] was 4.4 dB at 6 GHz, with F_{mf} = 1.5 dB. Recalculating with F_{mf} = 1.5 dB, the noise factor becomes approximately 3.4 dB. According to [8], the lowest possible noise factor can also be estimated. For a diode with a cutoff frequency around 500 GHz, a noise factor of approximately 3.5 dB should be obtained at 5 GHz. Using F_{mf} = 0.5 dB, and with the addition of a filter loss of 0.4 dB, we obtain f = 2.9 dB, which agrees with earlier estimates. According to unpublished results, a single-channel noise factor of /145
2 dB in the S-band was measured at MIT in 1966 using point-contact diodes in a balanced mixer. The input filter's damping, which in this case was 3 dB, was subtracted from the result of the measurement. F_{mf} was 0.5 dB. The surprisingly low result must be viewed with a certain reservation until it is verified through publication. Measurements in [2] show that the noise factor with an open image-frequency circuit, in comparison to the bandwidth case, improves approximately 0.6 dB, for example, from 8.5 to 7.9 dB. In this case, point-contact diodes were used. Therefore, it can be seen as likely that Sb diodes with lower conversion losses should be able to exhibit corresponding improvement of approximately 1 dB in keeping in agreement with the above calculations.

In terms of the use of mixer diodes at very high frequencies, where today the point-contact diode (over 50 GHz) is used exclusively, one can expect that the Sb diode, through refined production techniques, will be usuable almost as high up in the frequency spectrum. A 50 GHz Sb diode has a junction diameter of approximately 3 μ. It should be technically possible to reduce this value to 0.3 μ, which would approach the diameter of the very smallest points (0.1 μ) in point-contact diodes. An attempt to estimate the function of future Sb diodes over 50 GHz can be done based upon the way in which the conversion losses in point-contact diodes typically vary with high frequencies, as is indicated in [9]. Figure 27 shows an approximate average-value curve for measured conversion losses (the broken portion of the curve represents an extrapolation which has some support from the measurements presented in [10]).

If it is assumed, then, that the conversion losses in a

ORIGINAL PAGE IS
OF POOR QUALITY

Sb diode could be approximately 1 dB lower and t is set approximately equal to one, the noise factor F can be calculated (using F_{mf} = 0.5 dB at the lower frequencies and F_{mf} = 1.5 dB at the higher frequencies, where a higher mf is required). The result obtained, which can be viewed as an approximate forecast for 1985, is shown in figure 28.

/146

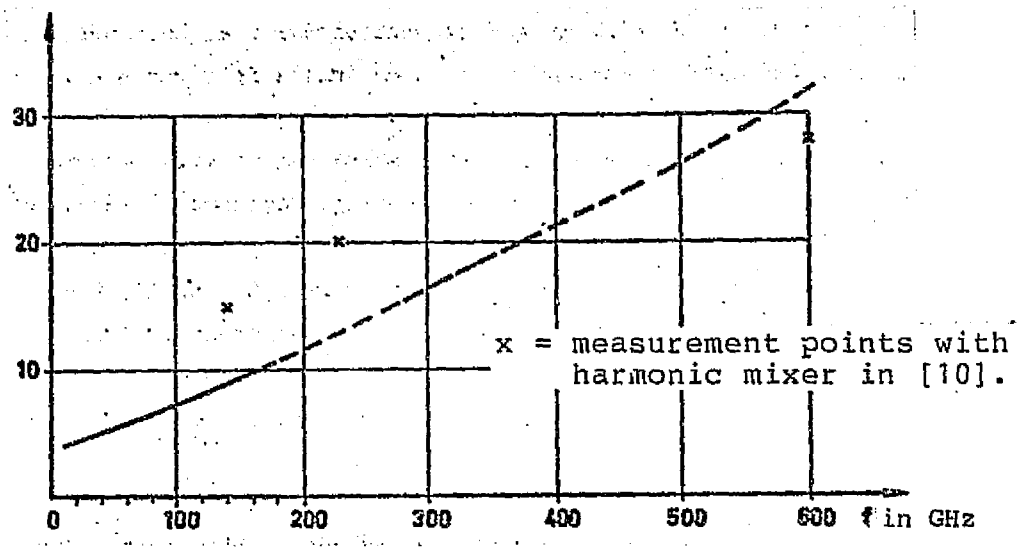


Figure 27

/147

For frequency conversion at very high frequencies where a local oscillator is not feasible, the so-called harmonic mixer can be utilized, in which the mixer crystal produces a sufficiently high LO-frequency through harmonic generation. In this case, however, the conversion losses and the noise factor will be significantly higher than in the previously discussed fundamental mixer. It should be noted that a laser can be utilized as a local oscillator at extremely high frequencies. Experiments with point-contact diodes have been done in which frequencies as high as 30,000 and 60,000 GHz (10 and 5μ wavelength respectively) have been mixed [11, 12]. These cases have used diodes of the metal-metal type in which a metal point with a diameter of approximately 1,000 A is pressed against a metal plate. Tunneling of electrons through a thin, "natural" oxide layer between them produces a nonlinear I-V characteristic and the possibility for mixing.

The possible future improvements in flicker noise given in figure 25 are very hard to predict. It can be speculated, however, that minor improvements in these values can be expected.

Noise factor for broadband mixers with Sb diodes.

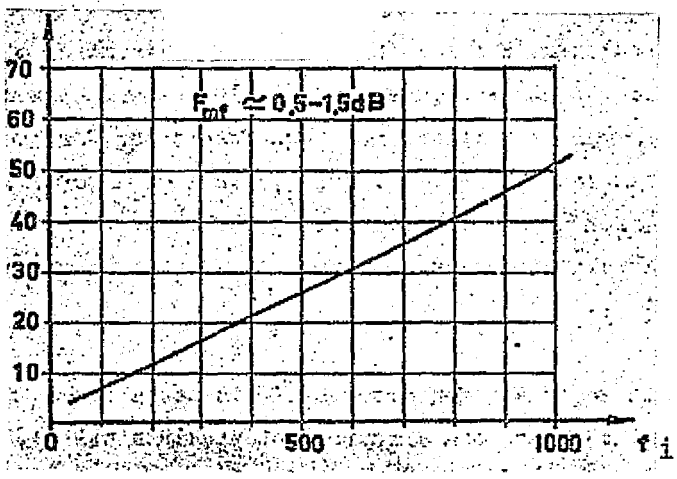


Figure 28. Performance estimate for 1985.

Maximum Power Level for a Mixer Diode

After the noise factor, the next most important characteristic of a mixer is its durability for high signal power, partially in terms of the saturation level and partially concerning the significantly higher level at which the diode's characteristics are permanently worsened.

The saturation level. With increasing signal power, the medium-frequency power will in the end decrease in relation to the desired linear relationship to the signal. The signal level which produces such a reduction or compression of 1 dB is usually called the saturation level. Saturation usually occurs when the signal power exceeds by 0.3 to 0.5 times the local oscillator power. Therefore, increased LO-power is needed to increase the saturation level. This, however, leads to increased conversion losses (both L_2 and L_3, discussed earlier will increase), and, as a result, to an increased noise factor. /148

Comparing point-contact and Sb diodes shows that with increasing LO-power, the latter maintains its lower conversion losses within a much larger range than the point-contact diode. For example, [3] reports that a 2 GHz Sb diode can normally be driven with 20 mW LO-power to 1 mW, and that the increase in the noise factor is then approximately 0.5 dB, while the increase in the saturation level is 13 dB, or to approximately +10 dBn.

Thus, it is possible to contruct a mixer with a saturation level of 20 to 40 mW, but as the level increases, one must accept an increased noise factor, which, therefore, exceeds the

ORIGINAL PAGE IS
OF POOR QUALITY

earlier discussed minimum values. It should be noted here that the so-called back diode, which, since its noise characteristics are completely comparable to those in the Sb diode, has to this point not been fully discussed, has a much smaller dynamic range. This is because the diode begins to draw current with a polarity opposed to the detection direction, at very low voltages. One possible advantage in this context is that the back diode mixer reaches maximum sensitivity at very low local oscillator power (several tenths milliwatts).

The mixer diode's power durability. Since mixers are often used in radar receivers where they are subject to stray pulses through the TR-tube, the mixer diode's durability for short high-frequency pulses (so-called spike leakage) is of special interest. Therefore, a testing method was developed early [1] using a pulse which was much shorter than the diode's thermal time constant. It was predicted, then, that only the pulse's energy content (measured in erg) would be deciding in terms of how the diode was affected. The test pulse was obtained from a charged artificial lead and its energy was chosen so that all of the "weaker" diodes were destroyed or soon worsened by one pulse and could be sorted out, while normal diodes were not markedly affected. Attempts have shown that if the diodes are subject to a large number (10^5 to 10^6) of such test pulses, they can only withstand approximately ten times lower peak energy [5].

However, practical experience from radar stations has shown /149 that the pulse test described here is not very valuable for judging a diode's durability for spike leakage. Very recent investigations [13] have shown instead that with high-frequency pulses lasting 2 to 50 ns, it is the power level during the pulse which determines how the diode is affected. This is probably because the diode is destroyed by dielectric breakdown in the epitaxial junction, which is caused by the high-frequency voltage and, consequently, by the high-frequency power. This means, among other things, that in radio stations where the high-frequency pulse has a very short climbing time, so that the TR-tube does not have time to damp its first part, diodes are destroyed particularly easy, even if the stray pulse is very short.

To increase a diode's durability for very short pulses at high high-frequency power, it is necessary to construct the diode so that, at a high power level through the injection of minority charge carriers, it quickly achieves the lowest possible high-frequency impedance, and in this way prevent a high voltage from building up. Special point-contact diodes with just this quality are beginning to be produced [13]. One such diode for 16 GHz withstood up to 50 W during a 5 ns-long high-frequency pulse, while its durability for the energy of a discharge pulse can go up to 10 erg. The corresponding figures for a 16 GHz Sb diode of GaAs are 20 W and 10 erg, respectively, and for a normal point-contact

diode (1 N78) 7.5 W and 2 erg. These numbers underscore the possibility of developing special "spike resistant" mixer diodes and since this development has produced tangible results in a short time, one can expect further major improvements, possibly also in connection with special Sb diodes. A mixer diode could possibly be able to resist up to several hundred watts of spike power. However, if this development does not come about, mixer diodes can still be effectively protected by a combination of an SM-tube and a diode limiter, which will eliminate the spike effect.

Concerning the power durability of point-contact diodes with pulses longer than 0.1 μs, [5] presents the continuous curve which is reproduced in figure 29. In this case, the diode is destroyed through thermal over-loading. Considering that a point-contact diode for 16 GHz is destroyed with approximately 1 W in-fed CW power, but that a special power-resistant diode and a Sb diode can withstand up to 4 W, one should be able to estimate the maximum powers which the latter type of diode can withstand at various frequencies by moving the continuous curve in figure 29 to a four-times-higher power level. The power durability given above for CW power is applicable to the naturally existing misadaption at high power. If high-frequency power is adapted to the diode, a X-band Sb diode will withstand 0.8 W and a point-contact diode will withstand 0.6 W, when the diodes are direct-current loaded with 10 K. In contrast, at 10 ohm load, the Sb diode can only withstand 0.16 W and the point-contact diode 0.8 W [14]. /150

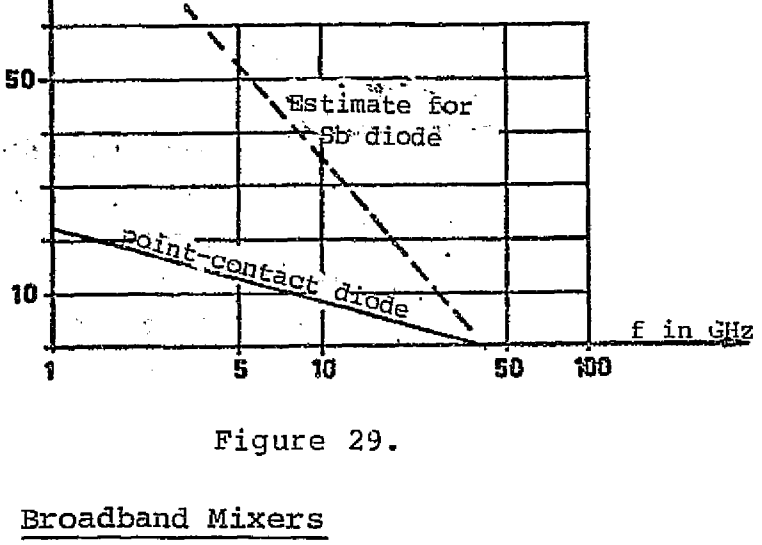


Figure 29.

Broadband Mixers

With normal point-contact diodes in coaxial circuits, it has been possible for a long time to build balanced mixers with octave bandwidth up to the X-band. (In this case, the signal and image frequencies have the same source impedance.) By using Sb diodes with an intrinsic impedance of close to 50 ohm, and in which the normal capsule is replaced by a "beam-lead" design which results in low parasitic reactances, very large bandwidths can be obtained in combination with microstrip lines. One such commercial mixer /151

ORIGINAL PAGE IS
OF POOR QUALITY

is described in [15] which covers from 1.5 to 12.4 GHz with a noise factor of 6 to 7 dB, and a similarly integrated mixer for 0.01 to 18 GHz is mentioned in [16]. Detailed aspects of the construction of such broadband circuits are given in [17], among other places.

The continuing development within this range can hardly result in increased bandwidths for frequencies under 20 GHz, but can lead to bandwidths somewhat over an octave up to 100 GHz. Naturally, broadband construction is always a compromise with the simultaneously-increasing noise factor. Mixers with the lowest possible noise factor, having a reactive image-frequency termination, should obtain maximum bandwidths of 10%.

One type of mixer which is of interest in certain special systems is the double-balanced mixer, which suppresses the reception of the image frequency within a relatively broad band. Such mixers with 20 dB of image frequency suppression have been built with octave bandwidth between approximately 0.5 and 8 GHz [18]. Continuing development should produce octave bandwidth for such mixers up to approximately 30 GHz.

It should be emphasized here that the type of mixer discussed above can be expected to have great significance in combination with a broadband, low-noise pre-amplifier, for example a parametric amplifier. Here, the noise is amplified equally in the signal and the image frequency channels, and a simple balanced mixer should then be subject to approximately twice as large noise effects as a mixer with image-frequency suppression. In this case, the difference in the total noise factor will be near 3 dB. (It is assumed here that the useful signal is only found in one channel.) Naturally, instead of using a mixer with suppressed image frequency, a narrow-band filter can be placed before a normal mixer, but then a very small-band function, in which the filter must be tuned mechanically for carious frequencies, results.

The Tunnel Diode Mixer

During the first part of the 1960's, the development of the /152 tunnel diode mixer went on at the same time as the development of the tunnel diode amplifier. The former should have the advantage in possessing good stability and being operational without a circulator. One hoped to obtain noise factors as low as in an amplifier, and definitely lower noise than in a normal mixer diode.

The tunnel diode's nonlinear I-V characteristic can produce good mixing qualities and with sufficient modulation in the negative-resistance region, the conversion losses can be reduced.

Despite the development of a large number of tunnel diode mixers, the results obtained have not been competitive with normal

diode mixers and least of all with mixers using Sb diodes. (The highly-doped tunnel diode and the back diode comprise an exception to this, in that they produce nearly the same characteristics as a Sb diode, but do not exhibit any negative resistance.) To be sure, the tunnel diode mixer produced conversion losses around 0 dB, or low amplification, but the noise factor was about the same as in a good Sb diode mixer. Furthermore, the dynamic range became less and the power durability, which in the beginning was thought to be larger than with a point-contact diode, was barely in the same class as the Sb diode. Reducing the conversion loss only for the purpose of increasing amplification is no longer of great significance, since very low-noise, compact, integrated medium-frequency amplifiers can be built with transistors.

Even if there is little interest now in the tunnel diode mixer, it is conceivable that a mixer which is in some ways similar and which furthermore utilizes the superconductor phenomenon will be very significant at very high frequencies in the future. This superconductive mixer is based on the Josephson effect and is treated in the following section.

The Josephson Junction as a Local Oscillator and Mixer /153

As previously discussed, point-contact diodes in semiconductors can be replaced by metal diodes, and when the detection mechanism is exhibited in an intermediate oxide layer, the diode can be used for mixing in the sub-millimeter range. The Josephson junction is also of use for mixing from the millimeter to the sub-millimeter range and is of very special interest since it can function simultaneously as a local oscillator [19]. (The construction of the Josephson junction was described earlier in the section titled the Josephson oscillator.)

The presence of a mixing function in a Josephson junction can be most easily observed in the I-V characteristic in the form of step-by-step current increases with constant voltage at regular voltage intervals, when the junction is subject to power at a certain frequency f_S. The Josephson junction oscillates at a frequency determined by the voltage V_B according to the equality $f_{osc} = 2eV_B/h$. When V_B is such that $f_{osc} = f_S$, the first current stage produces an additional direct current, since the difference frequency is 0. Since the junction also functions as an effective harmonic generator, the frequencies nf_S are created, and with increasing V_B beats are exhibited at even intervals with the one harmonic after the other. The size of the nth current stage is according to [20] proportionate to Bessel's function of the order $n = J_n\frac{2eV_\sim}{hfs}$, where $V_\sim$ is the equivalent applied high-frequency amplitude. Figure 30A shows the appearance of the I-V characteristic for a Josephson junction without applied high-frequency power, and figure 30B shows the changes with radiation of a certain frequency. Thus, the observation of the current stages described

here make it possible to utilize the Josephson junction as a frequency analyzer, which according to [21] is of use up to at least 1000 GHz and possibly up to 8000 GHz.

An experiment is described in [19] which shows the usefulness of the Josephson junction for routine superheterodyne receiving. If V_B is adjusted so that the difference between the oscillation frequency f_{osc} and the receiving frequency f_s is equal to a resonance frequency f_o in a resonator which is coupled to the junction, the measurements indicate that a "medium- /154
frequency signal" will be excited at f_o. A mixer of this type can be of major significance for the millimeter and sub-millimeter ranges, especially if the problems creating an effective coupling (adaption) between the signal circuit and the junction, and between the difference-frequency circuit and the junction are overcome. In addition, the construction of point-contact junctions currently in use must be improved to allow for greater mechanical stability. One can assume that a mixer of this type will be developed into an extraordinarily sensitive receiving device for extra high frequencies. The necessary cryotechniques, which in many cases are a major disadvantage, contribute to the good low-noise characteristics of this mixer.

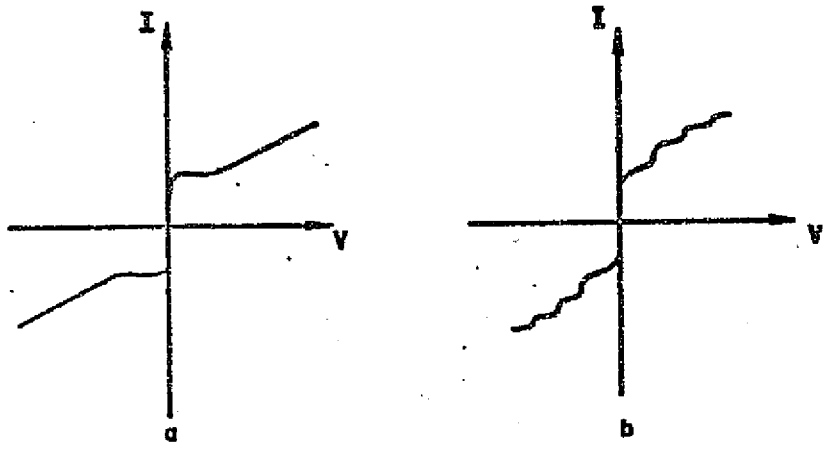


Figure 30.

