RCA Review

June 1979

Volume 40 No. 2

RCARCI 40(2) 101–238 (1979)

RCA Review, published quarterly in March, June, September and December by RCA Research and Engineering, RCA Corporation, Princeton, New Jersey 08540. Entered as second class matter July 3, 1950 under the Act of March 3, 1879. Second-class postage paid at Princeton, New Jersey, and at additional mailing offices. Effective January 1, 1978, subscription rates as follows: United States and Canada: one year $8.00, two years $14.00, three years $18.00; in other countries, one year $8.60, two years $15.20, three years $19.80. Single copies (except for special issues) up to five years old $3.00.

RCA Review

A technical journal published quarterly by RCA Research and Engineering in cooperation with the subsidiaries and divisions of RCA.

Contents

June 1979 Volume 40 Number 2 ISSN 0033-6831

A Temperature-Compensated Laser Module for Optical Communications

M. Ettenberg, D. R. Patterson* and E. J. Denlinger

RCA Laboratories, Princeton, N.J. 08540

Abstract—This paper describes a temperature-compensated hybrid laser module for fiber-optical communications. A simplified compensation technique is introduced that employs the temperature dependence of the emitter–base voltage of the silicon driving transistor to compensate for the temperature dependence of the laser threshold current. Such modules have been built that operate at ~25 Mbit rates while maintaining the output from a 50 μm core diameter, 0.25 NA graded-index fiber within 0.5 dB of 5 mW from −20 to 70°C. Over this range, the maximum power input is less than 375 mW. The module reliability and the use of lasers without prebias also are discussed in detail.

Introduction

The threshold current of a semiconductor injection laser is temperature dependent.[1] To maintain the laser light output constant with respect to temperature, optical feedback or thermoelectric temperature control is often employed. Optical feedback uses an optical detector near the laser to automatically adjust the current such that the laser light falling on the detector remains constant, while thermoelectric temperature control uses a thermoelectric element and a temperature sensor near the laser heat sink to maintain a constant heat-sink temperature. In this paper, we describe the properties and advantages of a simple laser driver circuit with built-in compensation for ambient temperature variations. The temperature dependence of the emitter-base voltage of the silicon

* Present address, Optical Information Systems, Exxon Enterprises, Elmsford, NY.

driver transistors is utilized as the compensating element, and therefore no additional circuit components are required. This design philosophy minimizes the power consumption and maximizes the circuit reliability due to the small number of circuit components.

The driver-current temperature compensation can be employed in two different circuit configurations, one that dc biases the laser just above threshold and another that modulates the laser above threshold without dc bias. In the first configuration, any modulation current necessary to increase the laser output appreciably above the threshold level is superimposed on top of the bias current, and this modulation current also may be temperature compensated to adjust for any changes in the laser slope efficiency (change of light output for a change of drive current) with temperature.

The other configuration, which we studied in some detail, uses no dc bias, and supplies the laser with a modulated current in excess of the laser threshold. The magnitude of the modulated current must be compensated for the temperature variation of the threshold current in order to maintain the peak light output constant. Although there are limitations on modulations rates ($\lesssim$100 MHz), this circuit provides the simplest possible driver with the least power consumption. Low-threshold lasers ($I_{th} \approx 50$ mA) have been incorporated into miniature hybridized driver modules of this type to maintain the light output within 0.5 dB of 5 mW from a mode-stripped, one-meter length of 0.25 NA, 50-μm-core-diameter graded-index fiber over a temperature range of -20 to 70°C. Over this range, the total power consumption of the module is less than 375 mW.

In this paper, we will discuss the advantages and limitations of this open-loop, temperature-compensation method, both in terms of initial performance and long-term reliability. Emphasis will be placed on the unbiased configuration employing a cw injection laser; this bias arrangement has not been extensively explored in present systems.

Principles of Operation

Many factors control the temperature dependence of the threshold current of injection lasers. A common approximation for this dependence, first developed for GaAs homojunction lasers,[2] is that the threshold current I_{th} is proportional to an exponential function of the temperature,

$$I_{th} = K \exp(T/\theta). \qquad [1]$$

Here, T is the ambient temperature in degrees centigrade, K is a proportionality factor which is the threshold current at 0°C, and θ is the

characteristic temperature of the laser in °C. While the characteristic temperature θ may be sensitive to many factors of laser design, for double-heterojunction cw lasers of AlGaAs, the main factor controlling θ has been found to be the bandgap step at the heterojunctions. In Fig. 1 we plot the ratio of the threshold current density at 70°C to that at 22°C ($J_{th}(70°C)/J_{th}(22°C)$) for AlGaAs lasers[4,5] as a function of heterojunction bandgap step ΔE_g, in electron volts (eV). As can be seen, this data is well fit by Eq. [1] if we assume that

$$\theta = 300\ (K/\text{eV})\ \Delta E_g(\text{eV}). \qquad [2]$$

The right hand axis of this graph gives the value of θ corresponding to the $J_{th}(70°C)/J_{th}(22°C)$ ratio. The bandgap step is usually set between 0.25 and 0.35 eV for cw AlGaAs injection lasers for optical communications, which sets the maximum range of θ between 90 and 150°C.

In the driver circuit, the temperature dependence of the emitter–base voltage (V_{BE}) of silicon bipolar transistors is a linear function of temperature, and has been used in the design of monolithic integrated circuits that require stable operation over large temperature ranges.[6] The

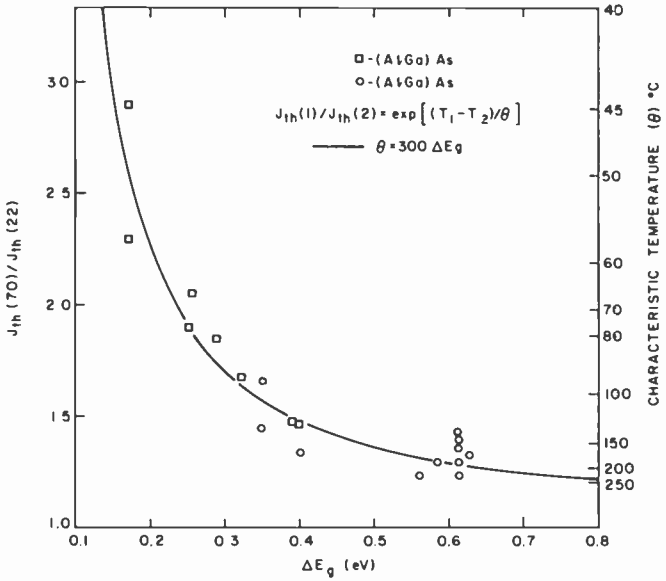


Fig. 1—Threshold current density ratio ($J_{th}(70°C)/J_{th}(22°C)$) as a function of bandgap step at the heterojunction. Data for AlGaAs double-heterojunction lasers (○), is from H. Kressel and M. Ettenberg[4] and data (□) is from Goodwin et al.[5]

magnitude of V_{BE} and its temperature coefficient depends on the transistor type, but is fairly constant within one transistor type.

One circuit configuration that uses the change in V_{BE} with temperature to compensate for the temperature change in laser threshold is illustrated in Fig. 2. This particular circuit is designed for unbiased laser operation and is designed to interface with circuits at TTL (Transistor-Transistor-Logic) rates (~25 Mbits/sec). The circuit is composed of two Schottky TTL gates followed by two transistors in an emitter-follower configuration, with the laser diode in the collector circuit of the second transistor. For this circuit, it can be shown that

$$I_{laser} \approx \frac{V_0 - 2V_{BE}}{R_2}. \qquad [3]$$

By differentiating Eqs. [1] and [3] with respect to temperature and setting $I_{laser} = I_{th}$, we can calculate the values of R_2 and V_0 that match the temperature dependences of the driver current and laser threshold at a given temperature. Specifically,

$$R_2 = \frac{-2\theta}{I_{laser}}\left(\frac{\Delta V_{BE}}{\Delta T}\right) \qquad [4]$$

and

$$V_0 = -2\theta\left(\frac{\Delta V_{BE}}{\Delta T}\right) + 2V_{BE}. \qquad [5]$$

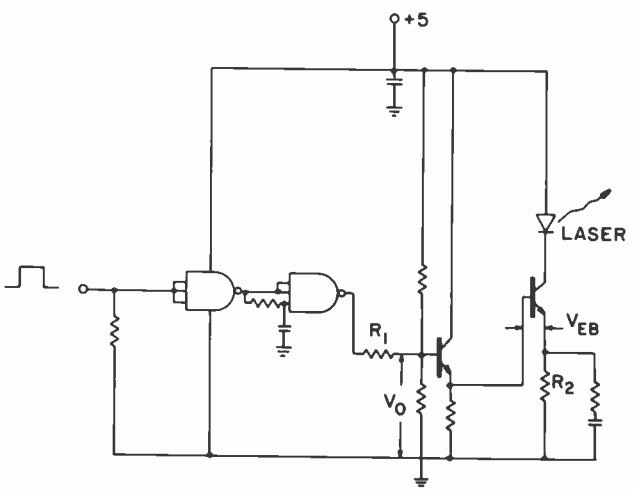


Fig. 2—Circuit design for temperature-compensated, unbiased operation of injection lasers.

Typical injection lasers have a threshold current between 40 and 200 mA and a value of θ between 70 and 200°C. For common high-frequency Si transistors, we have measured $\Delta V_{BE}/\Delta T = -1.9$ mV/°C and $V_{BE} \cong 0.67$V. Substitution of these values in Eqs. [4] and [5] set the range of V_0 at between 1.6 and 2.1 V and the value of R_2 at between 1.6 and 19 ohms. These are reasonable values for solid-state Si circuitry, and provide a minimum increase in power dissipation over and above that required to operate and modulate the laser in any other circuit configuration.

Using Eqs. [1], [4], and [5] to match the laser and circuit at 20°C, we plot in Fig. 3 the calculated current output of the circuit and the threshold current as a function of temperature from −40 to +80°C for the reasonable range of threshold currents at room temperature I_{th} and the threshold temperature sensitivities θ. As might be expected, the match over this large temperature range is best for lasers with the smallest threshold temperature dependence and lowest threshold current. These are generally desirable properties and our state-of-the-art lasers approach the best values illustrated in Fig. 3.

The rest of this paper will deal mainly with the operational characteristics of a hybrid module using this compensating circuit concept.

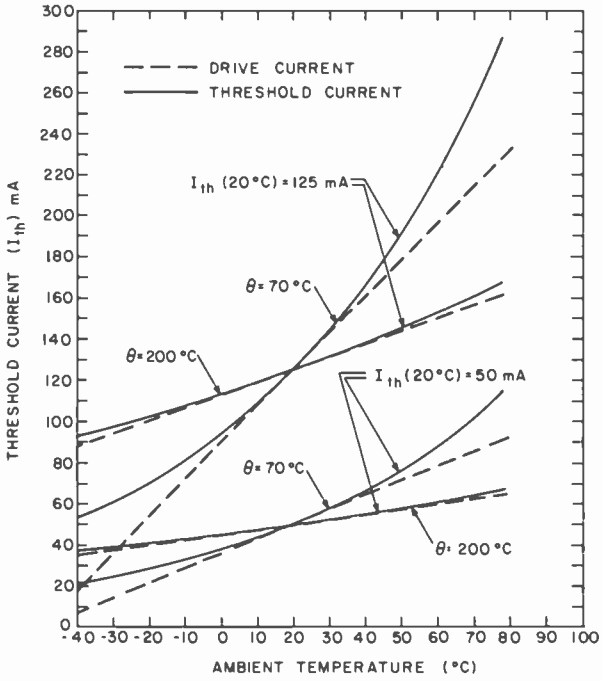


Fig. 3—Calculated output drive current and laser threshold current as a function of temperature for lasers and driver circuits matched at 20°C.

Laser Module Characteristics

Using basically the circuit illustrated in Fig. 2, we have constructed a laser module that contains a hybrid circuit with chip components, an injection laser, and a fiber pigtail. A picture of the hybrid module with the cover removed is shown in Fig. 4.

The injection laser used is a low-threshold, high-efficiency device (I_{th} at 22°C is ≈50 mA) described previously;[7] the fiber pigtail is a 1 meter length of mode-stripped graded-index fiber with a diameter of 50 μm and a numerical aperture of 0.25. The hybrid circuit elements (chip transistors, capacitors, and resistors) are mounted on an Al_2O_3 carrier containing the electrical interconnects. The laser, with prealigned fiber, is fastened into a slot in a BeO base block, and the entire circuit is placed over it. The BeO allows the module to be well heat sunk while still being electrically isolated. The four electrical leads in Fig. 4 correspond to circuit common ground, signal input, and two power supply leads. These two power supply leads allow the power to the driver circuit and to the laser to be independent. They may be connected together internally, reducing to one bias level requirement, as shown in Fig. 2.

In Fig. 5 we plot the peak optical power from the mode-stripped fiber pigtail and the total current input from a 5V power supply as a function of ambient temperature for a module operating at a 50% duty cycle and a 25 Mbit/sec rate (20 nsec wide pulses with a 40 nsec period). As can be seen, the light output remains constant at ~5 mW (7 dB) within ±0.5 mW over a temperature range of −20 to 70°C with a total maximum

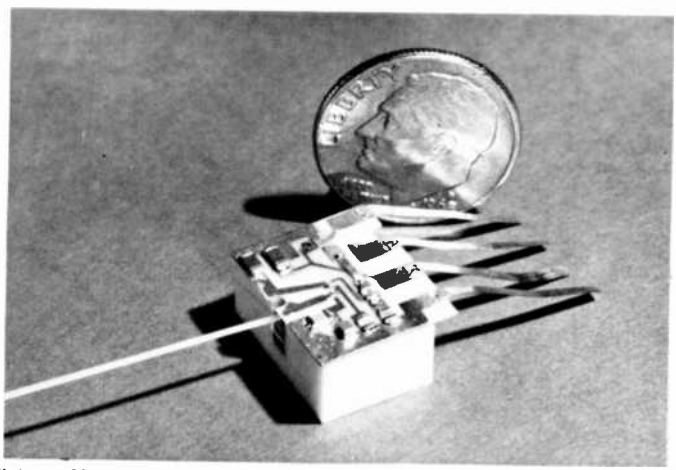


Fig. 4—Picture of hybrid laser module with its cover removed to reveal the circuit and a dime in the background.

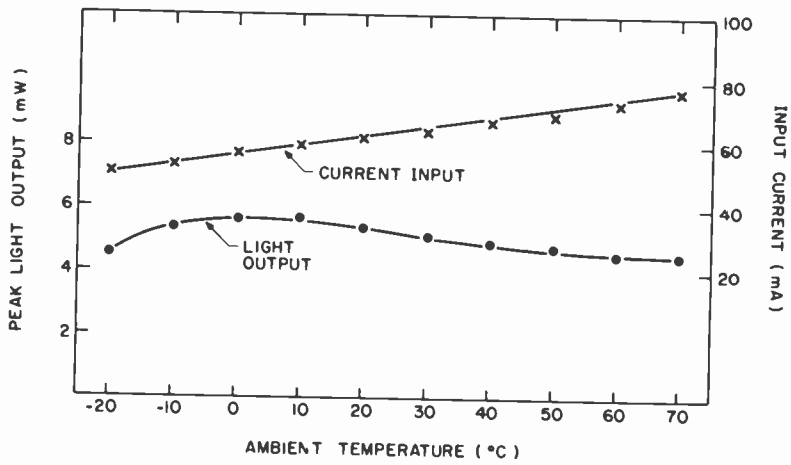


Fig. 5—Peak light output from the module illustrated in Fig. 4 as a function of ambient temperature. Power input is from a 5 V dc power supply, and output is at 0.82 μm from a 0.25-NA, 50-μm-core-diameter graded-index fiber.

power input of less than 375 mW. This relatively low power consumption is not achieved without some trade-offs in pulse shape and overall modulation capability.

In the particular module discussed, we are mainly limited in speed by the ~25 Mbit/sec limitation of the Schottky TTL interface. However, even with a faster interface such as ECL (Emitter Coupled Logic), there are limitations on the modulation capabilities of an injection laser operated without prebias in such a circuit configuration as discussed further below.

These limitations are associated with the operating characteristics of the circuit and the turn-on delays with unbiased operation of the laser. If a current pulse with a rise and fall time of τ_I is presented to the injection laser, the laser will turn on when the current reaches the threshold current (I_{th}). Thus, the start of the stimulated light emission process will be delayed from the start of the current pulse of magnitude I by $(I_{th}/I)\tau_I$. The second and more fundamental contribution to the laser turn-on is associated with the injected carrier or spontaneous lifetime; this is the time necessary for the carrier population to become inverted *after* the input current has reached threshold. The lasing delay (τ_d) for population inversion is given by:[8]

$$\frac{1}{\tau_d} = \frac{1}{\tau_{th}} \ln\left(\frac{I}{I - I_{th}}\right). \qquad [6]$$

For injection lasers $\tau_{th} \approx 2.5$ nsec[1] and for our low-threshold lasers (I_{th} ~ 50 mA) operating at 10 mW output, we require about 80 mA of drive

current; from Eq. [6], τ_d should thus be about 2.5 nsec. To give an example, we describe the situation for the use of injection lasers without prebias with TTL logic rates of about a 25-Mbits return-to-zero code. The result is illustrated graphically in Fig. 6. For TTL logic, the current pulses have rise and fall times of about 4 nsec. Thus, for the low-threshold laser discussed above, the delay associated with current rise-time (τ_I) is (50/80)4 = 2.5 nsec. The turn-on delay should thus be 2.5 nsec for the current to reach threshold and 2.5 nsec for the population to invert, resulting in a *total* turn-on delay of 5 nsec. This is inexact because of the variation in pulse shape and small differences in τ_{th} from laser to laser. However, we have studied a variety of lasers and modules and, in general, for lasers with I_{th} = 50 mA operated at ~80 mA, we observe as illustrated in Fig. 6b a 15-nsec-wide light pulse for a 20-nsec-wide input signal pulse. This is achieved only when delay effects within the circuit itself are minimized.

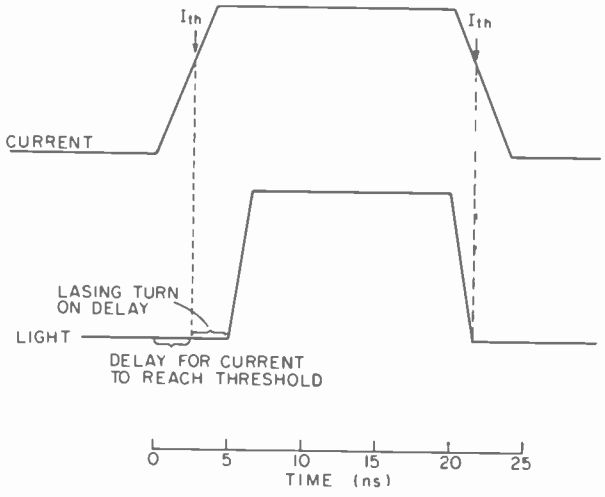
The 5-nsec delay can be reduced by a variety of means: first, by simply making the current rise time faster (although this may induce oscillatory resonances within the optical output); second, by distorting the width of the current pulse within the circuit to compensate for the delay; or, third, by driving the laser at a higher current. This last method may have a deleterious effect on laser operating life, and is therefore not preferred. Due to the delays associated with unbiased operation, which are at least a few nanoseconds, unbiased laser operation is limited to less than 100 MBits/sec in a return-to-zero format, regardless of the electrical interface.

Reliability

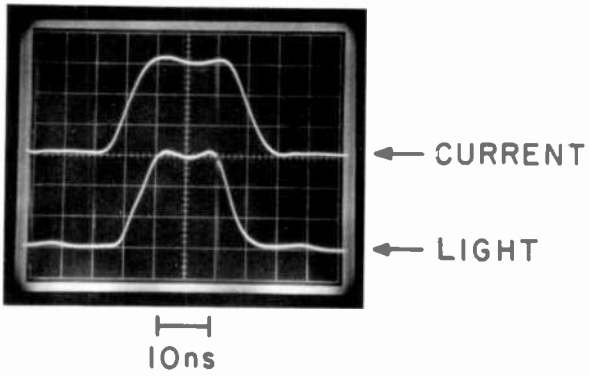
The final module characteristic of importance is reliability. Since the reliability of Si components and hybrid circuitry has been firmly established, we will concentrate our discussion on how the driving circuitry affects the reliability of the injection laser.

Injection laser failure is strongly temperature dependent, with an activation energy between 0.7 and 0.9 eV;[9,10] thus, if the laser module is to be used in an environment that is much above room temperature, the most reliable module would include thermoelectric cooling for laser temperature stabilization. This approach is rather costly in terms of power consumption, however, and thus is precluded for many remote applications. The other reasonable approaches are optical feedback and the open loop circuit compensation discussed here. These techniques can be compared in terms of reliability by studying the manner in which lasers degrade.

In Fig. 7, we plot the failure statistics for low-threshold lasers of the



(a)



(b)

Fig. 6—(a) Graphical representation of pulse input and light pulse output for an injection laser without dc bias and (b) oscilliscope trace of current input and light output during pulsed operation of unbiased laser.

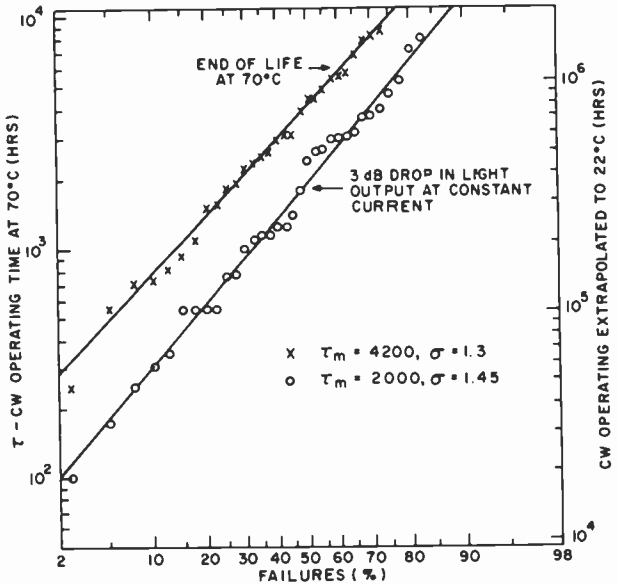


Fig. 7—Failure statistics for 40 low-threshold lasers operated cw at 70°C. Data is from paper by M. Ettenberg (Ref. [11]).

type used in this module. These are previously obtained results[11] of accelerated (elevated temperature) life tests. The lasers were operated cw at 70°C with ~5 mW output. Two failure points are indicated. The first is the point at which the laser degrades to half its initial output (3 dB) at constant current. The second (end of life) is obtained by increasing the current as the laser degrades, and is the time at which the laser can no longer emit ~1 mW of stimulated emission cw at 70°C, independent of current input. The first failure point thus corresponds to a constant-current temperature-stabilized drive, as described in this paper, and the second corresponds to an optical feedback drive. Both sets of failures follow similar log-normal statistics and extrapolate to room temperature with meantime-to-failure values of $\sim 5 \times 10^5$ and 10^6 hours, respectively. It would thus appear that the optical feedback provides up to twice the operating life of temperature compensation.

However, interpretation of these results in this manner is somewhat misleading. It was found that after significant degradation takes place, there are changes in the laser characteristics that may reduce their usefulness in fiber communication systems, regardless of the circuit approach used. In Fig. 8, we show the injection laser's lateral far-field pattern at 22°C before and after failure at 70°C. This (typical) result illustrates a significant degradation of the lateral modal pattern asso-

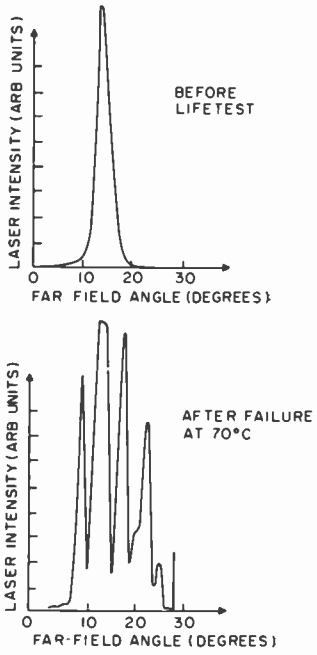


Fig. 8—Far-field distribution in the plane of the junction for low-threshold lasers before and after failure at 70°C.

ciated with gradual degradation; the coupling to fibers also will degrade as the modal pattern changes. Furthermore, we have found that self pulsations at high frequencies (>100 MHz) are observed after degradation. These pulsations are similar to those reported by Paoli[12] for degraded lasers, and may prematurely degrade the performance of optical fiber systems. The point is that even if the total light output of the laser can be maintained twice as long with optical feedback, that is not the sole criteria for adequate fiber system performance; therefore, even the factor of two advantage in operating life for optical feedback versus constant-current operation is quite optimistic. Furthermore, even though the reliability of injection lasers has been improved to the point where fairly reliable systems can be implemented, the laser is still about a factor of 1000 less reliable than comparable high-reliability Si components.[13] As the reliability of the injection laser improves, the relatively small potential advantages of optical feedback for reliability will be far outweighed by the simplicity of open-loop temperature compensation.

Finally, it should be noted that the unbiased temperature-compensated module described in this paper is operated at 50% duty cycle, so that the usable laser life may be extended by a factor of two over that

determined from cw life-tests, since it is only during operation that the laser degrades.

Conclusion

We have described an open-loop temperature-compensated drive circuit for injection lasers and have constructed such circuitry in the form of a miniaturized hybrid module mated to a fiber pigtail and a low-threshold laser for applications where size and power consumption are at a premium. The module maintains the light output within 0.5 dB of 5 mW from a mode-stripped graded-index fiber (50-μm core diameter, 0.25 NA) over a temperature range of -20 to 70°C.

Independent accelerated lifetests of the laser places a meantime-to-failure in the module at $\sim 10^6$ hours at room temperature. Analysis of the laser reliability data indicates that the complexity and power consumption introduced by optical feedback techniques will only marginally increase the laser life above that obtained with the temperature compensation circuitry described here.

References:

[1] For example see H. Kressel and J. K. Butler, *Semiconductor Lasers and Heterojunction Lasers*, Academic Press, New York 1977.

[2] J. I. Pankove, "Temperature Dependence of Emission Efficiency and Lasing Threshold in Laser Diodes," *J. Quant. Elec.*, **QE-4**, p. 119, April (1968).

[3] M. Ettenberg, C. J. Nuese, and H. Kressel, "The Threshold-Temperature Dependence of Double Heterojunction Lasers," *J. Appl. Phys.* (to be published).

[4] H. Kressel and M. Ettenberg, "Low Threshold Double Heterojunction AlGaAs/GaAs Lasers Theory and Experiment," *J. Appl. Phys.*, **47**, 3533, Aug. 1976.

[5] A. R. Goodwin, J. R. Peters, M. Pion, G. H. B. Thompson, and J. E. A. Whiteaway, "Threshold Temperature Characteristics of Double Heterostructure $Ga_xal_{1-x}As$ Lasers," *J. Appl. Phys.*, **46**, p. 3146, July 1975.

[6] L. P. Hunter, *Handbook of Semiconductor Electronics, Third Edition*, McGraw Hill, New York (1970).

[7] M. Ettenberg and H. F. Lockwood, "Low Threshold Current CW Injection Lasers," *Fiber and Integrated Optics*, **2**, 47, Nov. 1977.

[8] Jose E. Ripper, "Measurement of Spontaneous Carrier Lifetime from Stimulated Emission Delays in Semiconductor Laser," *J. Appl. Phys.*, **43**, p. 176, April 1972.

[9] R. L. Hartman and R. W. Dixon, "Reliability of DH GaAs Lasers at Elevated Temperatures," *Appl. Phys. Lett.*, **26**, p. 239, March 1975.

[10] H. Kressel, M. Ettenberg, and I. Ladany, "Accelerated Step Temperature Aging of $Al_xGa_{1-x}As$ Heterojunction Laser Diodes," *Appl. Phys. Lett.*, **32**, p. 305, March 1978.

[11] M. Ettenberg, "A Statistical Study of the Reliability of Oxide-Defined Stripe CW Lasers of (AlGa)As," *J. Appl. Phys.*, **50**, p. 1195, March 1979.

[12] Thomas L. Paoli, "Changes in the Optical Properties of CW (AlGa)As Junction Lasers During Accelerated Aging," *IEEE J. Quantum Elec.*, **QE-13**, p. 351, Feb. 1977.

[13] D. S. Peck and C. H. Zierdt, Jr., "The Reliability of Semiconductor Devices in the Bell System," *Proc. IEEE*, **62**, p. 185, May 1974.

The Adherence of Chromium and Titanium Films Deposited on Alumina, Quartz, and Glass—Testing and Improvement of Electron-Beam-Deposition Techniques

Martin Caulton, William L. Sked, and Francis S. Wozniak

RCA Laboratories, Princeton, N.J. 08540

Abstract—The adherence of standard electron-beam (EB) evaporated chromium-copper metallization on alumina substrates was found to be marginal, especially when the metallization patterns are both narrow (<10 mils) and thick (>7 μm). The adherences of chromium and titanium films were therefore measured under different evaporation conditions on alumina, quartz, and glass. In the tests the seed metal is plated with 2 or more mils of copper, and the force per unit width necessary to pull the chromium-metal film off the substrate is measured. It was found that although titanium, resistance-filament chromium, and sputtered chromium-rich films had a good adherence, it is necessary for EB evaporated chromium on alumina and quartz to add oxygen in a definite pressure range to produce a superior adhesion. The details of the testing technique and the peel strengths of chromium on alumina, quartz and glass under different deposition conditions are described. The addition of oxygen during the initial EB chromium evaporation produces an adhesion that is as strong as the materials used.

1. Introduction

The successful fabrication of integrated circuits depends on the adherence of metal patterns to underlying substrate material. In standard thin-film technology, adherence is accomplished by the use of a resistive reactive metal, such as chromium or titanium, that chemically interacts with the oxygen in the substrates (i.e., Al_2O_3). In the same vacuum run, the pure chromium deposited is covered by a seed layer of a good con-

ductor such as copper. Normally, if nominal requirements of circuit environment and pattern geometry are met, the adherence is not closely examined. This is also true for other combinations of materials, such as adhering layers Ti-Pt, Ti-Pd, or Mo-W deposited on Al_2O_3, BeO, SiO_2, etc. However, if the environment of the circuit is subject to high temperature and/or corrosive gases, including water vapor, the adhesion must be compared to some standard. In addition, if the patterns are narrow and/or thicker than 5 μm, then adherence may be marginal and again must be measured and compared to some standard. Last, if metallization patterns fail to withstand normal handling and processing, obviously something is wrong, and tests are necessary to quantify this failure so that it can be corrected. The testing method varies with the nature of the problem, as there is no one method that will satisfy all situations.[1,2]

In this paper, we discuss an adherence problem encountered with chromium-copper deposited by an electron beam (EB) system. The films (300-Å Cr, 3–5 μm of Cu) passed normal tests, i.e., tape, razor scraping, and handling with tweezers, that would at most damage the metal but not peel it. We discovered our adherence problem when, for a particular application, it was necessary to plate lines 3 mils wide to thicknesses of 10 μm or greater. It was then observed that a tweezer accidentally touching this line could cause complete removal. It was determined that this was not due to processing, but was caused by marginal adherence that only became apparent with thick narrow lines that could become "handles" for lift-off. Thinner and wider lines could not be as readily removed because they did not provide "handles". By reproducing the conditions for failures, a testing procedure was developed.

This paper describes (1) the simple testing procedure developed to quantify this marginal adherence, (2) the result of the testing procedure, and (3) deposition conditions that improve the adherence by orders of magnitude. In the study we investigated the adherence of chromium and titanium films on alumina (99.5% Al_2O_3), quartz (100% SiO_2), and glass. We found that addition of oxygen in a definite pressure range during the first half of the initial deposition of chromium, closely followed (with the oxygen removed) by pure chromium and the next metal, improved the adhesion strength by orders of magnitude.

2. Simple Quantitative Peel Test

As stated, the adhesion problem was first observed when the metal thickness was increased by electroplating to values greater than the nominal 5 to 7 μm normally used, and it was noticed primarily with narrow (<10 mil) lines. However, we found that even larger areas of metal

could be peeled with thicker plating. Our first step was to electroplate the alumina circuit with 1 mil or greater of copper, break the alumina, and then attempt to pull the two apart. It was found that the thick copper did not break, and the chromium-copper metallization peeled smoothly and easily from the substrate leaving clean white alumina behind. This technique is a "practical" measurement of a failure mechanism. Some other techniques in the literature[3,4] are more comprehensive, but do not apply directly to the problems described here. A technique similar to this is described in Ref. [5].

Fig. 1 outlines the technique. The basic procedure is to plate the evaporated Cr-Cu on the substrate with copper to a thickness of 1 or 2 mils [Fig. 1(a) and (b)]. This thick copper forms a handle with a cross-sectional strength greater than the weakest of the successive interfaces [Fig. 1(b)], evaporated Cu/plated Cu, Cu/Cr, Cr/alumina. It should not rip or tear before one of the layers peels apart. By sawing a 200-mil wide strip [Fig. 1(c)] and breaking it near the center, we could then determine the force necessary to separate the metal from the substrate [Fig. 1(d)]. It is also possible to scribe the metal strip on the plated substrate, and start the metal peeling with a pair of pliers.

Initially, we used the Instron* pull tester to give a calibrated graph of force versus time. Later we simplified the measurement by using a spring pressure gauge (clipped to pliers) to measure the force on the metal strip. The substrate is held down by hand. Forces on 10 to 200-mil (5.1 mm) wide strips were measured for different depositions. Tables 1–3 give a summary of peel strengths expressed in grams per millimeter for different substrates and conditions. These results will be discussed later; the data have been shown to be reproducible under many test conditions.

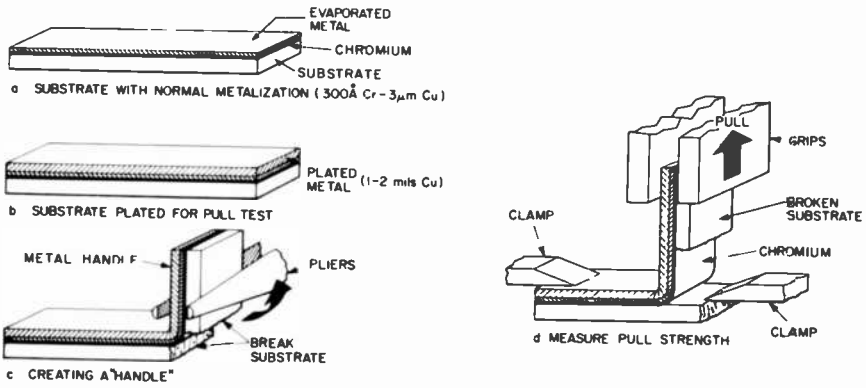


Fig. 1—Technique to create a pull system for the measurement of adhesion.

* Instron Corp., Canton, Mass.

The test technique was then simplified as follows:

(1) Plate the metallized substrate to a minimum thickness of 2 mils of copper in an acid copper bath. Thinner films (1 mil or less) are used for materials with lower adhesion strengths, i.e., glass and quartz.
(2) Cut off or delineate a 100- to 200-mil wide strip of substrate.
(3) Start the edge peel with tweezers and/or pliers so that a "handle" of metal can be formed. One can also do this by breaking the substrate strip in half and using half of the substrate as a handle [see Fig. 1(d)].
(4) Measure the peel strength with a gauge and normalize the peel force in g/mm. Different widths can be used (10 to 100 mils) to obtain a workable force that can be readily measured. The pulling force varies linearly with width except for substrates with very poor adherence.
(5) The measuring strip can be sawed off a large substrate that has been metallized to the edges and then plated. Adhesion can then be measured, while the rest of the substrate is left for circuit use. Fig. 2 is a photograph of a metallization being peeled off as described.

Workers elsewhere[1,2,5,6] have been using a measurement technique similar to the one described here. They have peeled their films at angles other than 90°. The adhesion strengths measured are much less than those described in Table 1, as their substrate material was SiO_2 (Tables 2–3) and the development of a "handle" of thick metal for cross-sectional

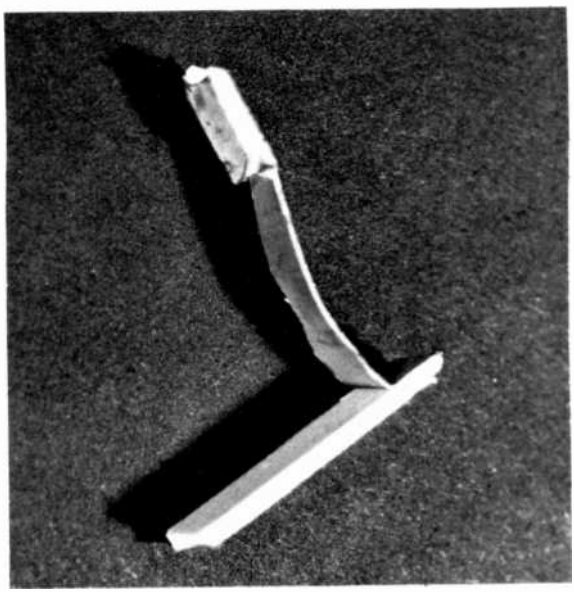


Fig. 2—Photograph of 200-mil wide substrate with electroplated handle.

Table 1—Peel Strengths for EB Evaporated Chromium Copper on Alumina Substrates for Various Test Conditions. Except Where Noted, Standard Evaporation Using MRC MARZ* Grade Cr 150-300 Å Thick. If Present, O_2 for Half Cr Deposition.

Test Condition	Metallization Number	Oxygen Pressure (Torr)	Peel Strength* (g/mm) I	G	Remarks
Standard	7	—	144		Clean Cr-Al_2O_3 Separation
Cleaned E-Gun	9	—	117		Clean Cr-Al_2O_3 Separation
High Starting Pressure (4×10^{-6})	10	—	54		Clean Cr-Al_2O_3 Separation
Screened E-Gun Filament	11	—	45		Clean Cr-Al_2O_3 Separation
High Starting Pressure	13	—	72		Clean Cr-Al_2O_3 Separation
UMC Chromium	18	—	180		Clean Cr-Al_2O_3 Separation
Fisher Cr	19	—	112		Clean Cr-Al_2O_3 Separation
Fisher Cr	20	—	90		Clean Cr-Al_2O_3 Separation
O_2	21	1×10^{-5}	143		Clean Cr-Al_2O_3 Separation
O_2 Not Shut Off	22	5×10^{-5}	5		Cr Separated from Cu
O_2	23	5×10^{-5}	+1180		60-mil Strip, Cu Broke
O_2	24	5×10^{-5}	+377		Ceramic Broke
O_2	25	3×10^{-5}	+337		Grip Slipped
Back Side Substrate (27)	27	4×10^{-5}	+1610		50-mil Strip
O_2	28	4×10^{-5}	224		Surface Roughness ~10 microinches
Back Side Substrate (29)	29	4.5×10^{-5}	+670		75-mil Strip, Cu Strip Separated
MRC Starter plug, min of 100 ppm O_2	30	4×10^{-5}	350		Surface Roughness ~10 microinches
O_2	34	No O_2	215		
Back Side Substrate (39)	39	4×10^{-5}	+440		Cu broke
O_2	40	5×10^{-5}	395		Irregular Separation
30-μm Cu Evaporation	44	7×10^{-5}	1620		
	85			50	No Plated Copper

* I—Instron Test; G—Spring Gauge.

Table 2—Peel Strength for EB Evaporated Chromium-Copper on Alumina, Quartz, and Glass for Various Test Conditions. Standard EB Evaporation Using MRC MARZ* Grade Chromium 150–300 Å Thick

Test Condition	Metallization Number	Oxygen Pressure (Torr)	Peel Strength* (g/mm) Alumina I	Alumina G	Quartz I	Quartz G	Glass I	Glass G	Remarks
Standard after O_2 run	49	?	402			2	45	22	200-mil strip
							29		500-mil strip
Standard	50	4×10^{-5}				3			200-mil strip
Standard	51	4×10^{-5}				4			
Standard	52	9×10^{-5}			7	6		28	500-mil strip I
									200-mil strip G (heavy Cr without O_2)
Standard	54	wet H_2O 4×10^{-5}			2	2	39		500-mil strip I
									200-mil strip G
Standard	54							22	500-mil strip
Henley Cr	61	No O_2			2	2	14		500-mil strip I
									200-mil strip G
Henley Cr	64	—					11		500-mil strip I
Henley Cr	65	—					11		500-mil strip I
Standard	71	2×10^{-4}		157		37		6	200-mil strip
Standard	73	3×10^{-4}		394		39		6	200-mil strip, quartz pulled away with Cu in 71 and 73

* I—Instron Test; G—Spring Gauge.

Table 3—Peel Strengths for Different Metal Depositions (No Oxygen Added During These Evaporations)

Description	Metallization Number	Peel Strength (g/mm) Alumina I	Quartz G	Glass G	Remarks
Cr-Cu Evaporation	S-020	+943			Glue Grip Failed (Metallization made 4/68)
Cr-Au Evaporation	S-041	386			Insufficient Cr (Metallization made 4/68)
Cr-Au Evaporation	S-045	+296			Copper Broke (Metallization made 4/68)
Sputtered Cermet-Cu*	35D	+817			Copper Broke
Sputtered Cermet-Cu* Backside 35D	b-35D	197			Clean Cermet-Alumina Separation
Sputtered Cermet-Cu*	39D		21		
Ti-Au, Different EB System	S-019	+898			Copper Broke
Ti-Cu	14	+735			Ceramic Broke
Ti-Cu	15	+1170			Ceramic Broke
Ti-Cu	53		11	13	

* Cermet: 80% Cr, 20% SiO_2 by Weight
I—Instron; G—Spring Gauge.

strength was not as important to them. T. T. Hitch[7] also describes adhesion tests for thick films for which he obtains peel strengths that can be compared to those quoted here, but the technique is different.

3. Deposition Techniques

Using the measurement technique described, we examined samples of metallization produced by different groups in our laboratories. In particular, we found that chromium-copper deposited by a resistance-heated-filament system had very good adherence. A 10-mil-wide line was peeled off with the alumina or sapphire sticking to it. This evidence and the examination of newer evaporations using resistance-filament systems (Table 3, metallizations S-045, S-020, etc.) are confirmation that good chromium-alumina adherence is possible. Measurements of sputtered cermet (20% SiO, 80% Cr)-copper (Table 3, number 35-D) also produced a very high pull strength (causing either breaking of the cermet or peeling). Our conclusion is that the filament-evaporation and sputtering systems allowed the achievement of an "ultimate" adherence, i.e., one as strong as the substrate material. The question of what is wrong with the electron beam (EB) deposited chromium is thus posed. All measurements of the chromium-copper depositions with EB systems (Table 1, metallizations 7, 9, 10, 11, 13, 18, 19, 20) showed pull strengths of less than 200 g/mm.

Our first efforts were to remove contaminants, to clean all exposed areas in the fixturing and substrate holders, and to screen the EB filament. This gave no improvement (Table 1, metallizations 9, 10, 11, 13); the pull strength varied from 40 to 150 g/mm. A 200-mil (5.1 mm) wide strip of copper was used as a standard. Next, to test gaseous contamination, we used a residual gas analyzer. The system proved to be very clean during evaporation (only traces of elemental gases in the 10^{-9} torr region were observed). However, when the chromium shutter was opened and deposition on the substrates began, a hydrogen peak would appear, sometimes as large as 10^{-6} torr. This peak was also observed during copper evaporations. The shutter opening was accompanied by a rise in the background pressure from 7×10^{-7} to 4×10^{-6} torr over a time span of several seconds.

In addition, a run of titanium-gold was made (Table 3, metallization S-019) in another EB system. The adhesion was superior–over 900 g/mm. Because of these good results, we ran titanium-copper (Table 3, metallizations 14, 15) and found superior adhesion, again greater than 1200 g/mm. It should be mentioned that in most of the experiments reported here, the substrate surface roughness was less than 2 microinches CLA. Some alumina had back sides with a partially polished surface of 10 microinches CLA.

4. Improvement of Adherence Using Oxygen

The better adherence of titanium as compared to that of chromium can be explained or implied by a better gettering of the hydrogen and other gases. We tried to improve the removal of the hydrogen by adding oxygen to the atmosphere during the beginning of the chromium deposition.[8] This turned out to be a relatively simple matter. By using a metering needle valve between an oxygen tank and an input port we could control the partial pressure of oxygen; and by use a shut-off valve the oxygen could be pumped out rapidly at the time desired.

4.1 Studies on Alumina

We introduced a partial pressure of oxygen into the system after the chromium source was outgassed and heated and just before the shutter was opened for deposition. With the O_2 pressure stable, we evaporated half of the chromium thickness with oxygen present and shut off the oxygen during the remainder of the chromium deposition. With a partial pressure of oxygen of 4×10^{-5} torr, the electron beam current fell about 20%. Adherence as a function of oxygen pressure was then tested. It was found that as the oxygen pressure exceeded 3×10^{-5} torr, the adherence increased markedly, so much so that the ceramic or the copper foil broke or tore off instead of the copper foil handle pulling away (Table 1, metallizations 23, 24, 27, 28, 39). In the table, + designates the value measured when the material failed (i.e., the copper or ceramic broke or tore) before the full peel strength was reached. Due to an error in metallization 22 (Table 1) the oxygen was not shut off and the copper peeled away from the chromium oxide, as might be expected. As oxygen pressure was increased to 8×10^{-5} and beyond, the alumina adherence started to fall (Table 2, metallizations 71, 73).

The question was raised of whether the copper-plating itself causes poor adherence. We checked this using metallization 85 (Table 1) in which the copper handle was formed by evaporating excess copper (30 μm). The typically poor pull strength without oxygen (50 g/mm) resulted without the metallization having experienced immersion in a plating bath.

The chromium generally used was MRC MARZ grade, specified very pure. The reason for the improved adherence with O_2 is not understood. One suggestion is that chromium oxide is more soluble in Al_2O_3 than is pure chromium. Another is that excessive hydrogen, seen in the residual gas analysis when the chromium source is heated, must be removed for better adherence. As observed, the chromium-plated tungsten filament in a resistance boat evaporation without the introduction of oxygen

achieves the same excellent adhesions as reported here. In order to see if a different supply of chromium would either (1) have more oxygen or (2) have less hydrogen so that an external supply of oxygen is not required, we studied chromium from other suppliers. The only one showing a measureable improvement (215 g/mm, metallization 34, Table 1) used a slug (melted to crucible size) prepared by MRC with analyses of 100 to 300 ppm of O_2. Other chromium sources showed varying results. It is concluded that oxygen should be added as an integral part of the EB evaporation of chromium for best adherence; this ensures a bond as strong as the materials taking part. The range of pressures recorded is that used in our particular vacuum system. The best range of pressures (as described in Table 4) for optimum adhesion will probably vary according to how well the ionization gauge measures the environment at the substrate, and will probably be different for each vacuum system, rate of deposition, and substrate temperature.

4.2 Studies on Quartz and Glass

We next examined the adherence of chromium to quartz and glass in preparation for metal depositions on SiO_2 films. The results were different from that of alumina in several ways. First, the adherence itself, 2 g/mm without O_2, is more than an order of magnitude lower than that of alumina. We were able to achieve an ultimate bond (quartz layers peeled apart with the chrome-copper strip) only at even higher oxygen pressures than those for alumina. This peel strength (40 g/mm) was only 2% that of alumina. As may be seen in Table 2, peel strengths of 2 g/mm (metallizations 49, 54) were not improved until an oxygen pressure of 9×10^{-5} torr was reached. When pressures as high as 5×10^{-4} torr were used, the material limit in adherence was achieved (Table 2, metallizations 71, 73). A pressure of 2×10^{-4} of oxygen appears to be less destructive (i.e., less oxidation of the fixtures) and capable of good adherence (Table 2, metallization 71). Note also that at these high pressures the bonds to glass and alumina begin to deteriorate.

A study of the glass-chromium bond is also given in Table 2. The peel strength of 22-25 g/mm was approximately constant for no oxygen up to 9×10^{-5} torr, but fell when higher oxygen pressures were used (Table 2, metallizations 71, 73). At the same time, the quartz bond was improving.

The chromium adherence to quartz and glass depended not only on the amount of oxygen present, but also on the rate of deposition of the chromium-oxygen layers. Deposition rates on the order of 100 Å/minute were satisfactory. With faster or slower depositions the adherence did not improve with the addition of oxygen. The variation of adhesion

strength with deposition rate was not evident with alumina. The deposition-rate dependence for adherence on quartz can be caused by stress, which changes with many factors, including deposition rate.[9]

It should be noted that neither our titanium-copper nor sputtered cermet-copper on quartz gave quite as good results as the excess oxygen run of the Cr-Cu (Table 3, metallizations 53 and 39-D) and (Table 2, metallizations 71, 73). However, the adherence of sputtered cermet or titanium is still an order of magnitude improvement over evaporation of Cr-Cu with little or no oxygen. The ultimate bond strength varies with the material, and the nature of the bond appears to be somewhat different.

4.3 Other Observations

Metallizations 28, 30, 40, 39-D (Tables 1 and 3) were all on the back side of substrates that were first evaporated with excellent adherence. These back side metallizations are all inferior to their initial side ones: 224/+1620, 350/+676, 395/+449, and 197/+817 g/mm. One speculates that the heating and exposure to a hot backing plate supplies enough contaminants to spoil the surface for future adherence. This was also noted in evaporations without O_2, but the adherence with O_2 is better. A support without a backing plate would be expected to give an improvement. Accordingly, an open structure for the fixture support, which holds the substrates at the corners or edges, was tested. Surprisingly, this did not always improve the adherence. The back side's lower adhesion was found to be correlated to the roughness of the surface; most of our substrates were highly polished on the top side (<2 microinches) and much rougher (~10 microinches) on the back. It is speculated that an improved cleaning procedure or changes in chromium thickness might improve the adhesion for rough surfaces. This was not investigated.

Metallization number 49 (Table 2) showed a higher than expected peel strength on alumina without oxygen, 402 g/mm. Interestingly, this run was the next one after an experiment in which an excess of oxygen was used, and the system, including the chromium remaining, was very oxidized. It is suspected that enough oxygen remained on the chromium to improve the adherence in the following run, but this is not known.

Oxygen is not a cure-all for all EB chromium evaporations, as noted with the glass bonds described in Table 2. In fact, it can degrade the bond, as noted in some experiments (not reported here) with magnesium titanate and garnet.

It is of interest to compare the bond strengths of our better adhering chromium-alumina depositions with those of the screened inks in thick film depositions. From reports by T. T. Hitch,[7] peel strengths of over

100 lb/in (1790 g/mm) have been achieved with high-adhesion thick-film inks. The tests were made using 40 mil wires soldered to patterned pads. The bonds broke at the high pulls, with ink remaining on both surfaces. The two tests cannot easily be compared; however, in the experiments reported here with peel strengths of over 1600 g/mm (a lower limit, since the material often broke apart), the metallization sometimes peeled cleanly taking chips of alumina with it. Initial attempts to measure pull strengths of thick films by our technique were unsuccessful because it was difficult to plate the inks. Thick films so plated tended to snap off, not allowing pull measurements. It should also be noted that the vacuum system evaporations were all on substrates with a surface roughness of 2 microinches or less, while the thick-films relied on a surface roughness of 20 microinches or greater.

5. Summary and Conclusions

We have described how a particular metallization problem led us to develop a technique to measure the adherence peel strength of thin films. This technique involves plating a strip of metal so that its cross-sectional strength is stronger than the pull force to be applied. The force to peel the copper-backed chrome metallization away from the substrate is then measured. Deposition techniques to improve the adherence of chromium films on alumina (Al_2O_3), quartz (SiO_2), and glass were also developed. These techniques use the addition of oxygen to the initial chromium being deposited on the substrate. One can only speculate as to the reason for the adherence improvement with oxygen, but the presence of hydrogen coming from the pure chromium materials during evaporation is noted.

Table 4 summarizes the maximum peel strengths achieved and measured with and without the additions of oxygen for EB chromium metal evaporation and the other techniques. It is concluded that there is a proper range of oxygen to be inserted in the vacuum system for EB chromium depositions on different materials.

The maximum peel strength measured on alumina and quartz are lower limits in many cases. The metallization with high adhesion peeled away from the alumina and quartz leaving no metal behind but often tearing chunks of alumina or quartz out of the respective substrates. These peel strengths appear to be limits of the substrate material properties rather than of the bond alone. The vacuum system top-side evaporations were all on substrates with a surface roughness of less than 1 microinch CLA. Peel strengths greater than 2000 gm/mm are reported in Ref. [5] for electroplated deposited metals on metal substrates.

The superior adhesion of titanium to alumina and quartz is evident

Table 4—Maximum Peel Strength for Different Conditions

Description	Oxygen Pressure Range (Torr)	Peel Strength (g/mm)
Chromium-Metal EB Deposition		
Chromium on:		
Alumina	—	220
Quartz	—	2
Glass	—	15
Chromium + O_2 on		
Alumina*	$(3–7) \times 10^{-5}$	1600
Quartz*	$(1–5) \times 10^{-4}$	40
Glass	$(0^{+}–9) \times 10^{-5}$	30+
Other Depositions		
Titanium-Metal EB Deposition on:		
Alumina		1200+
Quartz		11
Glass		13
Cr-Filament-Resistance Boat on:		
Alumina		1000+
Sputtered Cermet-Copper on:		
Alumina		900+
Quartz		21

* The maximum pull on alumina and quartz damages the substrate materials.

from the results of Table 4, but the peel strengths do not exceed that of the EB evaporation of chromium with oxygen. The same applies to a resistive-filament-evaporated chromium on alumina. The expected superior adhesion of sputtered material is also noted.

The implementation of the addition of oxygen turned out to be a simple matter, and it is now incorporated in an automatic fashion in the EB chromium deposition system.

Acknowledgments

We are pleased to acknowledge the cooperation of D. M. Hoffman in the use of the residual gas analyzer, as well as in the pertinent discussions that ensued. The authors also wish to thank T. T. Hitch and J. L. Vossen, Jr., of RCA Laboratories for helpful discussions. In addition, the cooperation of the members of the Microwave Technology Center at RCA Laboratories gave many inputs to the nature of the problem and its solution.

References

[1] D. Mattox, 1977 Delaware Valley Symposium, Adhesion of Thin Films, AVS (American Vacuum Society), Princeton, N.J., May 19, 1977.

[2] D. Mattox, "Thin Film Metallization of Oxides in Microelectronics" *Thin Solid Films*, **18**, p. 173 (1973).

[3] B. N. Chapman, "Thin Film Adhesion," *J. Vac. Sci. Technol.*, **11**, No. 1, p. 106 (1974).

[4] K. L. Mittal, "Adhesion Measurements of Thin Films," *Electrocomponent Science and Technology*, **3**, p. 21 (1976).

[5] O. J. Klingenmaier and S. M. Dobrash, "Peel Test for Determining the Adherence of Electrodeposition on Metallic Substrates," *Adhesion Measurement of Thin Films, Thick Films, and Bulk Coatings*, ASTM STP640, K. L. Mittal, Ed., American Soc. for Testing and Materials, 1978, pp. 369–390.

[6] K. Howard, IBM, Armank, N.Y., 1977 Delaware Valley Symposium, "Degradation of Adhesion in Cr-Au System," American Vacuum Society, Princeton, N.J., May 19, 1977.

[7] T. T. Hitch, "Reproducible Adhesion Test for Soldered Thick-Film Conductors," Final Report, Contract No. N00163-76-C-0287, prepared for Naval Avionics Facility, Indianapolis, to be published.

[8] D. M. Hoffman, "The Structure and Properties of Thin-Metal Films," *Proc. 28th Annual Symposium on Frequency Control*, 1974.

[9] Ibid, p. 86.

Doppler Servo Tracker With Frequency Synthesizer Feedback

M. I. Rozansky

RCA Missile and Surface Radar, Moorestown, N.J. 08057

Abstract—Linear servo theory indicates that a sampled-data type 2 feedback tracker can follow a ramp input precisely, with zero steady-state error at the sampling instants. It is shown in this paper that for the case of a doppler tracker employing a frequency synthesizer in the feedback, phase discontinuities exist that give rise to a steady-state error. A mathematical model that illustrates the source of the error is developed and analyzed for this implementation, and a technique is offered for extracting the uncorrupted data. The result confirms observations made on the GEOS-C satellite program.

Introduction

This paper discusses a particular implementation of the doppler frequency extractor circuit shown by the simplified block diagram in Fig. 1. The circuit is a sampled data servo characterized by a frequency synthesizer in the feedback path. Two integrations are performed on the frequency variable after discrimination, and for this reason the system is commonly labeled type 2. From the theory of linear sampled data control systems, we expect the steady-state frequency error at the sampling instants due to a constant velocity input (that is, the frequency ramp function input that arises from a range accelerating target) to be zero.[3] However, when the circuit was installed in the field for measurement data on the GEOS-C satellite,[1] it soon became apparent that just such an error did exist. When the tracker was locked onto the target, an error having the same profile as the target range acceleration was con-

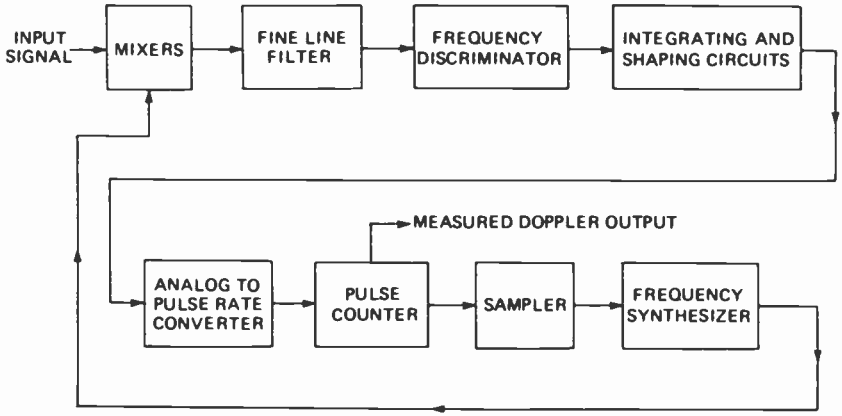


Fig. 1—Simplified block diagram of the C-band radar doppler tracking loop.

sistently observed. It will be shown that the key to the existence of this error lies in the generation of the feedback frequency by the synthesizer. The problem, therefore, is germane to any frequency or phase tracking loop with a similar implementation.

To derive the correct model for the performance of the tracker, it is necessary to introduce the phase variable in addition to frequency, and to include a differentiator and another integrator in order to convert between the two. The resultant circuit can no longer be properly classified as type two and, therefore, we will not be surprised at the existance of the error. The primary contributions of this paper are the tutorial development of the correct model of the tracker in terms of the Z-transform parameter, the derivation of expressions that agree excellently with observed results for the steady-state frequency error that occurs when tracking range accelerating and range jerking targets, and a scheme to avoid or minimize these errors.

Mathematical Model

In reducing the block diagram of Fig. 1 to a mathematical model, the pitfall to be avoided is the assumption that the loop may be modeled in frequency alone, irrespective of phase angle. The reason is that the circuit response reflects the more generalized concept of frequency as the derivative of phase angle. Consider, for example, a cw waveform that is constant in frequency but that undergoes a step discontinuity of δ in phase every T seconds,

$$f(t) = \cos(2\pi ft + n\delta),\ nT < t \leq (n+1)T \qquad [1]$$

for $n = 0,1,2,\ldots$. The circuit downstream from this signal will follow it

by smoothing out the discontinuity and "catching up" until the response is in accord with the current phase angle of the input. The average frequency of the response signal in radians per second during each interval is

$$\frac{\Delta\phi}{\Delta t} = \frac{2\pi fT + \delta}{T} = 2\pi f + \delta/T. \qquad [2]$$

If this same circuit were modeled in the frequency parameter only, the average frequency computed in any interval could never be any different from the constant instantaneous value, $2\pi f$. The servo loop in Fig. 1 must be modeled with this in mind. The integrating circuits that follow the frequency discriminator will respond to the average frequency during each pulse repetition interval (Eq. [2]) and the mathematics must correctly reflect this.

Let us investigate how the phase discontinuity arises in the tracker of Fig. 1. The ramp function of frequency versus time that is returned by a target with constant range acceleration is shown in Fig. 2. On the same plot, the stair-step frequency function versus time that is generated in the feedback path of the servo by the frequency synthesizer is shown. It is seen that by proper time alignment, the difference frequency (error) at the sampling instants can be made precisely zero. Instantaneous phase angles (the integral of frequency) corresponding to each of these frequency functions are plotted together in Fig. 3. In this paper, the constants of integration, which are not significant, are suppressed. The phase function due to the ramp is a quadratic, while that due to the stair-step is madeup of piecewise linear ramp segments. The difference between

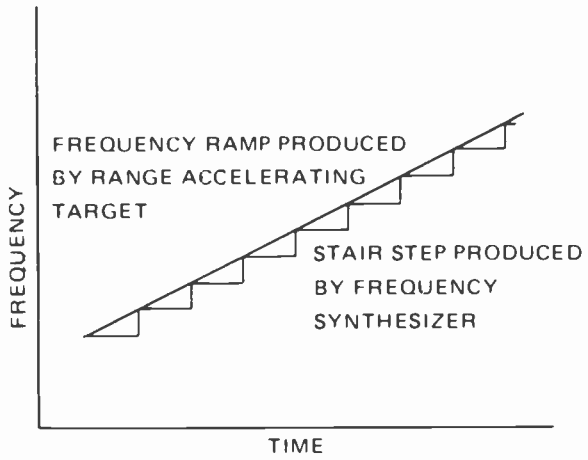


Fig. 2—Frequency functions.

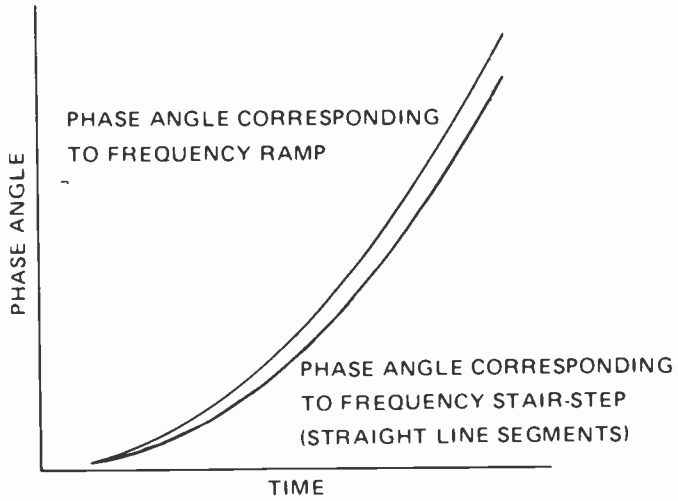


Fig. 3—Phase functions.

these produces a phase discontinuity at the sampling instants very similar to the term $n\delta$ in Eq. [1]. As mentioned, this phase discontinuity will translate into a frequency error. It is most important to appreciate that this error is not due to any limitation in the frequency synthesizer. That is, as suggested in Fig. 2, the error is not attributable to any quantization error in the setting of the frequency value, nor, as seen in Fig. 3, is the error due to any phase or waveform discontinuity in the synthesizer.

To properly account for the phenomenon described, it is necessary to model the frequency discriminator as a device that differentiates phase angle and then converts the resulting parameter to a dc signal. This further requires that instead of processing frequency, one must model the loop in terms of phase. Referring to Fig. 1, the error frequency out of the mixers is effectively sampled and held by the fine line filter. During this interval the phase angle, since it is the integral of frequency, increases linearly. This process is described at the loop input by simultaneously sampling frequency and phase and then, during the pulse repetition interval, integrating the former (i.e., converting it to phase), holding the latter, and combining the result. The feedback frequency is effectively sampled and held (zero order) by the synthesizer. Phase, then, is the integral of this frequency. The phase error is formed and input to the discriminator where, as indicated, it is differentiated and converted to an analog voltage representing frequency. Thereafter, the signal is integrated twice before it is sampled and fed back by the frequency synthesizer. Since only the steady-state performance is of in-

terest, the shaping circuits that are required for stability will be ignored. Also, the synthesizer actually "samples" the frequency a small interval δ prior to the arrival of the next radar pulse. This, however, is not a factor in this analysis and will also be ignored.