REFERENCES

1. Torrey, Whitmer, "Crystal rectifiers," Radiation Laboratory Series, Vol. 15, McGraw Hill, (1948). /155

2. O'Neill, H. J., "Image-frequency effects in a microwave crystal mixer," Proc. IEE, p. 2019-2024, (November 1965).

3. Hewlett-Packard application note #907, (May 1967).

4. Strum, P. D., "Some aspects of mixer crystal performance," Proc. IRE, p. 875-889, (July 1953).

5. Sylvania, Microwave product guide, (c. 1970).

6. Texas Instruments, "Planar Schottky barrier microwave mixers and varactors in gallium arsenide," House publication, (c. 1966).

7. Leichti, C. A., "Down-converters using Schottky barrier diodes," IEEE, ED, p. 975-983, (November 1970).

8. Barber, M. R., "Noise figure and conversion loss of the Schottky barrier mixer diode," Trans. IEEE, MTT, p. 629-635, (November 1967).

9. Bauer, Cohn, Cotton, Packard, "Millimeter wave semiconductor diode detectors, mixers and frequency multipliers," Proc. IEEE, p. 595-605, (April 1966).

10. Burrus, C. A., "Millimeter-wave point-contact and junction diodes," Proc. IEEE, p. 575-587, (April 1966).

11. Abrams, Gandrud, "Heterodyne detection of 10.6-μ radiation by metal-to-metal point contact diodes," Applied Physics Letters, p. 150-152, (August 1970).

12. Sokoloff, et. al., "Extension of laser harmonic-frequency mixing into the 5-micron regions," Applied Physics Letters, p. 257, (September 1970).

13. Anand, Y., Howell, C., "The real culprit in diode failure," Microwaves, p. 36-38, (August 1970).

14. Bayliss, Cabrera, Howe, "Why a Schottky-barrier? Why a point-contact?" Microwaves, p. 34-45, (March 1968).

15. "Multi-octave balanced mixers," Microwave Journal, p. 34, (September 1969). /156

16. "Spectrum analyzer measures absolute power-- 130 dBm to 33 dBm," Microwaves, p. 50, (November 1970MI

17. Van Wagoner, R., "Multi-octave bandwidth microwave mixer circuits," 1968 GMTT Intern. Microw. Symp. Digest, p. 8.

18. "Image rejection mixers," Microwave Journal, p. 38, (October 1969).

19. Longacre, A. Jr., "The Josephson function as a 100 GHz oscillator-mixer," AD-report 704906, (April 1970).

20. Grimes, Richards, Shapiro, "Josephson-effect far-infrared detector," JAP, p. 3905, (July 1968).

21. McDonald, D. G., et al, "Harmonic generation and submillimeter wave mixing with the Josephson effect," Applied Physics Letters, p. 121, (August 1969).

ORIGINAL PAGE IS
OF POOR QUALITY

The Detector Diode

It has already been pointed out in connection with the mixer /157
diode that the point-contact diode was the first semi-conductor device to be significant within microwave technology. Its function as a detector was improved greatly during the 1940's along with the refinement of semiconductor technology. The next notable step in the development occurred around 1961 in the form of the back and Sb diodes, which possessed increased current sensitivity and improved noise characteristics. At the same time, reproductability was increased greatly in the manufacturing processes.

A detector's characteristics are determined first and foremost by its I-V characteristic. For point-contact and Sb diodes, both of which consist of a more or less perfect metal-semiconductor junction, the current's dependency upon the voltage is described by the equality

$$I = I_s(e^{\alpha V} - 1),$$

where I_S = the saturation current = the mostly constant electron current from the metal to the semiconductor;

$\alpha = \frac{e}{nkT}$;

e = the electron charge;

k = Boltzmann's constant;

T = absolute temperature in ^{o}K;

and n = a figure of merit which, for an ideal diode, is one. 1.05 is typical for Sb diodes and 1.8 for point-contact diodes.

With V in volts and n = 1.0, at room temperature we have

$$I = I_s(e^{40V} - 1).$$

The size of I_S for a Sb diode is on the order of 0.01 μA.

The expression for the I-V chacteristic in a back diode is significantly more complicated, since in this case the size of the current is determined by the tunnel effect, as illustrated /158
in Figures 13-B and E. The peak current, which is shown in C and E of the same figure, has been reduced to almost zero by combining a highly-doped n-type substrate with a low-doped p-semiconductor. This produces a widening of the layer through which the electrons must tunnel, and, consequently, a reduction in the tunnel current. Figure 31 gives a qualitative comparison of the I-V characteristics of the three diodes discussed.

When the signal voltage V_S is low, the detector current is

ORIGINAL PAGE IS
OF POOR QUALITY

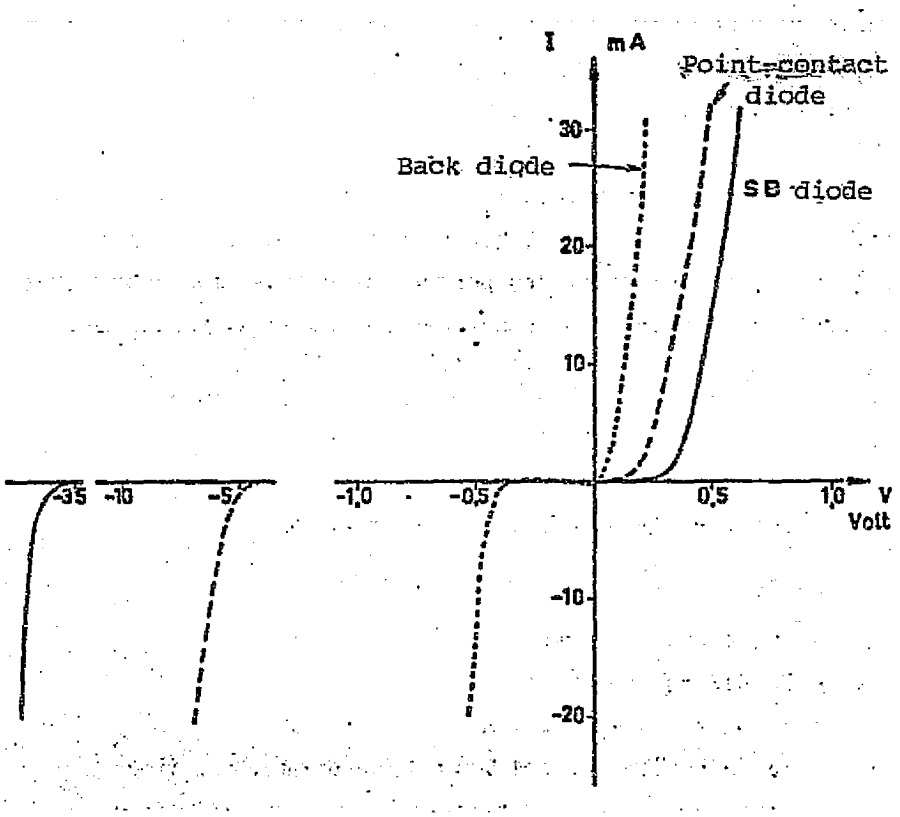


Figure 31.

proportionate to V_s^2 and, therefore, to the input power. In this case, the detector is said to have a quadratic characteristic and, for example, for a point-contact diode, the detector remains quadratic for currents up to 10 μA. At larger signal voltages, the diode functions practically as a current breaker which produces interruption during the negative half-period and short-circuits /159
during the positive half-period with a short-circuit resistance dependent upon the slope of the linear part of the I-V curve. For such a so-called linear detector, which is normally used for envelope-detection of an amplitude-modulated signal, consideration does not need to be given to the diode's noise chacteristics. This is so because the requirements for such a detector are restricted to: the largest possible linear region in the I-V curve, low series resistance, high back resistance, and high breakdown voltage. It can be quickly determined here that the Sb diode is superior to the other diode types in these regards. The requirements for the chacteristics of a quadratic detector are more complicated, since consideration must be given to the diode's intrinsic noise, and in the future, only this type of detector will be treated.

ORIGINAL PAGE IS
OF POOR QUALITY

One of the most important chacteristics of a small signal detector is the current sensitivity β. This is defined as the ratio of the increase in the short-circuit current to the input high-frequency power, and is written

$$\beta = \frac{\Delta I}{P_{IN}}.$$

For small signals in the active semiconductor junction, the following applies:

$$\beta = \frac{\alpha}{2} = \frac{e}{2nkT}.$$

In addition to the active junction or video resistance R_j, the detector has two unavoidable parasitic elements, the junction capacitance C_j and the diffusion resistance R_s. Because of these, β is reduced to a value which can be easily derived from the detector's equivalent circuit, shown in Figure 32. The following expression is then obtained:

$$\beta = \frac{\alpha}{2} \cdot \frac{1}{\left(1 + \frac{R_s}{R_j}\right)} \cdot \frac{1}{1 + \frac{\omega^2 C_j^2 R_j^2 R_s}{R_j + R_s}} = \beta_o \cdot \frac{1}{1 + \frac{\omega^2 C_j^2 R_j^2 R_s}{R_j + R_s}}.$$

It is generally true that β is dependent on the forward bias and reaches a maximum value when the I-V curve is most non-linear. Figure 31 shows that this maximum is obtained without forward bias in a back diode, but with approximately +0.3 V forward bias in a Sb diode. For a more complete calculation of the variation /160 in β with the forward bias V_d and the corresponding current I_d, the following expressions for R_j and C_j can be used:

$$R_j = \frac{1}{\frac{dI}{dV}} = \frac{nkT}{e(I_d + I_s)}$$

$$C_j = \frac{C_{jo}}{\left(1 + \frac{V_d}{V_b}\right)^{\frac{1}{2}}},$$

where V_b = the diffusion potential,
and C_{jo} = the junction capacitance with 0 V.

With low forward bias, when the derivitive of the I-V curve is small and, therefore, R_j is large, one can state

$$\beta = \frac{\alpha}{2} \cdot \frac{1}{1 + \omega^2 C_j^2 R_j R_s}.$$

If one assumes that the active diode surface has a radius r, then $C_j \sim r^2$, $R_j \sim 1/r^2$, and $R_S \sim 1/r$, which yields $C_j^2 R_j R_S \sim r$. Therefore, to maximize β at a certain frequency, r must be made very small. At the same time, consideration must be taken so that R_j does not become too big, which can lead to an unsuitably high time constant in the video circuit. Typical values for β_O for a point-contact, Sb, and back diode are respectively 12, 20 and 30 μA/μW.

To characterize a detector's noise characteristics when it /161
is followed by a video amplifier, one has originally defined a factor of merit M which comprises one component in the signal-noise ratio in the video amplifier's output. It can then be shown that [1]

$$\text{The signal-to-noise ratio} = M \frac{P_{IN}}{\sqrt{4kT_o B}}$$

where $$M = \frac{\beta R_j}{\sqrt{R_j + R_A}}$$

R_A = the equivalent noise resistance in the video amplifiers input, normally 1200 ohm;
B = the video amplifier's effective bandwidth.

This expression for M applies given that the noise from the detector's R_j is purely thermal (R_S is ignored). If an ideal transformer with conversion ratio n is in-coupled between the detector and the video amplifier, R_A in the expression for M will be reduced to R_A/n^2, which, with large n approaches zero. Therefore, a short-circuit value for M is sometimes defined as

$$M_k = \beta\sqrt{R_j}.$$

In cases where a certain direct current is applied to a forward-biased detector, the noise effect from R_j will increase to the value $t4kT_oBR_j$ because of the presence of fluctuation and flicker noise. (T is the normalized noise temperature.) The detector's figure of merit then takes the form of

$$M = \frac{\beta R_j}{\sqrt{tR_j + R_A}}.$$

One increasingly common dimension of a detector's sensitivity is the tangential sensitivity. This indicates the signal power level at which the signal-noise ratio, measured in voltage, is approximately 2.5 at the video amplifier's output. This yields

$$M\frac{P_{in}}{\sqrt{4kT_oB}} = 2.5 \quad \text{or} \quad P_{IN} = \frac{31{,}5 \cdot 10^{-11}\sqrt{B}}{M}. \qquad /162$$

ORIGINAL PAGE IS
OF POOR QUALITY

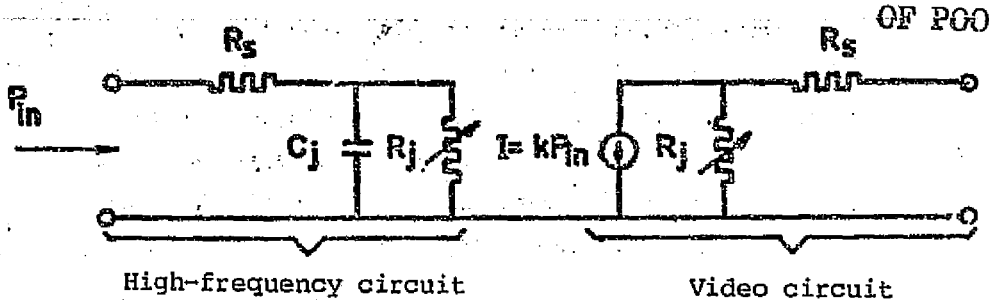


Figure 32.

With B expressed in MHz and P_{IN} expressed in dBm, the tangential sensitivity becomes

$$T_s = -35 - 10 \log M + 5 \log B_{MHz} \quad \overline{dBm},$$

which is normally stated for B = 1 MHz, and is then written

$$T_{s_{1MHz}} = -(53 + 10 \log \frac{M}{60}) \text{ dBm}.$$

[2] gives a somewhat different expression for the tangential sensitivity, which is used in the evaluation of point-contact and Sb diodes:

$$T_s = 10 \log \left[\frac{5\sqrt{4kT_o}}{\alpha} \frac{\sqrt{R_s + R_j}}{R_j} \sqrt{\frac{R_j + R'_A}{R_j}} \right]$$

$$+ 10 \log \left[1 + \frac{\omega^2 C_j^2 R_s R_j^2}{R_s + R_j} \right] + 5 \log B \text{ dBm}$$

where R'_A is the equivalent noise resistance representing both the video amplifier's internal noise and the flicker noise.

$$R'_A \simeq R_A + \frac{1{,}1 \cdot 10^5}{B} \log \frac{f_2}{f_1}$$ (when the diode current < μA and B is in Hz)

where the video bandwidth is defined as $B = f_2 - f_1$. It can be immediately seen that the second term in the expression for R'_A representing the flicker noise with normal R_A and B values produces an almost-ignorable increase in noise. For example, with $B \approx f_2 = 10^5$ Hz and $f_1 = 10$ Hz, an equivalent flicker-noise resistance of 4.4 ohm is produced. Then, it is also apparent that the earlier standard value of $R_A = 1200$ ohm must be reduced in order to utilize the Sb diode's low internal noise. If R_A is reduced by a transformer-coupling, so that $R_A \approx R'_A = 100$, then, according to [2], nearly the same sensitivity level is obtained as when $R'_S = 0$.

To compare the various detector diode characteristics, [3] /163
has set $R_A = 0$, so that the contribution of the video amplifier is not calculated, and calculated T_S for B = 1 Hz. The expression for T_S can then be simplified to

$$T_{s_{1Hz}} = 10 \log \frac{5 \cdot 10^3}{a} \sqrt{4kT} \frac{\sqrt{R_s + R_j}}{R_j} \sqrt{t_w\left(1 + \frac{K_n I_d}{f_v}\right)}$$

$$+ 10 \log \left\{ 1 + \frac{\omega^2 C_j^2 R_s R_j^2}{R_s + R_j} \right\} \text{ dBm}$$

where f_v = the video frequency.

As in the mixer diode, the normalized noise temperature from thermal and fluctuation noise is indicated by t_w. K_n is a flicker-noise constant approximately equal to 1.8 Gz/μA for a Sb diode and 4000 Hz/μA for a point-contact diode. Since t_w is approximately equal to one, we can set

$$t_w\left(1 + \frac{K_n I_d}{f}\right) \simeq t_w + \frac{K_n I_d}{f} = t_d.$$

T_d as a function of the video frequency is given in Figure 25.

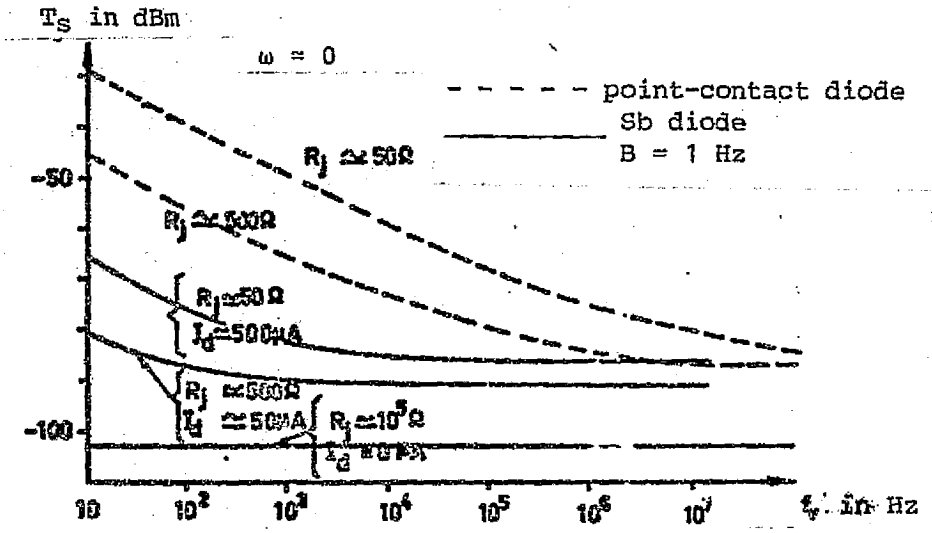
To calculate T_{s1Hz}, it is appropriate to state

$$T_{s_{1Hz}} = -78 - 10 \log \frac{R_j}{\sqrt{R_s + R_j}} + 10 \log n + 5 \log t_d$$

$$+ 10 \log \left\{ 1 + \frac{\omega^2 C_j^2 R_s R_j^2}{R_s + R_j} \right\} \text{ dBm}$$

With low high-frequencies ($\omega = 0$), maximum sensitivity is reached with maximum R_j, which without bias is approximately 10^5 ohm. When n = 1, $T_{s_{1Hz}} = -103$ dBm. However, it is practically impossible to adapt both the high-frequency and the video circuits to this high impedance, but the obtained value can be seen as an absolute minimum value. For a more realistic, adaptable R_j of /164
approximately 500 ohm, which is obtained with $I_d = 50$ μA, $T_{s_{1Hz}}$ becomes approximately -91 dBm, ignoring the flicker noise (this is applicable at sufficiently high f_v). In the construction of very broadbanded detectors, R_j must be reduced down to 50 ohm (commercial detector holders are now available with bandwidths of 10 MHz to 12.4 GHz), for which a current I_d of approximately 500 A is required. With $R_s = 10$ ohm, we obtain in the same way as before $T_{s_{1Hz}} = -86$ dBm. The contribution from the flicker noise = log t_d as a function of the video frequency f_v, is then taken from Figure 25 in the section titled "The Mixer Diode." The tangential sensitivity of today's Sb and point-

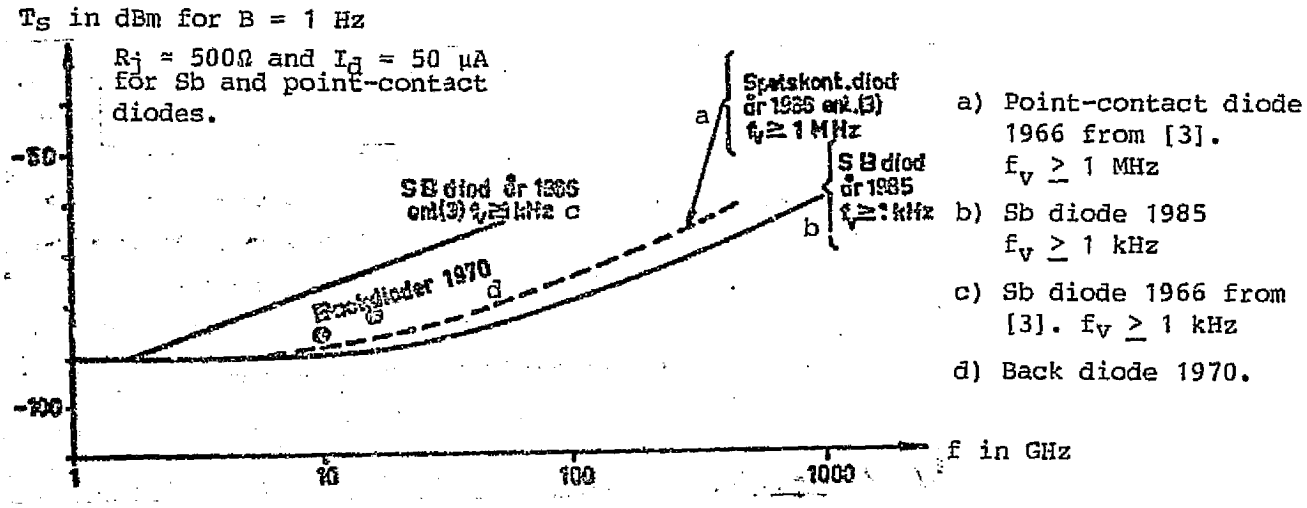
contact diodes have been calculated in this manner and are shown in Figure 33-A.