The S-plane model thus derived is shown in Fig. 4. The transfer function of the zero order hold is

$$H(S) = \frac{1}{S}(1 - e^{-TS}). \qquad [3]$$

In this form it is possible to convert to Z transforms and express the loop response directly. The result is

$$C(Z) = \frac{G_1G_3(Z)R_1(Z) + G_2G_3(Z)R_2(Z)}{1 + HG_3(Z)}. \qquad [4]$$

The frequency input is denoted by $R_1(Z)$ and the phase input by $R_2(Z)$. $G_1G_3(Z)$ denotes the Z transform of the product combination of

$$\left[\frac{1}{S^2}(1 - e^{-ST}) - \frac{T}{S}e^{-ST}\right] \cdot \frac{K_a}{S},$$

$G_2G_3(Z)$ denotes the Z transform of

$$\left[\frac{1}{S}(1 - e^{-ST})\right] \cdot \frac{K_a}{S},$$

and $HG_3(Z)$ the Z transform of

$$\left[\frac{1}{S^2}(1 - e^{-ST})\right] \cdot \frac{K_a}{S}. \qquad [5]$$

In each of these expressions, K_a is set equal to the product $K_dK_sK_iK$. These expressions are all found with the aid of the table of Z transforms given in Ref. [2] (p. 60) and the transform,

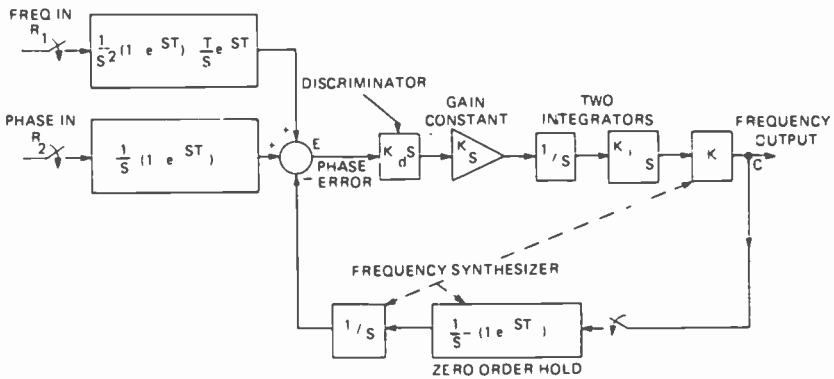


Fig. 4—S-Domain model of the doppler tracking loop.

$$1 - e^{-TS} \rightarrow 1 - Z^{-1} = \frac{Z-1}{Z}. \qquad [6]$$

The results are

$$G_1G_3(Z) = \frac{K_a[T^2(Z+1) - 2T^2]}{2(Z-1)^2}, \qquad [7]$$

$$G_2G_3(Z) = \frac{K_aT}{Z-1}, \qquad [8]$$

and

$$HG_3(Z) = \frac{T^2(Z+1)K_a}{2(Z-1)^2}. \qquad [9]$$

Linear Frequency Input

We are primarily interested in the steady-state frequency error, that is, the loop output denoted by C in Fig. 4 minus the input frequency. Eq. [4] gives $C(Z)$ as a function of the inputs $R_1(Z)$ and $R_2(Z)$. Consider first the case of an input frequency that is linearly increasing with time,

$$f = qt, \qquad [10]$$

where q is in Hz/sec. Input phase in cycles, being the integral of Eq. [10], is

$$\phi(t) = qt^2/2. \qquad [11]$$

Again resorting to the table of Z transforms in Ref. [2],

$$qt \rightarrow \frac{qTZ}{(Z-1)^2},$$

and

$$qt^2/2 \rightarrow \frac{qT^2Z(Z+1)}{2(Z-1)^3}. \qquad [12]$$

Using Eqs. (4), (7) through (9), and (12), the Z transform of the frequency error is

$$F(Z) = \frac{K_a \dfrac{T^2(Z+1) - 2T^2}{2(Z-1)^2} \dfrac{qTZ}{(Z-1)^2} + \dfrac{K_aT}{Z-1} \dfrac{qT^2Z(Z+1)}{2(Z-1)^3}}{1 + \dfrac{K_aT^2(Z+1)}{2(Z-1)^2}}$$

$$- \frac{qTZ}{(Z-1)^2}. \qquad [13]$$

Finally, the steady state value is available from the final value theorem, (p. 71 of Ref. [2]),

$$\lim_{t \to \infty} f(t) = \lim_{Z \to 1} \frac{Z-1}{Z} F(Z). \qquad [14]$$

After simplification, Eq. [14] reduces to

$$\lim_{t \to \infty} f(t) = \lim_{Z \to 1} \frac{qK_aT^3 - 2qT(Z-1)}{2(Z-1)^2 + K_aT^2(Z+1)} = \frac{qT}{2}. \qquad [15]$$

Eq. [15] indicates that, under the influence of an input frequency ramp (target having constant acceleration), the loop leads the input frequency by a constant amount $qT/2$ Hz. Since the input frequency is increasing by q Hz/sec, it follows that the loop response anticipates the input frequency by $T/2$ seconds.

The doppler frequency returned by a target having constant radial velocity v is*

$$f = \frac{2v}{c} f_t, \qquad [16]$$

where c is the speed of light and f_t is the transmitted frequency. Differentiating both sides of Eq [16] and letting $f = q$,

$$q = \dot{f} = \frac{2\dot{v}}{c} f_t, \qquad [17]$$

where the dots denote first derivative. We may determine the error in radial velocity corresponding to an error of

$$f = \frac{qT}{2} \text{ Hz}. \qquad [18]$$

Solving Eq. (16) for v and using Eqs. (17) and (18), we obtain

$$v = \dot{v}T/2. \qquad [19]$$

For an 87° elevation pass, $\dot{v}$ at the point of closest approach is approximately 50 meters/sec^2. Assuming a PRF of 160/sec,

$$v = 15.6 \text{ cm/sec} \qquad [20]$$

error in the measured target radial velocity.

One word of caution is required because of the modulo 2π nature of phase angle. The mathematical model used here assumes that the phase discontinuity cited in Fig. 3 translates into a positive frequency increment. That is, that the average frequency is always thereby increased.

* Eq. [16] is an approximation. Its use will introduce negligible error.

This is surely the case for small phase steps. However, as this discontinuity becomes greater than 180°, it is reasonable to expect that instead of "catching up", downstream circuits will retard, slowing down to accommodate the current phase angle. Average frequency will, in that case, be less than the instantaneous value. As the phase step approaches 360°, it is easy to see that no transient at all will be induced. Therefore, it is necessary to confirm that for the GEOS-C tracking loop this phase step is significantly smaller than 180°.

Again assume the input frequency is expressed as qt. The phase step seen by the discriminator will be the difference in the phase excursion for each pulse repetition interval (PRI) between the input and feedback signals. The phase excursion of the input is

$$\Delta\phi_i = \int_{nT}^{(n+1)T} qt\, dt$$
$$= qT^2n + (qT^2/2). \qquad [21]$$

Assume the case in which the feedback is following the input perfectly. Then at time nT the frequency out of the synthesizer is held at qnT, and the phase change during the next PRI is

$$\Delta\phi_f = (qnT)\, T = qnT^2. \qquad [22]$$

Therefore, the phase discontinuity at the next sample instant is

$$\Delta\phi = \Delta\phi_i - \Delta\phi_f = qT^2/2. \qquad [23]$$

Using Eq. [17], a typical maximum acceleration of 50 m/sec^2, a C-band transmission frequency of 6 Gigahertz, and a PRF of 160/sec,

$$\Delta\phi \approx 14°,$$

which is seen to satisfy the requirement and validate the model.

Quadratic Frequency Input

Consider now the case in which the target motion is characterized by constant jerk. Its doppler frequency will then appear as a quadratic,

$$f = bt^2, \qquad [24]$$

and the associated phase as a cubic,

$$\phi = bt^3/3. \qquad [25]$$

The corresponding Z transforms of Eqs. (24) and (25) are

$$bt^2 \rightarrow \frac{bT^2Z(Z+1)}{(Z-1)^3},$$

and

$$\frac{bt^3}{3} \rightarrow \frac{bT^3Z(Z^2+4Z+1)}{3(Z-1)^4} \qquad [26]$$

(notice that entry 1.04 in the Table of Z Transforms, p. 588 of Ref. [3] is incorrect). Inserting [26] in Eq. [4] to determine the steady-state loop error, as was done for the constant acceleration (linear doppler) case, leads to the conclusion that the frequency error becomes infinitely large under constant jerk. It is, however, possible to determine the rate at which the error increases. What is needed is the steady-state value of the first difference. But the first difference of any function expressed as a Z transform, $F(Z)$, is readily available as

$$(Z-1)\cdot F(Z). \qquad [27]$$

Then, inserting (26), (27), and Eq. [4] in Eq. [14], the final value of the first difference of the frequency error under constant target jerk is

$$\begin{aligned}
&\lim_{Z\to 1}\frac{Z-1}{Z}(Z-1)\\
&\times\frac{\dfrac{K_aT^2(Z-1)}{2(Z-1)^2}\dfrac{bT^2Z(Z+1)}{(Z-1)^3}+\dfrac{K_aT}{Z-1}\dfrac{bT^3Z(Z^2+4Z+1)}{3(Z-1)^4}}{1+\dfrac{K_aT^2(Z+1)}{2(Z-1)^2}}\\
&-\frac{bT^2Z(Z+1)}{(Z-1)^3}\\
&=\lim_{Z\to 1}\frac{(K_abT^4/3)2(Z+2)-6bT^2(Z+1)(Z-1)}{2(Z-1)^2+K_aT^2(Z+1)} \qquad [28]\\
&=bT^2.
\end{aligned}$$

Hence, the doppler tracking loop will lead the input frequency under constant jerk by an amount increasing at the rate of bT^2 Hertz per pulse repetition interval (PRI).

It was postulated that the true target doppler frequency equals bt^2. Equation (28) indicates that the frequency of the tracker, denoted by C in Fig. 2, increases as $bt^2 + bTt$ (hence a difference of bTt or bT^2 each PRI). At what time difference, X, shall we monitor the tracker to get the best estimate of the input frequency?

$$\begin{aligned}
&b(t-x)^2+bT(t-x)\stackrel{?}{=}bt^2\\
&bt^2+bt(T-2x)+b(x^2-Tx)\stackrel{?}{=}bt^2. \qquad [29]
\end{aligned}$$

Letting $x = T/2$ (that is, monitoring the loop frequency $T/2$ seconds prior to the arrival time of the radar return) gives the input frequency to within a constant error, $-bT^2/4$. This error is typically less than the acceleration

error (error due to a quadratic input) that would exist at the instant of the radar return if the loop were truly type two (Ref. [3], p. 269). The time shift $T/2$ is consistent with the value derived under the influence of a constant accelerating target producing the ramp input.

The error derived here may be re-interpreted in terms of target velocity. Starting from Eq. [16], we have

$$\dot{f} = \frac{2\dot{v}}{c} f_t;$$

hence

$$\Delta f = \frac{2\Delta v}{c} f_t, \qquad [30]$$

and

$$\ddot{f} = \frac{2\ddot{v}}{c} f_t. \qquad [31]$$

Differentiating Eq. [24] twice and substituting into Eq. [31] and setting Eq. [30] equal to bT^2 per PRI leads to

$$\Delta v = \frac{\ddot{v}T^2}{2}. \qquad [32]$$

For an 87° pass, maximum $\ddot{v}$ is approximately 0.3 meters/sec^3. For a PRI of 160, Eq. [32] gives the velocity error as equal to 0.0006 cm/sec per PRI. These results, of course, apply only so long as the acceleration induced by the jerk causes a phase discontinuity that is small compared with 180° as discussed previously.

Conclusion

Linear servo theory predicts that a type 2 sampled data servo can precisely follow a ramp input with zero steady-state error at the sampling instants. It has been shown that, in the case of a doppler tracker, if the feedback signal is developed by a frequency synthesizer, the type 2 designation is no longer preserved and a steady-state error will exist for a frequency ramp input that is induced by a range accelerating target. If the target has a jerk component, its doppler frequency will increase quadratically. In that case, the steady-state error will increase uniformly without bound.

The anomaly was seen to be due to step discontinuities in phase at the input to the frequency discriminator. This phenomenon would be entirely transparent if the tracking loop were modeled in frequency. To derive the correct response, it is necessary to model the loop in terms of

phase angle and to describe the discriminator as a device that converts the derivative of phase to a baseband signal.

It is shown that in tracking the GEOS-C satellite in an 87° pass, the expected doppler error will increase at a rate that peaks at a maximum value of 0.0006 cm/sec per PRI until the maximum error of about 15.6 cm/sec is reached. However, this error can be minimized by monitoring the servo loop doppler signal at $T/2$ seconds prior to the actual arrival time of the radar return, where T is the pulse repetition interval. The resultant doppler error will be less than that predicted for a legitimate type 2 servo loop under the stimulus of a quadratic input monitored at the radar return arrival time.

Acknowledgement

The author wishes to thank R. D. Mitchell for his encouragement and W. W. Weinstock and H. Urkowitz for their assistance in the preparation of this paper.

References:

[1] *Flight Test Data Analysis Program Radar Performance Parameter Data Final Report*, July 1970, Contract No. F04701-69-C-0131.

[2] Kuo, Benjamin C., *Analysis and Synthesis of Sampled Data Control Systems*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963.

[3] Tou, Julius T., *Digital And Sampled-Data Control Systems*, McGraw-Hill Book Co., Inc., New York, 1959.

An Optimized Grid Design for a Sun-Concentrator Solar Cell

A. R. Moore

RCA Laboratories, Princeton, N.J. 08540

Abstract—Two different collecting grid designs for a solar concentrator cell are considered: a simple linear grid pattern and an inverted square pattern. The grid dimensions of each pattern are optimized for maximum power output from the cell. The calculations show that the inverted square pattern is superior at all levels of concentration. Two methods of optimizing the square pattern are given, one of which allows the grid spacing to be a function of position across the cell face while the line thickness is kept constant. Detailed calculations are given for 100- and 300-sun concentrations.

1. Introduction

The design of the collecting-grid structure of a photovoltaic solar cell intended for use at high solar concentration levels is critical for maximum energy conversion efficiency. This is because the current density is so high that appreciable power is lost in the various parasitic resistances if the grid lines are too narrow or too widely spaced, while the grid shadowing of the sun is too large if the lines are wide and finely spaced. The problem clearly separates into two parts. The first is the choice of a grid pattern. There is no general mathematical method of predicting the best form, although Napoli et al.[1] showed that a cross-hatched pattern of fine grid lines will always be poorer than a straight simple line grid (called here the linear structure). Therefore the choice must be based on intuition and general understanding. The second part of the problem is to optimize the chosen pattern so that the spacing, width, and thick-

ness of the lines results in minimum power loss or maximum power output. For a concentrator cell, the cost of the metallization is not an important consideration. Hence cost optimization[2,3] need not be included. When optimization is complete, the optimized results for various patterns are compared and the best chosen. The purpose of this paper is to make that quantitative comparison between the simple linear grid (see Fig. 1a) and a more complicated pattern with inverted square symmetry (Fig. 1b).

The principles of grid optimization are well known.[1-3] There are three sources of parasitic resistive losses: spreading resistance in the surface sheet under the grid, line loss in the resistance of the grid line, and contact resistance loss at the metallization–semiconductor contact area. The other contribution to the power loss is the grid shadowing. It is assumed here that for small-area concentrator cells the collecting bus can always be located outside the illuminated area, in which case it can be made as wide/thick as is necessary to make its contribution to the losses negligible. One calculates the fractional power loss for each contribution, defined as the power lost in a given area divided by the available power in that area, then minimizes the total fractional power loss with respect to the independent variables. This results in a set of simultaneous equations whose solution gives the desired optimum geometry. It is necessary to include a constraint on the thickness of a grid line, stated as the ratio of width to thickness that is technologically practical. This ratio is commonly taken as four, but the formulas used allow any choice. The constraint avoids the mathematically correct but physically impossible solution consisting of an infinitely narrow line, infinitely thick but of finite conductance.

It is best to calculate the *power* loss $\int I^2 dR$ for each contribution[2,3] rather than the voltage drop times the total current in the area under consideration.[1] In the case of the spreading resistance and the line loss, the voltage drop method overestimates the power loss because the total current does not flow through the entire sheet or line. The difference will be reflected in the calculations to follow.

2. The Linear Grid

The structure of the linear grid is shown in Fig. 1a. It is the simplest grid structure and forms the basis of computations for more complicated patterns. We assume first that the front surface of the cell carrying the grid is uniformly illuminated by sunlight and second that the photogenerated current from the junction flows uniformly into the surface sheet and spreads laterally to be collected by the grid lines. It then flows through the grid lines to the collecting bus. The current flow pattern

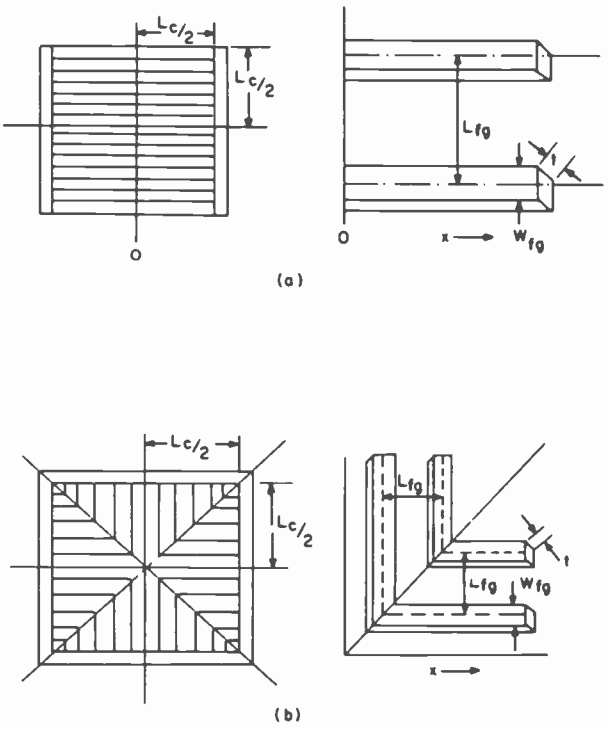


Fig. 1—(a) Linear grid configuration and (b) inverted square symmetry grid configuration.

between two adjacent grid lines is shown in Fig. 2. Two grid lines of half the cell length form the basic geometrical unit, the unit cell. The entire device is made up of repeated unit cells.

2.1 Spreading Resistance in the Surface Sheet

We begin at the symmetry line between the two grid lines. The current passing through the sheet and collected by the grid in a half-unit cell depends on the distance from the symmetry line as

$$I(y) = \frac{J_m y L_c}{2}, \qquad [1]$$

where J_m is the photocurrent density and L_c is the cell length. An element of resistance in the surface sheet is $dR = 2\rho_S dy/L_c$ where ρ_S is the sheet resistance. The power dissipated in that element is $dP = I^2(y)\, dR$. The total dissipated power in the half unit cell is

$$P_{dis} = (J_m^2\, L_c\, \rho_S/2) \int_0^{L_{fg}/2} y^2\, dy = \frac{J_m^2\, L_c\, \rho_S\, L_{fg}^3}{48}, \qquad [2]$$

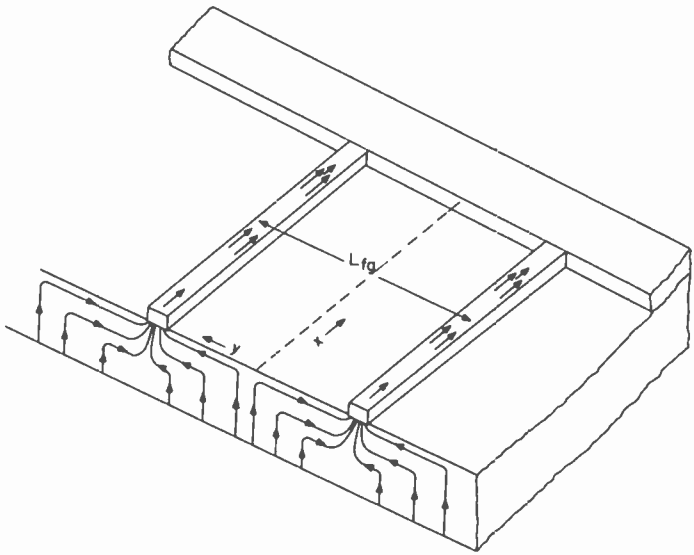


Fig. 2—Current flow pattern in the surface sheet and grid lines for the linear grid configuration.

The available power from a half unit cell is simply the current times the voltage, V_m,

$$P_{av1} = J_m \times \text{area of half unit cell} \times V_m = \frac{J_m V_m L_{fg} L_c}{4}. \quad [3]$$

The fractional power loss due to spreading resistance is therefore

$$F_{SR} = \frac{P_{dis}}{P_{avl}} = \frac{J_m \rho_S L_{fg}^2}{12 V_m} \quad [4]$$

2.2 Line Drop

The current in the grid line of width W_{fg} and thickness t for a full unit cell depends on the distance x along the line from the neutral symmetry position as

$$I(x) = J_m L_{fg} x. \quad [5]$$

An element of resistance in the grid line is $dR = (\rho_M dx)/(W_{fg} t)$ where ρ_M is the metal resistivity. The power dissipated in an element of the grid wire is $dP = I^2(x) dR$. The total dissipated power in a full width of the grid line, corresponding to a full unit cell, is

$$P_{dis} = \frac{J_m^2 L_{fg}^2 \rho_M}{W_{fg} t} \int_0^{L_c/2} x^2 dx = \frac{J_m^2 L_{fg}^2 \rho_M L_c^3}{24 W_{fg} t}. \quad [6]$$

The available power for the full unit cell is simply twice Eq. [3]. The fractional power loss due to line drop is therefore

$$F_{LD} = \frac{P_{dis}}{P_{avl}} = \frac{J_m \rho_M L_{fg} L_c^2}{12 V_m t W_{fg}}. \quad [7]$$

We now apply the constraint on the allowable width of the line mentioned earlier and contained in the definition $n = W_{fg}/t$. It is also convenient to express Eq. [7] in terms of the shadowing loss $S = W_{fg}/L_{fg}$. With these two substitutions the fractional line drop loss becomes

$$F_{LD} = \frac{n J_m \rho_M L_c^2}{12 V_m S^2 L_{fg}}. \quad [8]$$

2.3 Contact Resistance

Since the current is the same through each area element of the contact resistance, the voltage drop method is correct for this case. The voltage drop V_{Dc} across the contact resistance ρ_C is the current collected in one unit cell times the contact resistance of one full-width line in the unit cell. Thus

$$V_{DC} = \frac{J_m L_c L_{fg}}{2} \frac{2\rho_C}{L_c W_{fg}} = \frac{J_m \rho_C}{S}. \quad [9]$$

in terms of the shadowing loss S. The power dissipated in the contact resistance of one unit cell is

$$P_{dis} = J_m \frac{L_{fg} L_c}{2} V_{DC}. \quad [10]$$

The available power is again twice Eq. [3]. Thus the fractional contact loss is, using Fig. [9],

$$F_C = \frac{P_{dis}}{P_{avl}} = \frac{V_{DC}}{V_m} = \frac{J_m \rho_C}{S V_m}. \quad [11]$$

2.4 Shadowing Loss

The shadow fraction, as has already been mentioned, is simply the ratio of the line width to line spacing, $S = W_{fg}/L_{fg}$. This is also the fractional power loss. Thus

$$F_S = S = W_{fg}/L_{fg}. \quad [12]$$

2.5 Optimization

Eqs. [4], [8], [11] and [12] are the four components of power loss. The total fractional power loss, F_{SUM}, is then

$$F_{SUM} = F_{SR} + L_{LD} + F_C + F_S. \quad [13]$$

Setting the derivatives of Eq. [13] with respect to the independent variables L_{fg} and S equal to zero, we get two simultaneous equations in L_{fg} and S. When one is substituted into the other, we obtain

$$\frac{V_m}{J_m} = L_{fg}^3 \left[A + BL_{fg}^{1/2}\right] \quad [14]$$

and

$$S^2 = \frac{n\rho_M L_c^2}{2\rho_S L_{fg}^3}, \quad [15]$$

where $A = \dfrac{2\rho_C\rho_S}{n\rho_M L_c^2}$ and $B = \dfrac{2^{1/2}\,\rho_S{}^{3/2}}{3n^{1/2}\,\rho_M{}^{1/2}L_c}$.

If we multiply both ρ_S and ρ_M by the factor 3/2 and with $n = 4$, we get the solution of Napoli et al.[1] This reflects the over-estimate of F_{SR} and F_{LD} in the voltage drop method. Eq. [14] must be solved numerically, most easily by iteration. Table 1 gives the optimized grid dimensions and the various losses calculated from these equations for a 1-cm^2 cell at a concentration level of 100 suns (J_m = 3 ampere/cm^2). Both gold and silver metallization are considered for the "power" solution, i.e., the present calculations; for comparison purposes the "voltage drop" solution, i.e., the method of Ref. [1], is given, but only for gold metallization. The most important figure is the fraction of power remaining, 1-F_{SUM}. It is the deciding factor in any design.

Table 1—Optimized Grid Dimensions and Losses for 1-cm^2 Cell with Linear Grid Metallization at 100-Sun Concentration (L_c = 1 cm, J_m = 3 A/cm^2, V_m = 0.7 volts, $\rho_C = 10^{-4}$ ohm-cm^2, ρ_S = 100 ohms/square, n = 4, and $\rho_M = 2.5 \times 10^{-6}$ ohm-cm for Au and 1.6×10^{-6} ohm-cm for Ag)

Dimensions and Losses	Voltage-Drop Solution Au Metal	Power Solution Au Metal	Power Solution Ag Metal
L_{fg} (μm)	192	241	199
W_{fg} (μm)	16.1	15.3	12.7
t (μm)	4.0	3.8	3.2
F_S	0.0841	0.0714	0.0635
F_{LD}	0.0395	0.0327	0.0284
F_C	0.0051	0.0060	0.0067
F_{SR}	0.0197	0.0164	0.0142
F_{SUM}	0.148	0.127	0.113
1-F_{SUM}	0.852	0.873	0.887

Table 2—Optimized Grid Dimensions and Losses for 1-cm² Cell with Square Symmetry Grid Metallization (L_{fg} Constant) at 100-Sun Concentration (Parameters Are the Same as Table 1)

Dimension and Losses	Power Solution Au Metal	Power Solution Ag Metal
L_{fg} (μm)	192	178
W_{fg} (μm)	11.4	9.5
t (μm)	2.9	2.4
F_S	0.0596	0.0533
F_{LD}	0.0262	0.0226
F_C	0.0072	0.0080
F_{SR}	0.0131	0.0113
F_{SUM}	0.106	0.0953
$1\text{-}F_{SUM}$	0.894	0.9048

function of position, inasmuch as W_{fg} and t are tied to L_{fg} and S through Eq. [12] and the definition of n. But it is not practical to have a grid thickness that varies over the surface of the cell, because the grid thickness is built up by electrolytic plating.

A computational technique to eliminate the technologically impractical variable thickness but still retain the variable grid spacing is to eliminate L_{fg} in Eqs. [4] and [8] in favor of W_{fg} and t. The F_{SUM} is then

$$F_{SUM} = \frac{J_m \rho_S W_{fg}^2}{12\, V_m S^2} + \frac{J_m \rho_M L_c^2}{12\, V_m t S} + \frac{J_m \rho_C}{S V_m} + S. \qquad [21]$$

We then optimize with respect to the *single* variable S, regarding W_{fg} and, hence, t as constants and arbitrarily chosen but still related by $n = W_{fg}/t$. The optimized value of S is then given by the solution of

$$S^2 = (K_2/S) + K_1 \qquad [22]$$

where

$$K_1 = \frac{J_M}{V_m}\left[\frac{\rho_M L_c}{12\, t} + \rho_C\right] \text{ and } K_2 = \frac{J_m \rho_S W_{fg}^2}{6\, V_m}.$$

When S is known, F_{SUM} may be computed. We now divide the quadrant up into a number of "L" shaped segments following the square symmetry of the grid lines. Regarding the mean length of the line in any segment as L_c in Eqs. [9] and [10], we compute a weighted average of F_{SUM} over the quadrant, the weight factor for each segment taken as its fractional area. This weight factor assumes uniform illumination.

The procedure is then to take successive values of W_{fg} (and hence t), compute the weighted average of $1 - F_{SUM}$ and select that value which makes $\overline{1 - F_{SUM}}$ a maximum. The design will then automatically be optimized for the two variables S and L_{fg} with W_{fg} and t constant over

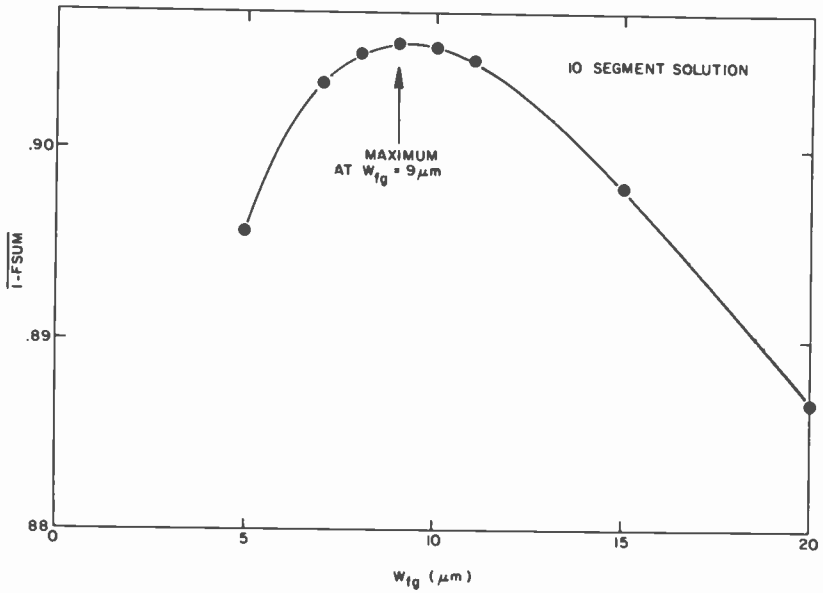


Fig. 4—Optimization of the weighted average of $1-F_{SUM}$ by variation of grid line width. Best choice of W_{fg} is 9 μm.

the cell. To find the variation of L_{fg} over the quadrant, $L_{fg}(x)$, we simply use Eq. [22] segment by segment, finding $\overline{S(x)}$ and hence $L_{fg}(x)$.

Fig. 4 shows the computed value of $\overline{1-F_{SUM}}$ as a function of W_{fg} for a cell under 100-sun concentration with the same dimensions and physical constants as used for the prior calculations given in Tables 1 and 2. Silver metallization has been assumed. The curve shown is for the quadrant divided into ten segments. A few trials have shown that this number provides adequate accuracy. Table 3 gives a brief summary of the grid dimensions and losses for the square symmetry grid optimized

Table 3—Optimized Grid Dimensions and Losses (10-Segment Solution) for Square-Symmetry Grid Metallization (L_{fg} Variable) with 100-Sun Concentration (Parameters Are the Same as in Tables 1 and 2 for Ag Metallization)

	Power Solution Ag Metal
L_{fg} (μm)	func. of x (See Fig. 3)
W_{fg} (μm)	9
t (μm)	2.25
$\overline{F_S}$	0.0526
$\overline{F_{LD}+F_C+F_{SR}}$	0.0419
$\overline{F_{SUM}}$	0.0945
$\overline{1-F_{SUM}}$	0.9055

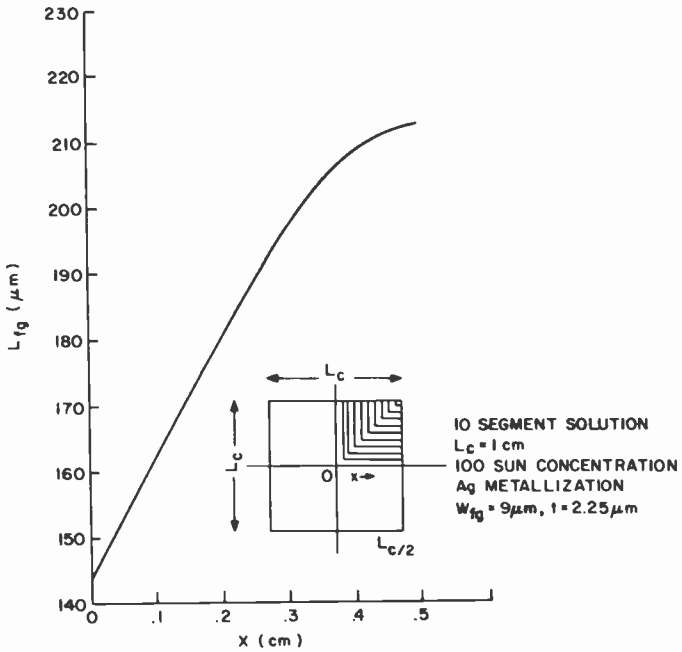


Fig. 5—Line spacing as a function of position, optimized square symmetry grid.

in this way. Fig. 5 shows the calculated optimum variation of grid spacing with x.

4. Discussion

A comparison of Tables 1, 2, and 3 shows primarily that the square symmetry grid is definitely better than the linear grid at 100-sun concentration. Indeed, this will be true at all concentration levels, the higher the level the more favorable to the square pattern. Comparing the worst case with the best, i.e., the voltage drop method and gold metallization in the linear pattern to the power loss method and silver metallization in the square symmetry pattern, shows an improvement of 6.3%. That is, a cell that had 16% conversion efficiency at 100 sun when metallized according to the former scheme would show 17% efficiency when metallized according to the latter. Perhaps a fairer comparison is linear versus square, both with silver metallization calculated by the power method. Here the advantage is 2.1% at 100 suns. At 500 suns and a cell size of 0.4 × 0.4 cm (data not shown), the gain would be 2.4%.

Comparison of the two different methods of optimizing the square symmetry pattern, constant line spacing versus variable line spacing

(Tables 1 and 3), does not afford much basis for choice. Both methods yield about the same value of $1 - F_{SUM}$ with a truly remarkable shadow fraction of only slightly over 5%. The weighted-average variable-line-spacing method has a potential advantage if we taken into account nonuniform illumination across the cell face. Then, it is possible to adjust the weight factors in the computer program to reflect any reasonable illumination (i.e., current density) variation. For example, much of the sunlight might be concentrated near the center of the image due to lens aberrations or defocusing. The weighted average method would optimally adjust the line spacing to be closer together where needed such that the overall power output is a maximum. Square symmetry in the image as well as in the metallization pattern would have to be assumed, however. Circular symmetry is also possible.