/165

ORIGINAL PAGE IS
OF POOR QUALITY

Figure 33A.



a) Point-contact diode 1966 from [3]. $f_v \geq 1$ MHz

b) Sb diode 1985 $f_v \geq 1$ kHz

c) Sb diode 1966 from [3]. $f_v \geq 1$ kHz

d) Back diode 1970.

Figure 33B.

A careful calculation of the tangential sensitivity when $\omega \neq 0$ is complicated by the fact that three terms in the previously-given expression for T_S vary with I_d, which means that for every ω, there is an optimum I_d value. However, these first become significant at very low f_v values and are ignored in the following estimate. In order to obtain the best possible sensitivity with a detector diode at high-frequency, C_j and R_S must be minimized. For this reason, the point-contact diode today is superior to the Sb diode at very high f (equal to $\omega/2\pi$), except at very low f_v, because of its significantly

lower C_j. This is treated in detail in [3]. As previously discussed in connection with mixer diodes, it is likely that Sb diodes in the future will be produced with C_j as low as in a point-contact diode; in other words, approximately 0.1 pF. Using the date given in [2] and [3], a corresponding R_S value of 5 ohm can be assumed. Significant reductions in t_d as a function of f_v and I_d can hardly be expected in the future. R_j, according to the expression given, is determined by I_d. Using these assumptions and setting I_d = 50 μA, T_S has been calculated as a function of f. The results are shown in Figure 33-B. Detectors with this performance level will probably be produced long before 1985, but further improvements cannot be expected after that, and so Figure 33-B shows a forecast for 1985.

The same figure shows the performance level for diodes in 1966, according to [3], and, for the most part, the data is the same in 1970. Observe the difference in the video frequency f_v for point-contact and Sb diodes. /166

Concerning the use of point-contact diodes at extremely high frequencies, it should be noted here that they are used for the detection of up to 3000 GHz, from both incoherent heat sources [4] and from gas lasers [5]. In [5], the tangential sensitivity has been measured at f_v = 100 Hz and 1 Hz bandwidth as -36 dBm with 1000 GHz, and -23 dBm with 3000 GHz. These results are somewhat worse than an extrapolation of the broken line in Figure 33-E gives [3]. The detector's time constant has been estimated at less than 1 ns.

It is necessary for back diodes to exhibit low R_j (≈ 100 ohm) and high tangential sensitivity ($T_{S_{1Hz}} \approx$ -89 dBm) without bias, so that the addition from the flicker noise is null. However, the sensitivity worsens with increasing high frequency more rapidly than in other diodes, due to the unavoidable increasing of C_j. According to the data in [3], one can estimate a worsening of approximately 3 dB already at 10 GHz in comparison to a point-contact diode (and a future Sb diode, ignoring the flicker noise). Several typical values for a back diode are shown in Figure 33-B. Thus, the back diode will continue to be of great interest in connection with the construction of broad-band detectors up to 10 to 20 GHz in which very low video frequencies are significant. However, the dynamic range will be less than in an Sb diode, as will be shown.

The Detector's Quadratic Range

The upper limit for a detector's quadratic range was defined in [3] as the power, P_{kv} at which the detector function deviates 0.3 dB from a pure quadratic characteristic. It has been shown for the three previously discussed diodes that with low ω values,

the following expression approximates P_{kv}:

$$P_{kv} = \frac{140}{\beta_o^2 R_v}\left(1 + \frac{R_s}{R_v}\right) \text{ mW.}$$ /167

With R_v = 100 ohm, a point-contact diode yields approximately -30 dBm, an Sb diode -24 dBm, and a back diode -28 dBm. The corresponding values for the dynamic range calculated from T_{s1Hz} are approximately 58 dB, 64 dB and 61 dB, respectively. The last value (for the back diode) is probably too high, since the calculation of P_{kv} produced an altogether-too-high power level. According to [3], measured P_{kv} values lie at -36 dBm. By choosing the average of the special measured value and the calculated value, P_{kv} = -32 dBm, the dynamic range of the back diode becomes approximately 37 dB. Thus, these approximate figures show that the Sb diode has the largest dynamic range (with low ω and high f_v). However, with f_v values under 100 Hz, the back diode is best in this respect. Major improvements of the dynamic range can hardly be expected for any diodes.

Concerning the resistance of detector diode to overloading, the discussion in the section titled "The Mixer Diode" is applicable to a certain degree. However, under normal conditions a detector diode is destroyed by pulses of several s in length, not by short spike-leakage. Therefore, it is completely natural to specify their power durability by a maximum peak-power level. For example, if the diode is adapted to a low power level, a point-contact diode with a range of 10 to 40 GHz will withstand approximately 100 mW of both peak and CW power. A X-band SB diode will withstand 200 mW CW power and close to 1 W peak power.

Less Common Types of Detectors

/168

In addition to the three previously discussed detector diodes, there are several less common types which can be very significant in the future. To this belong the Josephson detector and two others given in [3], whose function has so far been only scantly researched. They are called the thermoelectric detector (TED) and the space-charge-limited dielectric diode (SCLD).

The Josephson Detector

The Josephson junction, which was described in the chapter on generation and frequency conversion, can also be used as a broadband video detector up to very high frequencies. [6] shows how an alteration of the zero-voltage current by the effects of high-frequency radiation can be utilized toward this goal. Figure 34 illustrates schematically the way in which the I-V curve is changed by the effect of high-frequency radiation, and how the detector voltage varies with a pulsed signal. Such a detector can be very broadbanded. For example, niobium point-contact junctions have been used from 75 to 850 GHz. The maximum sensitivity of this material is around 150 GHz. A measurement of the detector's sensitivity at 75 GHz produced with 1 GHz amplifier bandwidth a value of approximately $5 \cdot 10^{-13}$ W ("noise equivalent power"). Converted to tangential sensitivity, still with 1 Hz bandwidth, this becomes T_s = -89 dBm. It should be observed that this high sensitivity is obtained without any special adaption of the signal to the detector, and at a frequency at which the niobium detector does not have maximum sensitivity. (Compare to figure 33B.) The measurements have also shown that the Josephson detector has a very low time constant, in each case less that 10 ns. According to [7], the detector's sensitivity can be increased an additional 20 dB by strongly coupling it to a resonator with a high Q-value, and by choosing the forward bias V_B so that the Josephson junction oscillates at the cavity's resonance frequency, f_o, or, in other words, so that $V_B = hf_o/2e$. At the "resonance" voltage, a current-increasing step in the I-V curve is achieved and the height of the step varies with the high-frequency power.

With an appropriate direct current through the junction, then,/169 a varying high-frequency power can be transformed into a corresponding detector voltage, in agreement with figure 34. The increased sensitivity can then be explained by the I-V curve's (dV/dI) being much larger than the zero-voltage step in figure 34. Of course, such a detector will be narrow-banded, but should be easily mechanically tuned.

The very good results already obtained with the Josephson detector indicate that it can be of great significance in millimeter and sub-millimeter wave techniques in the future in situations where the need for cooling with liquid helium does not get in the

way. The broadband type of detector will possibly produce tangential (T_s) values which lie 25 to 30 dB below the lowest curve in figure 33B. Resonance detectors can be expected to be 10 to 15 dB better. These values apply without regard to the flicker noise, of which very little is known.

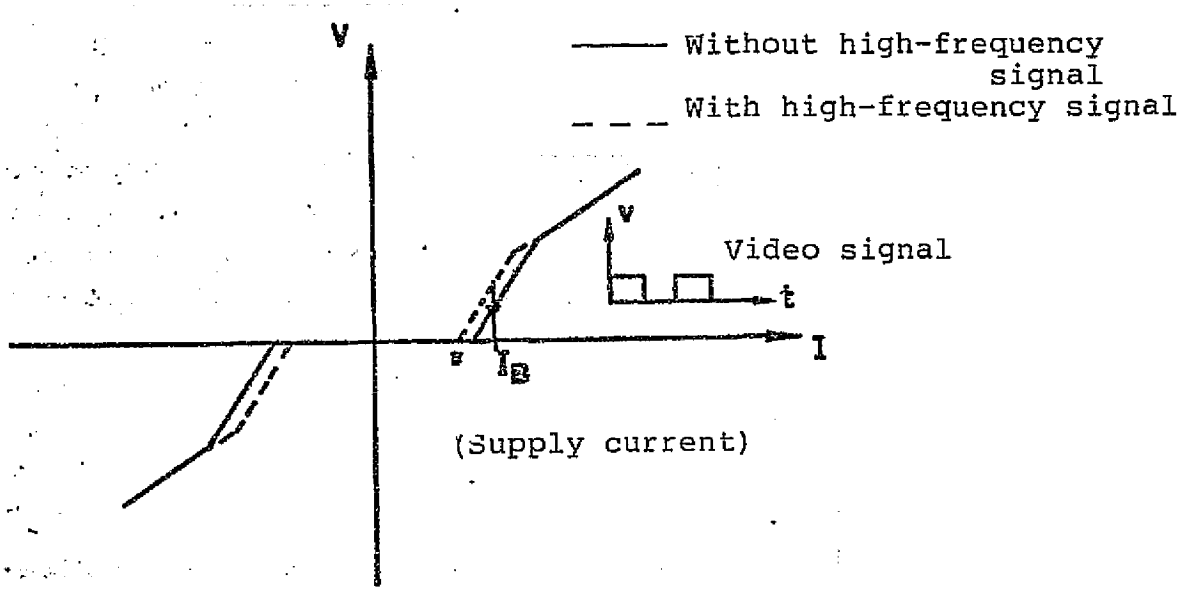


Figure 34.

ORIGINAL PAGE IS
OF POOR QUALITY

The Thermoelectric Detector

The TED consists of a very fine metal point which produces a pure ohmic contact with the center of the semiconductor's plane side in the form of a hemisphere. The high electric field strength impinging on the point-contact produces in the closest electrons a high energy which causes them to wander in the semiconductor, so that the contact receives a positive voltage. The TED functions as a very high, pure resistance on both the high-frequency and the video side. The output voltage is directly proportionate to the input power up to -5 dBm. The upper cutoff frequency is determined by the relaxation time of the excited electrons and is estimated as over 1000 GHz for silicon. There is no flicker noise. For a TED with a relatively low video resistance of 100 ohm, T_{s1Hz} is calculated as -59 dBm, and the dynamic range is 54 dB. The tangential sensitivity can be made comparable to normal detector diodes, but only at the cost of a greatly increased video resistance (up to 100 K), which leads to extremely small bandwidth in both the high-frequency and the video circuits. It is assumed that the TED's durability for high power levels is greater than in a point-contact diode. /170

In addition to the point-contact and the Sb diode, it is possible that the TED will be of great significance in the future for detection of sub-millimeter waves, particularly in combination with the need for low flicker noise.

The Dielectric Detector

The space-charge-limited dielectric diode, SCLD, consists of a highly resistive semiconductor between two ohmic contacts. Calculations of such a diode with 0.5 V forward bias and 500 ohm resistance has produced $T_{s1Hz} = -79$ dB and $P_{kv} = -6$ dBm; or, in other words, an unusually large dynamic range of 73 dB. The flicker noise characteristics are unknown, and it is also uncertain if the SCLD can be used at high frequencies. Experiments done so far have only covered several GHz. The future prospects for the SCLD are, therefore, very hard to evaluate.

Combinations of Pre-Amplifiers and Detectors /171

In the so-called straight receiver, consisting of a detector diode with a post-coupled video amplifier, the demand for a very broadbanded input of high-frequency circuit is of primary importance. The sensitivity is poor in comparison to a superheterodyne receiver, but its simple construction has the advantage of being less expensive. When special requirements for a very high sensitivity are combined with the monitoring of a broad frequency range, a broadband, low-noise amplifier can be in-coupled before the detector. A high-frequency signal can them be amplified 20 dB or more, with only an insignificant increase in the signal-to-noise

ratio. Therefore, an increase in the tangential sensitivity of the same order can be expected.

Both tunnel diode amplifiers and parametric amplifiers can be used, and will produce high-frequency bandwidths up to one octave. These combinations of low-noise amplifiers and detectors will probably receive increased utilization in the future through the application of integrated circuit technology, which will make the amplifiers both smaller and less expensive.

A complete examination of the noise conditions in a pre-amplifier-detector combination is presented in [8]. Through mixing in the detector, the noise amplified within the pre-amplifier's bandwidth B_1 will produce a power spectrum which is maximum at zero frequency and which then decreases lineally with increasing frequency, to a zero value at the frequency B_1. Furthermore, within the range from zero to $0.5\ B_1$, a constant noise addition, originating from the mixing of the amplified signal and the noise, is added. One can easily see, then, that the video amplifier will receive the whole noise spectrum if its bandwidth, B_2, is greater than or equal to B_1. Furthermore, one expression for the noise after the video amplifier can be derived for $0.5\ B_1 < B_2 < B_1$, and another when $0 < B_2 < 0.5\ B_1$. The tangential sensitivity for the whole receiving system can be written in the following manner if the detector is quadratic:

/172

$$T_{s_{B_2}} = \frac{F_1 \gamma k T_o B_1}{2} \left(1 + \sqrt{\frac{4}{\gamma}\left[\frac{2B_2}{B_1}\left(1 - \frac{B_2}{2B_1}\right) + D\right]}\right)$$

when $0{,}5\ B_1 \leq B_2 \leq B_1$

ORIGINAL PAGE IS OF POOR QUALITY

and

$$T_{s_{B_2}} = F_1 \gamma k T_o B_2 \left(1 + \sqrt{\frac{2B_1}{\gamma B_2}\left[1 - \frac{B_2}{2B_1} + \frac{B_1}{2B_2} D\right]}\right)$$

when $B_2 \leq 0{,}5\ B_1$

where

$$D = \left(\frac{2}{\beta \sqrt{R_j} F_1 G_1 \frac{B_1}{B_2}}\right)^2 \frac{F_2 + t - 1}{k T_o B_2}$$

and

F_1 = the pre-amplifier's noise factor,
G_1 = the pre-amplifier's amplification,
B_1 = the pre-amplifier's bandwidth,
F_2 = the video amplifier's noise factor,
B_2 = the video amplifier's bandwidth,
β = the detector's current sensitivity,

R_j = the detector's video resistance,
t_j = the detector's normalized noise temperature (normally, approximately 1),
$kT_o = 4 \cdot 10^{-21}$ W/Hz,
γ = the signal-to-noise power ratio in the video amplifier's output = 6.25 = 8 dB. (Is often otherwise given as a V-ratio, 2.5 = 4 dB.)

In most normal applications, $B_2 << B_1$ and the product F_1G_1 is seldom larger than one thousand. With $\beta\sqrt{R_j} = M_k$ = the diode's factor of merit, one can then state

$$T_{s_{B_2}} = 3{,}16 \cdot 10^{-7} \frac{\sqrt{B_2 F_2}}{G_1 M_k} \text{ mW.}$$

Thus, the tangential sensitivity is independent of F_1 and inversely proportionate to G_1. As has already been noted, it is, therefore, possible with a normal low-noise amplifier before the detector to improve the tangential sensitivity by 15 to 20 dB.

ORIGINAL PAGE IS
OF POOR QUALITY

REFERENCES

1. Torrey, Whitmer, "Crystal rectifiers," MIT Radiation Lab Series No. 15, McGraw-Hill, (1948). /173

2. HP Application Note 907, (May 1967).

3. Cowley, Sorensen, "Quantitative comparison of solid-state microwave detectors, IEEE Trans. MTT, p. 588-602, (December 1966).

4. Happ, V. H., et. al., "Der Kristalldetektor als Empfanger thermischer Strahlung im Gebiet von 100 - 1000 μ Wellenlange," Z. Naturforsch., p. 522, (1957).

5. Becklake, Cram, Prewer, "Spectral response of pointcontact diode in the submillimeter range," Electronic Letters, p. 601, (December 1968).

6. Grimes, Richards, Shapiro, "Josephson-effect far infra red detector," JAP, p. 3905, (July 1968).

7. Richards, Sterling, "Regenerative Josephson effect detector for far-infrared radiation," Applied Physics Letters, p. 394, (June 1968).

8. Lucas, W. J., "Tangential sensitivity of a detector video with r. f. preamplification," Proc. IEE, p. 1321, (August 1966).