Table 4 shows a comparison of linear and square grid designs based on a cell that has actually been built and tested in fairly large quantity.[4] It is the 300-sun cell used in the RCA 300-watt experimental concentrator system. These back-surface field cells have demonstrated an efficiency of 15.5% at approximately the 300-sun concentration level. The first column of Table 4 shows the optimum linear grid and the respective losses calculated according to the voltage drop method. The actual cells closely match these calculations.[4] The second column gives the optimum square symmetry grid dimensions and losses calculated by the power method for the same gold metallization resistivity assuming constant grid spacing over the cell. The improvement in $1 - F_{SUM}$ amounts to 5.9%. Thus a cell that had 15.5% conversion efficiency would improve to 16.4% with the optimized gold grid design. If silver metallization were used with the square design, the efficiency would increase to 16.6% (column three).

Table 4—Comparison of Linear and Square-Symmetry (L_{fg} Constant) Metallizations with 300-Sun Concentration (L_c = 0.564 cm (0.25 cm² Area Illuminated, Square Equivalent), J_m = 10 A/cm², V_m = 0.7 Volts, $\rho_C = 10^{-4}$ ohm-cm², ρ_S = 100 ohms/square, n = 4, and $\rho_m = 2.5 \times 10^{-6}$ ohm-cm for Au and 1.6×10^{-6} for Ag)

Dimensions and Losses	Linear Grid Voltage-Drop Solution (Au Metal)	Square Grid Power Solution (Au Metal)	Square Grid Power Solution (Ag Metal)
L_{fg} (μm)	113	111	102
W_{fg} (μm)	11.9	8.5	7.1
t (μm)	3	2.1	1.8
F_S	0.105	0.0768	0.0700
F_{LD}	0.0456	0.0291	0.0246
F_C	0.0136	0.0186	0.0205
F_{SR}	0.0228	0.0145	0.0123
F_{SUM}	0.187	0.139	0.127
$1\text{-}F_{SUM}$	0.813	0.861	0.873

The advantage of the square symmetry grid and the power method is clear and, since the grid is made by photolithography and electroplating, no increase in cost is involved.

Acknowledgment

I would like to thank L. Napoli and G. A. Swartz for useful discussions. The original idea for the inverted square symmetry grid is due to L. Napoli.

References:

[1] L. S. Napoli, G. A. Swartz, S. G. Liu, N. Klein, D. Fairbanks, and D. Tamutus, "High-Level Concentration of Sunlight on Silicon Solar Cells," *RCA Review*, **38**, p. 76, March 1977.

[2] D. Redfield, "Procedure for Minimizing the Cost Per Watt of Photovoltaic Systems," *RCA Review*, **38**, p. 463, Dec. 1977.

[3] A. R. Moore, "Optimum Shape of the Bus Bar on Solar Cells of Arbitrary Shape," *RCA Review*, **38**, p. 486, Dec. 1977.

[4] G. A. Swartz, L. S. Napoli, and N. Klein, "Silicon Solar Cells for Use at High Solar Concentrations," Extended Abstracts, *Proc. Int. Electron. Dev. Conf.*, p. 226, Wash., D.C., 1977.

Optimum Grid Design for a Nonuniformly Illuminated Sun-Concentrator Solar Cell

A. R. Moore

RCA Laboratories, Princeton, N.J. 08540

Abstract—When a sun-concentrator type solar cell is operating in a practical optical system, the illumination is not uniform across the active region of the cell. In order to properly optimize the dimensions of the current-collecting grid for minimum resistive and shadow power loss, the local variation in photo-current must be taken into account. For the efficient inverted square-grid pattern now being used on concentrator cells, a method is presented for performing the optimization when the light distribution has square symmetry but is not otherwise restricted. The method is an adaptation of the optimization procedure for the simpler linear grid. It allows the grid wire spacing to be varied from the cell center to the bus edge, thus adapting it to the local current density, while the grid thickness and line width are kept constant. Detailed results are presented for a square- and a trapezoidal-shaped distribution. The value of tailoring the grid to the light distribution is demonstrated.

1. Introduction

In an accompanying paper[1] two different collecting grid designs for a solar concentrator photovoltaic cell were considered, a simple linear grid pattern and a more complicated inverted square pattern (see Figs. 1a and 1b). The grid dimensions of each pattern were optimized for maximum power output from the cell. The calculations showed that the inverted square pattern was superior. A method of optimizing the square pattern was given that allowed the grid spacing, L_{fg}, to be a function of position across the cell face while the grid line thickness t remained constant. All the calculations in that paper, however, were made under

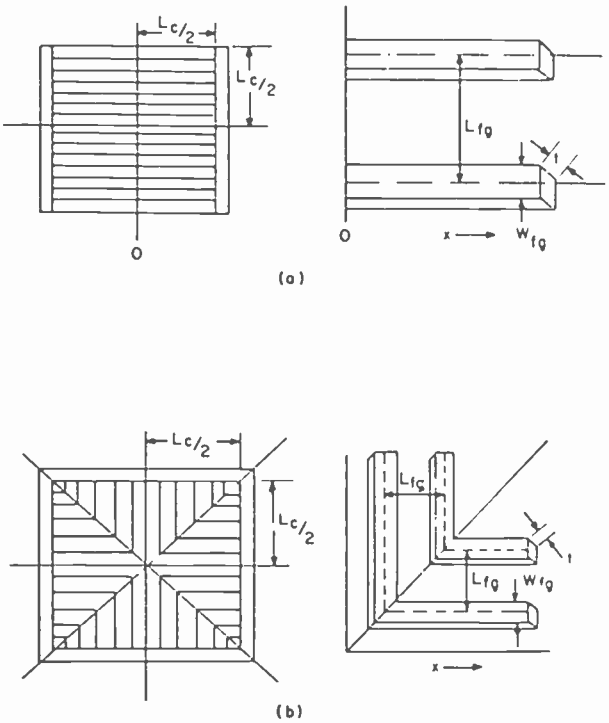


Fig. 1—Linear (a) and inverted square (b) grid patterns. The current collecting bus occupies all four sides of the cell in (b) and only two sides in (a).

the assumption that the illumination was uniform across the active area of the cell. In this paper, we show how those results can be modified to take into account nonuniform, but symmetric illumination.

In a solar cell designed for one sun (a flat plate array), it can be assumed that the solar illumination is uniform across the cell face. In a concentrator system there are a number of reasons why this is not likely to be the case. The optical properties of the lens will clearly be important in determining the actual illumination distribution. Another factor is tracking error. Also, in multiple lens arrays, lens-to-cell alignment errors must be considered. For all these reasons, it is normally desirable to design the system so that the light spot is significantly smaller than the active area. The result is that the illumination is no longer uniform from one edge of the cell to the other. In a small cell with the collecting bus along the outer edge, as in the inverted square design discussed above, concentrating the light near the center with a peripheral region having little or no illumination must inevitably lead to increased grid losses. The problem is how to design the grid so that these losses are minimized.

The method of optimizing the inverted square grid assuming uniform illumination, given in Ref. [1], is an adaptation of the linear grid formulation. We briefly repeat that method here using the same definitions of symbols so that the required modifications for nonuniform illumination can be understood.

The total fractional power loss (F_{SUM}) for a uniformly illuminated linear grid structure consists of the sum of the fractional losses (assumed small) due to spreading resistance (F_{SR}), line drop (F_{LD}), contact resistance (F_C), and shadowing (F_S):

$$F_{SR} = J_m \rho_S L_{fg}^2/(12\ V_m) \quad [1]$$

$$F_{LD} = n J_m \rho_M L_c^2/(12\ V_m S^2 L_{fg}) \quad [2]$$

$$F_C = J_m \rho_C/(V_m S) \quad [3]$$

$$F_S = S = W_{fg}/L_{fg} \quad [4]$$

where

L_c = active device dimension, assumed square and not including the collecting bus.
W_{fg} = width of the fine grid line.
L_{fg} = spacing between fine grid lines, center to center.
n = W_{fg}/t, where t is the grid thickness.
ρ_S = sheet resistance of surface sheet under the grid (ohm/square).
ρ_M = resistivity of the metallization material (ohm-cm).
ρ_C = contact resistance (ohm-cm^2).
J_m = current density from the cell at the maximum power point and at the stated illumination level.
V_m = cell voltage at the maximum power point and at the stated illumination level.
S = shadow fraction.

Eliminating L_{fg} in Eqs. [1] and [2] in favor of W_{fg} and t using Eq. [4] gives

$$F_{SUM} = \frac{J_m \rho_S W_{fg}^2}{12 V_m S^2} + \frac{J_m \rho_M L_c^2}{12 V_m t S} + \frac{J_m \rho_C}{S V_m} + S. \quad [5]$$

We then optimize with respect to the single variable S, regarding W_{fg} and t as constants (related by $n = W_{fg}/t$). Thus, after setting $\partial F_{SUM}/\partial S = 0$, we find S as the solution of the equation

$$S^2 = \frac{K_2}{S} + K_1 \quad [6]$$

where $K_1 = \frac{J_m}{V_m}\left(\frac{\rho_M L_c^2}{12t} + \rho_C\right)$ and $K_2 = \frac{J_m \rho_S W_{fg}^2}{6 V_m}$.

For purposes of calculation, a quadrant of the cell area is divide into equal-width "L" shaped segments following the square symmetry of the grid lines (see Fig. 3). The number of such segments is chosen so that each will contain several fine grid lines. Ten segments are convenient, giving adequate accuracy in the calculation. Eqs. [6] and then [5] are applied to each of these segments successively. This means substituting λ_j, the mean line length in the jth segment, for L_c in Eqs. [6] and [5]. The result is a set of values F_{SUM_j} where j runs from 1 to 10. We now compute a weighted average of F_{SUM} over the quadrant, the weight factor for each segment taken as its fractional area, i.e.,

$$\overline{F_{SUM}} = \sum_j F_{SUM_j} \frac{A_j}{A} = \sum_j F_{SUM_j} \frac{A_j}{L_c^2/4}. \qquad [7]$$

The final step is to take successive values of W_{fg} (and hence t), compute the weighted average of $1\text{-}F_{SUM}$ and select that value of W_{fg} that makes $\overline{1\text{-}F_{SUM}}$ a maximum. The design will then be optimized for the two variables S and L_{fg}, with W_{fg} and t constant over the cell.

2. Application of Linear Case To Nonuniform Illumination

To apply the foregoing method to a nonuniformly illuminated cell, it is necessary to rederive the four basic fractional power losses (Eqs. [1], [2], [3], [4]), taking into account the variation of the photogenerated current density along the grid lines. Since these equations are derived from the linear grid formulation, we neglect at this stage the variation in the photogeneration rate normal to the grid lines. This is permissible because the grid lines are very close together, so that the current density varies little between any two parallel lines. The perpendicular variation across the cell face is taken into account later. In any case, symmetry of current density is assumed around the 45° line in the quadrant, i.e., $J(x,y) = J(y,x)$.

2.1 Spreading Resistance in the Surface Sheet

Fig. 2a shows two adjacent grid lines of half the cell length, a unit cell. The photogenerated current density J coming into the surface sheet is taken as $J(x)$. The current to a length dx of a grid line from one half the unit cell depends on the position y normal to the grid line as

$$I(x,y) = yJ(x)dx. \qquad [8]$$

We now define the definite integrals

$$I_1 \equiv \int_0^{L_c/2} J(x)dx \qquad [9]$$

$$I_2 \equiv \int_0^{L_c/2} J^2(x)dx. \qquad [10]$$

An element of resistance in the surface sheet of width dx is $dR = \rho_S dy/dx$. The power dissipated in the element of resistance is $dP = I^2(x,y)dR$. The total dissipated power for the half unit cell is

$$P_{dis} = \rho_S \int_0^{L_{fg}/2} y^2 \int_0^{L_c/2} J^2(x)dxdy = \frac{I_2 \rho_S L_{fg}^3}{3 \cdot 2^3}. \qquad [11]$$

The available power is the cell voltage times the total current from the half unit cell

$$P_{avl} = \frac{V_m L_{fg}}{2} \int_0^{L_c/2} J(x)dx = \frac{V_m L_{fg} I_1}{2}. \qquad [12]$$

The fractional power loss due to spreading resistance is, therefore,

$$F_{SR} = \frac{P_{dis}}{P_{avl}} = \frac{\rho_S L_{fg}^2 I_2}{12 V_m I_1}. \qquad [13]$$

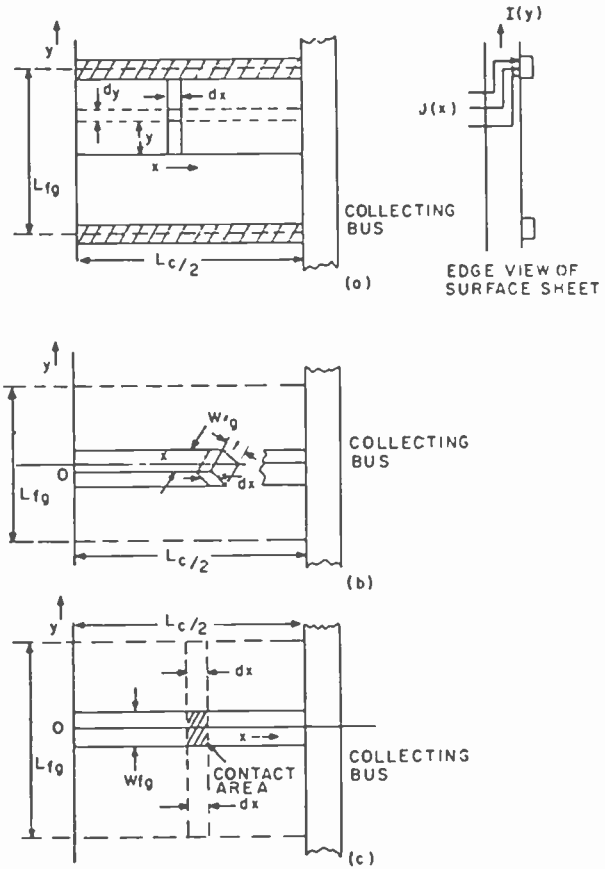


Fig. 2—Detail of grid wire collection needed for (a) spreading resistance, (b) line drop, and (c) contact resistance calculations.

2.2 Line Drop

Referring to Fig. 2b, the current in the grid wire line for a full unit cell depends on the distance x along the line from the neutral symmetry position 0 as

$$I(x) = \int_0^x L_{fg} J(x')dx' = L_{fg} I_1(x), \qquad [14]$$

where we define the cumulative integral, a function of x,

$$I_1(x) \equiv \int_0^x J(x')dx'. \qquad [15]$$

An element of resistance in the grid line of resistivity ρ_M is $dR = \rho_M dx/(W_{fg}t)$. The power dissipated in the element of resistance in the grid wire is $dP = I^2(x)dR$. The total dissipated power

$$P_{dis} = \frac{L_{fg}^2 \rho_M}{W_{fg}t} \int_0^{L_c/2} I_1^2(x)dx = \frac{L_{fg}^2 \rho_M I_3}{W_{fg}t}, \qquad [16]$$

where the definite integral

$$I_3 \equiv \int_0^{L_c/2} I_1^2(x)dx. \qquad [17]$$

The available power from the full unit cell is

$$P_{avl} = V_m L_{fg} \int_0^{L_c/2} J(x)dx = V_m L_{fg} I_1. \qquad [18]$$

The fractional power loss due to line drop is therefore

$$F_{LD} = \frac{P_{dis}}{P_{avl}} = \frac{\rho_M L_{fg}}{W_{fg}t V_m} \frac{I_3}{I_1} \qquad [19]$$

2.3 Contact Resistance

To calculate the contact resistance contribution, we can use the voltage-drop method, because the current that flows through any given contact area contributes to power loss only at that area.

Referring to Fig. 2c, the voltage drop due to the contact resistance in one strip dx of a full unit cell is the current in one strip times the contact resistance of that strip when connected to the grid wire:

$$V_{DC} = L_{fg} J(x)dx \frac{\rho_C}{W_{fg}dx} = \frac{L_{fg} J(x) \rho_C}{W_{fg}}.$$

An element of dissipated power from the current in the strip dx is the voltage drop times the current in that strip:

$$dP = V_{DC} L_{fg} J(x) dx = \frac{L_{fg}^2 \rho_C J^2(x) dx}{W_{fg}}.$$

The total dissipated power

$$P_{dis} = \frac{L_{fg}^2 \rho_C}{W_{fg}} \int_0^{L_c/2} J^2(x) dx = \frac{L_{fg}^2 \rho_C}{W_{fg}} I_2. \quad [20]$$

The available power is already given in Eq. [18]. The fractional power loss due to contact resistance is then

$$F_C = \frac{P_{dis}}{P_{avl}} = \frac{L_{fg} \rho_C}{W_{fg} V_m} \frac{I_2}{I_1}. \quad [21]$$

2.4 Shadowing

The photogenerated current distribution along a grid line pair has no influence on the shadow fraction which remains, as in Eq. [4], $F_S = W_{fg}/L_{fg}$.

2.5 A Simple Example: $J(x) = J_m$

In Sections 2.1–2.4 above we have rederived the fractional loss equations for the linear grid structure on the assumption that the photogenerated current density varies as $J(x)$ along a grid line. The results are expressed in terms of three definite integrals over the distribution, I_1, I_2, and I_3. To get a feel for how this works out, consider the simple case in which $J(x)$ is a constant, J_m. This is the uniform illumination case, the fractional losses for which were given at the start as Eqs. [1], [2], [3], and [4]. In the new formulation, with $J(x) = J_m$,
from Eq. [15]

$$I_1(x) = \int_0^x J_m dx' = J_m x;$$

from Eq. [9]

$$I_1 = \int_0^{L_c/2} J_m dx = J_m L_c/2;$$

from Eq. [10]

$$I_2 = \int_0^{L_c/2} J_m^2 dx = J_m^2 L_c/2;$$

from Eq. [17]

$$I_3 = \int_0^{L_c/2} I_1^2(x) dx = \int_0^{L_c/2} J_m^2 x^2 dx = \frac{J_m^2 L_c^3}{3 \cdot 2^3}.$$

When these values of the integrals I_1, I_2, and I_3 are used in the new formulas for fractional losses (Eqs. [13], [19], and [21]), the original fractional loss formulas for uniform illumination are obtained. As an example, calculate F_C for $J(x) = J_m$. From Eq. [21],

$$F_C = \frac{L_{fg}\rho_C}{W_{fg}V_m}\left(\frac{I_2}{I_1}\right) = \frac{L_{fg}\rho_C}{W_{fg}V_m}J_m.$$

But $S = W_{fg}/L_{fg}$, giving $F_C = J_m\rho_C/(V_mS)$, which agrees with Eq. [3]. The point is that I_1, I_2, and I_3 are constants, completely determined once $J(x)$ and L_c are given. They do not interfere with the optimization with respect to either S or W_{fg} because they do not contain them.

3. Optimization With Respect To *S*

Following the procedure used before, we eliminate L_{fg} in favor of W_{fg} and t in Eqs. [12], [18], and [21] and formulate the F_{SUM}:

$$F_{SUM} = \frac{I_2\rho_S W_{fg}^2}{12I_1V_mS^2} + \frac{I_3\rho_M n}{I_1W_{fg}V_mS} + \frac{I_2\rho_C}{I_1V_mS} + S. \qquad [22]$$

We then optimize with respect to the single variable S by taking $\partial F_{SUM}/\partial S|_{W_{fg}} = 0$. This leads to the optimum value of S as the solution of

$$S^2 = \frac{K_4}{S} + K_3 \qquad [23]$$

where $K_3 = \left(\frac{I_2}{I_1V_m}\right)\left(\frac{I_3\rho_M}{I_2t} + \rho_C\right)$ and $K_4 = \frac{I_2\rho_S W_{fg}^2}{6V_mI_1}$.

Note the similarity of this equation to Eq. [6].

4. Averaging Over The Quadrant

As before, the cell quadrant is divided into ten segments following the square symmetry of the grid lines, as shown in Fig. 3a. Fig. 3b shows the jth segment, enlarged for clarity. Eqs. [23] and [22] are then applied to each of the segments successively. The integrals I_1, I_2, and I_3 must be evaluated for each segment.
For the jth segment

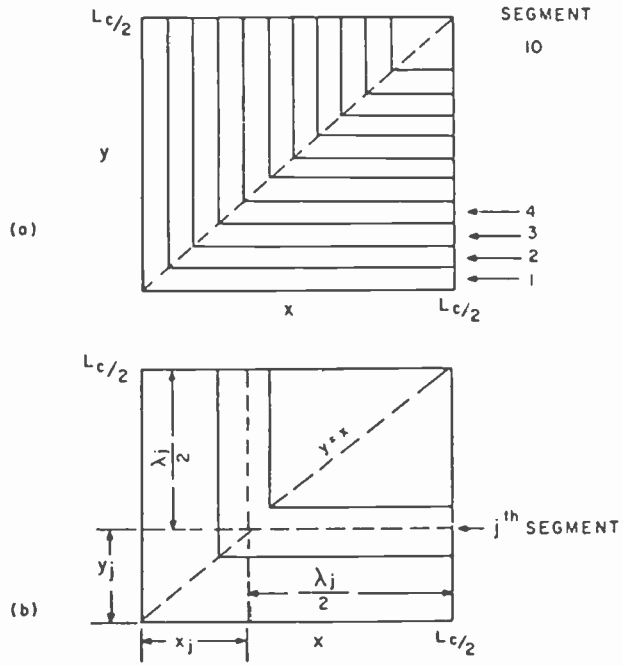


Fig. 3—(a) Division of cell quadrant into ten segments and (b) detail of the *j*th segment.

$$
\begin{aligned}
I_{1j}(x) &= \int_{x_j}^{x} J(x,y_j)dx \\
I_{1j} &= \int_{x_j}^{L_c/2} J(x,y_j)dx \\
I_{2j} &= \int_{x_j}^{L_c/2} J^2(x,y_j)dx \\
I_{3j} &= \int_{x_j}^{L_c/2} I_{1j}^2(x)dx
\end{aligned}
\qquad [24]
$$

Notice that the function J now depends on y through $J(x,y_j)$. Thus the variation of J in the y direction is taken into account as we move from segment to segment. Also, the lower limit of integration on x is x_j, since the integrals run from the 45° symmetry line outward to the cell edge $L_c/2$. The result of this part of the calculation is a set of values of the integrals that are used to evaluate the set of F_{SUM_j}. Again following the original method, we compute a weighted average of F_{SUM} over the quadrant, but the weight factor for each segment is now taken as the fractional power contributed by that segment. Since the voltage is the same for all segments and for the entire cell, we actually use the fractional current:

$$\overline{F_{SUM}} = \frac{\sum_j F_{SUM_j} I_{1j}}{\sum_j I_{1j}}. \qquad [25]$$

For uniform illumination, $J(x,y) = J_m$, Eq. [25] reduces to Eq. [7]; i.e., in this special case, the fractional power equals the fractional area.

The weighted average of $1\text{-}F_{SUM}$ is optimized with respect to W_{fg} by computing for a range of W_{fg} and finding the value of W_{fg} that maximizes $\overline{1\text{-}F_{SUM}}$. The design is completed by applying Eq. [23] with the optimum W_{fg} to each segment, determining $S(x_j)$ and hence $L_{fg}(x_j)$, the variation of grid spacing over the quadrant.

5. Examples

As examples, we have made complete calculations for the two different current distributions shown in Fig. 4. The first, a square distribution, was chosen for its simplicity. The second distribution (trapezoidal) was chosen because it approximates the actual light distribution found on a concentrator system with square lenses operating somewhat out of focus. In both cases the integrals of Eq. [24] can be performed analytically. This is not a requirement of the method. however. For complicated distributions more accurately representative of practical cases, the integrals can be performed numerically. The only requirement of the method is that the distribution be symmetrical about the lines $y = \pm x$, $y = 0$, and $x = 0$. The symmetry requirement about the X and Y axes could be relaxed if one were willing to make separate computations for each quadrant. This type of approach is needed when considering off-center illumination, but will not be pursued here.

Table I shows the several cases calculated. They have in common a normalization adjustment that makes the total photogenerated current

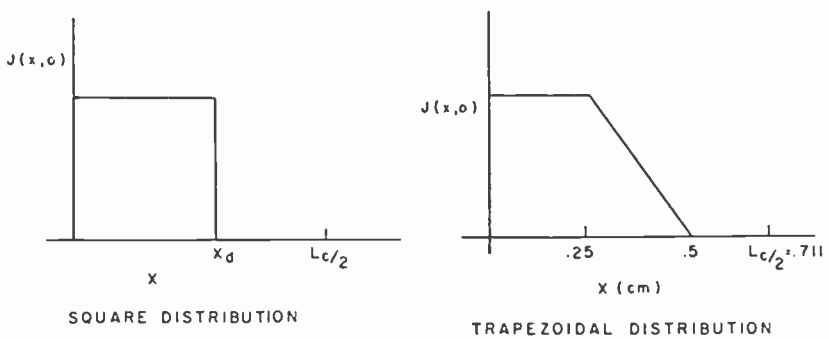


Fig. 4—Light intensity (current density) distribution for square and trapezoidal distributions.

Table 1—Calculated Results for Several Different Types of Grid (In All Cases, Total Current = 5 A, $\rho_C = 10^{-4}$, $\rho_S = 100$, $\rho_M = 1.6 \times 10^{-6}$, $n = 4$, and $V_m = 0.7$)

	Distribution					
	Uniform		Trapezoidal	Square		Equivalent Square
Case	A_1	A_2	B	D_1	D_2	D_3
L_c (cm)	1.00	1.42	1.42	1.42	1.42	1.42
L_c^2 (cm^2)	1.00	2.00	2.00	2.00	2.00	2.00
x_d	0.5	0.711		0.5	0.25	0.381
$J(0,0)$ (A/cm^2)	5	2.5	8.6	5	20	8.6
Sun (0,0)	166	83	287	166	664	287
$\overline{1\text{-}F_{SUM}}$	0.882	0.906	0.824	0.849	0.725	0.807
F_{SUM}	0.118	0.103	0.176	0.151	0.275	0.193
S	0.0662	0.0583	0.0994	0.0850	0.144	0.106
Opt. W_{fg} (μm)	10	12	16.5	15	17	16

equal to 5 amperes in each case, as well as common values of ρ_C, ρ_S, ρ_M, n, and V_m. The comparison is therefore on the basis of cell size and light distribution. Uniform distribution is regarded as the limiting case of a square distribution in which x_d equals $L_c/2$. The highest value of $\overline{1\text{-}F_{SUM}}$ (minimum losses) is obtained when the light is uniformly distributed over the largest area, keeping the current density at the center as low as possible (case A2 in Table 1). Conversely, the lowest value of $\overline{1\text{-}F_{SUM}}$ (maximum losses) is found when the light is most strongly concentrated near the center (case D2). A simplified approach to the trapezoidal distribution is obtained by defining an "equivalent square" (case D3). An "equivalent square" is a square distribution whose dimensions are determined by the requirements that the total current and the current density at the origin be the same as the corresponding trapezoidal distribution (case B). Since the results for these cases are similar, the "equivalent square" can often be substituted for the actual distribution, with considerable saving in computation time.

Fig. 5 shows the computed optimum variation of L_{fg} with x for several of these cases. For all cases in which the light is confined to an area less than the cell area, L_{fg} rises at the illumination edge, abruptly for the square distribution and more gradually for the trapezoidal. This is an expression of the principle that where there is no light, no grid is needed. However, since a purpose of leaving a dark band around the edge of the illuminated region is to allow for tracking error, it would be a mistake to follow the mathematical solution slavishly. In practice, there would at times be light in the outer region. In a practical design it would be desirable to terminate the $L_{fg}(x)$ curve somewhere near the illumination edge and simply extend it horizontally to the bus edge. This would do no harm as long as the distribution is centered (on target) and would result in minimum degradation when it is not.

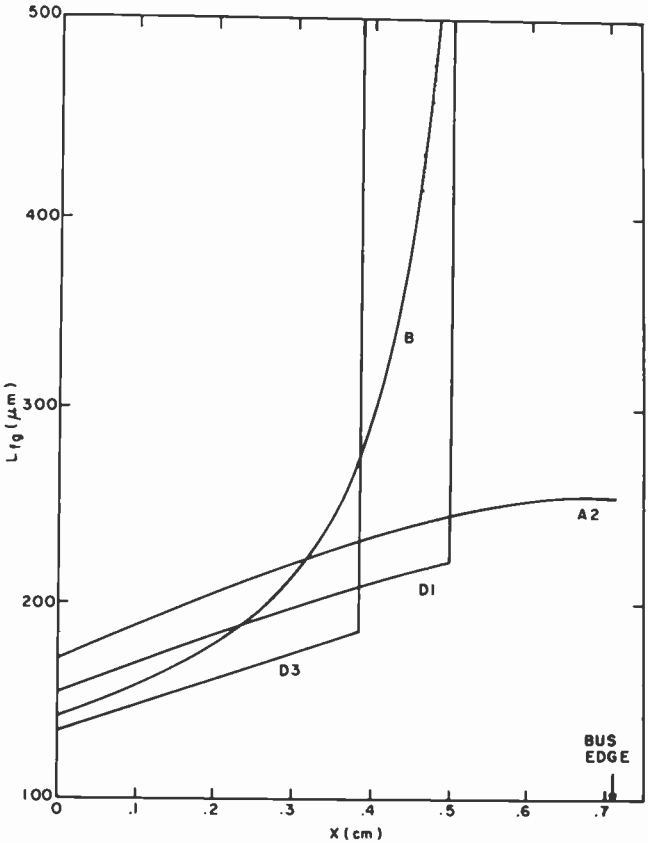


Fig. 5—Optimized grid spacing as a function of x, the distance from the cell center, for several of the cases given in Table I.

The trapezoidal distribution (B) and its "equivalent square" (D3) have similar $L_{fg}(x)$ curves. After the off-center modification suggested above they would be more closely similar. This is the basis of the simplification already mentioned.

Finally, in Fig. 6 we show the value of optimizing the grid design for the expected light distribution. The figure postulates two series of square light distributions. In both, the light is initially spread over the entire active area ($x_d = L_c/2$) and then is imagined to be gradually confined ($x_d \rightarrow 0$), the total flux remaining constant. In the figure, $L_c/2$ has been chosen as 0.711 cm, a 2 cm^2 cell. The difference between the two series is that in the first the grid has been reoptimized at each position of x_d, while in the second the optimization has been made only at $x_d = L_c/2$. Both curves show serious degradation of $\overline{1\text{-}F_{SUM}}$ as the current density

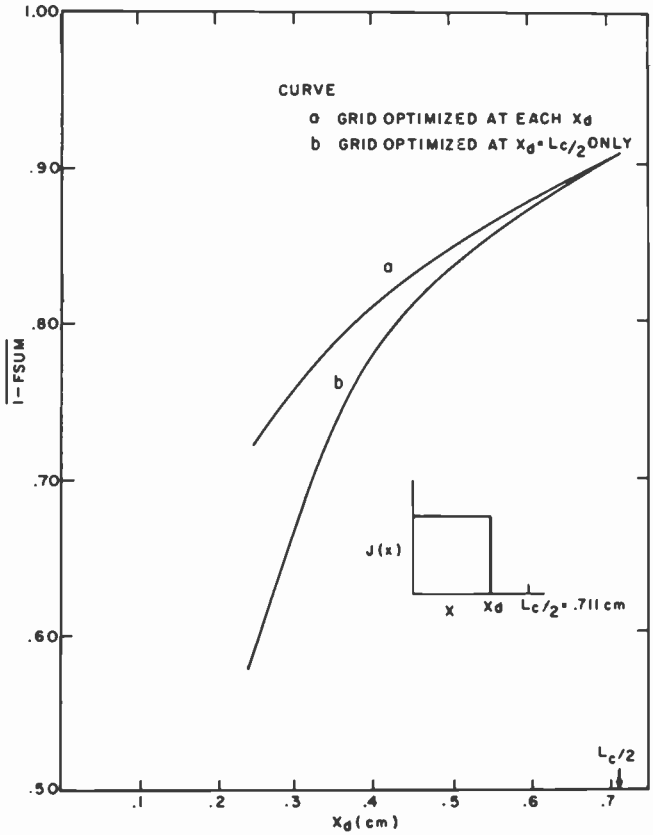


Fig. 6—Effect of changing the spot size on the fraction of delivered power, under two types of optimization: (a) grid optimized at each value of x_d and (b) grid optimized at x_d = 0.711 cm only (uniform illumination). Total current remains constant.

rises, but the second incompletely optimized curve suffers far more. The general conclusion is that the light should be spread over as large an area of the cell as possible. When it is necessary to confine the light due to tracking error, etc. the grid should be optimized to the distribution. No doubt this result could have been obtained intuitively, but the present calculation provides a quantitative method for accomplishing this.

Acknowledgment

I would like to thank D. Redfield for his critical reading of the manuscript, and L. Napoli and G. Swartz for useful discussions.