Electronic Switching

The Breaker Diode

General Information on the High-Frequency Breaker /174

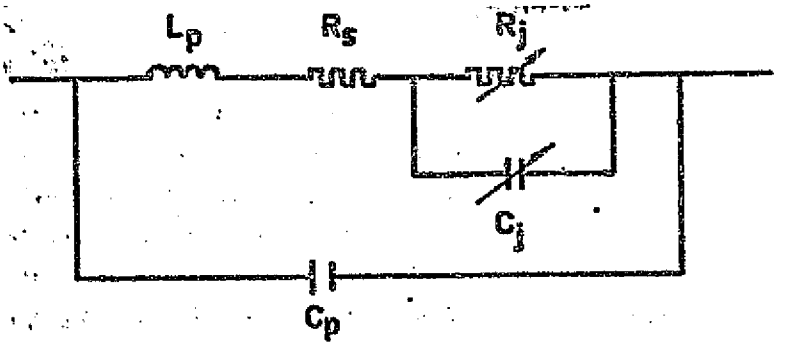
Series-coupled or shunt-coupled semiconductor diodes are used for electronic breaking or switching of high-frequency energy in a transmission line. The impedance of these diodes can be altered from a very high to a very low value by using a control voltage. The damping from a series-coupled diode with a total impedance of $Z_D = R_D + jX_D$ in a wave-guide with the characteristic impedance Z_o can be written as

$$\alpha = 10 \log \left((1 + \frac{R_D}{2Z_o})^2 + (\frac{X_D}{2Z_o})^2\right) \text{ dB.}$$

For a shunt-coupled diode with a total admittance of $Y_D = G_D + jB_D$ and $Y_o = 1/Z_o$, we obtain

$$\alpha = 10 \log \left((1 + \frac{G_D}{2Y_o})^2 + (\frac{B_D}{2Y_o})^2\right) \text{ dB.}$$

If one wishes to further study the size of the damping changes obtained with various types of diodes, one must be familiar with their equivalent circuits and the way in which the active junction's impedance is altered with the control voltage. There are two basic types. Type I consists of PN and Sb diodes, and type II consists of PIN diodes. Even though PN and Sb diodes are produced in different manners, their I-V and C-V curves are fairly similar, and so they both can be represented by the equivalent circuit in Figure 35. The PIN diode has highly-doped P and N sides and between them an undoped, high-ohmic semiconductor junction (the I refers to the word "intrinsic"). Its equivalent circuit is reproduced in Figure 36.

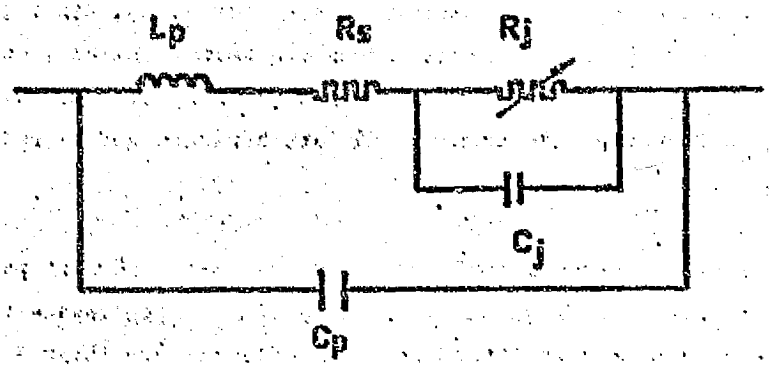


ORIGINAL PAGE IS
OF POOR QUALITY

FIGURE 35

When the bias in the diode increases from a negative value up to somewhat over 0 V, R_j in Figure 35 is large and can be ignored, while C_j increases rapidly near 0 V. With further increases in the forward bias, R_j decreases rapidly and produces an effective short-circuiting of the active layer. In a breaker, the extreme /175 value of the junction impedance is utilized; or in other words,

a low capacitance or a short-circuit. However, the reactance variations in between can be utilized in an analog phase shifter to produce a continual alteration of the phase in the reflected signal. However, the effective value of the junction capacitance is dependent upon the signal at large amplitudes, and in this way, the phase is also amplitude-dependent.



ORIGINAL PAGE IS
OF POOR QUALITY

FIGURE 36

Despite its similar equivalent circuit (Figure 36), the PIN diode has completely different characteristics. Because of the relatively large thickness of the I junction, C_j is very low, and furthermore, is dependent upon the forward bias at high frequencies. Near 0 V, R_j is large (typically several kiloohm), but decreases /176 inversely proportionate to the control current. Typically, R_j is around 1 ohm at 10 to 100 mA forward current. As the forward bias becomes negative, R_j will increase and can usually be ignored along with C_j. Since R_S is especially low in a PIN diode, the diode is able to produce very high impedance with low losses. The size of R_j is completely independent of the signal amplitude over the diode, but the amplitude must naturally be limited so that the diode's breakdown voltage is not exceeded.

To produce a good diode breaker, the lowest possible line damping in the front portion and the highest possible insulation in the back portion are sought. In principle, this requires a small R_S value and a large change in the active junction's impedance. If the good breaker function is to be unchanged within a large frequency band, the parasitic elements L_p, C_p (and C_j) must be small. If these are not small, or if the frequency is so high that the attendant reactances become large, good breaking can still be maintained within a narrow band by switching the diode between series and parallel resonances. With negative bias and the appropriate frequency, C_j, L_p and R_S will build a series-resonance circuit (according to Figures 35 and 36). With forward bias, when the active junction is short-circuited, L_p, C_p and R_S can create a parallel-resonance circuit at the same frequency. The result is then a resistive change from approximately R_S and up to $(\omega L_p)^2/R_S$. The damping changes can then be easily calculated with the previously-given expression for α.

The determination of which type of diode to use in a breaker

will be based first of all on the desired performance in terms of power durability and switching speed. The PIN diode is the most appropriate for high power levels (over 1 W), since it has a high breakdown voltage and a large cross-sectional surface which can absorb high currents. Increased power durability requires increased I-junction thickness, and this increases the diode's switching time. By overdriving the control voltage so that the charging and discharging currents are large, the switching time can be reduced very significantly. As an example of this, it is mentioned that diodes with the breakdown voltages 1000 V and 100 V typically /177
have a switching time from low to high impedance of 6 μs and 8 ns, respectively. These figures are in approximate agreement with the general rule stating that the product of the maximum power ($\sim V_B^2$) and the switching speed (~inverted switching time) is approximately constant. Particularly thin PIN diodes have produced high-frequency pulses with rise and fall times less than 1 ns. These short switching times are normally only obtained with point-contact, Sb and, possibly, varactor diodes.

It should be observed that special filtration problems occur in connection with extreme short-pulse modulation. If the upper portion of the control pulse's spectrum corresponds to the high-frequency band, an undesirable leakage occurs between the signal and the video circuits which cannot be halted by any filter construction. Sufficient uncoupling between the two circuits must then be obtained with some form of balanced bridge connection.

To summarize, it can be said that PN and Sb diodes are used today for signal powers under 1 W, and can easily produce switching times shorter than 1 ns. PIN diodes are normally used for very high powers (up to 1 MW in peak power) and with longer switching times, but they can be designed as desired to produce about the same function as PN and Sb diodes.

Because of its very good performance and its large dimensional flexibility, PIN diodes during recent years have become even more predominant in all types of high-frequency breakers. Since nearly all electric control is now done by digital methods, there is no longer any great interest in analog phase shifters. Its only application is when the use of another type of diode, the varactor, is necessary. The coupling characteristics of so-called amorphous, glass-like semiconductors have been studied from several different directions in the past few years. However, it does not seem to be able to compete with the PIN diode's characteristics, except in terms of price, and especially not at high frequencies. Therefore, only the PIN diode will be treated in detail in the following sections.

High-Frequency Breakers with PIN Diodes /178

General Characteristics

The charges induced in the PIN diode's I-layer with forward bias consist of holes from the p-side and electrons from the n-side. After a certain lifetime τ, the holes and the electrons recombine, and this charge loss is replaced in an equilibrium state by the diode current I_o. The charge stored in the diode can then be written as

$$Q_d = I_o \tau.$$

If the diode current is modulated with a frequency ω, we can then write

$$I_d = I_o + \frac{dQ_d}{dt} = \frac{Q_d}{\tau} + j\omega Q_d.$$

ORIGINAL PAGE IS
OF POOR QUALITY

This yields

$$Q_d(\omega) = \frac{I_d \tau}{1 + j\omega\tau}.$$

The absolute value of Q_d (ω) is almost independent of ω up to the value $\omega_o = 1/\tau$. In other words, the modulation of the current has as much influence as a corresponding alteration of the direct current. At very high ω-values, 20 log $Q_d(\omega)$ will decrease by approximately 6 dB per octave.

Thus, at frequencies under $f_o = 1/2\pi\tau$, the PIN diode operates as a normal pn diode rectifier, but at several octaves above this frequency, it appears as a linear resistance the size of which is determined by the direct current through the diode. Based on the construction of the diode, τ can vary between 0.03 μs and 3 μs, and therefore, f_o can vary from approximately 50 kHz to 5 MHz. The desirable τ-value depends on what switching speed is needed and the ratio of the high-frequency current to the control current. A small τ produces a small stored charge Q_d, making rapid breaking possible. In practice, a switching time of 0.1 τ is usually used. However, Q_d must be larger than the charge which the high-frequency current of amplitude I_1 draws out during one half period. This requirement means that

$$I_o \tau > \frac{I_1}{\pi f} \text{ or } \frac{I_1}{I_o} < \pi f \tau .$$

In the context of switching high power at low frequency, τ must be increased so that the control current I_o does not become too large. (Normally, I_o can control high-frequency currents thousands of times higher.) /179

According [1], the PIN diode's R_j in Figure 36 can be written at low frequencies as

$$R_{j_{lf}} \approx \frac{I_o \delta kT}{eI_o} \approx \frac{48}{I_o(mA)}.$$

At high frequencies (>> f_o) the resistance becomes

$$R_{j_{hf}} = KI_o^{-x}.$$

Typically, K = 13 and x = 0.87 with I_o in mA. Changing the current from 1 μA to 10 mA decreases the resistance from approximately 10 K to 1 ohm. If the parasitic element is ignored, this diode, shunt-coupled to a 50-ohm lead and using the expression for α given earlier, will produce a damping change from several hundred dB to 28 dB. With a series coupling, the damping values become 0.1 dB and 40 dB.

Power Durability

A simple analysis of the temperature rise in a PIN diode, according to [2], gives the following result in those cases where the length of the high-frequency pulse is comparable to the diode's thermal time constant:

$$T = T_s + P_D\theta_{js}\frac{1 - e^{-\frac{t_p}{\tau}}}{1 - e^{-\frac{t_r}{\tau}}},$$

ORIGINAL PAGE IS
OF POOR QUALITY

where T_S = the temperature of the heat sink,
P_D = the power gain in the diode,
θ_{js} = the thermal resistance between the active junction and the heat sink,
t_p = the length of the high-frequency pulse,
t_r = the time between two high-frequency pulses,
and τ = the diode's thermal time constant.

With a long high-frequency pulse or a CW signal, the effect /180
of the exponential term can be ignored after a period of approximately 3τ, and then we obtain

$$T_{max} = T_s + P_D\theta_{js}.$$

If the high-frequency pulses are much shorter than τ, the temperature rise will be somewhat higher than the above formula indicates. This case is treated completely [3].

Since the impedance in a breaker diode is always much larger or much smaller than the transmission line's characteristic impedance, it is clear that the gain P_D is only a very small part of the total signal power in the lead P_S. In general, the con-

tribution of the control current to P_D can be completely ignored.

In a shunt-coupling, assuming as before Z_o = 50 ohm, and that the PIN diode can be represented by a resistance R_{min} = 1 ohm or R_{max} = 10 K, we then obtain

$$\frac{P_D}{P_s} \approx \frac{4R_{min}}{Z_o} = 0{,}08 \text{ or } \frac{P_D}{P_s} \approx \frac{Z_o}{R_{max}} = 0{,}005.$$

For a series-coupled diode, this becomes

$$\frac{P_D}{P_s} \approx \frac{R_{min}}{Z_o} = 0{,}02 \text{ or } \frac{P_D}{P_s} \approx \frac{4Z_o}{R_{max}} = 0{,}02.$$

The corresponding expressions for the maximum signal power become:

$$P_{s\,max} = \frac{Z_o}{4R_{min}} \frac{T_{max} - T_s}{\theta_{js}} \text{ or } P_{s\,max} = \frac{R_{max}}{Z_o} \frac{T_{max} - T_s}{\theta_{js}},$$

$$P_{s\,max} = \frac{Z_o}{R_{min}} \frac{T_{max} - T_s}{\theta_{js}} \text{ or } P_{s\,max} = \frac{R_{max}}{4Z_o} \frac{T_{max} - T_s}{\theta_{js}}.$$

[4] gives the following typical data for high-power type PIN diodes. With reverse bias, these have $1/\omega C_j \ll R_j$, so that in order to obtain good insulation, the effect of C_j must be compensated by an outer shunt inductance.

/181

Table 8. Typical data for high-power PIN diodes.

Diode type	C_j in pF	$R_s + R_j = R_{min}$ in Ω at 100 mA	R_j in Ω at −100 V	θ_{js} C°/W
A	3	0.5	10000 (1 GHz)	2
B	1	0.8	8000 (3 GHz)	10
C	0.5	1.7	3000 (6 GHz)	20

When $T_{max} - T_S$ = 150 C, the following maximum signal powers are obtained with CW operation in a 50-ohm line. Maximum gain in the diode is obtained with forward current, which, therefore, determines the power limit.

Table 9. Maximum CW power for PIN diodes.

Diode type	Frequency in GHz	$P_{s\,max}$ i kW	
		Shunt diode	Series diode
A	1	1.9	7.5
B	3	1.6	6.2
C	6	0.6	2.2

Observe that it is normally very difficult to acquire a

ORIGINAL PAGE IS
OF POOR QUALITY

perfect heat sink for a series diode. Therefore, the values in the series-diode column are unrealistic.

By increasing the foward current, R_j can be further reduced, and therefore, P_{smax} can be increased. Around 1 A, the series resistance will be approximately one-half as large as at 0.1 A. Technological improvements in the future should also cut the thermal resistance in half, at least for type B and C diodes. Increasing T_{max} to approximately 200 C might also be possible in the future without changing the diode's lifetime. With regard to these improvements, PIN diodes in the future will be able to break the following CW powers:

Frequency GHz	$P_{s\ max}$ for a shunt diode in kW
1	5
3	8
6	3

These data are plotted in Figure 37 and an approximate curve /182 has been drawn through them. This curve, labelled 1, gives an approximate forecast for the year 1985.

It should be observed in connection with high-power switching that the conversion from low to high ohmic conditions must be done very rapidly, even if there are no special requirements for the switching speed in general. Otherwise, the diode will be destroyed during the time it has a value of approximately Z_o, when a large part of the incident signal power is absorbed.

[3] has already been referred to for a general discussion of temperature rises in a diode during pulse operation. For an evaluation of the pulse-power durability of a special diode, in general, one can begin with the thermal resistance during pulse operation, θ_p, given by the manufacturer. θ_p is given as a function of the pulse length. First, the average temperature T_m of the diode junction is calculated in the same manner as CW power. Then, the maximum temperature during the pulse

$$T_P = T_m + P_{DP}\theta_p,$$

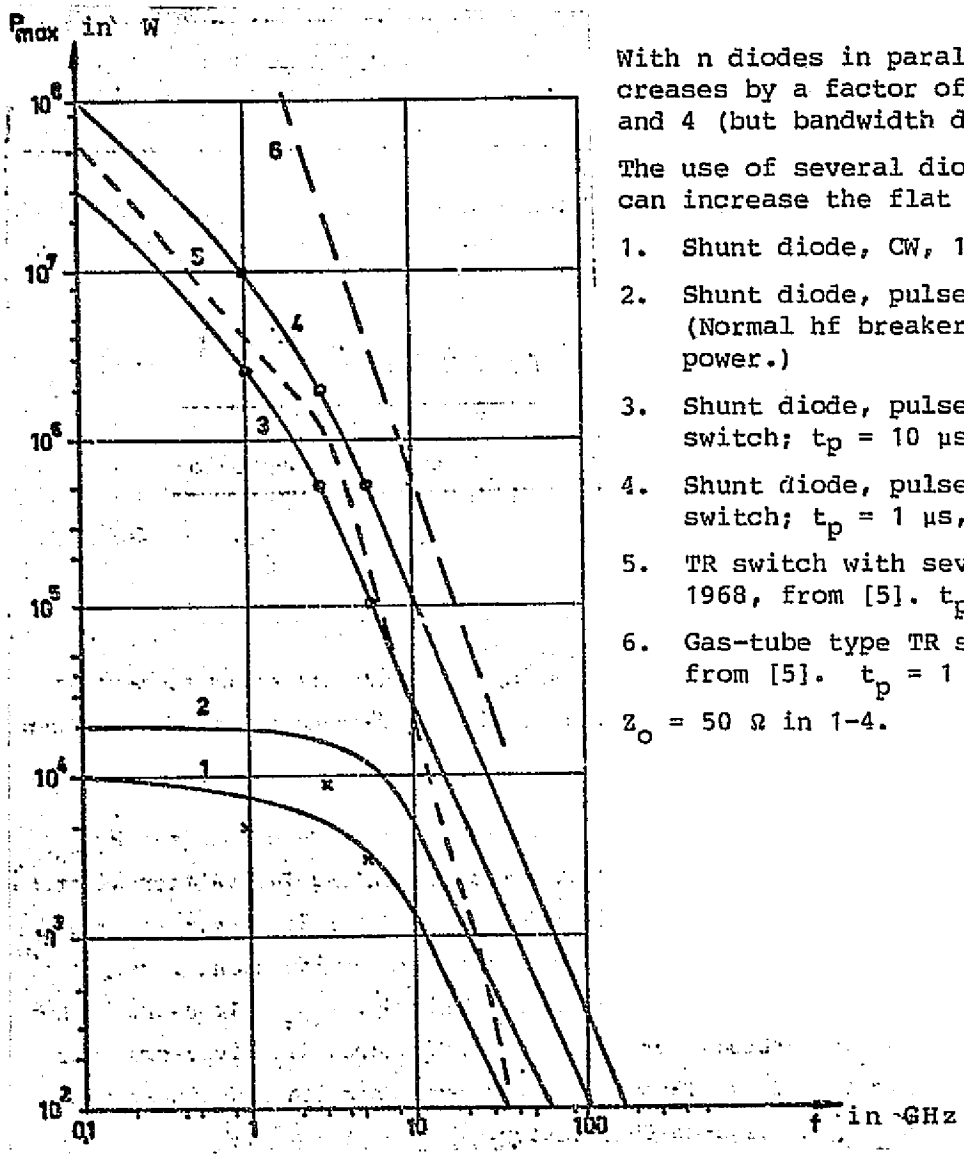
where P_{DP} is the average value of the diode gain during the pulse. The applicable θ_p-values for the previously-discussed high-power diodes are given in Table 10.

According to Figure 38, in a forward biased diode, we have

$$P_{DP} = (R_s + R_j)(2I_p)^2 = 4R_{min}I_p^2;$$

$$P_{SP} = I_r^2 Z_o, \text{ where } P_{SP} = \text{the available pulsed signal power.}$$

With the aid of the formulas in Figure 38, the maximum signal



With n diodes in parallel, P_{max} increases by a factor of n^2 for 1,2, and 4 (but bandwidth decreases).

The use of several diodes in series can increase the flat portion of 2.

1. Shunt diode, CW, 1985.
2. Shunt diode, pulse, 1985. (Normal hf breaker for low average-power.)
3. Shunt diode, pulse, 1985, as TR switch; t_p = 10 µs, u_f =0.001.
4. Shunt diode, pulse, 1985, as TR switch; t_p = 1 µs, u_f = 0.001.
5. TR switch with several diodes, 1968, from [5]. t_p = 1 µs.
6. Gas-tube type TR switch, 1968, from [5]. t_p = 1 µs.

Z_o = 50 Ω in 1-4.