References:

A. R. Moore, "An Optimized Grid Design for a Sun Concentrator Cell," *RCA Review*, **40**, p. 140, June 1979.

Response of Diazoquinone Resists to Optical and Electron-Beam Exposure

Michael Kaplan and Dietrich Meyerhofer

RCA Laboratories, Princeton, N.J. 08540

Abstract—The responses of diazoquinone/novolak resist systems to optical and electron-beam radiation are compared. The "intrinsic" sensitivity to radiation is determined by measuring the rate of destruction of the diazoquinone sensitizer with radiation. Under optical exposure, a quantum efficiency of 0.30 ± 0.03 is measured for two different sensitizers and various resist formulations. This quantity is constant over the wavelength region of 350–450 nm where only the photoactive site absorbs. Under electron irradiation, we measure *G*-values (number of molecules destroyed per 100 eV of energy absorbed) of 7–8 for undiluted sensitizer films. The optical quantum efficiency and the *G*-value can be compared in terms of the energy required to destroy one sensitizer site. This energy is 12.5 eV for electron beam exposure and 10.5 for optical exposure. In resist films, the result depends on how the absorbed energy is partitioned between the sensitizer and the resin. If it is assumed that the energy is absorbed equally in both parts, then the *G*-value can be as much as a factor of 2 higher in the resist than in the sensitizer.

These resist systems are positive-working for both optical exposure and electron irradiation at low doses. The solubility of the films changes more rapidly with energy absorbed for electron irradiation than for optical irradiation, if only the energy absorbed by the sensitizer is taken into account. The reverse is true if all the energy is included. Practical exposure values required for lithographic applications are 3–4 mJ/cm^2 optical dose and 7–20 μC/cm^2 electron dose at 10 kV.

1. Introduction

In recent years, a number of papers have been published relating to the optical and electron-beam exposure characteristics of diazoquinone based resists,[1–4] and there are important commercially available resists

incorporating these species.[5] As microlithography tends toward smaller and smaller dimensions, these positive resists increase in importance because they provide higher resolution than the commonly used negative resists.

These resists[6,7] consist of two parts, a novolak resin with the addition of a diazoquinone derivative as sensitizer or inhibitor. The resin provides the necessary film-forming qualities and the resistance to liquid and gaseous etchants. It is very soluble in aqueous alkali. The sensitizer provides the photochemical behavior. Under ultraviolet irradiation, it changes from a species that is insoluble in alkali to one that is very highly soluble. It can therefore be used to delineate lithographic patterns all by itself.[8] When added to the resin, it causes the resin to become much less soluble. Again, as the sensitizer is photochemically degraded, the solubility of the resist mixture increases dramatically, thus allowing the generation of lithographic patterns.

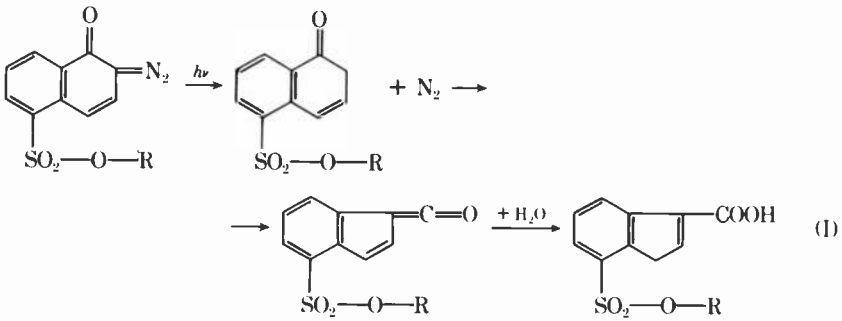
While these resists have mainly been used with optical exposure, they behave in a very similar manner with electron-beam exposure.[3–4] This is true of the sensitizer material by itself, but it is known[9] that the chemical degradation of the sensitizer material after electron-beam exposure does not necessarily lead to the same products as after optical exposure. Also, electron bombardment causes additional effects, such as cross-linking, in both the sensitizer and the resin.

In this paper, we compare the response of some resists of this type to optical and to electron-beam exposure and attempt to determine the mechanisms involved. The first step is to investigate the basic sensitivity, namely the rate at which the diazoquinone part of the sensitizer molecules is degraded by the optical or electron-beam irradiation, taking into account that the absorption mechanism is quite different in each case. The second step consists of determining the solubility parameters as functions of exposure and relating them to the quantity of sensitizer degraded.

2. Response of the Sensitizer to Radiation

The photochemical reactions of the diazoquinone moiety have been studied for many years.[10] Only recently have all the steps of the Süs reaction (I) been demonstrated, and it has been shown that the final product is the 3-indene carboxylic acid derivative, not the 1-indene

carboxylic acid:[11–13]



(I)

The starting material is not soluble in aqueous base, but the product, the carboxylic acid, is very soluble. Thus the sensitizer alone acts as a positive-working photoresist.

The dynamics of this reaction have been followed by studying the evolution of N_2,[14] the bleaching of the ultraviolet absorption of the diazo species,[1] and the change of the infrared absorption bands as a function of the UV exposure.[13,15] This last study showed that, under UV irradiation, "the diazide functional group (2110 cm^{-1}) and the ketone functional group in the naphthoquinone structure (1610 cm^{-1}) both decrease with increasing light exposure. Vibrational modes associated with a carboxyl carbonyl group (1700 cm^{-1}) and an indene ring structure (960 cm^{-1}) increase with light exposure," [15] in support of (I).

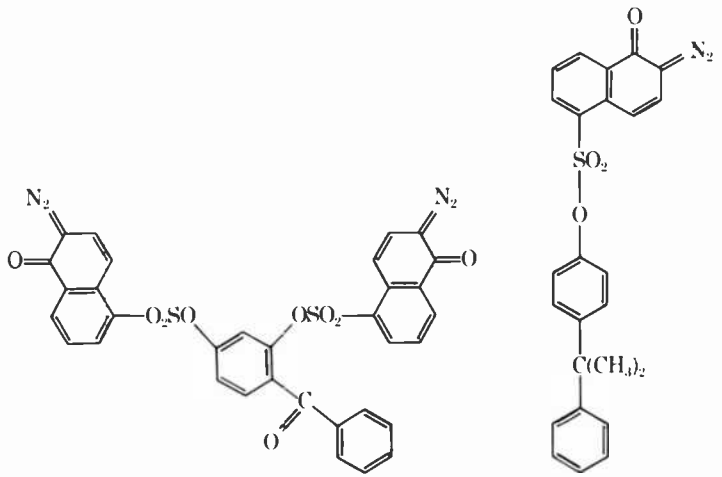
Less is known about the chemical reaction following electron bombardment. In measurements analogous to those following optical exposure, Zanzucchi and Barton showed that the diazide and ketone functional groups disappeared in the same way with exposure.[15] However, no new vibrational modes developed, which makes it very unlikely that any indene carboxylic acid is being formed.

Degradation of the sensitizer can also take place thermally. Dill and Shaw[16] showed that the degradation process and the resulting products are different from those of the photochemical process. In films of AZ1350J resist, they found a strong absorption at 405 nm after thermal degradation, which does not appear after optical exposure.

3. Experimental

3.1 Materials

Two diazoquinone sensitizers were investigated, A and B:



A B

Sensitizer B is monofunctional; sensitizer A is bifunctional. To study the photo- and electron-sensitivity of these sensitizers, we dissolved them in 2-methoxyethyl acetate and prepared films from solutions by conventional spin-coating techniques. For electron-beam exposure, conducting substrates were used, either 200 nm of evaporated Ni on glass or transparent plates of tin oxide-coated quartz. The relatively low solubility of these sensitizers limited the film thickness to about 100 nm.

To form lithographic resists, the sensitizers were mixed with various amounts of a cresol-formaldehyde novolak resin, Alnovol 429K of an average molecular weight of $M = 5000$, and dissolved in 2-methoxyethyl acetate. For most of the experiments, films of about 1-μm thickness were used. They were oven-baked, in air, at 70°C for 2 hours. This removes most of the solvent and does not affect the sensitizer properties.[16,17]

The resists formulated by mixing sensitizer A and the resin will be referred to as resists A, those with sensitizer B as resists B, and concentration will be indicated by the weight percent of sensitizer in the mixture. To compare resist A and resist B, we calculate the density of active sites n_o in mixtures containing 10% sensitizer, assuming a film density of 1 g/cm^3:

for A, $n_o = 1.8 \times 10^{20}$ cm^{-3};
for B, $n_o = 1.35 \times 10^{20}$ cm^{-3}.

The optical absorption of neat sensitizer B in the region of photosensitivity is shown in Fig. 1 (unexposed film). Films of this sensitizer on quartz were prepared by spinning from 2–6% solution in 2-methoxyethyl acetate. Assuming the density of the film to be 1 g/cm^3, the peak height at 405 nm corresponds to an absorption cross-section of 2.8 ×

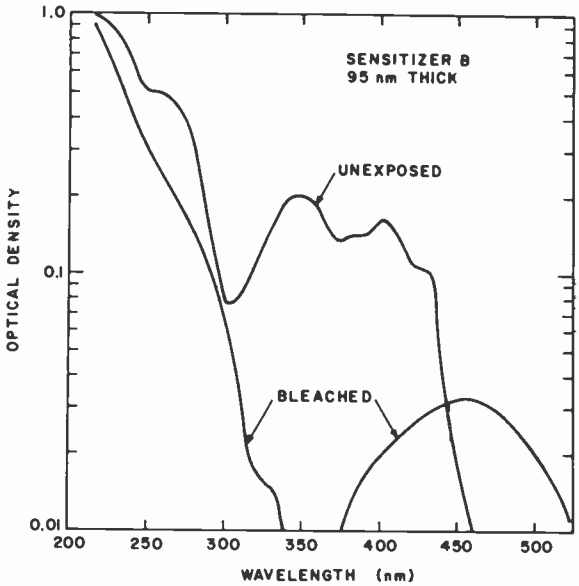


Fig. 1—Absorption spectrum of a 95-nm thick film of sensitizer B; curves are shown for the film as spun and after fully bleaching it by a long UV exposure. The shape of the curves at small optical densities was confirmed on much thicker films.

10^{-17} cm^2 (molar absorptivity of 7300 l/mole·cm). For comparison, we also measured the absorption of a solution of the sensitizer in 2-methoxyethyl acetate. A very similar spectrum resulted in the 300–500 nm range, with the 405 nm peak and absorption edge shifted by about 4 nm to shorter wavelengths. The height of that peak was 8200 l/mole·cm. Below 300 nm, the absorption of 2-methoxyethyl acetate is much larger than that of the sensitizer.

Similar measurements on sensitizer A showed spectra that differed in only minor details from those of B. Peak absorptivities are the same within experimental uncertainty if the two absorbing sites per molecule are taken into account.

The resists formed by combining the sensitizers with the resin were studied next. The absorption spectra were found to be very similar to those of the neat sensitizers, except that the absorption of the resin is added. The resin has a sharp absorption edge at 300 nm and is essentially transparent at the longer wavelengths. Thus, the optical properties above 300 nm are not strongly affected by the resin.

More detail about the optical properties of the sensitizer below 300 nm can be obtained by comparing the absorption of the unexposed film of Fig. 1 with that of a film that has been completely bleached by a long exposure with UV light.[1] The difference in optical density between the

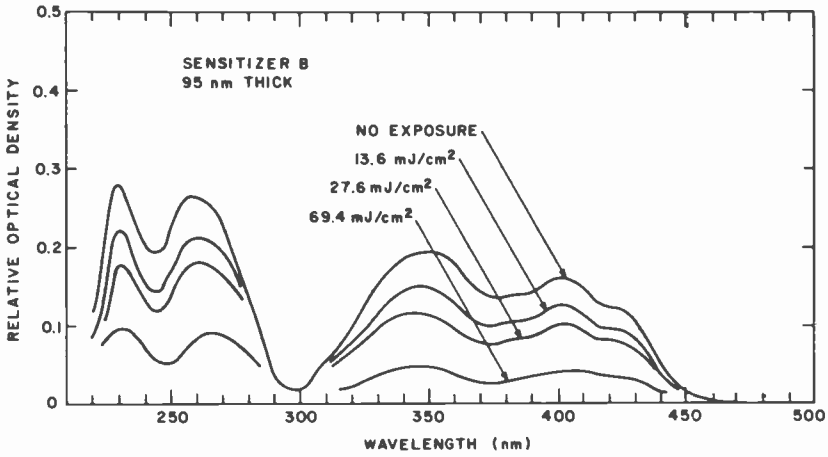


Fig. 2—Absorption spectra of the photosensitive part of the film shown in Fig. 1, obtained by subtracting the absorbance of the fully bleached film. The four curves are for four different degrees of exposure with 404.7 nm light.

two curves, obtained directly from a dual beam spectrophotometer, is shown in Fig. 2 for the same sensitizer film as Fig. 1. This shows that the absorption of the sensitizer below 300 nm consists of two parts: a bleachable part, with two more peaks at 220 and 260 nm, and a non-bleachable broad absorption band. During exposure, the entire photosensitive spectrum decreases uniformly without changing shape, as can be seen in Fig. 2, which means that all four peaks can be assigned to the diazoquinone part of the sensitizer molecules. Fig. 1, with the more sensitive scale, shows that the only decay product in evidence has a weak absorption band centered at 450 nm (1500 l/mole·cm).

These results are very similar to those obtained by Ilten[1] on Shipley AZ-1350 photoresist. His data do not show a peak at 220 nm, but only a broad plateau. Otherwise, the details are the same, confirming the identical nature of the photoactive center.

3.2 Electron Beam Exposure

Because we were interested in studying the optical properties and dissolution rates of these materials, it was desirable to expose rather large areas of the films. Conventional SEM focused-electron-beam exposure is not practical for this purpose. An apparatus was assembled that enabled us to uniformly irradiate an area up to several centimeters in diameter (Fig. 3). This flood exposure device consists of the electron gun and the view section of a display storage tube (RCA 7183A) with appropriate potentials applied to the various electrodes. Minor nonuni-

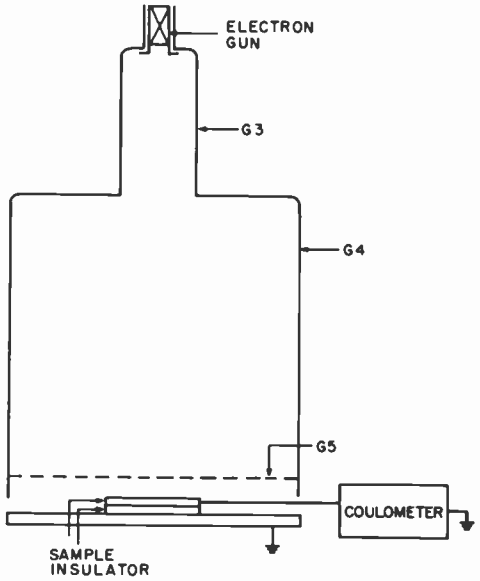


Fig. 3—Cross-section of the electron-beam flood-exposure apparatus.

formities in the flux are averaged out by rotating the sample under the electron beam. We typically operate at 10 kV and with currents of from 1 to 5 μA over a 7 cm diameter, so that an exposure of 5 $\mu C/cm^2$ requires about 5 minutes. A coulometer in the circuit measures the total exposure received by the sample by integrating the current. Measuring the total charge eliminates uncertainties due to current fluctuations. The Cr/Ni or tin oxide coating of the substrate serves as a means to remove the incident charge from the sample.

In order to make multiple exposures of various dosages on the same sample, we used grounded metal masks in contact with the sample. In this way, the effect of varying the exposure level could be investigated under otherwise identical conditions.

3.3 Solubility Rates

While, as was mentioned in the introduction, the sensitizers act as inhibitors and cause the unexposed resist films to be relatively insoluble in the alkaline developers, some finite solubility (erosion) always remains. It is, therefore, the difference between the solubility rate of the exposed film and the erosion rate that is of importance for lithographic applications.

When a film of sensitizer or resist is exposed optically, the intensity

of the exposure decreases with depth in the film due to absorption. For electron-beam exposure, the energy absorbed increases with depth due to the slowing down of the exposing electrons. Consequently, the development rate must vary with depth for both cases. In addition, even the unexposed resist films and the pure resin films generally show an increasing solubility rate with depth in the film.[16]

For this reason, we determine the solubility rates of all resist films by measuring the thickness of the films as a function of time in the developer. Since the solubility varies greatly with preparation of the films and with developer strength and temperature, we always prepare the films in the same manner (approximately 1-μm thick and baked 2 hours at 70°C) and use the same developer solution (Shipley AZ-303A diluted with six volumes of water and held at 24 ± 1°C).

For measuring the thickness, we use a simple *in situ* monitor that measures the reflectivity of a He-Ne laser beam.[17] The solubility rate as a function of depth is then determined by the slope of the thickness-versus-time curve. Fig. 4 shows two examples of such curves. For both films, the development rate increases by about a factor of three from the surface to the interior. This variation may be due to a nonuniform quantity of solvent remaining in the film or to oxidation of the surface layer. It is probably not related to swelling of the film, since one never

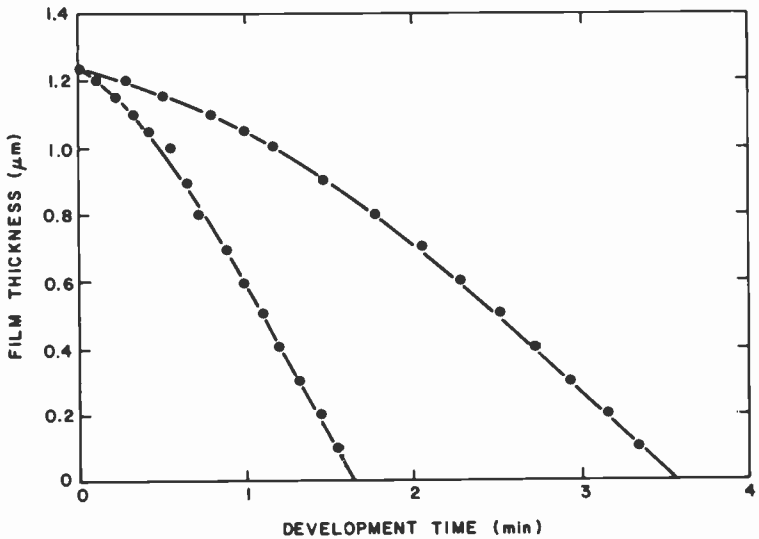


Fig. 4—Thickness of film of resist B (7.7% sensitizer) as a function of time in the developer. The upper curve is for an unexposed film, the lower one for a film exposed at 404.7 nm to 20 mJ/cm^2. In the latter case, the energy absorbed decreases about 30% from the top of the film to the bottom. The films were baked 2 hours at 70°C.

observes an increase of film thickness at the start of development. Also, these curves are very similar to those reported by Dill and Shaw[16] using a much milder and slower developer, which would be expected to produce more swelling (AZ developer diluted with one part of water that has a normality equivalent to 0.10 normal NaOH compared to 0.24 for our developer).

In order to compare the solubility rates of different films, we use the value at a fixed depth of 0.7 μm below the film surface, since the change in rate with depth is relatively small in that region.

4. Sensitizer Degradation (Bleaching)

4.1 Degradation by Optical Exposure

The effect of UV irradiation on the absorption of the photoactive part of the sensitizer molecule is shown in Fig. 2. Since all four peaks decrease uniformly, the bleaching process may be followed quantitatively by measuring the absorption, or rather the transmission, at the same wavelength that is used for the bleaching. Due to the very nature of the absorption, the light intensity (and therefore the density of bleached molecules) varies with depth in the film. These relationships are derived in Appendix 1 where it is shown that, for the case of negligible background absorption, a simple relationship exists between the transmission of the film, T, and the exposure, E, at frequency ν

$$\frac{1-T}{T} = \frac{1-T_o}{T_o} \exp\left(-\frac{\sigma\phi E}{h\nu}\right). \qquad [1]$$

The absorption cross-section, σ, is obtained from the sensitizer concentration and the initial transmission, T_o; ϕ is the quantum yield of the photochemical reaction and is a measure of the inherent sensitivity of the sensitizer to optical radiation.

We measured the bleaching behavior of both sensitizer and resist formulations in the 330 to 450 nm band. In that range, the absorption of the fully bleached material may be neglected and Eq. [1] should apply. The light sources used were a He-Cd laser at 441.6 nm and a high pressure mercury arc with 8-nm-bandwidth interference filters at 404.7 and 365.0 nm. In all cases, care was taken that both the incident and transmitted beam had a uniform power density over the sample area investigated. The light intensity was measured with a photodiode calibrated with a thermopile. An example of experimental data on a single film of sensitizer and the fitted straight line are shown in Fig. 5. Using the known photon energy and the previously measured absorption cross-section (assuming a film density of 1 g/cm^3), a quantum yield of 0.31 is calcu-

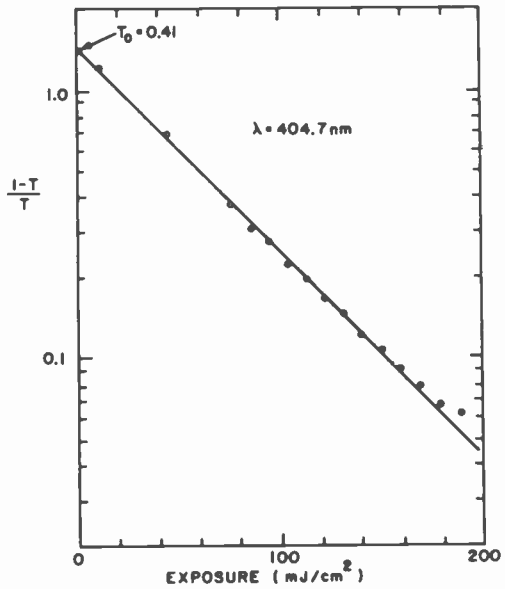


Fig. 5—Bleaching of a film of sensitizer B with 404.7 nm light according to Eq. [1]. The straight line fitted to the experimental points has a slope of 17.6 cm^2/J.

lated. The same value is obtained at the other wavelengths, as is expected for the absorption of a single center. The average value of all the measurements was 0.30 ± 0.03. The investigation could not be extended to the higher energy band because the high absorption of the nonbleachable part masks the bleaching process.

The same kinds of results were also obtained on two films of resist B, one containing 7% and the other 20% concentration of sensitizer. At 404.7 nm, where the resin does not absorb, the measured quantum efficiency was the same as for pure sensitizer, showing that the photosensitive center is not affected by the resin matrix.

We also investigated the bleaching of sensitizer A and resist A, with similar results. The measured absorption cross-sections and quantum efficiencies are summarized in Table 1.

We can compare these results with those reported in the literature on related materials. Volodarskii et al.[14] measured the quantum efficiency by determining the N_2 produced by exposure of the material in solution. His values ranged from 0.13 to 0.26 for a variety of diazonaphthoquinone sulfonic acids. Two groups have reported on the bleaching of a Shipley AZ-1350 resist, which is known to contain a sensitizer similar to that in A.[18] Ilten and Sutton[1] used the same technique as we did and obtained values of quantum efficiency of 0.15, 0.23, and 0.16 at 365, 405, and 435

Table 1—Quantum Efficiencies of Photo-Bleaching

Material	Sensitizer Fraction	Wavelength (nm)	Absorption Cross-Section* (10^{-18} cm^2)	Quantum Efficiency*
Sensitizer A		442	3.2	0.28
Resist A	12%	405	24	0.33
Resist A	12%	365	27	0.32
Sensitizer B		442	2.5	0.27
Sensitizer B		405	28	0.31
Resist B	20.1%	405	29	0.34
Resist B	7.2%	405	28	0.33

* Experimental uncertainty is ±5%

nm, respectively. However, the parameter values they report do not completely agree with the experimental data they show, and the approximation they use for Eq. [1] introduces some error, which prevents detailed comparison with our results. Dill and coworkers[2] measured the bleaching of AZ-1350J, which is believed to contain the same diazoquinone moieties as those in AZ-1350. They made no assumption about the density of photoactive sites so that no absolute values of absorption cross-section and quantum efficiency were determined. However, if one takes the ratio of their measured quantities A and C, the result is proportional to the quantum efficiency. From their Table 1, A/C varies as 1.0:0.78:0.62 at 365, 405, and 435 nm, respectively. This spread in values is larger than expected and does not agree with our data which show relative independence of wavelength. Paramonov and Prokhotskii[19] reported an efficiency of 0.56 on a resist similar to resist B, a surprisingly high value.

4.2 Degradation by Electron Beam Exposure

Electron beam exposure also causes bleaching of these materials, i.e., destruction of the diazoquinone portion of the sensitizer molecules. However, in contrast to optical exposure, energy is absorbed not only by the diazoquinone moiety but throughout the film, and some of this energy may be transferred nonradiatively to the labile site. The efficiency of bleaching thus depends not only on the amount of energy absorbed in the film, but also on how far the absorbed energy can be transferred within the film relative to the separation of the sensitive sites.

This means that one cannot define a unique parameter that describes the efficiency of degradation, as in the case of optical exposure. Instead, we use the more operational quantity, the G-value.[20] The G-value is defined as the number of diazoquinone moieties destroyed per 100 eV of energy absorbed in the film. Analogous to the photochemial quantum yield ϕ, the G-values are a measure of the inherent sensitivity of the

materials to electron-beam exposure, regardless of our ability to develop an image. Since the rate of sites being destroyed decreases during the exposure as their number is reduced, but the energy absorbed remains approximately constant, the G-value also varies. We thus restrict our interest to G_o, the initial G-value when all sensitizer molecules are active.

In order to calculate the G-value, the number of sensitizer molecules bleached is determined from the change in optical density at the 405 nm peak, as in the case of optical exposure. For this measurement, tin oxide coated substrates were used. The energy absorbed is estimated from the measured incident dose and the energy absorption characteristics of the resist. The details of the derivation are presented in Appendix 2 and the resulting equation for calculating the G-values is

$$G_o = 100\, a\, \zeta F w \rho / M. \qquad [2]$$

Here, a is the slope of a plot of the logarithm of the relative optical density (D/D_o) versus the dose (J/cm^3) absorbed in the film, w is the weight fraction and M the molecular weight of the sensitizer, ζ is the number of absorbing sites per sensitizer molecule (1 or 2), ρ the density of the film, and F is the Faraday. The subscript on G indicates that we have taken the limiting G-value as the absorbed dose approaches zero.

A typical plot made according to this methodology is shown in Fig. 6. The data were taken on about 20 different samples. The amount of scatter in the data is typical of the behavior of most of the films. Values

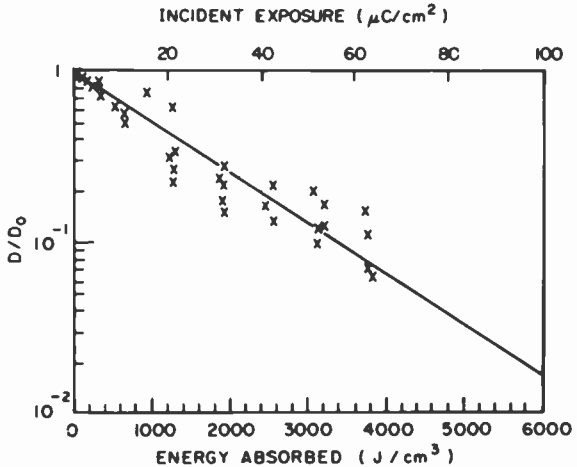


Fig. 6—Typical bleaching data under electron-beam exposure, Resist B (7.7% sensitizer). No more than 2 points derived from one film.

of a with uncertainties of 4–12% can be determined from the fitted straight line. The results for both kinds of sensitizer and for various resist formulations are shown in Table 2.

Table 2 shows a monotonic decrease in the G_o-values as resin is added to neat sensitizer. This is expected on the basis of the increasing dilution of the sensitizer. We can compensate for this dilution by considering the partition of absorbed energy between the two species. Since both sensitizer and resin are composed of essentially the same atoms, the amounts of energy absorbed by the two constituents are to first order proportional to their respective weight fractions. The column labeled G_o/w in Table 2 is the number of sites destroyed per 100 eV absorbed *by the sensitizer moiety*. We refer to it as the "compensated" G-value, and it exhibits a different behavior from G_o. Adding resin results in an *increase* in these compensated G-values, at least up to a point. The uncompensated G-values are of interest in quantifying the inherent sensitivity of these materials for use as resists; the compensated G-values tell us something about the mechanism of the process. The following conclusions are drawn from Table 2:

(1) For a given absorbed dose, both neat sensitizers are about equally destroyed.
(2) For a given absorbed dose, resist formulations having comparable sensitizer concentrations have a comparable amount of the sensitizer destroyed.
(3) Adding resin to either sensitizer results in a monotonic *decrease* in the efficiency of site destruction per unit dose *absorbed by the resist*.
(4) Adding resin to sensitizer B results in an initial *increase* in the efficiency of site destruction per unit dose *absorbed by the sensitizer*. For resist A, there is a similar initial increase. This increase implies either that energy is transferred from the resin to the sensitizer or that diluted sensitizer makes more efficient use of whatever energy it has absorbed.

4.3 Comparison of Optical and Electron Beam Exposure

We have seen that the optical quantum efficiency and the G-value are not equivalent, particularly for resist formulations. We can, however, compare the optical and electron exposures for one of the neat sensitizers. We consider the rate of sensitizer destruction as a function of energy absorbed in the sensitizer film. An example is shown in Fig. 7. The *initial* rates of change for both kinds of irradiation are proportional to the absorbed energy and differ by about 20%. In other words, while the absorbed electron energy required to destroy one site is 12.5 eV, the ab-

Table 2—G-Values of Sensitizers and Resists

Material	w	C (10^{20}/g)	D_o/d (μm^{-1})	a (10^{-3} cm^3 J^{-1})	G_o	G_o/w
Sensitizer A	1.0	17.8	1.88	0.280 ± .016	7.96 ± 0.45	8.0 ± 0.5
Resist A	0.136	2.42	0.271	0.562 ± .032	2.17 ± 0.12	16.0 ± 0.9
Sensitizer B	1.0	13.6	1.81	0.338 ± .014	7.34 ± 0.30	7.3 ± 0.3
Resist B	0.201	2.73	0.330	0.496 ± .027	2.16 ± 0.12	11.0 ± 0.6
Resist B	0.144	1.96	0.235	0.642 ± .056	2.00 ± 0.18	13.9 ± 1.2
Resist B	0.0773	1.05	0.126	0.682 ± .025	1.14 ± 0.042	15.0 ± 0.5
Resist B	0.0271	0.37	0.0446	0.462 ± .053	0.273 ± 0.031	10.0 ± 1.1

w: weight fraction of sensitizer in film
C: concentration of sensitizer molecules
D_o: initial optical density
d: film thickness
a: slope of log (D/D_o) versus energy absorbed per unit volume

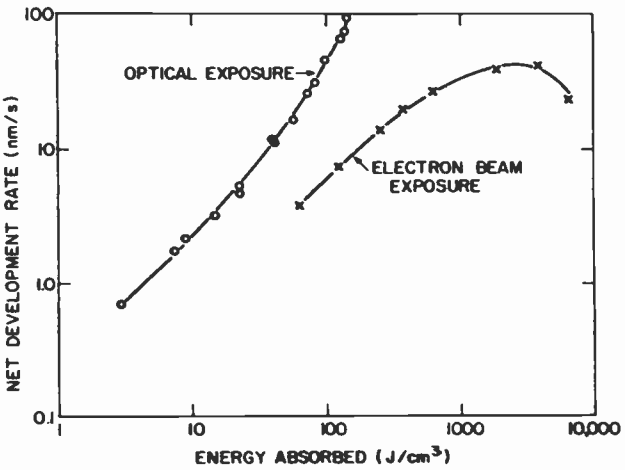


Fig. 9—Comparison of the effect of optical and electron-beam exposure on the net development rates of films of resist B (7.7% sensitizer). The net development rate is the difference between the solubilities of the exposed and unexposed film. The energy density absorbed is calculated at a depth of 0.7 μm below the original film surface, which is the point at which the solubility rates are measured.

One can characterize these curves by two parameters, $D_{1/2}$, the dose at which half the original film remains in the unexposed area, and Γ, the slope of the curve. $D_{1/2}$ is a practical measure of the resist sensitivity, while Γ represents the contrast. The parameter values of the six curves are listed in Table 3.

The resist formulations behaved very similarly under electron-beam exposure (Fig. 11), even though more of the electron energy must be transmitted to the sensitizer at the higher sensitizer concentrations, as discussed above. In contrast to this, the curves for optical exposure differ much more significantly. The contrast increases with sensitizer concentration in resist B and, at the same time, $D_{1/2}$ decreases. Part of this difference can be related to the fact that, under optical exposure, no cross-linking takes place, an effect that slows down the development-rate increase in the case of electron exposure.