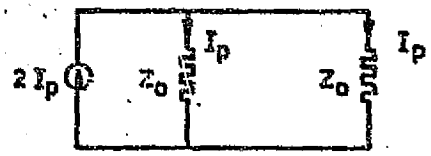
FIGURE 37. Forecast curves for breaker diodes.

power in pulse operation can now be calculated. The use of the /184
PIN diode as a TR-switch, as shown in Figure 39, is of special interest. In this case, the diode is very low-ohmic during the high-frequency pulse and is not subject to any power in its high-impedance condition. According to [4], a maximum temperature of 300°C is then allowable.

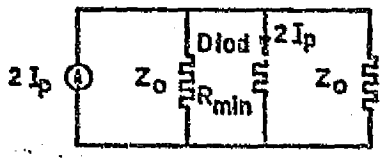
ORIGINAL PAGE IS
OF POOR QUALITY

Pulse length in µs	Thermal resistance in °C/W diode type A	B	C
> 10^5	2.5	10	25
10^3	0.4	1	3
10^2	0.1	0.4	1
10	0.015	0.05	0.2
1	0.002	0.007	0.02

Table 10.



$P_{SP} = Z_o I_p^2$



IF $R_s << Z_o$

$P_{DP} = R_{min} (2 I_p)^2$

Figure 38.

The expressions

$$25 + u_f P_{DP} \theta_{js} + P_{DP} \theta_p = 300;$$

$$P_{DP} = 4 R_{min} I_p^2;$$

$$P_{SP} = Z_o n^2 I_p^2;$$

where u_f = the utilization factor
and n = the number of parallel-coupled diodes

yield:

$$P_{SP_{max}} = Z_o n^2 \frac{275}{4 R_{min} (u_f \theta_{js} + \theta_p)}.$$

Assuming a pulse length of 10 μs, $U_f = 0.001$ and $Z_o = 50$ ohm, the maximum peak powers obtained are shown in Table 11, using type A, B and C PIN diodes, with R_{min} set at 0.1, 0.3 and 0.5 ohm, respectively.

Table 11. Maximum peak power for PIN diodes with 10 μs-pulses. /185

Diode type	$P_{SP\ max}$ in MW, t_p = 10 μs, u_f = 0.001 Year 1970	Year 1985
A (L-band)	2	2.5
B (S-band)	0.2	0.5
C (C-band)	0.02	0.1

Technological improvements can reduce θ_{js}, θ_p, and R_{min}, and yield the forecast values for 1985 which are also shown in Table 11. Table 12 shows the maximum signal power with a pulse of length 1 μs.

Diode type	$P_{SP\ max}$ in MW, t_p = 1 μs, u_f = 0.001 Year 1970	Year 1985
A (L-band)	9	10
B (S-band)	0.8	2
C (C-band)	0.1	0.5

Table 12.

Observe that the large S_j in these high-powered diodes with reverse bias must be compensated with a shunt inductance, so that in a TR-switch they produce a very small receiving band. Still, the bandwidth should generally be somewhat larger than in a conventional gas-filled TR tube. A large number of shunt diodes in parallel couplings are often used to connect very high pulse powers. The number diodes n can increase the allowable peak power by a factor of n^2. At the same time, bandwidth and the input damping is reduced, but the receiving losses increase. Two curves, labelled 3 and 4 in Table 37, show the estimated maximum peak power for 1985.

The above maximum power calculations are only applicable when the diode is used as a TR-switch. If a shunt diode is used for in and out-coupling of high peak power, the diode, which with reverse bias is in a high-impedance transmission state, will be subject to high-frequency voltages. It can be estimated that the /186 high-frequency voltage's peak-to-peak value will not exceed the breakdown voltage V_B, with which the reverse bias is assumed to be 0.5 V_B. With this assumption, the corresponding maximum power becomes

$$P = \frac{V_B^2}{8Z_o}.$$

ORIGINAL PAGE IS
OF POOR QUALITY

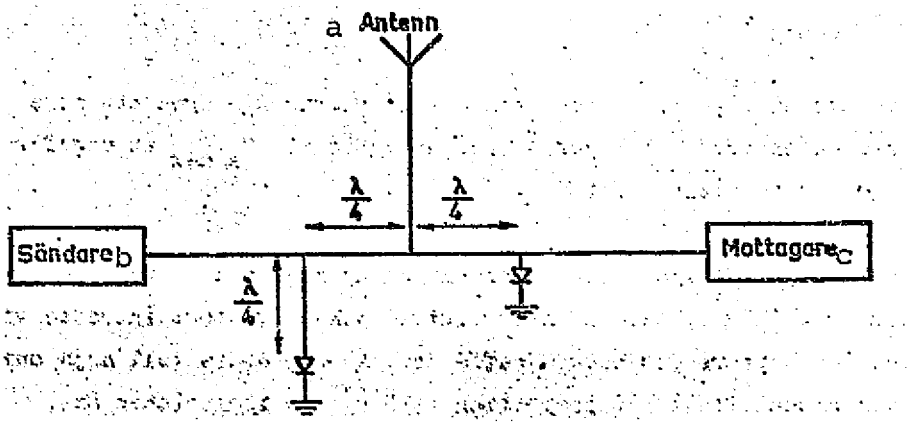
In the high-power diodes previously discussed, V_B is on the order of 1000 V, so that Z_o at 50 ohm produces a maximum power of approximately 2.5 kW. This power is 10 to 1000 higher than the maximum powers in the case of SM couplings. (A reduction of Z_o can naturally increase P by up to ten times.)

However, [4] indicates that the high-power diodes discussed here are able to modulate hundreds of volts in the forward bias region, without producing any current or worsening the high-impedance condition. In the same way, the high-frequency amplitude during the negative half-period can be allowed to exceed V_B without having the diode destroyed by an uncontrollable surge. Therefore, the high-frequency amplitude can be allowed to be larger than 0.5 V_B. In the best case, it is possible that a doubling of the amplitude could be allowed, which would increase the maximum power four times, up to a level of 10 kW.

It is somewhat uncertain whether the overdriving of the diode described here can be allowed. If so, the diode temperature and the average power must definitely be kept very low. In [10], both measurements and experiments show that the diode losses increase in the high-impedance condition with increasing power gain, and that this gain increases rapidly when the high-frequency amplitude is controlled in the forward bias region of the diode. (However, a sufficiently short high-frequency pulse does not have time to increase the temperature.) As a compromise, it is therefore assumed here that overdriving will, at most, allow a doubling of the peak power level.

Naturally, the construction of diodes with thicker I-junctions will produce higher V_B values. At the same time, however, heat dissiptation is worsened and the series resistance and the switching time is increased. Increasing the doping level can increase /187
the maximum value of the field strength which produces breakdown, but this has the disadvantage of increasing losses. Diodes in the future which will withstand high peak power in both the low-resistance and high-resistance condition can probably produce V_B values up to 3 or 4 kV. This should then allow maximum peak powers of up to 20 kW, with Z_o = 50 ohm. This power level has been used somewhat arbitrarily in Figure 37 to indicate an absolute limit of approximately 3 GHz. Above this, the power levels changes with frequency are drawn parallel to the other curves. The resulting curve, labeled 2, represents the possible level of development in 1985 for a high-frequency breaker operating with pulsed high-frequency power. It is interesting to observe how insignificant the increase in signal power becomes with the transition from CW to peak power (curves 1 and 2). This can be interpreted to mean that with Z_o approximately 50 ohm, the diode will be exerted equally in terms of power gain and the risk of breakdown.

Reducing Z_o decreases curve 1 and increases curve 2.



a) antenna
b) transmitter
c) receiver

FIGURE 39.

In general, the series coupling of several diodes can increase the total V_B significantly, but such a configuration always has difficult problems with effective heat dissipation. In addition, in all applications where the average power is relatively high, a high-impedance shunt diode must be well cooled. An upper temperature limit of 150 C is assumed in [4]. As was previously discussed, leak current will otherwise be exhibited which will lead to high input damping or even to the destruction of the diode.

The best result obtained in 1968, [5], with a TR-switch consisting of a large number of breaker diodes, have been drawn for comparison in Figure 37 (the broken curve labeled 5). The broken curve labeled 6 in the same figure indicates the maximum power level for a gas-filled TR tube [5]. Therefore, by 1985, a few (two or three) diodes in parallel can replaces the TR tube, even at the highest power levels. The extrapolations of the curves 1-4 down to 100 W and up to high frequencies are relatively uncertain. If one assumes that the junction capacitance must decrease in proportion to increased frequency (mainly due to the reduction of the area of the diode), the series resistance R_s will increase proportionate to the frequency, and the thermal resistance will increase proportionate to f. (Compare with the avalanche diode oscillator [9].) Since the signal power /188

$$P_{s\,max} \sim \frac{1}{R_s \theta_{js}},$$

we have $P_{s\,max} \sim f^{-1,5}$.

ORIGINAL PAGE IS
OF POOR QUALITY

At very high frequencies, the current displacement (Kelvin effect) will increase the exponent for f. In Figure 37, it has been assumed that P_{smax} is approximately proportionate to f^{-2}.

The earlier discussed use of n-shunt diodes in parallel couplings with TR-breakers is naturally also of great interest for regular breakers for high CW powers. In this case as well, the maximum signal power increases in proportion to n^2, and the insulation increases with n^2. Because of the increased losses, however, the input damping must also be increased. This disadvantage can be counteracted by reducing Z_o, which would otherwise still be necessary to avoid breakdown. A breaker with several diodes in a coaxial circuit usually takes the form outlined in Figure 40-B. With reverse bias, the diodes are capacitive and they create a lowpass filter with the coaxial third wire (Figure 40-B). In this manner, maximum bandwidth is obtained for a certain total diode capacitance. With forward bias, the diode's low impedance produces a very large reflection, and consequently, large insulation. With very high /189
signal powers, several diodes are placed around the coaxial circuit like the spokes of a wheel.

PIN diodes for breaking very low power levels can be designed with very small C_j values without R_s becoming too large. If the diode is not encapsuled in the normal manner, but is directly coupled in a "chip" or "beam lead" design in a microstrip line, C_p can almost be eliminated and L_b will be very low. When the C_j and L_p values in shunt diodes have been further adapted to produce a link in a lowpass filter, very broadbanded breaker circuits are obtained. An example of this is a breaker which covers a band from 0.1 to 18 GHz with 0.5 dB input damping and 20 dB insulation (with several diodes, the values 0.8 and 80 dB, respectively, can be obtained) [6, 7]. These breakers withstand signal powers of 1 to 2 W CW and 50 to 100 W pulsed [8], and have a coupling time of 10 to 20 ns. Any major improvement of this type of breaker will hardly be possible. However, breakers for higher frequencies with bandpass characteristics will certainly be developed. Bandwidths up to 30% should be obtained at 30 GHz, but around 100 GHz, this value will have decreased to approximately 5%.

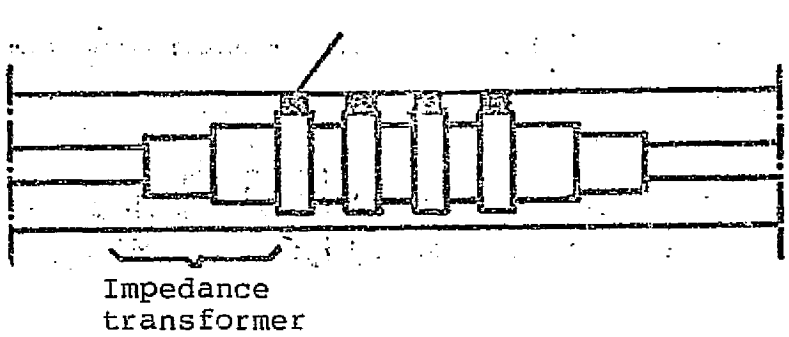


FIGURE 40A.

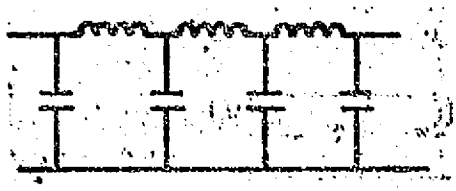


FIGURE 40B.

Attenuators with PIN Diodes

Because of its electrically-controllable high-frequency resistance, PIN diodes also lend themselves to the construction of rapidly-variable attenuators or modulators. In these cases, the attenuator is almost always required to have a low reflection factor at all damping values. This problem is most easily solved by using a circulator in which the gate between the input and the output is loaded with a PIN diode parallel to a terminator. In this way, the input remains adapted despite the varying diode resistance which determines how large a part of the input power is reflected to the output. If a matched diode pair terminates two outputs in a 3 dB-hybrid, the input arm will also remain free of reflection at all levels of diode impedance, since the signals reflected from the diodes are added only to the fourth output arm. A particularly effective attenuator is achieved by interconnecting two 3 dB-hybrids in which each connecting circuit between them is equipped with a shunt-coupled diode, as in Figure 41-A. Both reflection and transmission attenuation is utilized in this case. /190

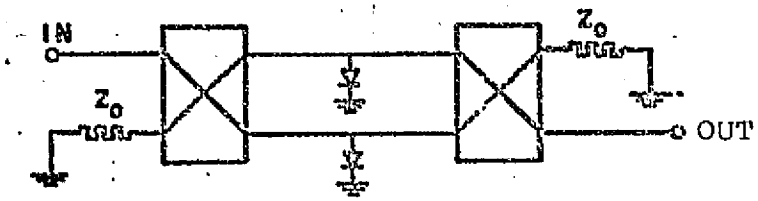


FIGURE 41A.

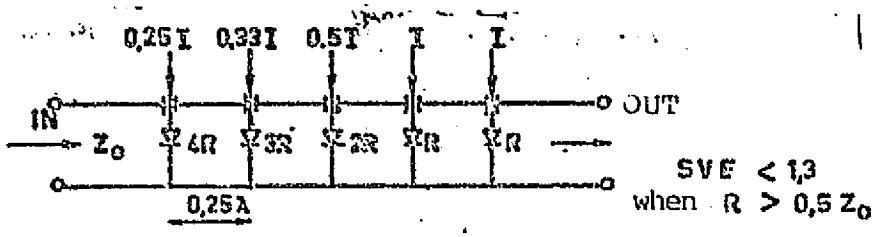


FIGURE 41B.

ORIGINAL PAGE IS
OF POOR QUALITY

A number of shunt-coupled diodes over a transmission line with an interior distance of 0.25 λ, can produce high attenuation and low reflection simultaneously if the control current to each diode is adjusted so that the outermost diode has high resistance and the others have successively less resistance (see Figure 41-B).

Attenuators for low frequencies (10 to 500 MHz) usually consist of three PIN diodes coupled in a π or T configuation. By appropriately varying the control current to the respective diodes, reflection can be held low with variable damping. /191

In evaluating the PIN diode's power durability as an attenuator, one must rely upon the worst case, in which a single diode terminates a line and must, therefore, be able to absorb all of the signal power. If the diode lies in parallel with a terminator, it will absorb at most one-half the signal power. The use of several diodes, as shown in Figure 41-B, increases the maximum allowable signal power. One of today's durable PIN diodes can dissipate up to 10 W, so that the attenuator in Figure 41-A should withstand approximately 40 W. Based on curve 1 in Figure 37, the future durability of PIN diodes is calculated for cases in which the diode resistance is equal to the characteristic impedance. This yields the following values.

Frequency in GHz	Maximum gain in W
1	50
10	10
40	1
100	0.1

Therefore, it will be possible to build attenuators for hundreds of watts of CW power up to the X-band with a few diodes.

Concerning the attenuator's bandwidth, it can be mentioned that the design shown in Figure 41-A covers a 1 to 4 GHz bandwidth with SVF $\leq$ 1.6 [1]. At present, octave bandwidths are obtained up to 18 GHz (8 to 18 GHz). The power durability for these broadband attenuators is only W CW, or 100 W during a 1 μs pulse with d.c. = 0.001. Maximum damping is 65 dB. The diode's parasitic elements will decrease the bandwidth as frequency increases over 20 GHz.

Phase Shifters with PIN Diodes

The use of electrically-controlled, very rapid phase shifters has become very significant in the past few years in the construction of group antennae. Phase shifting is achieved in connection with very high power levels in a ferrite material, but /192

at lower frequencies, a phase shifter with PIN diodes offers several advantages, including faster switching time and reciprocity. The following discussion is limited to PIN diode phase shifters.

The most common phase shifter, in which a switching diode determines the length of the line through the signal passes, is non-disbursive. This is seen in Figure 42-A, which also shows how the phase difference, $\Delta\varphi$, will increase with the frequency at the diode's connection. Figures 42 B-D show the principle designs of several different types of phase shifters using shunt diodes.

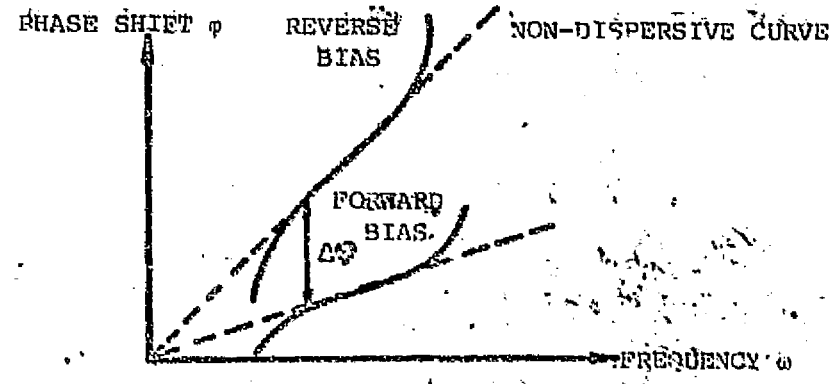


FIGURE 42A.

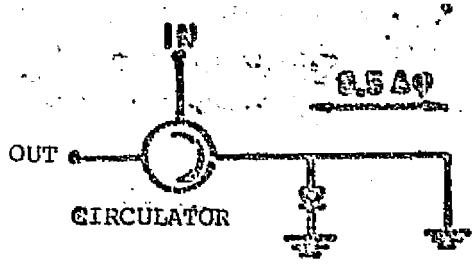


FIGURE 42B.

In the disbursive phase shifter, whose qualities are characterized in Figure 43-A, the phase difference $\Delta\varphi$ is constant within a large frequency range. Here, the signal's time delay is the same in the two states of the phase shifter. Two designs of this phase shifter are shown in Figures 43-B and 43-C. The variable susceptance in Figure 43-B can be achieved by coupling a varactor or a PIN diode between the two different reactive states. [2, Chapter 10] shows that the phase shift resulting from the capacitance difference ΔC can be written as /193

$$\Delta\varphi \approx 2 \operatorname{arctg} Z_0\omega_0\Delta C \left[1 - \left(\frac{\Delta\omega}{\omega_0}\right)^2\right],$$

where $\Delta\omega = \omega - \omega_0$.