As we have seen, it is difficult to compare optical and electron-beam exposure values directly, whether in terms of incident dose or in terms of absorbed energy. A more fundamental comparison of solubility rates can be based on the fraction of sensitizer molecules destroyed. We show such comparisons for two resist formulations in Figs. 12 and 13. The optical exposure curves show a strongly increasing solubility rate with exposure, especially for the formulation having the greater initial sensitizer concentration. Both curves of electron-beam-exposed resists show

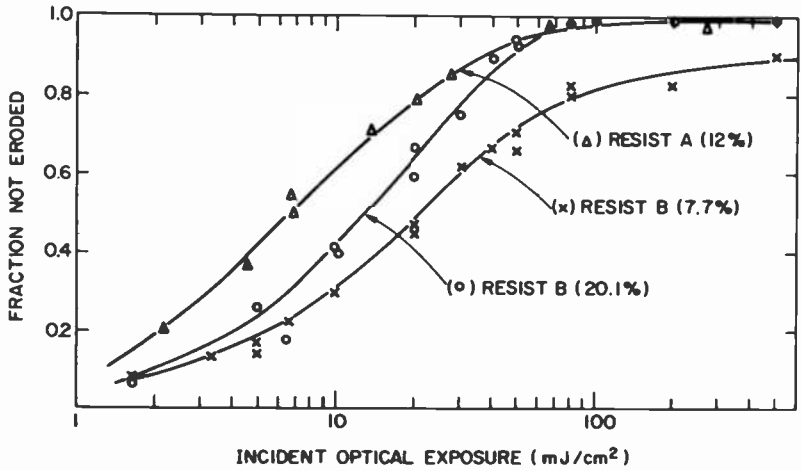


Fig. 10—Fraction of the resist film remaining in the unexposed area at the point where the exposed films are developed away completely as a function of the optical exposure incident on the film. Film thickness is 1.0 μm, exposure at 404.7 nm, and the developer is AZ-303A diluted with six parts of water. Three different resist formulations are shown.

initial increases of solubility rate that are larger than for the optical exposure. This confirms that even the initial radiation-chemical processes are different for the two cases. Hiraoka and Gutierrez[9] have

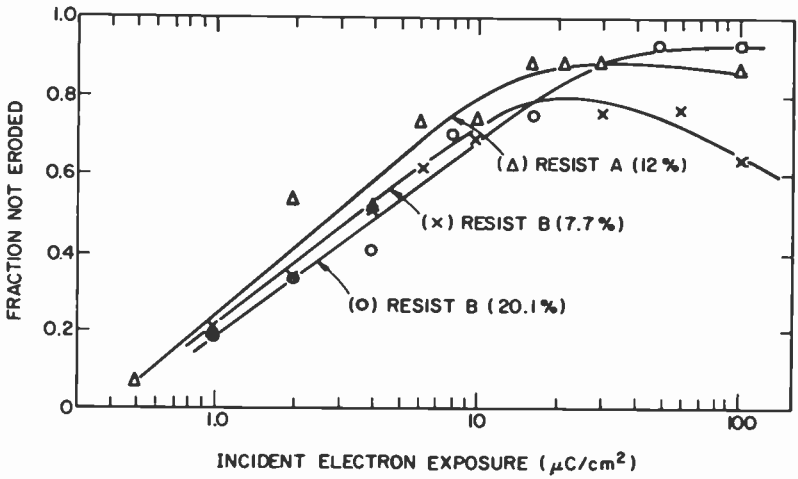


Fig. 11—The same information as in Fig. 10 for electron-beam exposure of the three resist formulations. Accelerating potential is 10 kV and the exposure is in terms of charge incident on the film.

Table 3—Resist Parameters for Optical and Electron-Beam Exposure

Resist	Sensitizer Concentration	Optical Exposure $D_{1/2}$ mJ/cm²	Optical Exposure Γ	Electron Exposure $D_{1/2}$ μC/cm²	Electron Exposure Γ
A	12%	6.6	0.65	3.0	0.55
B	20.1%	13	0.76	4.4	0.49
B	7.7%	21	0.61	3.6	0.49

postulated that there are two different decay modes of the sensitizer produced by electron beam irradiation; one is analogous to the optical mode (I), the other is initiated by the cleavage of the molecule at the sulfone link. They propose that the latter reaction leads to products that also increase the solubility of the resin. This reaction leaves the diazoquinone part of the molecule intact and so the quantity of sensitizer destroyed by it is not recorded by the UV absorption measurement. This feature could explain the differences in initial slopes of the optical and electron beam curves of Figs. 12 and 13, since the abscissa actually represents the fraction of diazoquinone moiety destroyed, as measured by UV absorption, rather than of the total number of sensitizer molecules.

At higher levels of sensitizer destruction, the electron-beam exposed curves saturate and then decrease, as discussed previously. It can be seen that this reduction of solubility relative to the optically exposed material is considerably larger for the material with the higher sensitizer concentration. This would not be expected if the decrease were solely due to cross-linking in the resin. Fig. 8 shows that, at the peak of the solubility

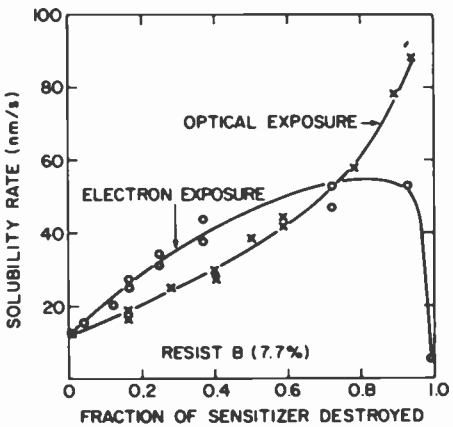


Fig. 12—Solubility rates as a function of the fraction of sensitizer destroyed. Exposure by light and electron beam are compared for films of resist B (7.7% sensitizer) developed in AZ-303A developer diluted 1:6.

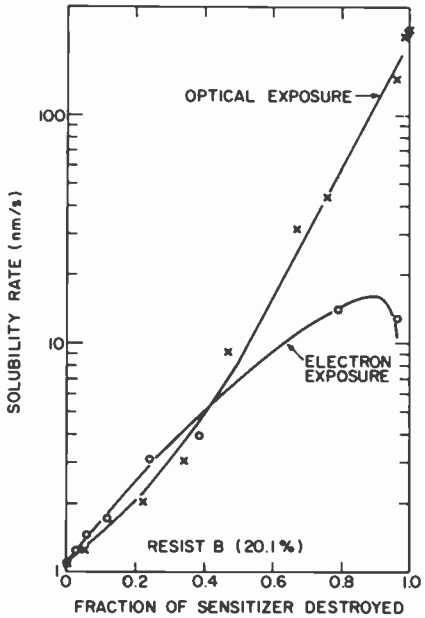


Fig. 13—Same data as in Fig. 12 for resist B with 20.1% sensitizer concentration. Because of the much larger range of solubility rates, the data are plotted logarithmically.

(about 5000 J/cm^3 absorbed), the solubility of the higher concentration resist is still much lower than that of the resin. Some other effect must be causing this discrepancy and it must be related to the large density of sensitizer reaction products.

6. Conclusions

We have compared the effect of optical and electron-beam radiation on the destruction of sensitizer and on the solubility rate changes that this destruction has produced. In neat sensitizers, the destruction of the diazoquinone portion was found to require about the same amount of energy to be absorbed in the film (10–12 eV per photoactive unit for either radiation). For optical irradiation, this also holds in resist films in the 350–450 nm band. In the electron-beam case, the radiation is absorbed by the resin as well as by the sensitizer. From the measured G-values, one concludes that some of the energy absorbed by the resin must be transferred to the sensitizer, since the efficiency of destruction is enhanced.

Since the only photochemical process is by reaction (I), the solubility of optically exposed resists increases monotonically with exposure until

the fully bleached state is reached. The increase grows with increasing initial sensitizer concentration in the resist. Electron exposure produces additional reactions that complicate the comparison. It appears to produce other modes of sensitizer destruction that enhance the solubility increase at low exposures. But it also causes cross-linking of the resin and, possibly, additional effects due to sensitizer decay products that inhibit the solubility, thus rendering the resists insoluble at very high doses.

For lithographic applications, the interesting exposure range is at relatively low doses. Here, we find significant differences among three resist formulations under optical exposure, but almost no differences under electron-beam exposure. Changing developer strength is not important, since ratios of solubility of exposed to unexposed resists are independent of developer. The development time can thus be chosen on the basis of practical considerations.

Note Added in Proof

We have recently measured the volumetric density of a film of resist A (12% sensitizer) to be 1.15 g/cm^3. We can assume similar values for resist B and for the resin. We cannot measure the density of the pure sensitizers. If we assume them also to be in the range of 1.15 g/cm^3, we find the following changes in the quantities discussed (all by about 15%): increases in n and n_0, decrease in σ, increase in ϕ, increases in the energy and energy density absorbed from the electron beam, but almost no change in G_0. These adjustments do not cause any qualitative changes in the comparisons between optical and electron-beam exposure. The difference in energy required to destroy one sensitizer molecule is somewhat increased.

Acknowledgments

We wish to thank L. A. Barton for providing the materials used in this study, E. J. Gavalchin and E. S. Poliniak for assistance with the experiments, and D. L. Ross for many useful discussions and for review of the manuscript.

Appendix 1—Exposure and Bleaching

Consider a resist film consisting of a sensitizer and a resin. The photosensitive site(s) of the sensitizer molecule absorbs the incident radiation. A fraction ϕ of the molecules that have absorbed a quantum of radiation degrade and form various reaction products. ϕ is defined as the quantum

efficiency of the process. The optical absorption of the reaction products at the exposure wavelength is smaller than that of the sensitizer molecules (bleaching). We make the following definitions:

d: film thickness (x direction is normal to the film)
n_o: initial concentration of sensitizer molecules per unit volume
$n(x,t)$: concentration of undegraded sensitizer molecules
I_o: incident light intensity of frequency ν
$E = I_o t$: exposure
$I(x,t)$: light intensity within the film
$T(t) = I(d,t)/I_o$: transmission of the film, the quantity measured
σ: optical absorption cross-section of the unreacted sensitizer at frequency ν
σ_r: cross-section of the degraded sensitizer at ν
α_b: absorption coefficient at ν due to the resin

We neglect reflections at the interfaces, which are small and can be taken into account in a straightforward manner if desired.

The differential equations governing absorption and bleaching, respectively, are[2]

$$-dI = I[n\sigma + (n_o - n)\sigma_r + \alpha_b]\, dx, \qquad [3]$$

$$-dn = \phi n \sigma (I/h\nu)\, dt. \qquad [4]$$

When the nonbleachable background absorption can be neglected (σ_r and α_b small) Eqs. [3] and [4] can be integrated to obtain an analytical form for the film transmission

$$\frac{1 - T}{T} = \frac{1 - T_o}{T_o} \exp\left(-\frac{\sigma\phi E}{h\nu}\right), \qquad [5]$$

where $T_o = T(t = 0)$. This equation allows comparison with experiment over the full range of bleaching.

When the background absorption can no longer be neglected, a closed form solution of Eqs. [3] and [4] can only be obtained for the *initial* change of transmission with exposure[2]

$$\left.\frac{dT}{dE}\right|_{t=0} = \frac{\sigma\phi}{h\nu} T_o(1 - T_o) \frac{n_o(\sigma - \sigma_r)}{\sigma n_o + \alpha_b}. \qquad [6]$$

When we need to know the energy absorbed at a particular depth in the film, we must determine the distribution of unexposed sensitizer molecules rather than their total number. For the case where σ_r and α_b are negligible, we obtain

$$n(x,t) = \frac{n_o \exp\{\sigma n_o x\} \exp\left\{-\dfrac{\phi\sigma}{h\nu} I_o t\right\}}{1 + (\exp\{\sigma n_o x\} - 1) \exp\left\{-\dfrac{\phi\sigma}{h\nu} I_o t\right\}} \qquad [7]$$

and, by definition, the absorbed dose e per unit volume is

$$e = -\int_0^t \frac{\partial I}{\partial x}\, dt = \frac{h\nu}{\phi}(n_0 - n). \qquad [8]$$

When the background absorption becomes significant, the absorbed dose must be determined by numerical integration.

Appendix 2—Energy Absorption During Electron Exposure—*G*-Value

To calculate the electron energy absorbed in a polymer film, we make use of the semi-empirical dose/depth relationship given by Everhart and Hoff.[21] According to these authors, the rate at which energy is lost by a high energy electron as it penetrates a film is given by the equation

$$\frac{d(\epsilon/\epsilon_0)}{d(x/R_G)} = \lambda\left(\frac{x}{R_G}\right) \qquad [9]$$

This equation is presumed valid for all materials and for electron energies in the kV range. Here, ϵ is the electron energy after it has traversed a distance x into a material in which its Gruen range is R_G; and ϵ_0 is the initial electron energy. The form $\lambda(x/R_G)$ is a polynomial in the reduced penetration depth and is given explicitly by the authors cited. The energy loss $(d\epsilon/dx)$ for electrons of 10 keV incident on a 1.0-μm-thick film of sensitizer or resist is shown in Fig. 14. The second curve represents the energy loss due to the fraction of electrons reflected back from the Ni

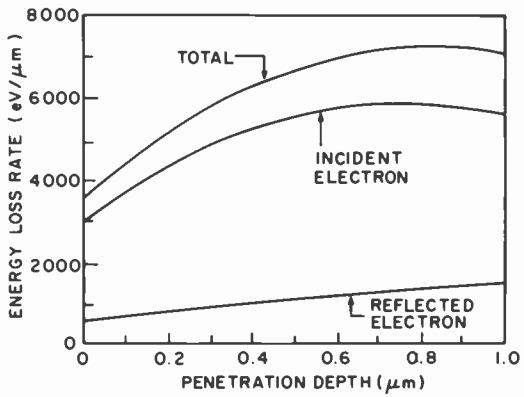


Fig. 14—Rate of energy loss of a 10-kV electron in a polymer film of 1.0 μm thickness on a chrome substrate. The losses of the incident and reflected electrons are shown separately.

substrate.[22] By integrating the film loss rate over the film thickness, one calculates the energy absorbed per electron incident on a 1.0-μm film: $\Delta\epsilon = 6200$ eV.

The rate of degradation of sensitizer molecules is related to $i(d\epsilon/dx)$, where i is the incident electron current density, and to the concentration of undegraded molecules. However, this relationship is not as simple as in the case of optical exposure, since the amount of energy absorbed by each undegraded molecule varies with their concentration.

The sensitivity of radiochemical reactions is generally expressed as the initial number of molecules degraded per 100 eV absorbed, the G-value,

$$\left.\frac{dn}{dt}\right|_{t=0} = \frac{i}{|q_o|}\frac{d\epsilon}{dx}\frac{G}{100}, \qquad [10]$$

where i is the incident electron current and q_o the electron charge. Integrating over the film thickness, we obtain

$$\left.\frac{dN}{dt}\right|_{t=0} = \frac{i}{|q_o|}\Delta\epsilon\frac{G}{100}. \qquad [11]$$

N is the number of undegraded sensitizer molecules per unit *area* of the film. It is proportional to the optical density, $D = -\log T$, where there is no background absorption:

$$N/N_o = D/D_o. \qquad [12]$$

N_o,D_o are the initial values of the parameters. Experimentally, we observe an exponential dependence on exposure or on the total energy absorbed,

$$D/D_o = \exp(-ai\,t\,\Delta\epsilon/d), \qquad [13]$$

so that

$$\frac{1}{N_o}\left.\frac{dN}{dt}\right|_{t=0} = \frac{1}{D_o}\left.\frac{dD}{dt}\right|_{t=0} = -ai\Delta\epsilon/d. \qquad [14]$$

Combining this with Eq. [11], we obtain

$$G = 100\,aN_o|q_o|/d. \qquad [15]$$

The initial number of sensitizer sites in the film of thickness d is given by

$$N_o = \zeta N_A w\rho d/M, \qquad [16]$$

where ζ is the number of active sites per molecule, N_A is Avogadro's number, w is the weight percent of sensitizer, M the molecular weight, and ρ the density of the film. Then Eq. [15] becomes

$$G = 100\, a\, \zeta F w \rho / M, \qquad [17]$$

where $F = N_A |q_o|$ is the Faraday.

References:

[1] D. F. Ilten and R. J. Sutton, "Spectrophotometric Determination of Photoresist Photosensitivity: II. Comparison of the Theoretical Model with Experimental Studies on AZ-1350," *J. Electrochem. Soc.*, **199,** p. 539 (1972).

[2] F. H. Dill, W. P. Hornsberger, P. S. Hauge, and J. M. Shaw, "Characterization of Positive Photoresist," *IEEE Trans. Electr. Dev.*, **ED-22,** p. 445 (1975).

[3] B. Broyde, "Exposure of Photoresists: II. Electron and Light Exposure of Positive Photoresist," *J. Electrochem. Soc.*, **117,** p. 1555 (1970).

[4] J. M. Shaw and M. Hatzakis, "Performance Characteristics of Diazo-Type Photoresists Under Electron-Beam and Optical Exposure," *IEEE Trans. Electr. Dev.*, **ED-25,** p. 425 (1978).

[5] For example, AZ resists manufactured by Shipley Company, Inc., Newton, Mass.

[6] J. Kosar, *Light Sensitive Systems*, Chap. 7.4, John Wiley and Sons, Inc., N.Y., N.Y. (1965).

[7] W. S. DeForest, *Photoresist Materials and Processing*, McGraw-Hill Book Co., N.Y., N.Y. (1975).

[8] See, e.g., M. P. Schmidt and O. Süs, U.S. Pat. 3,046,115 (*Chem. Abstr.*, 58, 12484a, 1963).

[9] H. Hiraoka and A. R. Gutierrez, "Electron Beam-Induced Reactions of Orthonaphthoquinone-Diazide-Sulfonyl Derivatives in Phenolic-Type Resins," *J. Electrochem. Soc.*, **126,** p. 860 (1979).

[10] O. Süs, *Liebigs Ann. Chem.*, **556,** p. 65 (1944).

[11] G. N. Rodionova, Yu. G. Tuchin, N. P. Protsenko, and R. D. Erlikh, "Spectroscopic Detection of Atoms During the Photolysis of Naphthoquinone Diazides," *Zs. Vses. Khim. Obshchest.*, **18,** p. 355 (1973).

[12] R. D. Erlikh, N. P. Protsenko, L. N. Kurkovskaya, and G. N. Rodionova, "Photolysis of O-Naphthoquinonediazides. Structure of Substituted Indenecarboxylic Acids," *Zs. Vses. Khim. Obshchest.*, **20,** p. 593 (1975).

[13] J. Pacansky and D. Johnson, "Photochemical Studies on a Substituted Naphthalene-2,1,Diazooxide," *J. Electrochem. Soc.*, **124,** p. 862 (1977).

[14] I. L. Volodarskii, B. I. Belov, and V. V. Kozlov, "Diazo Compounds. **XXI.** Light Sensitivity of 2,1-naphthoquinone-1-diazide and Its Isomeric Monosulfonic Acids," *Zh. Obshch. Khim.*, **35,** p. 2071 (1965).

[15] L. A. Barton and P. Zanzucchi, private communication.

[16] F. H. Dill and J. M. Shaw, "Thermal Effects on the Photoresist AZ1350J," *IBM J. Res. Devel.*, **21,** p. 210 (1977).

[17] D. Meyerhofer, to be published.

[18] J. Pacansky and J. R. Lyerla, "Photochemical Decomposition Mechanisms for AZ-Type Photoresists," *IBM. J. Res. Develop.*, **23,** p. 42 (1979).

[19] A. I. Paramonov and Yu. M. Prokhostkii, "Determination of the Primary Quantum Yield of Photodecomposition of O-Naphthoquinonediazides," *Zh. Nauchn. Prikl., Fotogr. Kinematogr.*, **20,** p. 219 (1975).

[20] A. Charlesby, *Atomic Radiation and Polymers*, Pergamon Press, Oxford (1960), p. 20.

[21] T. E. Everhart and P. H. Hoff, "Determination of Kilovolt Electron Energy Dissipation vs Penetration Distance in Solid Materials," *J. Appl. Phys.*, **42,** p. 5837 (1971).

[22] T. E. Everhart, "Simple Theory Concerning the Reflection of Electrons from Solids," *J. Appl. Phys.*, **31,** p. 1483 (1960).

The Effect of Rain on Satellite Communications Above 10 GHz

R. S. Engelbrecht

RCA Laboratories, Princeton, N.J. 08540

Abstract—Rain-incuced attenuation, scattering and depolarization can significantly affect the performance of future satellite communications systems in the 12/14 or 18/30 GHz frequency bands. In this paper, the pertinent rainfall statistics are reviewed and then applied to the prediction of earth–space attenuation distributions as a function of operating frequency, geographical location, and path-length (i.e., antenna elevation angle). For ground-site locations in the U.S., the fraction of total time during which rain-attenuation exceeds 3 dB can range from about 10^{-3} (short 11-GHz paths in low rain-rate regions) to almost 10^{-1} (long 30-GHz paths in high rain-rate regions). To maintain communications service during periods of increased attenuation, the system must operate in some temporary diversity mode. Various basic diversity methods are evaluated and shown to provide substantial improvements in service availability.

1. Introduction

At frequencies above 10 GHz, rain is known to cause significant absorption, scattering, and depolarization and, hence, becomes an important design parameter in future satellite communications systems. Its impact on systems cost is obvious, since uninterrupted service availability may, in severe cases, require switching to duplicate ground facilities at another location—facilities that stand idle 99% or more of the time. Complete knowledge of the spatial and temporal rainfall statistics is therefore highly desirable for each potential earth station site. But a detailed set of local measurements requires years to obtain, and as yet no general theory exists that permits extrapolation of data from

established sites to unchartered locations. In their absence, the systems planner must supply a certain amount of intuition and use design concepts that are reasonably tolerant to errors in judgment.

In this paper, we review the procedure for estimating the probability of excessive rain attenuation, scattering, and depolarization as a function of microwave frequency, geographical location, and path length. We then evaluate the available options for maintaining service when excessive rainfall occurs. Where possible, the emphasis is on underlying physical principles rather than specific empirical results that may or may not be broadly applicable. Although the physical picture is rather sketchy at present, it will hopefully become sufficiently rounded with time to achieve general validity, requiring only a modest amount of local experimental verification.

To predict the effect of rain on any proposed satellite communications system above 10 GHz, we must address the following questions, either directly or implicitly:

(1) At the desired frequency of operation and in an incremental volume in space, what is the instantaneous attenuation as a function of raindrop density and drop size? How is rain density (a volume function) related to measured rain-rate (a surface vector)? How are depolarization and scattering related to raindrop size, ellipticity, and canting angles?

(2) What are the point rain-rate statistics at the desired ground station location? If not available, how can they be estimated?

(3) What are the path-integrated rain statistics for the desired earth-space link? How can they be estimated from the point rain statistics and other known characteristics of rainfall?

(4) Combining (1), (2), and (3), what are the expected attenuation statistics for the proposed earth-space link? The depolarization? Interference from other sources via rain scattering? What are the temporal characteristics, e.g., the minimum and average duration of excessive attenuation? The number of occurrences per year? What is the probability of simultaneous strong attenuation on an alternate nearby path?

(5) How can we maintain the desired level of service availability in the presence of heavy rainfall, using one or several of the following diversity methods:

(a) Transmitter power diversity (how much, how often, and for how long during each occurrence?)

(b) Frequency diversity (can we switch to lower frequencies at which rain effects are negligible (e.g., 4/6 GHz)?)

(c) Bandwidth diversity (can we trade sufficient spectrum for loss in signal level?)

(d) Time diversity (for services not requiring real-time operation, can we use temporary data storage during periods of heavy rain?)
(e) Earth-space path diversity (what is the optimum spacing and relative orientation of duplicate ground stations?)
(f) Spacecraft antenna-gain diversity (can we aim adaptive spot beams at ground stations experiencing heavy rainfall?)

For each of the schemes in (5), what is the cost of achieving the desired level of service availability? Tolerance to errors in estimated rainfall statistics? Conservation of transmitter energy and frequency spectrum? Degree of required system cooperation and coordination?

Let us now examine each of these questions in turn.

2. Point Attenuation, Phase Change, and Depolarization Versus Point Rain-Rate

Rain is usually measured as the rate (mm/hr) at which water is collected by a horizontal surface near the ground. For each rain-rate, there exists a typical distribution of arriving drop velocities and diameters[1] (Table 1). From these, the instantaneous distribution of drop sizes in a volume of space above (and near) the rain gauge can be deduced as a function of the observed instantaneous rain-rate. Using the Mie scattering theory for spheres[2] and the known complex refractive index of water[3] (Fig. 1), the attenuation (absorption plus scattering) suffered by a plane electromagnetic wave can then be computed for each frequency of interest. Fig. 2 shows the resulting attenuation gradient (dB/km) versus rain-rate (mm/hr) with frequency as a parameter.[4] For future satellite systems

Table 1—Measured Raindrop Terminal Velocities and Drop-Size Distributions at Ground Level (Ref. [1]) (Laws and Parsons Drop-Size Distributions for Various Precipitation Rates)

Drop Diameter (cm)	Rain Rate (mm/hour)									Terminal Velocity $\left(\frac{m}{sec}\right)$
	0.25	1.25	2.5	5	12.5	25	50	100	150	
	Percent of Total Volume									
0.05	28.0	10.9	7.3	4.7	2.6	1.7	1.2	1.0	1.0	2.1
0.1	50.1	37.1	27.8	20.3	11.5	7.6	5.4	4.6	4.1	3.9
0.15	18.2	31.3	32.8	31.0	24.5	18.4	12.5	8.8	7.6	5.3
0.2	3.0	13.5	19.0	22.2	25.4	23.9	19.9	13.9	11.7	6.4
0.25	0.7	4.9	7.9	11.8	17.3	19.9	20.9	17.1	13.9	7.3
0.3		1.5	3.3	5.7	10.1	12.8	15.6	18.4	17.7	7.9
0.35		0.6	1.1	2.5	4.3	8.2	10.9	15.0	16.1	8.35
0.4		0.2	0.6	1.0	2.3	3.5	6.7	9.0	11.9	8.70
0.45			0.2	0.5	1.2	2.1	3.3	5.8	7.7	9.0
0.5				0.3	0.6	1.1	1.8	3.0	3.6	9.2
0.55					0.2	0.5	1.1	1.7	2.2	9.35
0.6						0.3	0.5	1.0	1.2	9.5
0.65							0.2	0.7	1.0	9.6
0.7									0.3	

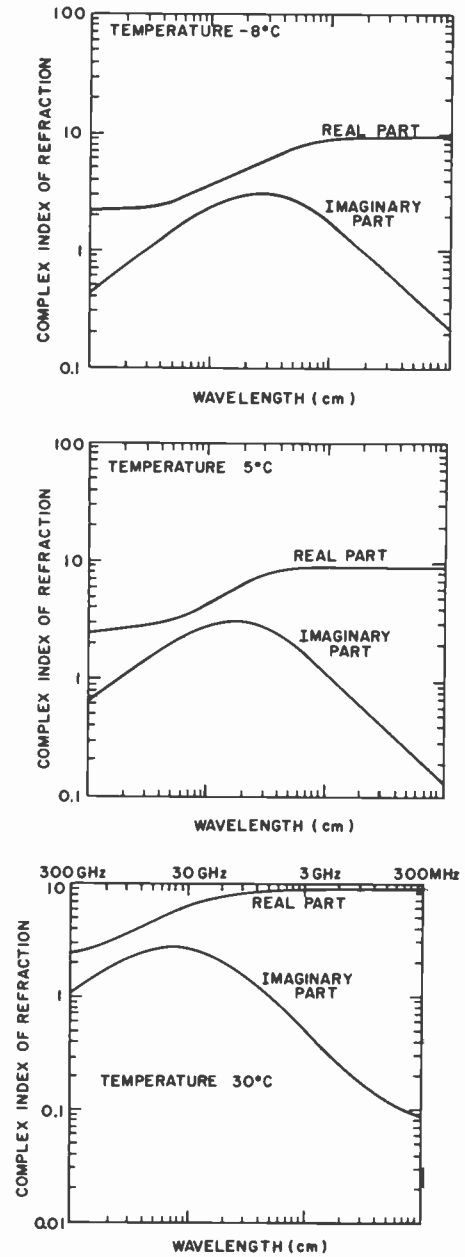


Fig. 1—Complex refractive index of water for temperatures from −8 to +30°C and for frequencies between 300 MHz (λ = 100 cm) and 300 GHz (λ = 0.1 cm) (Ref. [3]).

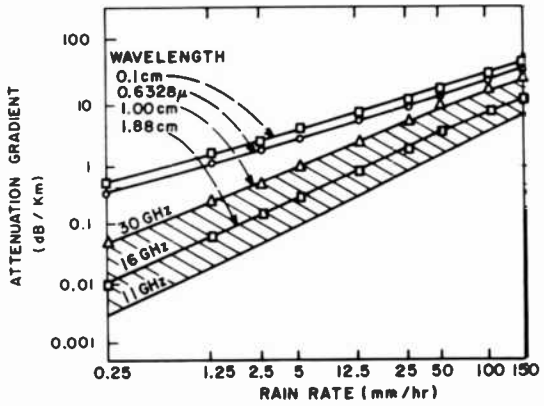


Fig. 2—Computed attenuation gradient α (dB/km) near ground level versus rain-rate at ground level.

(12/14 GHz and 18/30 GHz), the attenuation gradient ranges from less than 0.1 dB/km for 0.25 mm/hr rain rates (just measurable drizzle) to about 30 dB/km for 150 mm/hr rain rates (heavy downpour). In this frequency range, the computed values are given to good approximation (within 0.3 dB/km) by a simple linear relation

$$\alpha \cong \left(\frac{f-5}{135}\right) R, \qquad (10\ \text{GHz} \leqslant f \leqslant 30\ \text{GHz}) \quad [1]$$

where α = attenuation gradient, dB/km, f = frequency, GHz, and R = rain-rate, mm/hr. The phase change gradient of the wave is obtained (Fig. 3) from the same computations. Between 10 and 30 GHz, it ranges from 100 to 180 degrees/km for 150 mm/hr rain rates.

Actually, many raindrops are oblate spheroidal (flattened) rather than spherical and therefore produce slightly different amounts of attenuation and phase change for different polarizations of the wave. This effect has been calculated[5] using the drop size and velocity distributions of Table 1 and observed drop eccentricities.[6] The largest differences occur between the two cases where the electric field vector is either parallel or perpendicular to the raindrop's axis of rotational symmetry and are shown in Fig. 4. From these, the depolarization experienced by any linearly polarized wave can be calculated.[7] As one would expect from symmetry arguments, depolarization is zero for the two E-field orientations used in deriving Fig. 4 and maximum for an E-field oriented at 45° between them (or, identically, for circularly polarized waves). However, since raindrops are not lined up uniformly in space but exhibit

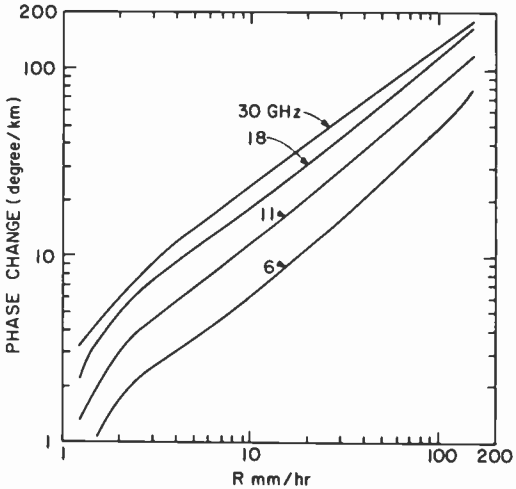


Fig. 3—Computed phase-change gradient (degr./km) near ground level versus rain-rate at ground level, for frequencies from 6 to 30 GHz (Ref. [4]).

a distribution in their canting angles,[8] the ideal case of zero depolarization is not observed in practice. The range of depolarization that may be expected is illustrated in Fig. 5 for different frequencies, as a function of total attenuation of the wave. It is of interest to note that in the 4/6 GHz band, depolarization cross-talk can become significant even when attenuation is modest (say, 3 dB), but at 11 GHz and higher, the attenuation must reach 15 dB or more before the depolarization cross-talk is of any consequence. It, therefore, appears that rain attenuation will be the limiting factor at these higher frequencies.

In addition to these degradations of a wanted microwave signal, rain also produces interference from unwanted sources via scattering. The total energy loss experienced by a plane wave (Fig. 2) is the result of both absorption and scattering. The contribution due to scattering alone is shown in Fig. 6. Below 10 GHz the energy loss due to absorption is at least ten times that due to scattering. At 30 GHz and above, the two effects are comparable (at least to 300 GHz; in the visible range, absorption becomes negligible compared to scattering). The interference due to scattering between systems (e.g., satellite and terrestrial) sharing the same frequency, of course, is largely determined by the geometry and power levels of each particular case. However, calculations based on typical models[9] show that scattering can significantly degrade the performance of low-noise receivers, particularly when the two interfering antenna beams share a common volume in space.