ORIGINAL PAGE IS
OF POOR QUALITY

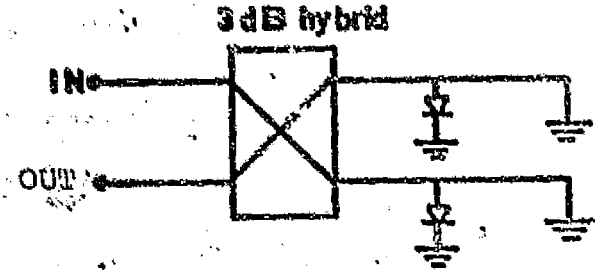
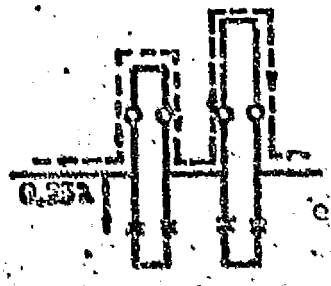


FIGURE 42C.



o = shunt diode with, for example, reverse bias.
x = shunt diode with the opposite bias.

FIGURE 42D.

/194

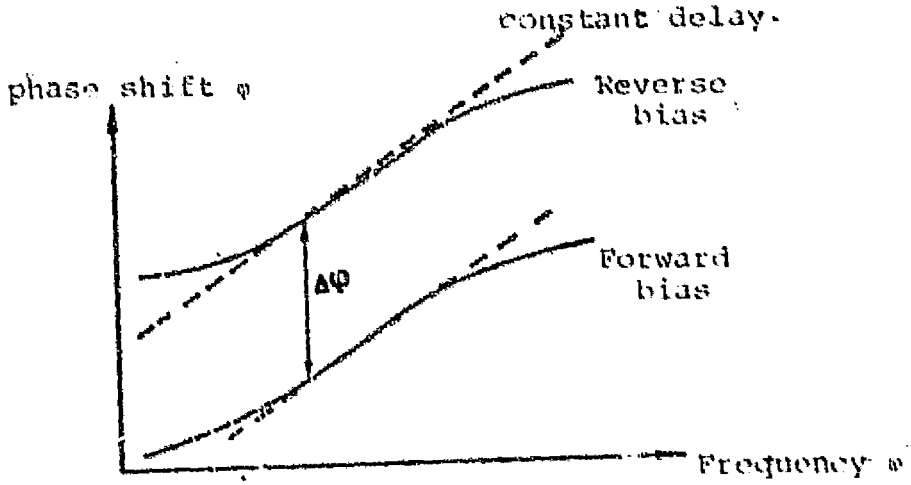
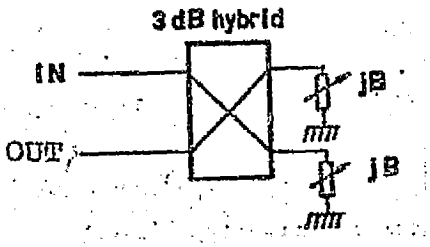


FIGURE 43A.



For example, jB = Breaker

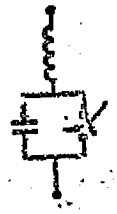


FIGURE 43B.

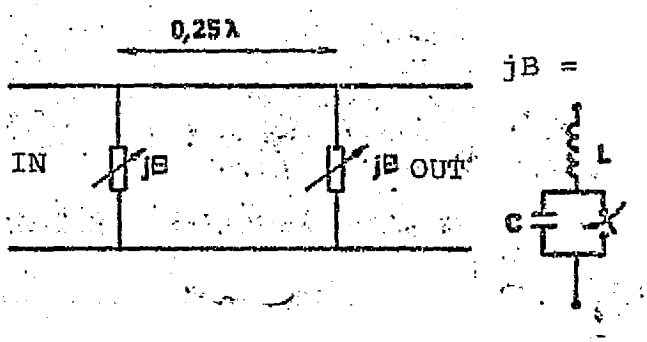


FIGURE 43C.

A frequency change of 10% is then permitted with a phase shift of 90° without changing $\Delta\varphi$ more than 0.5°.

The phase shifter in Figure 43-C consists of a periodic reactive loaded transmission line, of which the figure shows one unit. By connecting each diode to the main line through a transforming line section, greater design freedom is achieved. Naturally, this is aimed maintaining constant impedance while obtaining the desired phase shift. A more complete discussion of this problem is given in [2, Chapter 10]. /195

The Phase Shifter's Power Durability

For a phase shifter with a diode of the type presented in Figure 42-B, the maximum signal power $P_{s\ max_1}$ is achieved in a short-circuit condition determined by the maximum value of the high-frequency current's effective value I_{max}:

$$P_{s\ max_1} = \frac{I_{max}^2 Z_o}{4}.$$

With reverse bias, when the diode's impedance is assumed to be infinitely large, the maximum signal power is determined by the size of the allowable high-frequency voltage over the diode. As earlier, if one assumes that the high-frequency voltage's maximum top-to-top value will equal the breakdown voltage V_B, we obtain

$$P_{s\ max_2} = \frac{V_B^2}{32 Z_o \sin^2 0{,}5\varphi}.$$

ORIGINAL PAGE IS OF POOR QUALITY

I_{max} is determined by the power gain in the diode and is dependent upon both the diode's series resistance and the character of the signal power (pulsed or CW). If a maximally-loaded diode is desired with both forward and reverse bias, the following is chosen:

$$Z_o = \frac{V_B}{2\sqrt{2}\, I_{max} \sin 0{,}5\varphi}$$

$$P_{s\ max} \approx \frac{I_{max} V_B}{8\sqrt{2} \sin 0{,}5\varphi}.$$

In earlier breakdown calculations of the power durability of switching diodes in lines with Z_o-values of 50 ohm, it was accepted that the maximum signal power for SW signals was determined by $P_{s\ max_1}$. The forecast curve obtained in this manner, labeled 1 in Figure 37, can also be applied to a diode in a phase shifter. /196

In the same way, an expression for maximum peak power for breaker diodes with high peak power and low average power can be obtained which agrees closely with $P_{s\ max_2}$. For an overdriven breaker, one can state:

$$P_{b\ max} \approx \frac{V_B^2}{4Z_o}.$$

This yields $P_{s\ max_2} \approx \frac{P_{b\ max}}{8 \sin^2 0{,}5\varphi}$.

With a normal phase shift value, a so-called bit with $\varphi = 22.5^\circ$, we obtain $P_{s\ max_2} = 3\ P_{bmax}$. In this case, therefore, the values in curve 2 in Figure 37, multiplied by a factor of 3, can be used as a forecast for a phase shifting diode producing $\varphi = 22.5^\circ$. Observe that when $\varphi = 180^\circ$, $P_{s\ max_2}$ decreases to $0.13\ P_{b\ max}$.

A phase shifter of the type outlined in Figure 42-D, in which the diodes are included in a continuously-matched filter structure, can withstand [2, Chapter 10] the following powers:

$$P_{s\ max_1} = I_{max}^2 Z_o;$$

$$P_{s\ max_2} \approx \frac{V_B^2}{8Z_o} \approx \frac{P_{b\ max}}{2}.$$

This should mean that the forecast curve for CW operation will produce values approximately four times higher than curve 1 in Figure 37. However, for pulse operation with a low average power, values only one-half as large as curve 2 in the same figure are permitted. This appears to indicate that the diode will withstand four times more CW power than pulsed power, which is obviously unfeasible. The explanation is that the diode in the /197
forward bias condition withstands the high CW power, but would not be able to convert to the high-resistive state without breakdown occurring. Decreasing Z_o from 50 ohm to 25 ohm will then cut the maximum CW power in half and double the maximum peak power. This results in the forecast curve labeled 2 in Figure 37.

Observe that the same problem develops if one calculates a 180^o bit in the phase shifter discussed earlier. When Z_o = 50 ohm, $P_{s\ max_2} = (P_{b\ max})/8$, i.e., the maximum peak power will be one-fourth of the maximum CW power. Changing Z_o from 50 ohm to 25 ohm produces the same durability for CW and peak power, and yields a maximum power level which is half that in curve 1.

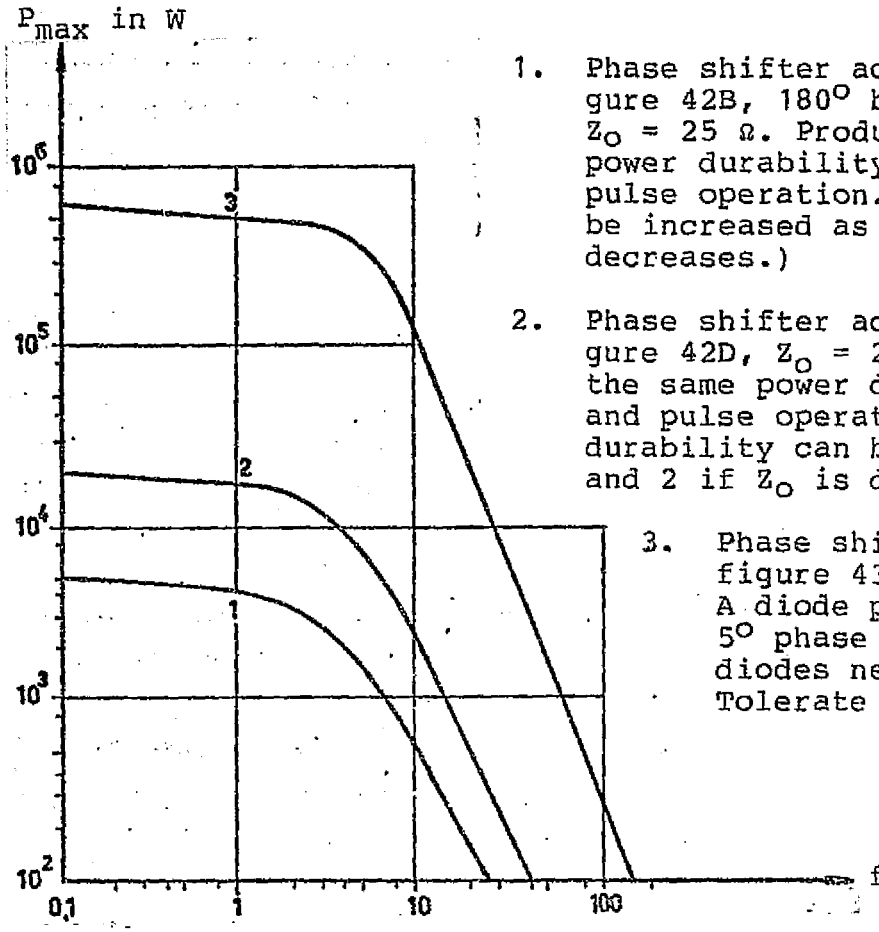
To summarize, it can be said that phase shifters, according to Figure 42-B and 42-D, will be dimensioned for power levels lying between 0.5 and two times the values indicated by curve 1 in Figure 37, in which Z_o = 25 ohm. These forecast values are shown in Figure 44.

If Z_o is decreased under 25 ohm, the maximum peak power level can be increased in proportion to $1/Z_o$.

The phase shifter shown in Figure 43-C, in which the diodes are directly n-coupled to the transmission line, do not withstand higher peak powers than other types of phase shifters. But, if each diode is coupled to the main line through a short transmission line, and is closely followed by a short-circuit, as in Figure 42-B, the power durability can be increased significantly if each pair of diodes contributes with only a small phase difference. This type of phase shifter is treated in detail in [11]. If one assumes that each pair of diodes must produce at least a 5^o phase shift, so that the total number of diodes does not become unmanageably large (in this case, 72 diodes or 180^o phase shift), the peak-power durability will be nine times larger than with 22.5^o phase shift per pair of diodes, according to measurements in [11]. The large number of diodes also produces high durability for CW power.

A forecast curve has been estimated for the phase shifter /198
shown in Figure 42-B which lies three times higher than curve 2 in Figure 37. Further multiplication by a factor of approximately 9 then produces a maximum peak power which lies approximately 30 times higher than the values in curve 2. This is presented as the upper forecast curve in Figure 44.

ORIGINAL PAGE IS
OF POOR QUALITY



1. Phase shifter according to figure 42B, 180° bit with a diode Z_o = 25 Ω. Produces the same power durability in CW and in pulse operation. (Peak power can be increased as the phase shift decreases.)

2. Phase shifter according to figure 42D, Z_o = 25 Ω, producing the same power durability in CW and pulse operation. (Peak power durability can be increased for 1 and 2 if Z_o is decreased below 25Ω.)

3. Phase shifter according to figure 43C, Z_o = 50 Ω. A diode pair produces only 5° phase shift (i.e. 72 diodes needed for 180°). Tolerate pulsed power.

Figure 44. Forecast curves for diode phase shifters.

Finally, some examples will be given here of data obtained to date with diode phase shifters:

High-power phase shifters in coaxial or waveguide arrangements have been built for frequenciws between 0.1 and 10 GHz, normally with 10% bandwidth. The phase shift is usually 22.5 to 180° (but sometimes 360°), and the power durability is usually several dozen kilowatts with pulse lengths of 10 to 100 μs and approximately 10 to 100 W CW. The input damping lies around 3 to 5 dB and the switching time between 0.1 and 10μs. Construction is mostly of the type shown in Figure 43-C, so that the power is divided among several diodes within each phase /199

step (bit). One such phase shifter for the C-band with 72 diodes withstood up to 100 kW peak power and produced a damping of 1 to 2 dB. For further information, see [5, page 119].

There are also several constructions of integrated phase shifters, usually designed for low power levels. In this context, diode pairs are frequently used in a branching formation (series diodes), functioning as converters to various lengths of a microstrip line. With eight diode-pairs, producing a phase shift of 22.5 to 360°, the damping in the L-band was 4 dB. A similar 4-bit phase shifter for the S-band (2.3 GHz) produced a lower input damping: 2.4 dB. Texas Instruments has estimated the price for such a phase shifter, including the drive circuit, at approximately 300 kroner when manufactures in lots of 1000.

A construction in accordance with Figure 43-C for the X-band using a microstrip arrangement of 24 shunt diodes withstood up to 3 kW peak power with a utilization factor of 0.001. The phase shift was 22.5 to 360°, the input damping was 2 ± 0.5 dB and the bandwidth 0.5 GHz, [9].

ORIGINAL PAGE IS
OF POOR QUALITY

PIN Diode Limiters

If a PIN diode is to be used as a passive limiter and a very low level of leakage pulses (so-called spike leakage) is required, its I-junction must be made significantly thinner than in a breaker diode. The charges induced in the I-layer the high-frequency field will rapidly create a direct current which will decrease the diode resistance. In this way, the power transmitted by the limiter will remain at a nearly constant level of several dozen milliwatts, despite increases of 40 to 50 dB or more in the applied power. /200

In addition, the diode capacitance must be kept low so that the input damping does not become too large nor the limiter's bandwidth too small. In comparison to a breaker diode, a limiter diode will then have significantly less volume and, consequently, will have lower power durability. This is particularly applicable to high-frequency pulses which are much shorter than the diode's thermal time constant (approximately 100 μs). Thus, the heat dissipation can be ignored, and the diode's power durability can be estimated from the equation

$$P_{DP} \approx \frac{\Delta T H}{t_p}$$

where P_{DP} = the allowable power gain in the diode;

t_p = the length of the high-frequency pulse;

ΔT = the allowable temperature rise (approx. 100°C);

H = the diode's heat capacity.

H is the product of the volume, the density and the specific heat, and if the diode diameter is D cm and the I-junction's thickness is W cm, we obtain

$$H = 1.56\ D^2 W \quad Ws/^oC.$$

The I-junction's capacitance is obtained with $\varepsilon = 12$:

$$C_j = 0.834 \frac{D^2}{W}\ pF.$$

From this we obtain

$$H = 1.9\ C_j\ W^2 \quad Ws/^oC.$$

The permitted power gain can then be written /201

$$P_{DP} \approx \frac{190 C_j W^2}{t_p} \quad [W]$$

where C_j is given in pF;

W is given in cm;

and t_p is given in s.

[12] gives experimentally determined maximum values for W at various frequencies. They have been determined so that the spike leakage is insignificant. These values, and the typical C_j values for various frequencies (taken from the chapter on breaker diodes) are summarized in Table 13, in which P_{DP} with T_p = 1 μs has also been calculated.

Table 13. Data for PIN-type limiter diodes.

f GHz	W_{max} cm	C_j pF	P_{DP} at t_p = 1 μs W	R_{min} Ω	P_{SP}
0.1	$2.7 \cdot 10^{-3}$	30	41,500	0.05	10 MW
1	$1.0 \cdot 10^{-3}$	3	570	0.5	14 kW
3	$0.5 \cdot 10^{-3}$	1	48	1.5	0.4 kW
10	$0.2 \cdot 10^{-3}$	0.3	2	5.0	5 W
30	$0.1 \cdot 10^{-3}$	0.1	0.2	15.0	0.2 W

The maximum signal power in the line can be written in the same manner as for breaker diodes:

$$P_{SP} = P_{DP} \frac{n^2 Z_o}{4 R_{min}} .$$

According to [12], 0.7 ohm is a typical value for R_{min} at 0.3 GHz. Since $R_{min} \sim W/D^2$, $R_{min} \sim 1/C_j$. With the assumption that R_{min} can be decreased in diodes in the future, it is assumed that R_{min} will equal 0.5 ohm at 1 GHz. R_{min} and P_{SP} have been calculated from this and are included in the table, with n = 1 and Z_o = 50 ohm. (It should be observed that if a 30 pF-diode at 0.1 GHz is replaced by 30 one-pF-diodes, the same P_{SP} is obtained.)

The maximum signal power in pulse operation shown in Table 13/202 has been presented as forecast curve 1 in Figure 45. However,

ORIGINAL PAGE IS
OF POOR QUALITY

the values from 10 GHz have been increased significantly, since these thin diodes are able to dissipate heat during the pulse. The PIN diode limiter can apparently handle very high powers up to the S-band, but above the X-band its power durability becomes poor, even if several diodes are used. At high frequencies, it is therefore necessary to increase the thickness of the I-junction and, as a result, the power durability. The leakage pulse which then occurs must be handled by a supplementary rapid limiter with lower power durability, e.g. a varactor diode. The maximum peak power which can be limited will then be less than for a controlled breaker diode of maximum size, represented by curve 4 in Figure 37. It can be estimated that PIN-varactor combinations will in the future be able to be represented by a power-frequency curve which lies somewhere between curve 4 in Figure 37 and the values calculated above for a single PIN diode limiter. The highest power values obtained in 1968 in pulse operation with $T_p = 1$ μs for a limiter with a diode in a waveguide will help in understanding the position of the forecast curve. The 1968 values, according to [13], are drawn in Figure 45 and labeled "according to [13]." Naturally, the waveguide has much higher impedance than 50 ohm, and therefore, the limiter diode can handle proportionally higher powers. A feasible forecast curve for $Z_o = 50$ ohm, number 2 in Figure 45, can be assumed to closely coincide with the values in [13].