It should be emphasized that a direct one-to-one correspondence

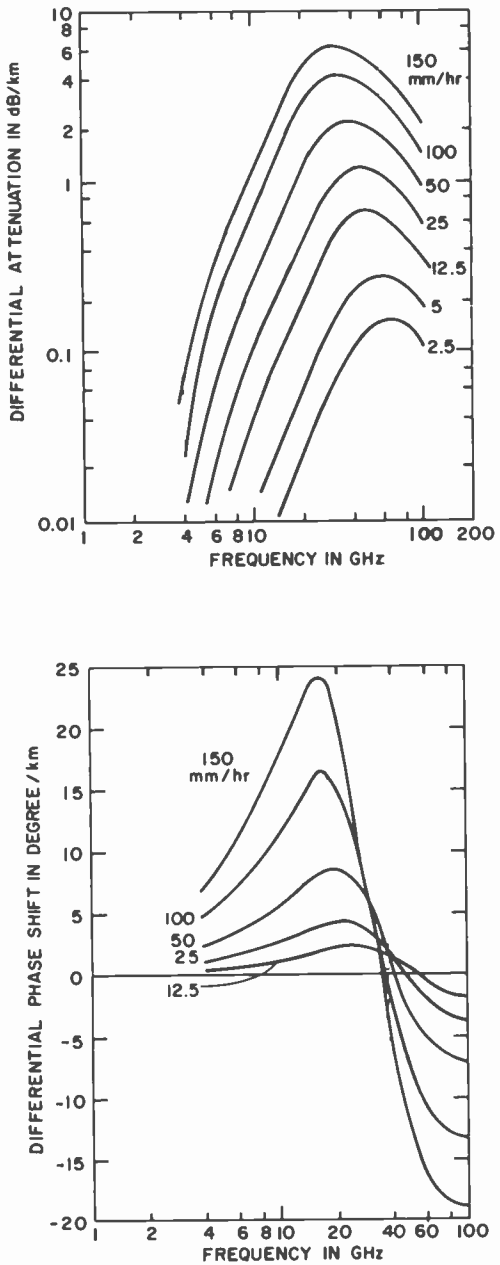


Fig. 4—Computed differential attenuation and phase-change gradients (dB/km and degr./km, respectively), due to raindrop oblateness, between waves polarized parallel or perpendicular to raindrop symmetry axis (Ref. [5]).

between observed rain-rate and simultaneous microwave propagation effects, as described above, exists only in that volume and during that time interval for which the rainfall is uniform and identical with that measured at the rain-gauge. In heavy showers, accompanied by gusting

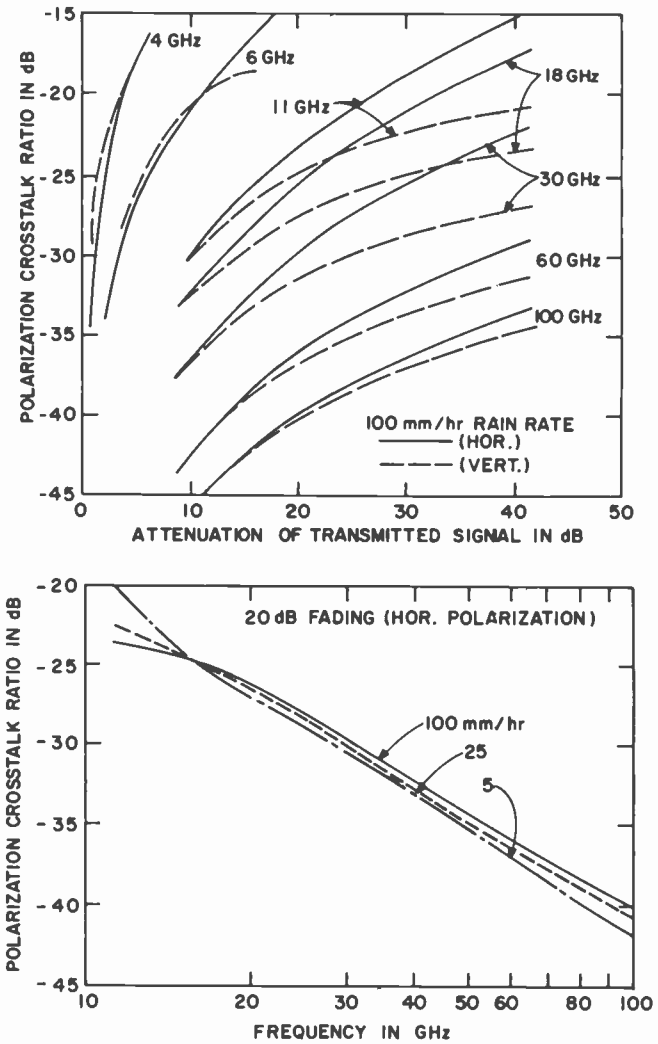


Fig. 5—Computed depolarization (dB crosstalk) for horizontal and vertical polarization versus signal attenuation at 100 mm/hr rain-rate (top), and for different rain-rates at 20 dB signal attenuation level (bottom), due to raindrop canting. Depolarization of circularly polarized waves is about 10 dB stronger than for horizontal polarization (Ref. [7]).

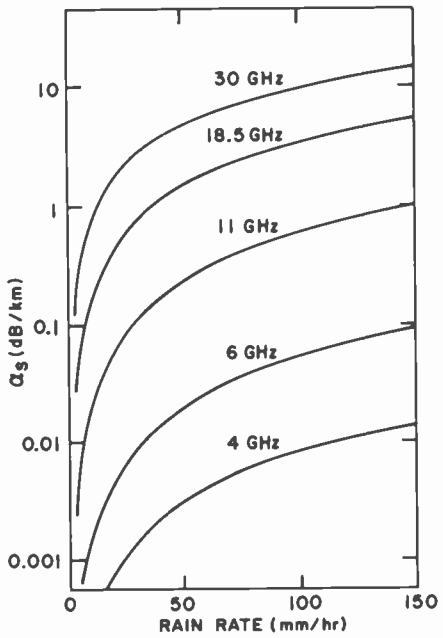


Fig. 6—Computed scattering loss component (dB/km) of attenuation gradient shown in Fig. 2 (Ref. [24]).

winds, the spatial and temporal coherence of rain may extend only over meters and seconds. Even light rain is rarely uniform over the entire length of a microwave link. The relations between total path attenuation (or depolarization, or scattering) and measured rain-rate at one point along the path are therefore, at best, statistical in nature. We will consider the statistics of rainfall next and see how they can be used to predict the overall propagation characteristics of an earth-space path.

3. Point Rain Statistics

Point rain data can be obtained from the records of many weather stations located throughout the U.S. and abroad. These records usually give the depth of water accumulated versus elapsed time, either in graphical or tabular form,[10] and must be differentiated to yield the rain-rate. Since rain varies rapidly with time, the raw data contain too much information to be generally useful and are therefore processed to reveal their long-term statistical characteristics. The most important of these is the cumulative distribution function (CDF), which shows the fraction of time a certain rain-rate is not exceeded. As an example, rain-rate CDF's of 20 U.S. locations are given in Fig. 7.[11]

Since truly instantaneous rain measurements are impossible to achieve and since tabulations involve discrete time slots, rain-rates are invariably averaged (i.e., integrated) over some nonzero intervals. Obviously the use of long integration intervals introduces smoothing, since brief heavy rain-rates tend to be missed and brief dry periods filled in. For the CDF's of Fig. 7 the integration interval is one minute. The effect of using different intervals is illustrated in Fig. 8. Here the CDF of Washington, D.C., is shown for rain-rates averaged over 1 minute, 30 minute, and 2 hour intervals. Log-log scales are used to accentuate the low rain-rate regime. We note that the fraction of time a *heavy* rain-rate is exceeded

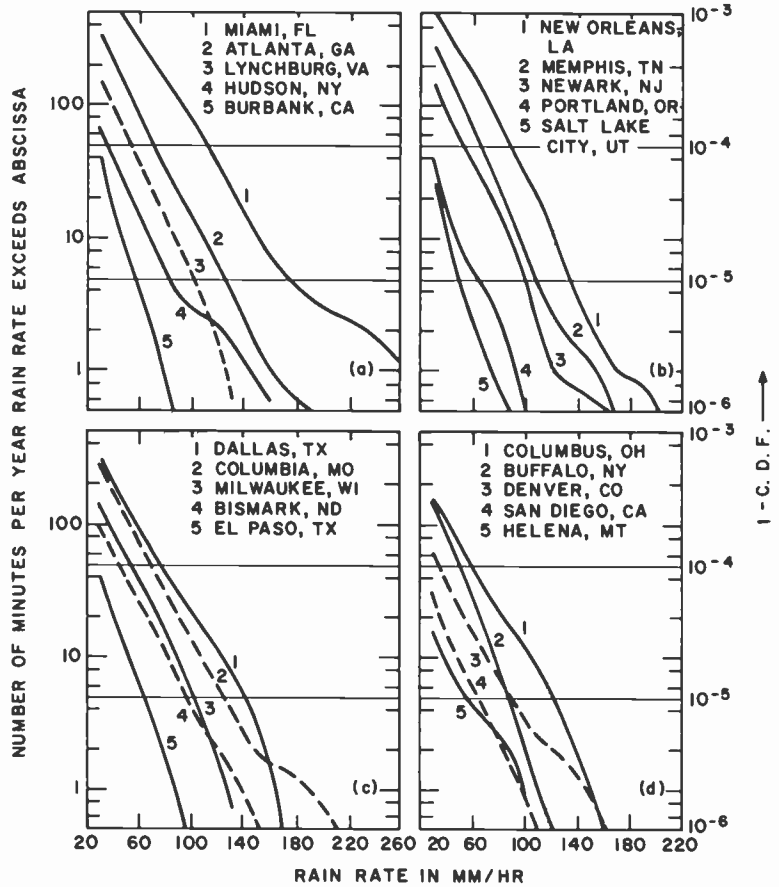


Fig. 7— Cumulative distribution function (C.D.F.) of rain-rates for 20 U.S. locations during the period 1966 to 1970; based on one-minute integration intervals (Ref. [11]).

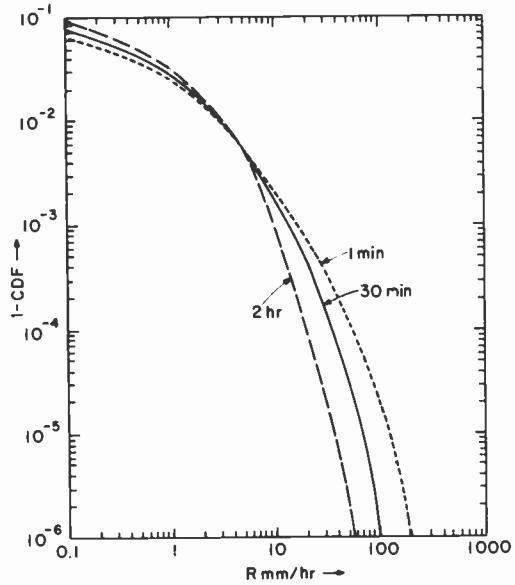


Fig. 8—Rain-rate cumulative distribution of Washington, D.C., for different integration intervals (Ref. [18]).

decreases with longer intervals, as expected. For *low* rain-rates (below a few mm/hr), it *increases* with longer intervals. This is consistent with the requirement that the long-term accumulated water depth be the same for all cases (about 1100 mm each year for Washington). With extremely long integration intervals (many years) the CDF would, of course, approach a step function (100% to 0%) at the long-term mean rain-rate (about 0.13 mm/hr for Washington).

In satellite communications we are mainly interested in rain-rates exceeding 10 mm/hr, for which the CDF becomes flatter with decreasing integration intervals. Even the one-minute intervals used in Fig. 7 cannot be considered to be instantaneous, as illustrated by Fig. 9. From it we deduce that the fine structure of heavy rain extends down to one second, perhaps even less. These data were obtained using a fast integrating rain gauge (1.5 seconds intervals). Very few direct measurements of this type are available, and most records permit resolutions of only one minute or longer intervals. By suitable normalization of the data, the dependence of rain-rate CDF on integration intervals has been shown to be similar for several locations in the eastern U.S.,[12] but its validity for other climatic regions has not been established.

As yet, no general method exists that permits an accurate extrapolation of measured rain characteristics to other geographical locations or

other time spans. Detailed measurements of the various important statistics exist only for a few locations (in the U.S., notably N.J.) and it would be highly useful if these could be applied or converted with confidence to other areas, particularly those with different climates (e.g. the West Coast). However, a good deterministic theory of rain generation and rainfall is not yet available, let alone its application to the prediction of localized rain occurrences. On the other hand, attempts to describe the statistical properties of rainfall by a few general parameters are beginning to show some results. For instance, there is evidence that the conditional rain-rates (CDF during raining time only) may be log-normally distributed.[13,14] This appears plausible if one assumes that the instantaneous rain-rate is controlled by a large number of independent

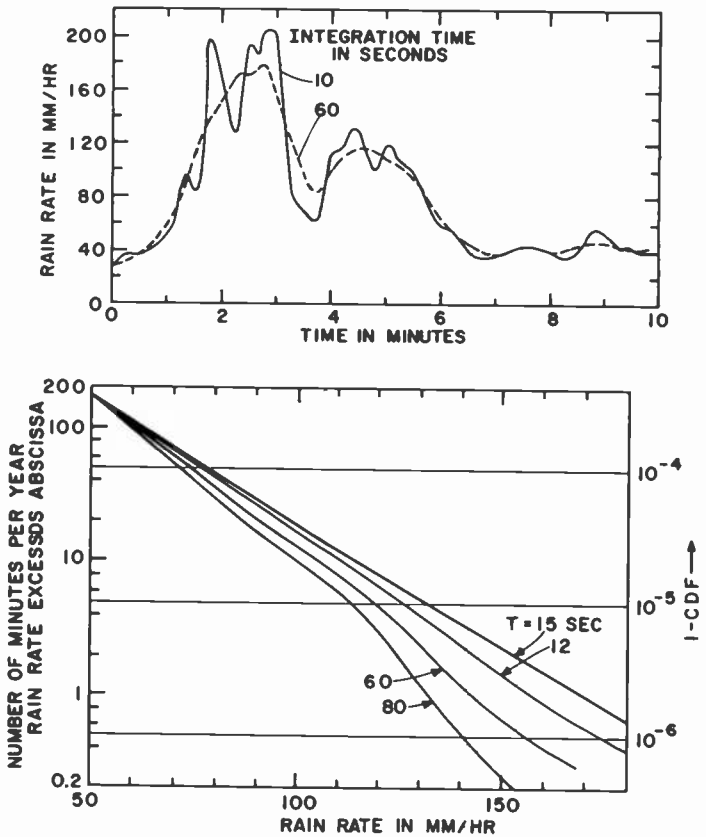


Fig. 9—Rain-rate versus time and cumulative distribution for different integration intervals (Ref. [11]).

multiplicative causes. Under this assumption, the point rain-rate CDF is given by

$$P(R \geq r) \simeq \frac{P_0}{2} \cdot \text{erfc}\left[\frac{\ln(r) - \ln(R_m)}{\sqrt{2}\, S_R}\right] \qquad [2]$$

where

$P(R \geq r)$ = probability that rain-rate R exceeds r (mm/hr),
P_0 = probability that rain is falling at the point of interest,
erfc = complementary error function,
R_m = median value of point rain-rate during raining time,
S_R = standard deviation of ln (R) during raining time.

Validation of this hypothesis is difficult because we are dealing with the extreme tail of the distributions, and very long time bases (many years) are required to obtain stable statistics.[15] Nevertheless, some good correlations have been obtained as illustrated by Fig. 10.

Until a suitable general theory emerges, our knowledge of the dependence of rainfall statistics on geographical location remains sketchy. Some comparisons may be drawn between such regional characteristics as annual depth of rainfall, rain-rates at a given probability level, or

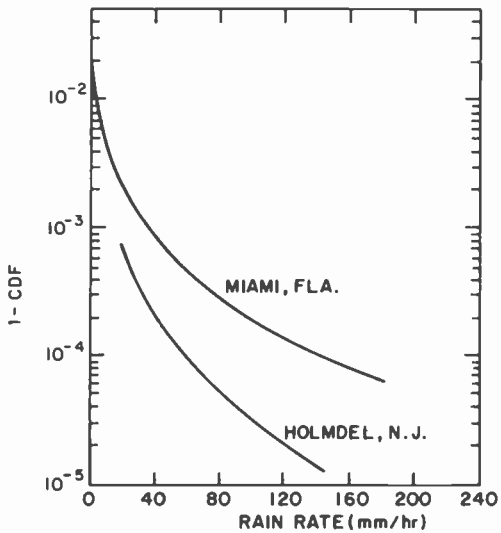


Fig. 10—Lognormal cumulative distribution (Eq. [2]), showing good fit to measured rain-rate distributions (2-second integration intervals) (Ref. [14]).

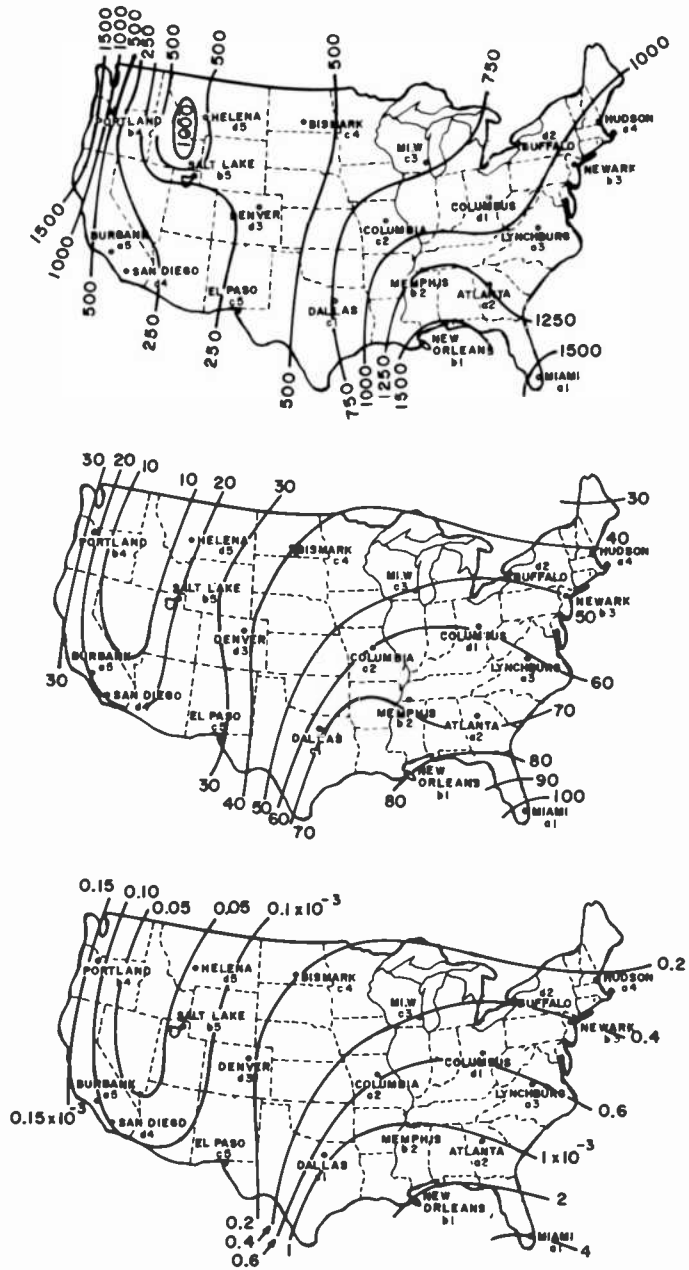


Fig. 11—Regional rainfall characteristics: average annual rainfall in mm (top); rain-rates (mm/hr) at C.D.F. = 10^{-4} level (center); and C.D.F. levels at 20 mm/hr rain-rate (bottom).

probabilities for a fixed rain-rate, as illustrated in Fig. 11 for the U.S. But aside from similar overall East-West tendencies, there are major differences. For example (using the curves of Fig. 7), Newark, N.J., shows a tenfold higher probability for a given rainrate than Portland, Ore., yet both have similar annual rainfalls (about 1000 mm). Also, Portland and El Paso, Texas, are seen to have nearly identical rain-rate distributions, but their annual rainfall differs greatly (1000 mm versus about 200 mm). Even local geographical features such as mountain ranges or bodies of water are known to cause significant differences. In the investigation of potential ground-station sites for future satellite links, therefore, no substitute exists at present for detailed evaluation (or acquisition) of local rain statistics. But one fact is clear—it is the statistics of short term ($\leq$1 minute) rain rates that are most important, not annual rainfall.

In addition to the cumulative distribution of rain-rates, it is important to know how long a given rain-rate, once it occurs, is likely to persist. Some of this information is contained in the so-called Rainfall Intensity-Duration-Frequency diagrams published by the U.S. Weather Bureau,[16] an example of which is shown in Fig. 12. From it, we see that (in the New York Area) a 30 minute shower for which the rain-rate continuously exceeds 50 mm/hr occurs, on the average, once every two years (the "return period"). Unfortunately, the data contain only showers of relatively long duration and high rain-rates, omitting the most fre-

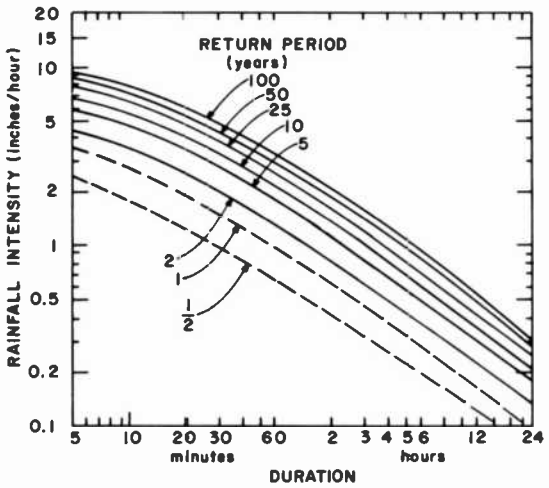


Fig. 12—Rainfall intensity-duration-frequency diagram for New York, N.Y. The return period is the average interval between rainfall occurrences whose intensity or duration exceeds the indicated values (Ref. [16]).

quent ones. Thus, comparison of the probabilities implied in the diagram (i.e., duration divided by return period) with the CDF curve for New York from Fig. 7 indicates that the average duration of heavy showers is, at most, a few minutes. It would therefore be desirable to obtain the necessary duration data directly from the rain-rate CDF dependence on integration time (Figs. 8 and 9), but again this will be possible only after a more general theory of rainfall has become available.

4. Path-Integrated Rain Statistics

As already mentioned, the instantaneous rain-rate measured at a point can be considered uniform only over a restricted area near that point. This area, and the corresponding volume above it for which the rain density is uniform, decreases with increasing rain-rate and is, in general, much smaller than that traversed by an earth-space propagation path. To obtain the total amount of rain (and hence the total attenuation) encountered over such a path at any instant in time, the contributions from many incremental volumes must be summed. However, the construction and operation of a dense rain-gauge network is usually not feasible and we must therefore seek to establish general statistical relations between path-average rainfall and measured point rain-rates.

From a rain-gauge grid covering a 12 × 12 km area in New Jersey,[17] the CDF's of instantaneous path-average rain-rates were obtained for path lengths from zero to 10.4 km, and are shown in Fig. 13(a). Fig. 13(b), in turn, illustrates the CDF's of point rain-rates as a function of integration interval. Comparing the two, we notice that for rain-rates from 50 to 150 mm/hr, the CDF's are roughly the same for the following pairs of integration interval (Δt) and path length (l):

l, km	Δt, min
5.2	5
7.8	11
10.4	14

This suggests that the point rain-rate integration interval and the instantaneous path-average rain-rate are related through a characteristic velocity

$$v = \frac{l}{\Delta t} \simeq 0.7 \text{ to } 1.0 \text{ km/min.} \qquad [3]$$

Measurements of instantaneous path-average rain-rates over 50 km distances in Ohio[18] have shown similar agreement with point rain-rates

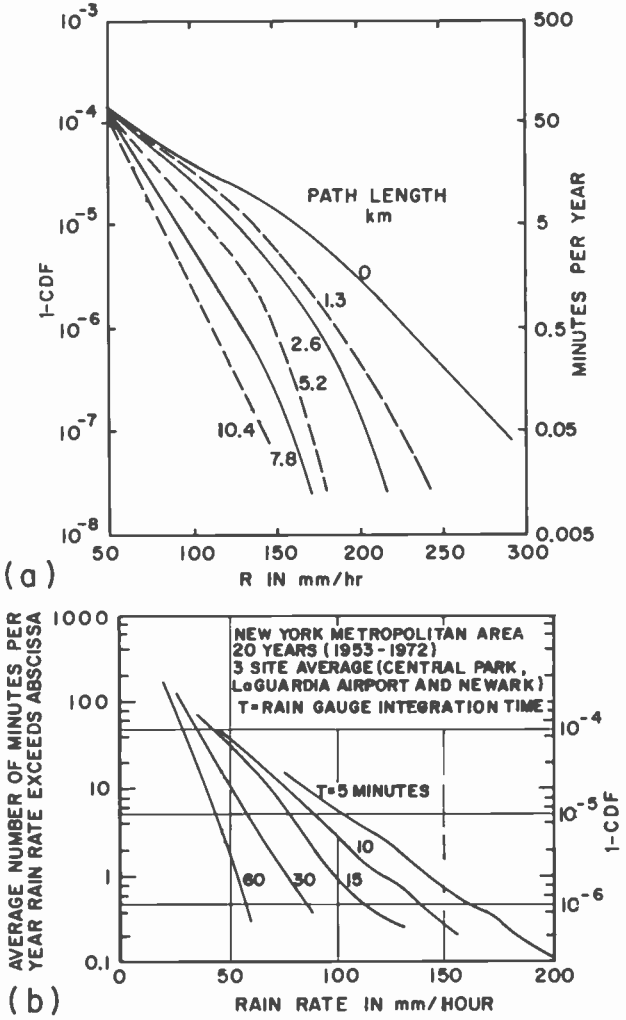


Fig. 13—(a) Cumulative distribution of instantaneous path-average rain-rates for paths of different lengths in New Jersey (Ref. [17]) and (b) cumulative distribution of point rain-rates in New York area for different integration intervals (Ref. [12]).

integrated over 60 minute intervals, corresponding to $v \simeq 0.8$ km/min. It is tempting to seek a connection between this velocity and that of convective air masses aloft (which travel at similar speeds) by assuming that rainfall is ergodic in time and space. However, as yet v has not been shown to depend significantly on path orientation relative to prevailing

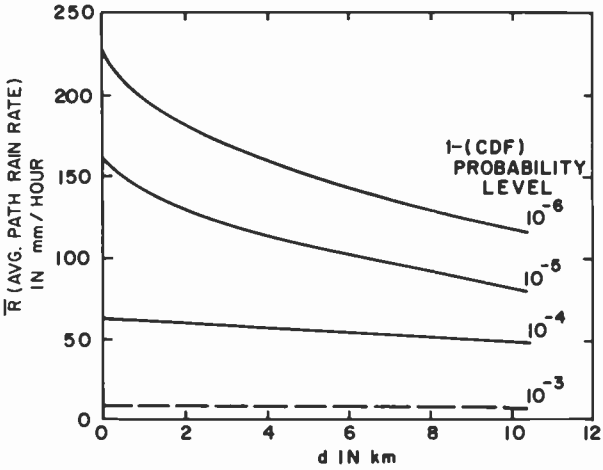


Fig. 14—Instantaneous path-average rain-rate $\bar{R}$ in New Jersey versus path-length for several cumulative distribution levels, from data shown in Fig. 13a (Ref. [25]).

wind direction or meteorological frontal movements, as would be expected if such a direct relation exists.

The curves of Fig. 13a can be redrawn to better show the path-average rain-rate dependence on path length for given probability levels, and this is shown in Fig. 14. From it, one can estimate the approximate size of rain cells, ranging from less than one km for 200 mm/hr rain-rates to several km for 50 mm/hr rain-rates. Direct measurements of the size of heavy showers by radar[19] are in good agreement with these numbers, as seen in Fig. 15. Rain cell sizes of a few km are also consistent with the earlier observation that the average duration of heavy showers is a few minutes, and that rain cells are apparently translated at velocities of about 50 km/hr.

5. Earth-Space Attenuation Statistics

We have discussed the three individual parameters (frequency, point rain statistics, and path length) that influence microwave propagation above 10 GHz, and can now look at their combined effect on an actual communications path. The single most important statistic of such a path is its attenuation CDF, which gives the fraction of total time during which the received signal strength can be expected to drop below some minimum useable level. If this fraction exceeds the service availability objectives, the system must be equipped to operate in some alternate mode (usually, some form of diversity) for the duration of excessive rainfall.

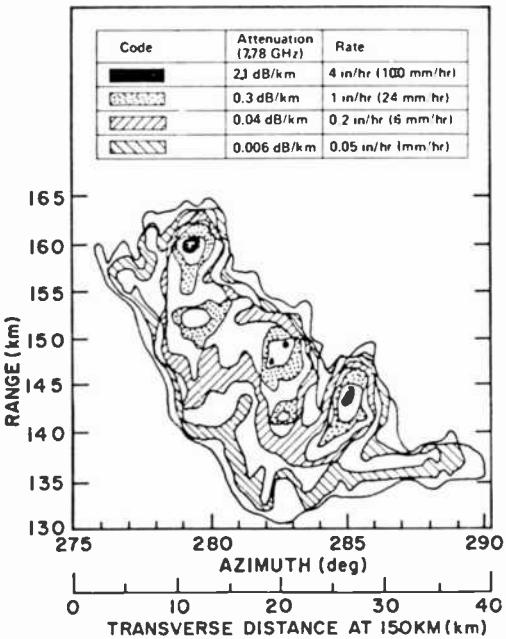


Fig. 15—Radar map of rainfall over New England showing typical rain-cell sizes for various rain-rates (Ref. [19]).

The actual alternate mode chosen depends, in part, on additional path statistics—the expected duration of excessive attenuation, the number of occurrences per year, the probability of simultaneous high attenuation on an alternate path, etc. Also, signal depolarization and interference via scattering from other sources in the same frequency band are important in many cases, but these can generally be expected to correlate strongly with path-attenuation.

After frequency, geographical location, and path length of a particular communications link have been selected, its path-attenuation CDF can be obtained using the graphs shown in Fig. 16. The figure is arranged to show the following relations:

(1) instantaneous point rain-rate CDF (first quadrant)
(2) path-attenuation CDF (second quadrant)
(3) path-attenuation versus path-integrated rain-rate (third quadrant)
(4) path-integrated rain-rate versus instantaneous point rain-rate (fourth quadrant).

To construct the path-attenuation CDF in the second quadrant, the appropriate curves for rain-rate CDF (based on geographical location,

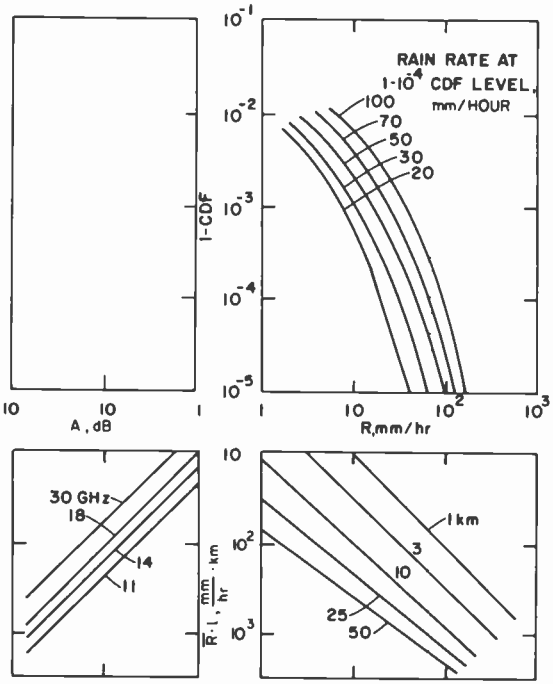


Fig. 16—Four-quadrant diagram for graphical construction of path-attenuation cumulative distribution as a function of geographical location (cf., Fig. 11), path-length (km), and frequency (GHz).

as discussed below), path length, and frequency are selected, and points from the CDF axis are successively reflected through these curves in the R, $\overline{Rl}$, A, and CDF directions. For instance, a communications link at a geographical location with index 50, a path length of 10 km, and a frequency of 14 GHz, has an attenuation exceeding 9.5 dB at the 10^{-3} probability level and 27 dB at the 10^{-4} probability level.

In the construction of Fig. 16, we have made the following assumptions:

(a) Rain-rates in the 10^{-3} to 10^{-5} probability range follow similar tendencies at all geographical locations (see the curves of Fig. 7) and can be characterized by a single index number (e.g., rain-rate at the 10^{-4} probability level). From a map such as Fig. 11, this index and hence the appropriate rain-rate curve in the first quadrant of Fig. 16 can be selected for the particular geographical location of interest. (Note again that Fig. 11 shows only rough regional tendencies, based on less than 30 measuring stations in the U.S., and that strong local variations can be expected to exist near high mountains, large bodies of water, etc.)

(b) Over a path of length l (km), the instantaneous path-average rain-rate CDF is approximately the same as the point rain-rate CDF obtained at a representative point along the path using an integration interval Δt (min), where $l/\Delta t \simeq 0.8$ km/min (see Eq. [3]). (Note again that this equivalence has been tested only for a few geographical locations, e.g., New Jersey and Ohio.) Also, it is assumed that the relation between point rain-rates of different integration intervals at a given probability level is everywhere the same as that measured for New Jersey (see Fig. 13b).