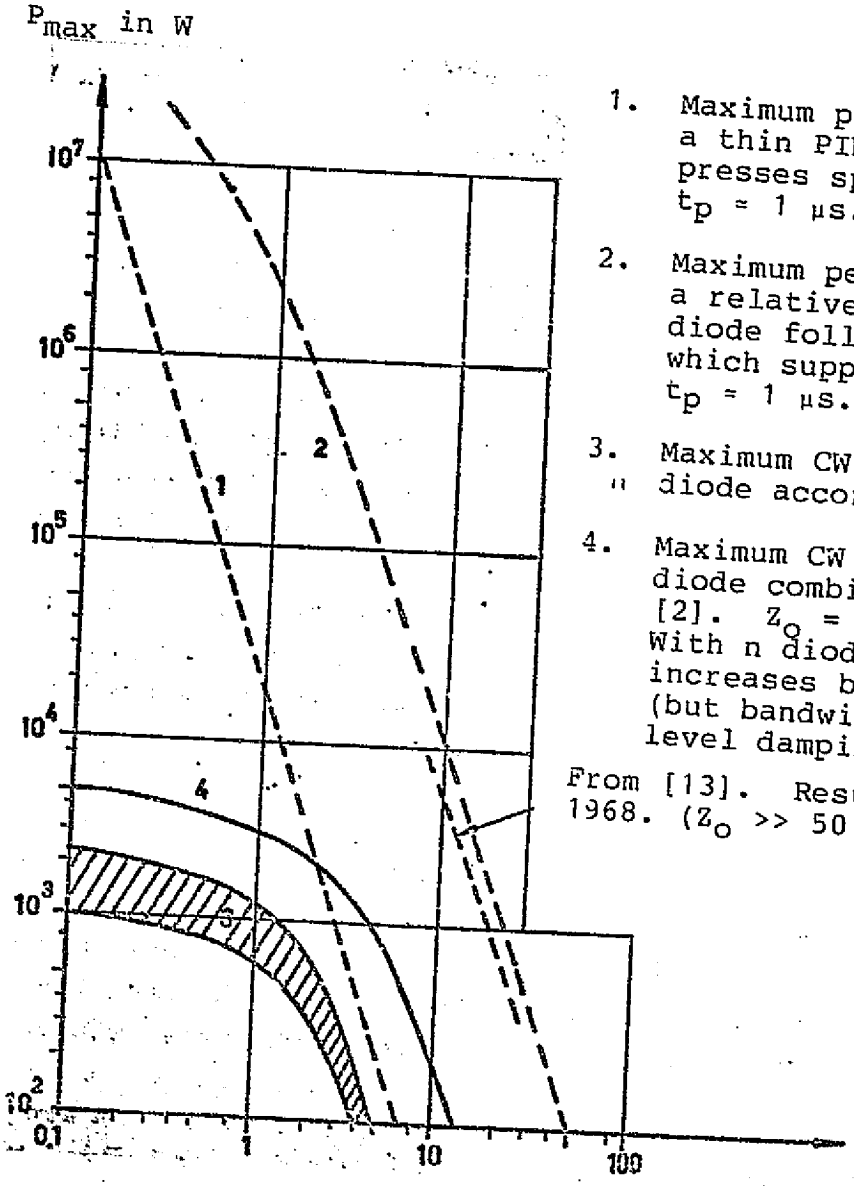
Curve 2 can also be assumed to represent the so-called half-active PIN diode limiter, in which a loose-coupled detector diode before the PIN diode detects the high-frequency pulse and emits a control signal to the PIN diode. (In this case, a following stage with a varactor diode is also required.)

The durability of a PIN limiter for CW signals is estimated in principle in the same way as for a breaker. Using earlier defined notation, we obtain for a shunt diode

$$P_{s\ max} = \frac{Z_o}{4R_{min}} \frac{T_{max} - T_s}{\theta_{js}}.$$

In contrast to the estimation of $P_{s\ max}$ which was done for the breaker diodes, there are no known data upon which to base a calculation for the limiter diode. An approximation can still be obtained if one assumes as earlier that R_{min} in the limiter diode will be twice as large as in the breaker diode. If one further assumes [12] that θ_{js} is twice as large in a limiter as in a breaker, the forecast curve for a limiter's power durability should be obtained by dividing the values in curve 1, Figure 37 by at least a factor of 4. Because of the great uncertainty of this estimate, the CW power durability for the limiter diode is indicated in Figure 45 by the lined region, 3, in which the mentioned calculation factor varies between 50 and 5. In order

/204



/203

1. Maximum peak power in 1985 for a thin PIN diode which suppresses spike leakage. t_p = 1 μs.

2. Maximum peak power in 1985 for a relatively thick limiter diode followed by a varactor, which suppresses spike leakage. t_p = 1 μs.

3. Maximum CW power 1985 for a " diode according to [1].

4. Maximum CW power 1985 for a diode combination according to [2]. Z_o = 50 Ω, ($P_{max} \sim Z_o$). With n diodes in parallel, P_{max} increases by a factor of n^2, (but bandwidth decreases and low-level damping increases.)

From [13]. Result of waveguide, 1968. ($Z_o >> 50$ Ω).

Figure 45. Forecast curves for PIN diode limiters.

ORIGINAL PAGE IS
OF POOR QUALITY

for the maximum CW powers not to be as large as the calculated maximum peak power (which, of course, is impossible), the reduction factor must be increased to 50 at frequencies over 4 GHz.

Curve 4 in Figure 45, which indicated the maximum CW power for a relatively thick PIN diode followed by a varactor diode, has been derived from Figure 37 using about the same method described for the construction of curve 2. The curve has very simply been laid in between curve 1 in Figure 37 and curve 3 in Figure 45.

For a very short description of results obtained to date with the utilization of several diodes in coaxial structures, one can say, based on [13], that curve 1 in Figure 45 represents the current situation very well. The maximum power at 10 GHz is somewhat higher (approximately 100 W) and at 0.1 GHz is somewhat lower than in this curve. The state of the art concerning waveguide limiters has already been discussed and is drawn in Figure 45 according to [13]. One particularly interesting result is mentioned in [14], in which a limiter which withstood 8 kW at 9 GHz consisted of a relatively large, homogenous cylinder of pure silicon. (However, the spike leakage was fairly large.) Naturally, the silicon body was contacted on both sides, but an intrinsic PIN diode was not created. Limiters of this type, along with PIN diodes, should be of great significance in applications at frequencies over 10 GHz. This is due, among other things, to the difficulties in producing a thin, abrupt I-junction.

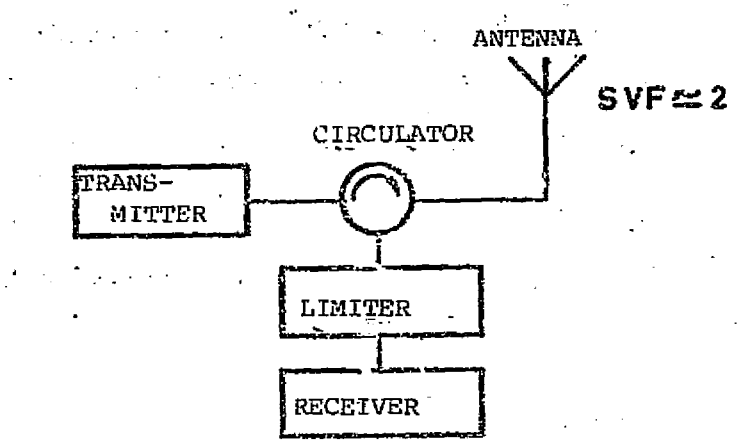


Figure 46.

It should be mentioned here that it has become very common /205
to combine a normal gas-filled TR-tube with a post-coupled rapid limiter diode. In this case, the TR-tube handles all the peak power and the diode eliminates the leakage pulse. The power durability is then determined by the TR-tube and coincides, therefore,

with curve 6 in Figure 37.

The bandwidth for a high-power limiter is normally approximately 10%, and in a low-power limiter ranges from 10% up to an octave. As previously mentioned, the use of several diodes in parallel leads to higher power durability but also to decreased bandwidth. At those power levels shown in the forecast curves in Figure 45, the bandwidth will not be much greater than 10%. The limiter's recovery time after the end of the high-frequency pulse varies with the thickness of the diode from several hundredths to several tenths ns.

When a passive limiter is used to protect a receiver in a radar station, as shown in Figure 46, the transmitting power in principle will exceed the limiter's power durability by a value corresponding to the insulation of the insulator, i.e. around 20 dB. However, in practice, the antenna's SVF is seldom less than two, so that the power reflected to the limiter will only be 10 dB lower than the transmitting power.

REFERENCES

1. HP application note 922, "Applications of PIN diodes." /206

2. Watson, H.A., Microwave semiconductor devices and their circuit applications, McGraw-Hill, Chapter 9, (1969).

3. Mortenson, K. E., "Analysis of the temperature rise in PIN diodes caused by microwave pulse dissipation," Trans. IEEE, ED, p. 305, (March 1966).

4. Unitrode application note M:-70-1, "General information on using Unitrode PIN diodes in RF applications."

5. Ryder, Brown, Forest, "Microwave iode control devices part 2, Duplexers, switchers and phase shifters," Microwave Journal, p. 115, (march 1968).

6. General Microwave data sheet, (1970).

7. HP data sheet 5082-3040.

8. "Semiconductor devices survey, Diode switches," Microwaves, p. 76, (April 1967).

9. Ward, C. S., "A 3-kW microstrip and waveguide diode phase shifter at X-band," G-MTT 1970 International Microwave Symposium, p. 351, (May 1970).

10. Rose, Haine, "On the conductance of PIN junctions at high microwave fields," European Micro ave Conference, p. 540 (September 1969).

11. White, J. F., "High power PIN dio controlled, microwave transmission phase shifters," Trans. IEEE, MTT, p. 233 (March 1965).

12. Brown, N. J., "Design concepts for high-power PIN dio limiting," Trans. IEEE, MTT, p. 732, (December 1967).

13. Ryder, Brown, Forest, "Microwave diode control devices, Part 1." Microwave Journal, p. 57, (February 1968).

14. Mortensou, White, "X-band room-temperature bulk-effect, high-power limiter," ISSCC Digest, (1967).

ORIGINAL PAGE IS
OF POOR QUALITY

Other Signal Manipulation

Since the above title encompasses a particularly large area /207
of electronics, this can in no way be a complete treatment of the adherent function principles and solid-state components. Rather, a number of examples will be given which illustrate newer methods of signal manuipulation within the microwave range.

Generally, the greatest interest in signal manipulation today is oriented toward data processing techniques. In this case, the signals are binary numbers and counting pulses which are handled by very complicated circuits, consisting of a very large number of electronic switching devices. These switches can change very rapidly between on and off states and normally consist of bipolar or FET transistors, although tunnel diodes and rapid breaker diodes can also be used. These semiconductors produce a large number of different gates which comprise the building blocks of the digital circuit. They can also be combined into flip-flops and reply-circuits for a static or dynamic type of memory function. Data processing today is seldom done faster than permitted by the highest incoming switching frequency on the order of 100 MHz. The adherent data signals can, therefore, not be included in the microwave range, but data processing is mentioned here because it is in the midst of a rapid development toward higher processing speeds. It is very likely that the earlier-discussed MOSFET type of transistor, which produces oscillations in the 40 to 60 GHz range, will increase future data processing frequencies far into the microwave range. A further developed monolithic-circuit technology will also produce such miniaturization that the signal's transit time between the various circuits can be reduced sharply. Computer memories in the future will probably consist of flip-flops of MOSFET transistors. It is possible that the utilization of the relatively newly discovered so-called magnetic bubbles will be a price-competitive alternative. According to [1], the price could be as much as ten times lower. (The price per bit in MOSFET structures is 0.03 cent.)

It should be mentioned here that very rapid breakers of the /208
same type used in data processing are now also used in connection with so-called sampling techniques for producing directly on an oscilloscope a microwave oscillation. Schottky-barrier diodes, which are also part of digital frequency meters, are normally used. These and similar diodes are also significant in the generation and measurement of extremely small pulses as, for instance, in the determination of discontinuities in a waveguide by studying how the pulse is reflected (time domain reflectometer). Step recovery, or SR diodes, are particularly suitable for producing pulses which have rise and decay times of approximately 0.1 ns.

Gunn diodes with wandering domain can be utilized in several

ways to easily generate complicated forms of pulses or series of pulses with a certain arrangement. The current through the diode is proportionate to the geometric cross-sectional surface or to the doping in that area through which the domain passes. For example, a pulse code is obtained if an evenly-doped diode has a number of more or less deep crevices or if the doping level is varied in a corresponding fashion. Controlling the domain, and therfore the current, through a number of capacitive, connected electrodes along the direction of the domain's diffusion can achieve the same result. The domain's growth time, i.e., the current pulse's rise time, is approximately 0.1 ns in GaAs. The domain's speed of motion is 10^7 cm/s. A 100 μ-wide inhomogenity will then produce a current pulse lasting 1 ns.

The use of varactors and YIG balls for electronic tuning of oscillators has already been discussed in connection with avalanche and Gunn diodes. The same tuning devices will certainly also be of great significance in the construction of electronically-tunable filters. In all these active circuits, negative impedance converters or NIC circuits will be common at increasingly high frequencies. They can be described as active four-poles using transistors as the amplifying element. An impedance connected /209
to one of the pole-pairs produces its inverted value at the other pair. For example, this is one way to obtain a miniaturized "integrated inductance." A NIC circuit and a condenser produce a positive reactance equal to $j\omega C$. NIC circuits can also produce negative resistance which can be utilized in electronically-tunable filters to increase their Q-values by a factor of approximately 10. This negative resistance can be held constant within a fairly large frequency range (at present, approximately 50% bandwidth at 1.5 GHz). Future NIC circuits will be built as high up in frequency as transistors will allow. Still, negative-resistance diodes should in principle be able to replace transistors in these circuits. Naturally, all of the active circuits of the type described here can only be used at low signal levels.

Signal manipulation techniques in the future will probably utilize to an increasing extent acoustic waves in piezoelectric crystals, and both acoustic waves and spinning waves, i.e., magnetoelastic waves in ferromagnetic material. Previous work has been mainly aimed at the utilization of volume waves in these materials, but surface waves should become dominant in the future. Their advantage is that the signal is easily accessible at the surface of the crystal and can be conducted by some type of microstrip line. Since an elastic wave diffuses approximately 5^5 times slower than an electromagnetic wave, the wavelength and consequently the dimensions of the circuit will decrease in comparison to coaxial circuits by a factor of 10^{-5}. In this way, extreme miniaturization of the signal manipulation circuits will be possible. As discussed in the chapter on the electroacoustic amplifier, difficulties in producing the necessary circuits occur

at high frequencies. In that chapter, the highest possible frequencies for surface waves were given as approximately 10 GHz. Since this is the maximum frequency for the electroacoustic converter which is always needed, the same maximum value applies in general for signal manipulation with surface waves.

The following signal manipulation functions can be easily achieved with pure acoustic waves: /210

1. Delay (with or without amplification).
2. Frequency filtration.
3. Pulse compression or expansion.
4. Coding and de-coding.
5. Spectral analysis, Fourier transformation, etc. [2, 3].

The use of magnetoelastic waves in yttrium-iron-granite (YIG) for signal manipulation is described in [4]. It is especially interesting that the signal delay in a YIG rod can be changed with an external magnetic field from, for example, several tenths of a microsecond to four microseconds. The input damping is on the order of 40 dB. The delay is normally frequency-dependent, but this dependency can be reduced by suitably designing the magnetic field. A non-disbursive delay which can be varied mechanically from 1 to 4 μs is easy to produce. (Within 10% bandwidth at 2 GHz, the disbursion is less than 0.1 μs.)

If the magnetic field is changed while a delayed signal exists in a spin-wave condition, the signal frequency can be lowered.

Magnetoelastic waves are also very suitable for pulse compression. For example, a compression ratio of 200 has been reached at 1.5 GHz with pulses lasting 0.5 μs.

The use of magnetoelastic surface waves can, in addition to the previously mentioned functions, also produce non-reciprocal wave diffusion. The amplification of these waves is possible in principle but so far has been insignificantly researched.

REFERENCES

1. Karp, H. R., "Magentic bubbles- a technology in the making," /211 Electronics, p. 83, (September 1969).

2. White, R. M., "Surface elastic waves," Proc. IEEE, p. 1238, (August 1970).

3. Yoder, M. M., "Challenge of the seventies: Systems exploitation of the multifunctional propoerties of praetersonics technology," 1970 Proc. Electronic Components Conference.

4. Smith, A. B., "Microwave magnetoelastic devices," 1970 Proc. Electronic Components Conference.

Integrated Microwave Circuits

The Integration of Circuits and Active Semiconductor Devices /212

Since 1960, so-called microelectronics have developed very rapidly simultaneously with semiconductor technology. The need for greatly miniaturized, dependable and inexpensive active circuits was particularly great within computer technology and, therefore, development was mainly oriented toward the adherent digital or logical devices. Since active semiconductors and other circuit devices are usually built as indivisible units, the name "integrated circuits" is used.

Around 1964, the use of integrated circuit techniques within the microwave range began. At these high frequencies, the signals must be transmitted in as loss-free a transmission line as possible, in which the necessary reactances can also be created for the desired resonance circuits. The transmission pattern for such distributed circuits is transferred by photo-lithographic methods to a thin metal film on the upper side of a ceramic base plate. Since the transmission thickness is increased by electrolytic coding, so that the losses are minimized, the desired pattern is uncovered through chemical etching. In this manner, a number of different transmission lines can be produced. Figure 47 shows the cross-section of a number of such lines (the direction of transmission is at a right angle to the plane of the paper). The microstrip lines [1] in Figure 47-A is so far the most frequently used type of line. The substrate material is normally aluminum oxide or sapphire, with $\varepsilon \approx 10$. Since the wave diffusion is not of a pure TEM type, the wavelength will be approximately 40% of the free wavelength. In the coaxial-strip line in Figure 47-B [2] the substrate lacks a ground layer, but is instead encircled by a waveguide-like metal tube. The advantage with this construction is that the variations in the substrate's ε or thickness only insignificantly effects the characteristic impedance, and furthermore, the transition from a normal coaxial line can easily be made with low SVF (≈ 1.02). The effective ε value is approximately 1.6, so that dimension reduction in relation to a normal coaxial line is insignificant. The slot lines [3] in Figure 47-C and the planar strip lines [4] in 47-D have about the same characteristics as the /213 microstrip lines, but are less dependent on variations in the substrate thickness and can more easily produce lines with high characteristic impedance. A connection to a ferrite element can easily be arranged to produce non-reciprocal couplings, for example, insulators and phase shifters. It should be pointed out here that microstrip lines can be built up in a ferrite substrate and then can also be easily manipulated to produce these circuit functions.

The dielectric strip line in Figure 47-E, which cannot be produced in the same way as the other types of lines in Figure 47, has been suggested to be especially appropriate for very high

frequencies (millimeter waves) [5].

Instead of obtaining the desired reactances with open or short-circuited shunt lines of suitable length, these reactances can be created by concentrated impedance devices. The same techniques used in producing microstrip lines can be utilized to create inductances in the form of horizontal spirals and condensers, consisting of two metal films separated by a thin oxide layer. Since these impedance elements must be much smaller than a quarter-wavelength, their dimensions will be very small at frequencies near 10 GHz. This places high demands on the photo- /214
resist process and the substrate's surface smoothness. Normally, optic plane positioned sapphire is utilized in which the practical upper cutoff frequency will lie between 10 and 15 GHz.

If the substrate consists of a high-resistive semiconductor, the circuits can be manufactured in the normal way and the required active elements, both transistors and diodes, can be directly produced at the desired place through the various doping processes. These so-called monolithic circuits are now very common in transistor amplifiers for low frequencies, but within the microwave range the manufacturing problems are still not completely solved. In general, the circuit losses will be too large at up to 10 GHz. However, at higher frequencies where both the dimensions and the losses per wavelength are small, these monolithic circuits will represent an economically attractive alternative when mass produced. Integrated circuits for frequencies over 50 GHz will probably have to be made with monolithic techniques using gallium arsenide as the substrate material.

However, in the large majority of integrated circuits for the microwave range, the active devices are produced separately and then connected to the transmission pattern. These are then called integrated hybrid circuits.

Since most integrated circuits to date are made with microstrip lines or concentrated devices, these will now be discussed in more detail with regard to obtainable Q-values and cutoff frequencies.

According to [6], the figure of merit for a microstrip line (which has a linewidth w much larger than substrate thickness h) can be written

$$Q = 0{,}63\, h_{cm} \sqrt{\sigma} \sqrt{f_{GHz}},$$ where σ = the metal's conductivity in $(\Omega m)^{-1}$.

For example, for copper we obtain /215

$$Q = 4780\, h_{cm} \sqrt{f_{GHz}}.$$

ORIGINAL PAGE IS
OF POOR QUALITY

This expression applies to co pletely smooth metal surfaces and a loss-free dielectric.

The line damping can be written

$$\alpha = \frac{27.3}{Q} \text{ dB}/\lambda g$$

with the wavelength $\lambda g \approx \frac{30}{f_{GHz}\sqrt{\varepsilon}}$ cm.

Still with w >> h, the characteristic impedance

$$Z_o \approx \frac{377}{\sqrt{\varepsilon}} \frac{h}{w}.$$

With the normal substrate values h = 0.64 mm and ε = 9.6, at 10 GHz we obtain

$$Q = 960;$$

$$\alpha = 0.028 \text{ dB}/\lambda g = 0.029 \text{ dB/cm}.$$

For comparison it can be mentioned that the highest measured Q-value in a microstrip line with Z_o - 23 ohm in the X-band is 738. The practically obtainable Q-value is approximately 20% lower than the theoretical maximum. This corresponds to an increase in the α-value of 0.009 dB/λg, which originates mainly from the dielectric losses. These losses are indicated by tgδ, which normally is frequency independent, so that the dielectric damping per λg is also independent of frequency.

According to the previously-given expression, Q is proportionate to $h\sqrt{f}$, but to obtain high Q-values, this products cannot be increased without limit. There is a cutoff frequency at which point large radiation losses arise, due to the fact that a surface wave is excited. This wave diffuses in the substrate /216 independent of the strip line structure. According to [6], the cutoff frequency is

$$f_{max} = \frac{7.5}{h_{cm}\sqrt{\varepsilon - 1}} \text{ GHz}.$$

The maximum Q-value is therefore

$$Q_{max} = \frac{35\,800}{\sqrt{f_{GHz}}\sqrt{\varepsilon - 1}}.$$

ORIGINAL PAGE IS
OF POOR QUALITY

F_{max} as a function of h and Q_{max} as a function of f is reproduced in Figures 48 and 49, respectively.

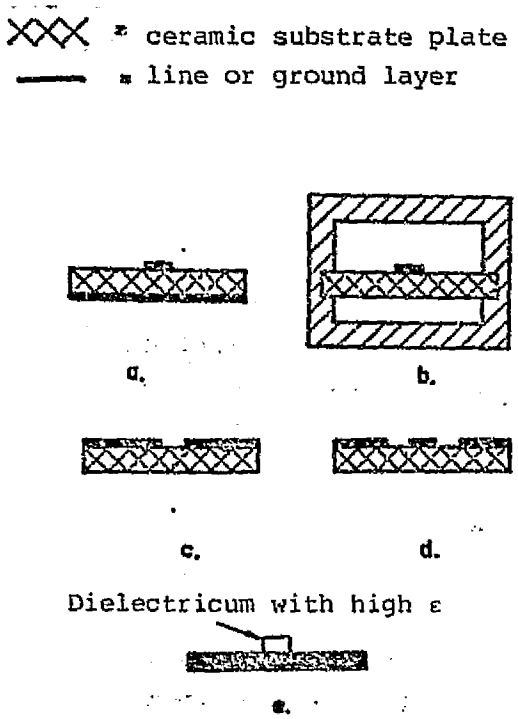


Figure 47.

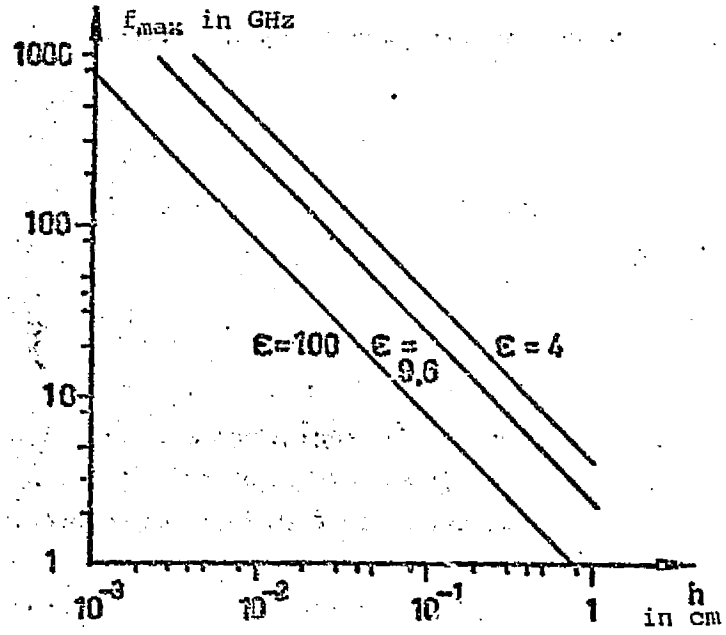


Figure 48.

Since the derived Q_{max}-value applies when w >> h, i.e., for /218 lines with very low characteristic impedance, it can be of interest to estimate how Q_{max} will decrease or α will increase with increasing impedance. The result of this calculation [6], which is based on [7], is presented in Figure 50. According to this, α has doubled (and Q reduced by half) first at an impedance of 70 ohm (w/h = 0.45). This means that the Q-values obtainable with practical linewidths does not lie so far under the calculated Q_{max}-value.

There is one more factor which can reduce Q in relation to the calculated maximum value: radiation losses caused by line discontinuities. The largest radiation losses occur in an open line. For example, with an effective ε-value of about 4, a short-circuited line will have approximately ten times lower radiation losses than an open line. [6] shows that the radiation losses from an open line will be as large as the normal resistive losses when Z_o = 10 ohm with ε = 9.6, or Z_o = 80 ohm with ε = 4. When the ε-value is much over 10, the radiation losses caused by discontinuities can be ignored, but even a. the normal value of 9.6, the calculated Q_{max}-value is reduced by as high a factor as 0.5. The lower cutoff curves in Figure 49 indicate the Q-values /219

ORIGINAL PAGE IS
OF POOR QUALITY

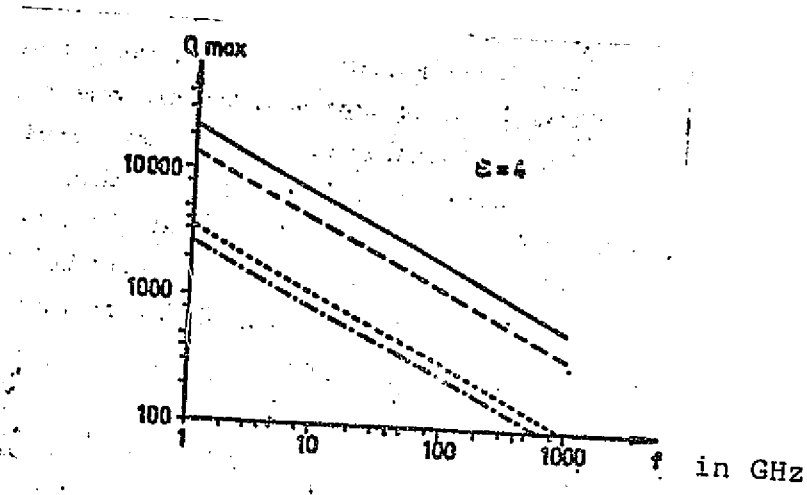


Figure 49A.

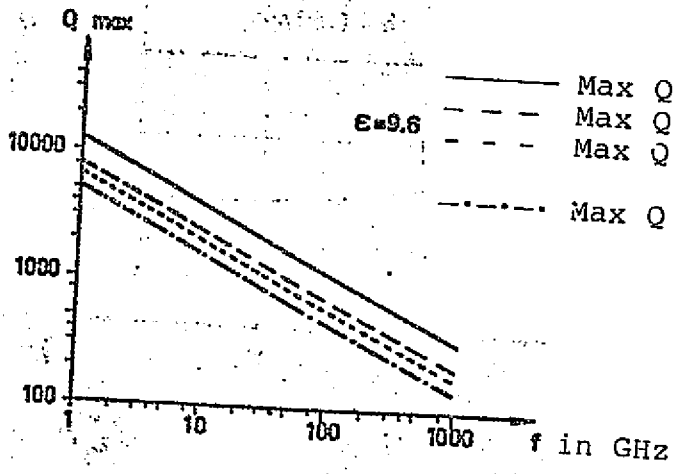


Figure 49B.

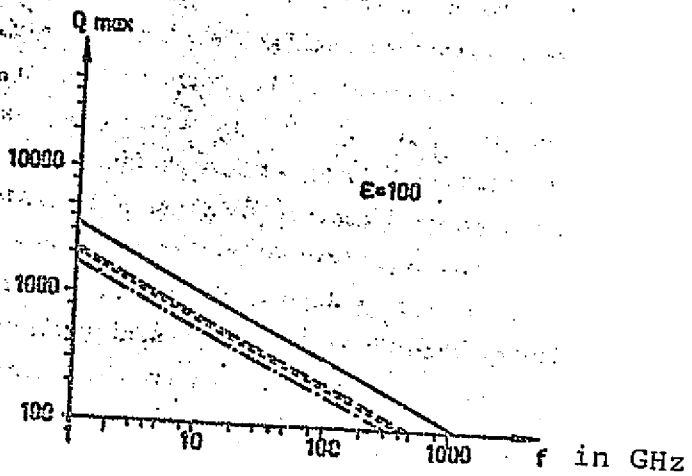


Figure 49C.

ORIGINAL PAGE IS
OF POOR QUALITY

reduced by either high Z_0 or high radiation. If an approximate value of 0.01 dB/λg for dielectric losses is added to the previous losses, the curves situated lowest in Figure 49 are obtained.

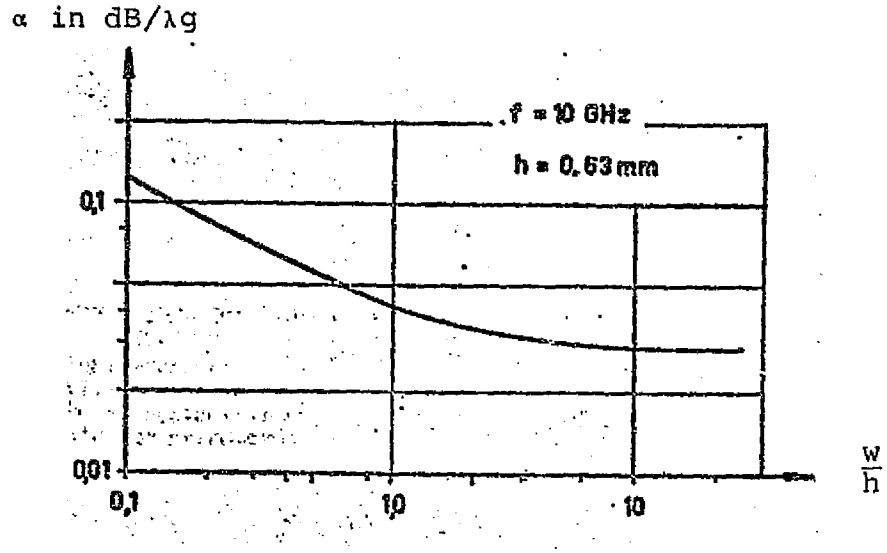


Figure 50.

By using U-form resonators, the radiation losses can be reduced so that the Q-value is increased 55% in relation to a straight open line [8].

In general, the Q-value in a microstrip circuit can be increased through coupling with a YIG ball (frequency magnetically-tunable), or to a dielectric resonator with high ε (for example, Q = 12,000 at 3 GHz, [9]). Superconductive circuits can also become useful in this context.

The Q-values obtained with concentrated circuit devices are given in [10]. The condensers are at 1 pF and made of SiO_2. The inductance device consists of a simple coil with a value of about 1 nH and a planar spiral of 4 nH. The Q-values are shown in Figure 51. For comparison, the Q_{max}-values for distributed circuits are also drawn in the same figure (for $\varepsilon = 9.6$).

Maximum Frequency for Microstrip Lines and Concentrated Impedance Devices

According to Figure 48, a substrate with a thickness of several dozen microns can be used for frequencies up to 1000 GHz. Figure 49-B indicates that a substrate with $\varepsilon = 9.6$ should produce sufficiently high Q-values (> 100) at this frequency. Even with an ample safety margin, one can then assume that circuits equipped with microstrip will be useful up to 200 GHz. A substrate thickness of ten or so microns should be relatively easy to produce

through "sputtering" on a metal ground layer. The adherent line widths of 1 to 30 μ should be easily produced with the necessary precision through future photoresist techniques. (Line widths of 0.3 μ have already been manufactured with electron-beam, electron-resist processes.) In general, line widths of approximately 10 μ should be suitable for connecting future active semiconductor components in "beam-lead" arrangements. /220

It is likely that semi-insulated GaAs can also be used as a substrate in monolithic constructions up to 200 GHz. So far, such monolithic circuits have been built for up to 100 GHz [11], where the damping in a 50 ohm line is approximately 0.25 dB/λg for a 0.1 mm-thick substrate. Monolithic techniques will probably be applied at frequencies over 20 to 30 GHz, particularly in connection with the manufacturing of large series for which the price will be very low.

Concentrated circuit devices are of special significance under 10 GHz, where, in relation to distributed circuits, they lend themselves to often-desirable miniaturization. Figure 51 indicates an upper frequency range of approximately 15 GHz, at which point the Q-value begins to be unfeasibly low.

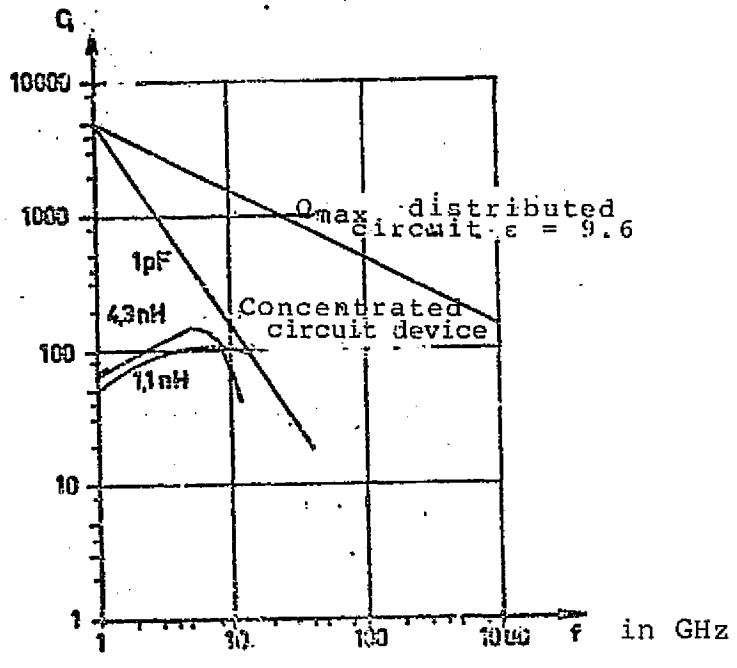


Figure 51.

ORIGINAL PAGE IS OF POOR QUALITY

C-3

The Power Durability of Microstrip Lines

Ceramic substrates with thicknesses of 0.5 to 1 mm can withstand maximum peak powers of approximately 10 kW. In the adherent OSM-type coaxial junctions, 8.5 kW will produce breakdown. These numbers show that the microstrip line will withstand significantly higher power levels than a single active semiconductor device would be likely to generate. /221

Some General Views on the Future Prospects for Integrated Circuits

As is evident from the above discussion, integrated circuit techniques produce extraordinary possibilites for the miniaturization of microwave circuits. The use of high-ε substrates in distributed circuits is of special significance, but the very best results have been achieved with concentrated circuits. The fact that the active semiconductor devices can be used in these circuits in their natural size contributes to miniaturization. In order to be n-coupled in a coaxial or a waveguide circuit, a semiconductor chip which normally has a volume of 0.1 mm^3 or less must be mounted in a capsule up to one thousand times as large. Capsulation also protects the semiconductor from atmospheric influences, but in connection with integrated circuit techniques, a complete system function is contained in one sealed capsule. Also, the so-called "beam-lead" type of small semiconductor "capsules" can be used. In the future, these miniaturization techniques should make microwave circuits for low power levels at least one hundred times lower in weight and volume than today's coaxial and waveguide circuits. The subsequent cost reduction is hard to evaluate, since it will depend on the manufacturing technique, the size of the series and the level of complexity in the circuit. Price reductions of up to one hundred times might very well be possible. One can also expect that these integrated circuits will have a very long lifetime and high operational dependability.

The advantages of integrated circuit techniques from the point of view of the circuit functions has been discussed several times earlier. The lowered parasitic reactances caused by the semiconductor's in-coupling in the circuits and the use of concentrated circuit devices with low reactance-frequency derivitives /222 is particularly significant for the construction of very broadband circuits.

Based on the rapid development of integrated microwave circuits [10, 11, 12, 13] and the characteristics outlined above, one must assume that practically all military microwave equipment for low power levels will be integrated before 1985, at least within frequency ranges which are common now. The possibility exists in principle to expand the frequency range upward to at least 200 GHz.

REFERENCES

1. Schneider, M. V., "Microstrip lines for microwave integrated circuits," Bell System Technical Journal, p. 1421, (May-June 1969). /223

2. Engelbrecht, West, "Microwave integrated circuits," Bell Laboratory Record, p. 328, (October-November 1966).

3. Mariani, E. A., et. al., "Slot line characteristics," Trans. IEEE, MTT, p. 1091, (December 1969).

4. Wen, C. P., "Coplanar waveguide," Trans. IEEE, MTT, p. 1087 (December 1969).

5. Hardeman, L., "Tech. sessions highlight mm-wave trends and low-noise amplifier techniques," Microwaves, p. 64, (August 1970).

6. Vendelin, G. D., "Limitations on stripline Q," Microwave Journal, p. 63, (May 1970).

7. Pucel, Masse, Hartwig, "Losses in microstrip," Trans. IEEE, MTT, p. 342, (June 1968) and p. 1064, (December 1968).

8. Roberts, Easter, "Microstrip resonators having reduced radiation loss," Electronic Letters, p. 191, (April 1971).

9. Day, W. R., "Dielectric resonators as microstrip-circuit elements," Trans. IEEE, MTT, p. 1175, (December 1970).

10. Caulton, Sobol, "Microwave integrated-circuit technology-a survey," Trans. IEEE, MTT, p. 292, (December 1970).

11. Mao, Jones, Vendelin, "Millimeter-wave integrated circuits," IEEE Journal Solid-State circuits, p. 117, (June 1968).

12. IEEE Journal Solid-State circuits, "Special issue on microwave integrated circuits," (June 1968).

13. Johnson, K. M., "Recent advances in microwave integrated-circuit solid-state source design," IEEE Journal Solid-State circuits, p. 119, (June 1970).

ORIGINAL PAGE IS OF POOR QUALITY