(c) The dependence of attenuation gradient on frequency and rain-rate is that given by Eq. [1]. Hence, the total path attenuation is

$$A(\text{dB}) = \int_0^l \alpha(x)dx = \left(\frac{f-5}{135}\right) \cdot \int_0^l R(x)dx = \left(\frac{f-5}{135}\right) \overline{R}l \quad [4]$$

where $\overline{R}$ = path-average rain-rate in mm/hr, l = total path length in km, and f = frequency in GHz. Note that this simple relation between total attenuation and path-average rain-rate is valid only over that frequency range for which the dependence of attenuation gradient on rain-rate (Eq. [1]) is approximately linear.

For terrestrial (horizontal) communications links, the path-attenuation CDF's predicted from Fig. 16 agree reasonably well with published observations. For earth-space links, however, additional uncertainties arise because their path traverses the less explored upper regions of the rain-carrying atmosphere. This affects both the path-length and the attenuation gradients:

(d) The slant path-length through rain of an earth-space link, for a given elevation angle, is directly proportional to the uppermost height at which rain is generated. This height varies from a few km for Nimbostratus clouds (light to medium rain-rates) to over 10 km for Cumulonimbus clouds (heavy rain-rates). Thus, it appears reasonable to assume a height of about 4 km for rain-rates under 10 mm/hr and about 7 km for rain-rates over 50 mm/hr. At 45° elevation, the corresponding path-lengths through rain can vary between about 6 and 10 km, and at 10° elevation between about 23 and 40 km.

(e) The relation between attenuation gradient, frequency, and rain-rate used in Fig. 16, (Eq. [1]) is strictly valid only for drop size distributions as observed near the ground. Since rain drops lose water by evaporation as they fall, their distribution tends towards larger sizes with increasing height. This would imply a higher attenuation gradient than given by Eq. [1] for the same rain-rate as measured on the ground. On the other hand, it appears well established[20] that rain generation occurs largely via supercooled (−10 to −20°C)

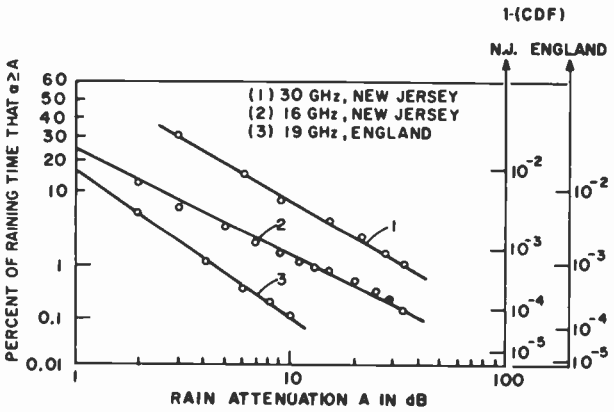


Fig. 17—Lognormal cumulative distribution (Eq. [5]) showing good fit to measured earth-space path-attenuation distributions (Ref. [13]).

droplets condensing on ice nuclei, which later melt as they fall through warmer air. At microwave frequencies, ice particles produce very little absorption (only scattering). It is therefore not clear how Eq. [1] must be modified for that portion of the propagation path that traverses the rain-generating region.

Measured path-attenuation distributions for earth-space links appear to be in reasonable agreement with those predicted from Fig. 16, assuming an upper height limit of about 7 km for heavy rain-rates. An example is given in Fig. 17 for two links operating in N.J. at an elevation angle of about 45° and frequencies of 16 and 30 GHz. These fit the predicted distribution from Fig. 16 if a slant path length of 10 km is used. (To convert the conditional probabilities of Fig. 17 to unconditional values, they must be multiplied by 5×10^{-2}, the probability of rainfall in N.J.)

Path-attenuations, like rain-rates, appear to be log-normally distributed, as seen from Fig. 17.[13] This is plausible because attenuation is a multiplicative process and a typical path contains a large number of independent, randomly varying segments of rain. A log-normal path-attenuation distribution obeys the relation

$$P(A \geq a) \simeq \frac{P_0}{2} \operatorname{erfc}\left[\frac{\ln(a) - \ln(A)_m}{\sqrt{2}\, S_A}\right], \qquad [5]$$

where

$P(A \geq a)$ = probability that path attenuation A(dB) exceeds a(dB),

P_0 = probability that rain is falling on path,
A_m = median value of path-attenuation during raining time,
S_A = standard deviation of ln (A) during raining time.

In addition to the total fraction of time a certain level of path-attenuation is exceeded, as given by its CDF, it is also important to know how long such fades (the usual term for temporary loss of signal) may be expected to persist, once they occur. Several measured distributions are shown in Fig. 18.[13] These indicate that fade duration, too, exhibits a log-normal distribution of the form

$$P[T(A) \geq t] \simeq \frac{1}{2}\,\mathrm{erfc}\left[\frac{\ln(t) - \ln[T_m(A)]}{\sqrt{2}\,S_T}\right], \qquad [6]$$

where

$P[T(A) \geq t]$ = probability that fade duration $T(A)$ exceeds t (min.),
$T_m(A)$ = median value of fade duration,
S_T = standard deviation of ln $[T(A)]$.

From the data of Fig. 18, $T_m(A) \sim 1.0$ to 2.5 minutes and $S_T \sim 1.0$ to 1.75. For a log-normal distribution, the mean value is given by $\overline{T}(A) = T_m(A)\exp(S_T/2)$, or $\overline{T}(A) \sim 2.4$ to 11.4 minutes in Fig. 18. The shorter durations correspond to heavy rain-rates (>20 mm/hr) and the longer to lighter rain-rates (<10 mm/hr), as inferred from fade-threshold levels and other information provided. These mean durations agree with the earlier observation that, on the average, heavy showers last only a few minutes (Fig. 12). For the data shown in Fig. 18, 10% of the fades lasted longer than about 15 minutes, 1% longer than about 80 minutes, and 0.1% longer than about 200 minutes.

Attenuation can change only as rapidly as rain can move into, or out of, a substantial segment of the propagation path. Thus, the time required for a small cell (0.5 km diameter, say) of intense rain to traverse the path (at 1.0 km/min, say) is about 30 seconds. While not much shorter than the median fade durations given in Fig. 18 (1 to 2.5 minutes), this nevertheless indicates that the onset of fades can be expected to be quite abrupt on occasion.

For certain applications, the number of fades in a given time (e.g., 1 year) may be more important than the distribution of their duration. By combining the information of Figs. 17 and 18, the number of occurrences per year for which the path-average rain-rate exceeds a given amount can be estimated for New Jersey. The result is plotted in Fig.

—Experiments of Distribution of Fade Duration

Location	Frequency (GHz)	Length Length (km)	σ_T	$\bar{\tau}$ (A) (min)	Fade Threshold (A dB)	Time Base
Palmetto, Georgia	18	5	0.44	2.4	36	November 1970-June 1971
Crawford Hill, NJ	16	Earth-Space	0.76	8	9	Dec. 2, 1967-Feb. 28, 1969
Crawford Hill, NJ	30	Earth-Space	0.76	11.4	9	Dec. 2, 1967-Feb. 28, 1969
Crawford Hill, NJ	30	Earth-Space	0.67	7.7	21	Dec. 2, 1967-Feb. 28, 1969
England	19	Earth-Space	0.53	4.4	5	June 6, 1968-May 31, 1970
Crawford Hill, NJ	18.5	6.4	0.65	7.2	5	June 23, 1967-Oct. 31, 1967
Crawford Hill, NJ	18.5	6.4	0.57	5.8	10	June 23, 1967-Oct. 31, 1967
Crawford Hill, NJ	18.5	6.4	0.63	5.1	20	June 23, 1967-Oct. 31, 1967
Italy	11	20	0.75	4.8	10	May-Oct. 1967

$\bar{\tau}$ (A) = Average fade duration, min

σ_T = Std. deviation of $\log_{10}(T) = 0.4343\, S_T$ (Eq. VI)

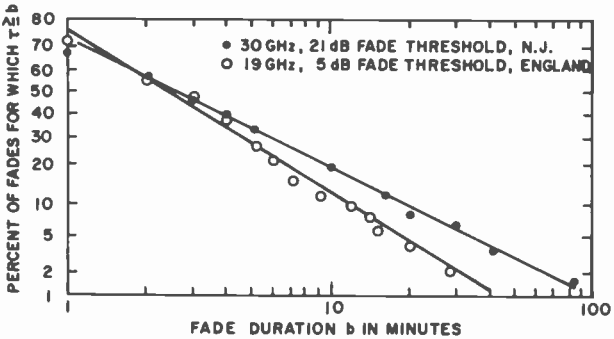


Fig. 18—Table of measured fade-duration distributions (top), and lognormal cumulative distribution (Eq. [6]) showing good fit to measured data (bottom) (Ref. [13]).

19. We note that path-average rain-rates of a few mm/hr are exceeded several hundred times per year, whereas rain-rates of 50 mm/hr are exceeded only about 10 times per year.

When excessive attenuation occurs on a given path, what is the probability that the same attenuation is also (simultaneously) exceeded on another, nearby path? Clearly, this depends on path separation and rain-rate. For path separations much less than the diameter of an average rain-cell, strong correlation between the two path attenuations is expected and the joint CDF for both is only slightly below that of either one. On the other hand, when the paths are separated by a much greater distance than the average rain-cell diameter, the two paths attenuations are essentially uncorrelated and the joint CDF would be the product of the individual CDF's. Since average rain-cell diameters are one km or less for heavy rain-rates and 10 km or more for light rain-rates, most practical cases will involve path separations between these extreme conditions and show partial correlation of their attenuations. An example

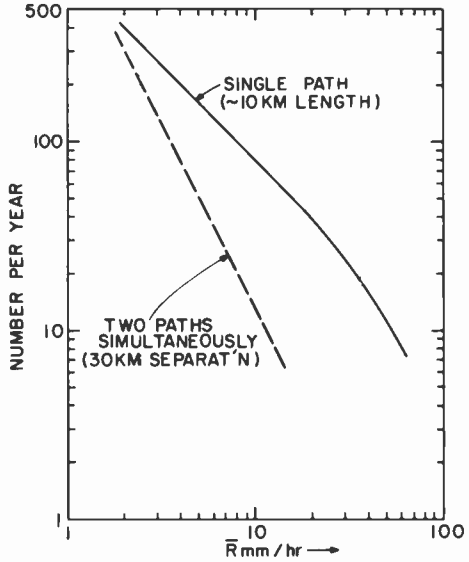


Fig. 19—No. of occurrences per year for which earth-space path-average rain-rate exceeds abscissa in New Jersey for single path (estimated from Figs. 17 and 18) and for two parallel paths simultaneously (Ref. [21]).

is shown in Fig. 20. As expected, the joint CDF's are only slightly less than the individual CDF's at the 10^{-2} probability level (corresponding to rain-rates of a few mm/hr) but are 10 to 100 times less when the individual CDF's are at the 10^{-3} probability level (corresponding to rain-rates of 20 mm/hr or more). Also, as expected, the largest differences between joint and individual CDF's occur for the largest path separations.

A detailed study of joint probabilities of simultaneous heavy rain-rates on two parallel paths[17] indicates that the joint probability decreases steadily with increasing path separations up to about 9 km, as illustrated in Fig. 21. In the range of 10 km to about 15 km path separation, the joint probability rises slightly and then decreases again for larger separations. Although the effect is small, it indicates that for path separations below about 20 km, an optimum choice is near 9 km.

As one would expect, the number of cases per year during which both paths suffer from simultaneous heavy rainfall is less than that for one path alone. Data obtained in New Jersey for two paths separated by 30 km[21] are shown in Fig. 19 and can be compared with the single path data. For very low rain-rates, the number of occurrences are nearly the same, but above 10 mm/hr the number of simultaneous occurrences is much lower than those on either path alone.

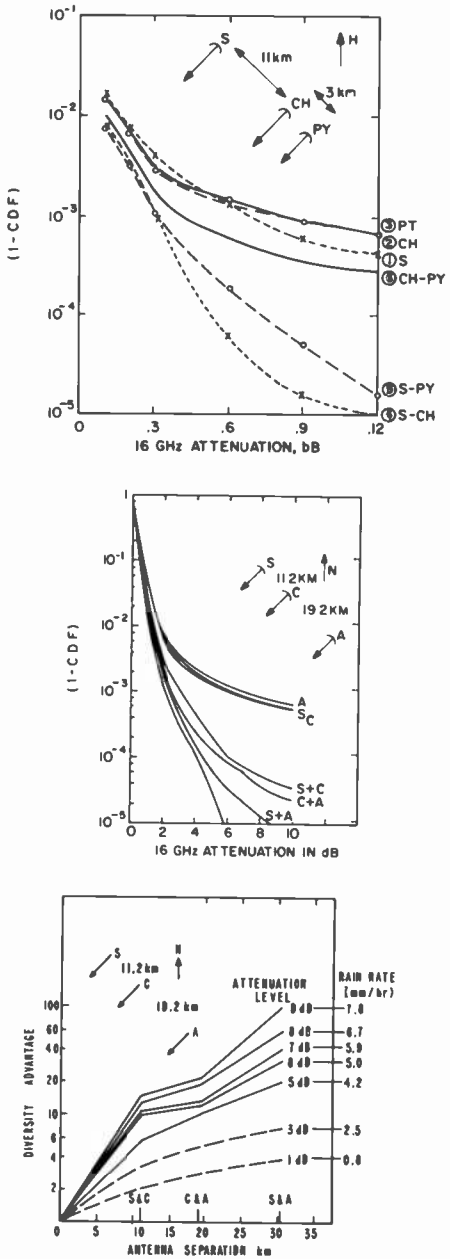


Fig. 20—Cumulative distribution of earth-space path-attenuation in New Jersey for single path and for two or three paths simultaneously. Bottom diagram shows dual-path diversity advantage versus path separation for different rain-rates (Ref. [24]).

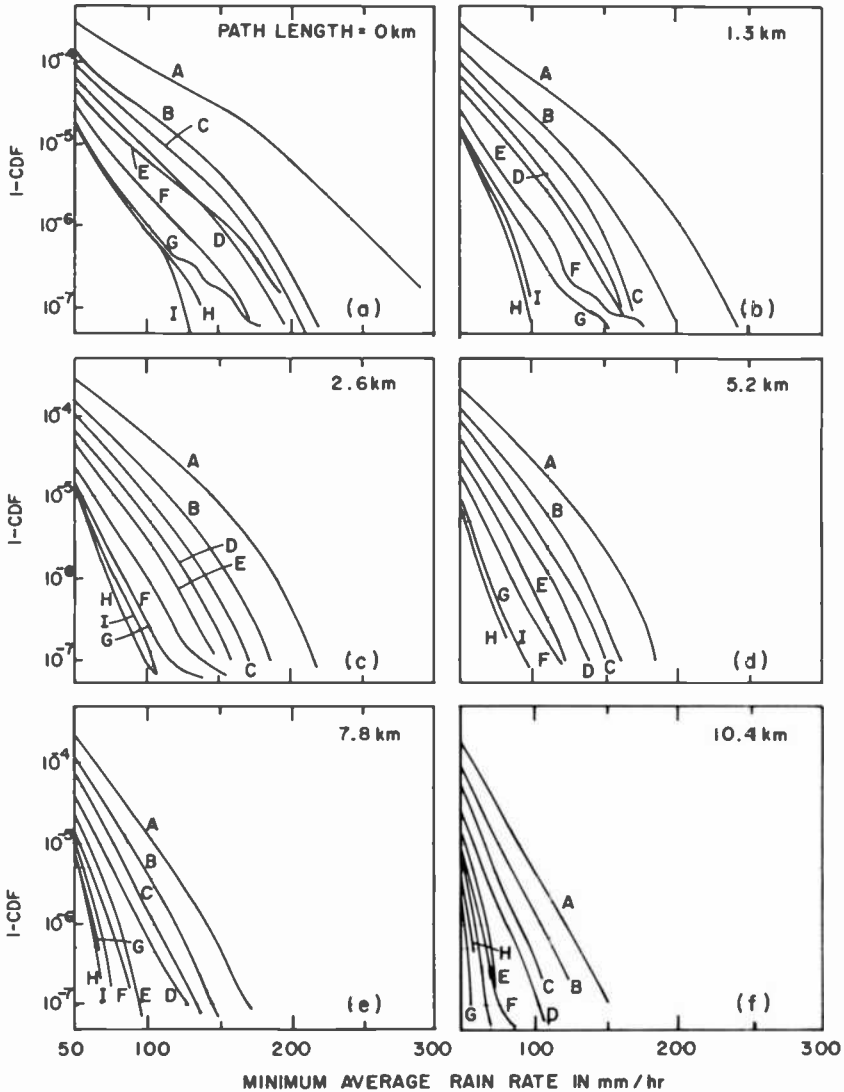


Fig. 21—Cumulative distribution of path-average rain-rates for both of two parallel paths as a function of path-lengths (a to f) and path separation (curves A to L) (Ref. [17])

6. System Diversity for Increased Service Availability

Using Fig. 16, the path-attenuation CDF's have been plotted for typical cases of interest (Fig. 22). The main effect of path-attenuation is a cor-

responding decrease in received signal strength; a secondary effect in down-links is increased system noise temperature. If we neglect the latter, the curves of Fig. 22 directly show the fraction of total time the signal power can be expected to fade below its minimum acceptable level versus the system's built-in clear-weather margin. For instance, a 3 dB margin will be exceeded about 1.3×10^{-3} of the time for a 11 GHz link with 10-km path length in a low rain-rate region, and about 7×10^{-2} of the time for a 30 GHz system with a 50-km path length in a high rain-rate region. These numbers may be adequate for certain applications, but they fall short of typical telecommunications objectives, which require

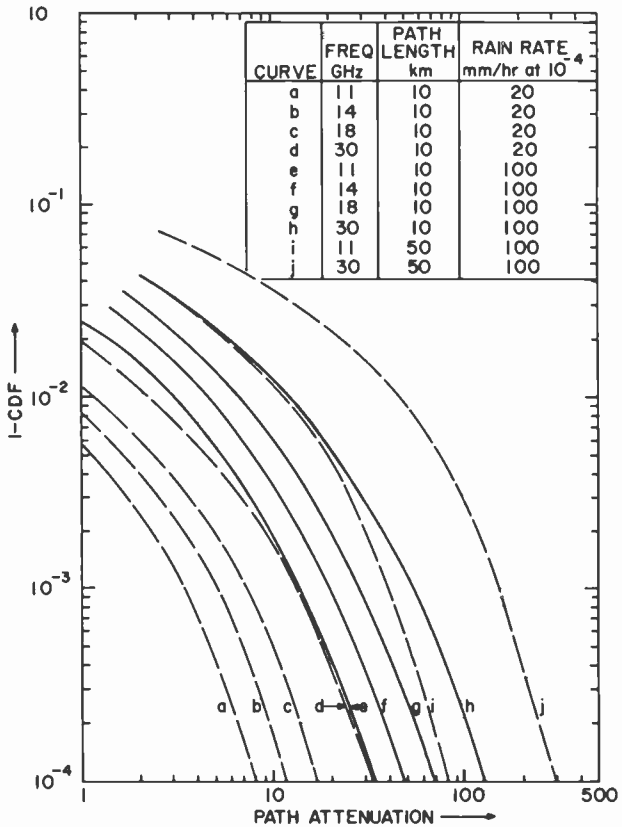


Fig. 22—Cumulative distribution of path-attenuation (dB) obtained from Fig. 16 for several combinations of frequency, rain-rate distribution, and path-length. A 10-km path-length corresponds to an earth-space antenna elevation angle of about 45° and a 50-km path-length to an elevation angle of about 8°, assuming rain generation at 7-km altitude.

that service be available for all except (at most) 10^{-4} of the total time. In general, it is impractical to meet such an objective by increased clear-weather margins (except perhaps for 11/14 GHz links in the lowest rain-rate regions). Rather, we must seek to increase service availability by switching the system into some temporary alternate mode of operation for the duration of excessive path-attenuation, i.e., we must provide for some form of diversity. We do not consider here, of course, temporary use of a non-satellite system, such as terrestrial microwave radio or millimeter waveguide.

In satellite communications, diversity can be provided by designing flexibility into one or more of the following systems parameters:

1. Transmitter Power
2. Microwave Frequency
3. Bandwidth
4. Transmission Delay (temporary storage)
5. Earth-Space Path
6. Spacecraft Antenna Gain

We will briefly discuss these diversity methods in turn to determine major advantages or limitations, and then attempt a comparison of their relative merits. For simplicity, we shall assume a 3-dB clear-weather margin and a service availability objective of 99.99% ($1-10^{-4}$) in all cases.

6.1 Transmitter Power Diversity

A few dB of rain-induced attenuation (beyond the clear-weather margin) can be readily overcome by a temporary increase of earth station transmitter power. However, when the attenuation exceeds about 10 dB, this method becomes progressively less attractive, since transmitters with very high peak power capability would be required. Another drawback, particularly for down-link applications, is the additional satellite transmitter energy required. For the low rain-rate cases of Fig. 22 (curves a to d), the peak power (dB above 3 dB clear-weather margin) and additional transmitter energy (dB above free-space case) required to reach a 99.99% service availability objective are (approximately) as shown in Table 2.

We must distinguish between the uplink and downlink cases. For uplinks, the transmitter power at each ground station need only be increased to compensate for its own path attenuation. For downlinks, the satellite transmitter must provide enough power to overcome the highest attenuation observed by any one of the ground stations it serves simultaneously. Thus, a satellite transmitter serving a very large number of (statistically independent) ground stations must operate near peak

Table 2—Peak Power and Additional Transmitter Energy Required in Power Diversity Systems

Fig. 22 curve	Peak Power (dB)	Additional Transmitter Energy (dB)			
		Uplink, per station	Downlink, 10 stations	Downlink, 100 stations	Downlink, ∞ stations
a	5	.003	.03	~.3	5
b	9	.01	.1	~1	9
c	14	.03	.3	~2	14
d	30	.4	3	~10	30

power all the time, requiring a correspondingly higher amount of energy.

6.2 Microwave Frequency Diversity

In this method, an earth-space link is switched to a lower microwave frequency at which rain-attenuation is negligible (e.g., 4/6 GHz) whenever its clear-weather margin is exceeded. Accordingly, ground stations and satellites using this method must be equipped for dual-frequency operation.

For a single-link-per-transponder system, the number of low-frequency transponders required to meet a given service availability objective is readily obtained. Let

p = probability of rain attenuation exceeding clear-weather margin for any one link

M = number of (statistically independent) links served simultaneously by satellite (one link per transponder)

N = number of low-frequency transponders in satellite.

The service availability (S.A.) is then (see Appendix):

$$\text{S.A.} = (1-p) + \sum_{n=1}^{N} \frac{(1-p)^{M-n} p^n (M-1)!}{(M-n)!\,(n-1)!} \qquad [7]$$

With S.A. = 99.99% as the objective, Eq. [7] was solved to give the required N versus M, with p as a parameter. The result is shown in Fig. 23. We note that for the cases illustrated in Fig. 22 ($p = 1.3 \times 10^{-3}$ to 7×10^{-2}), relatively few low-frequency transponders are required.

For systems in which several ground stations want to receive the same satellite transponder signal (as in broadcasting), the transponder must be switched to low frequency operation whenever the clear-weather margin is exceeded at one (or more) ground station. With:

m = number of (statistically independent) ground stations served simultaneously

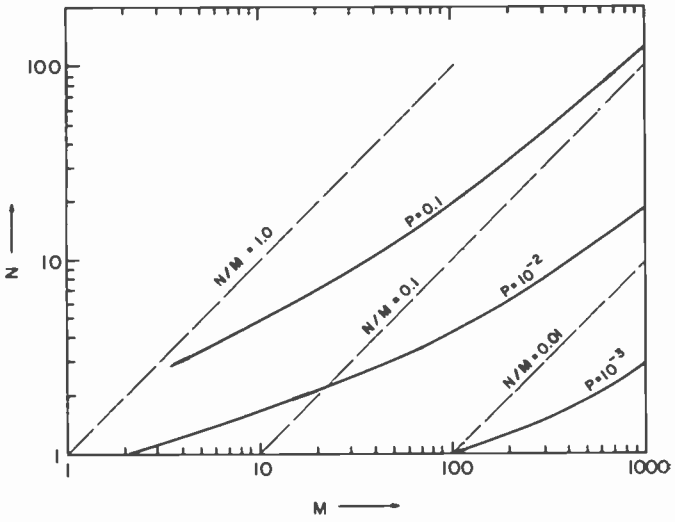


Fig. 23—Number of low-frequency transponders (N) required to yield 99.99% service availability in frequency-diversity system serving M independent earth-space links, for several probability levels (P) of the clear-weather margin being exceeded for any one link (Eq. [7]).

p = probability that clear-weather margin is exceeded at any one ground station

the probability that the transponder transmitter must be switched to low frequency is

$$P_m = 1 - (1 - p)^m \qquad [8]$$

In Fig. 24, P_m is plotted versus m for various p. We note that $P_m = 0.1$ for $m = 100$ when $p \simeq 10^{-3}$ (e.g., low rain-rate region, 10 km path-length), and for $m = 10$ when $p \simeq 10^{-2}$ (e.g., high rain-rate region, 10 km path length). For m or p much larger than given in these examples, low-frequency operation might as well be used continuously.

6.3 Bandwidth Diversity

In principle, bandwidth can be traded for signal power in accordance with the well-known channel capacity relation

$$C = B \log_2 \left(\frac{S + Q}{Q} \right) \text{ bits/sec}, \qquad [9]$$

where C is the channel capacity, B the bandwidth in Hz, S the signal

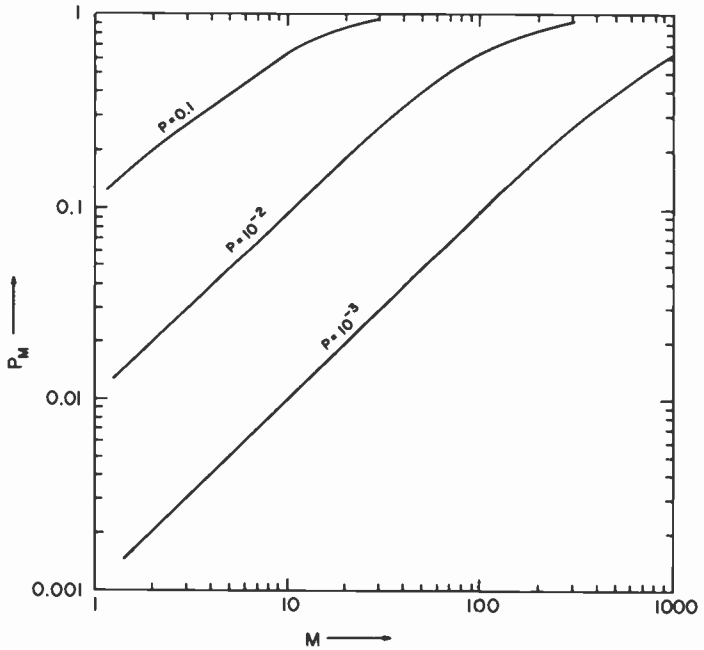


Fig. 24—Probability (P_m) that the clear-weather margin is exceeded for at least one of m (statistically independent) earth-space links for several probability levels (P) of the clear-weather margin being exceeded for any one link (Eq. [8]).

power, and Q is the noise power. For $S \gg Q$, a small percentage increase in bandwidth can effectively compensate for a much larger percentage decrease in signal power. But when $S \ll Q$, the tradeoff is very effective and the method becomes wasteful of the frequency spectrum.

In FM systems (with $S/Q \simeq 10$), wideband modulation has been used (in combination with frequency compression demodulation) to improve threshold performance,[22,23] but the decrease in required rf carrier power for a fixed output signal-to-noise ratio is about 6 dB at best. Bandwidth diversity (like transmitter power diversity) is therefore limited to the lower frequencies (11/14 GHz) and very low rain-rate regions.

6.4 Transmission Delay Diversity

When real-time operation is not mandatory (as in many data communications applications), service can be maintained by temporarily storing the data during the rain-induced fade and transmitting them after the fade has ended. The service availability for this case is

$$\text{S.A.} = 1 - P[T(A) \geq t] \cdot P(A \geq 3\,\text{dB}) \qquad [10]$$

where

$P(A \geq 3\,\text{dB})$ = probability that signal fade exceeds clear-weather margin of 3 dB (Eq. [5] or Fig. 22),

$P[T(A) \geq t]$ = probability that fade duration exceeds t minutes (Eq. 6 or Fig. 18).

For S.A. = 99.99% and the cases illustrated in Fig. 22, the duration t_{max} for which data storage is required can be estimated from Fig. 18, and are shown in Table 3.

We note that the required temporary storage capability ranges from less than one hour for the 11/14 GHz bands with path lengths of 10 km in low-rain-rate regions to several hours for path-lengths of 50 km in high rain-rate regions. The last column in Table 3 shows the approximate number of occasions per year for which temporary storage becomes necessary (in most cases, of much shorter duration than t_{max}).

Table 3—Temporary Storage Requirements in Delay Diversity Systems

Fig. 22 curve	$P(A \geq 3\,\text{dB})$	$P[T(A) \geq t)]$ for S.A. = 99.99%	t_{max} (from Fig. 18)	Occurrences/yr (from Fig. 19)
a	1.3×10^{-3}	7.6×10^{-2}	30 min	100
b	2.4×10^{-3}	4.2×10^{-2}	45 min	200
c	3.8×10^{-3}	2.6×10^{-2}	58 min	300
d	7.5×10^{-3}	1.3×10^{-2}	1.2 hrs	500
e	1.0×10^{-2}	1.0×10^{-2}	1.3 hrs	100
f	1.6×10^{-2}	6.3×10^{-3}	1.8 hrs	200
g	2.3×10^{-2}	4.3×10^{-3}	2.3 hrs	300
h	3.3×10^{-2}	3.0×10^{-3}	3.0 hrs	500
i	3.3×10^{-2}	3.0×10^{-3}	3.0 hrs	500
j	7.0×10^{-2}	1.4×10^{-3}	4.5 hrs	1000

6.5 Earth-Space Path Diversity

Of the various diversity methods, this one has probably received the greatest attention in the past.[21,24] Path diversity is obtained by providing for duplicate microwave facilities at some distance from the primary earth-station site, so that the two resulting earth-space paths show little correlation in their attenuation statistics.

Unfortunately, the improvement in service availability that can be obtained in this way is rather limited for reasonable path separations (<50 km) and the low path-average rain-rates associated with a 3-dB clear-weather margin (about 1 to 8 mm/hr for the cases of Fig. 22). Thus, assuming that the statistics obtained for New Jersey (Fig. 20) are valid at other locations, a service availability of 99.99% can be reached only

for 11/14 GHz systems with short paths (10 km) in low rain-rate regions.

Some further improvement is possible by using very large path separations (>100 km). However, the limit is reached when the two path-statistics become totally uncorrelated. The joint probability (of the clear-weather margin being exceeded on both paths) then becomes the product of the individual probabilities (i.e., their square for paths with identical statistics). This means that dual path diversity can provide a 99.99% service availability only if the individual CDF's are less than 10^{-2} at the clear-weather margin. Thus, for the assumed 3-dB margin, all systems above 10 GHz in high rain-rate regions or with paths much longer than 10 km are excluded. In these cases, the desired service availability can be achieved only by using triple (or even higher) path diversity configurations.

The main drawback of path diversity is the large expense of duplicate (or triplicate) microwave facilities needed for each ground station, and of the wideband ground transmission channels required to span the relatively large distances between these facilities.

6.6 Spacecraft Antenna-Gain Diversity

The spatial and temporal rain-induced attenuation variations in the earth's atmosphere could, in principle, be completely equalized by an adaptive spacecraft antenna. To do this, the angular antenna gain pattern on the satellite would have to mirror the existing attenuation pattern on the ground and change in synchronism with it.

In practice, of course, the spatial resolution of adaptive antennas is rather limited. Thus, at a given frequency, the required antenna area attached to the satellite increases more or less in direct proportion to the number of spatial resolution elements on the ground. An antenna capable of addressing individual 250 by 250 km regions would have about a ten times larger diameter than one that provides uniform coverage of the (contiguous) U.S. Similarly, the dynamic range of antenna gain (or EIRP, in the transmitting case) that can be achieved in any one angular direction is proportional to the overall antenna size. For the above case of 250 by 250 km resolution, a hundred-fold increase of antenna gain could be achieved, i.e., up to 20 dB of regional rain attenuation could be compensated on a dynamic basis.

To follow the temporal attenuation variations on the ground, the satellite antenna must be capable of directing movable spot beams toward the affected regions, either by physical motion of its structure or by electronically switching receivers (or transmitters) between a multiplicity of stationary feeds. Ideally, this system would operate by

continuously monitoring the received signal strength from each ground station (assumed to be spaced from its nearest neighbor by at least one resolution element) and adjusting the complex satellite antenna pattern instantaneously so as to maintain constant signal strength for all stations.

From the earlier example, it is clear that only relatively coarse regional variations can be compensated by an adaptive satellite antenna. Moreover, a 20-dB range in path-loss compensation would limit this method to low rain-rate regions, assuming a 99.99% service-availability objective. In an actual design, the achievable compensation range would be set primarily by constraints on spacecraft weight and electronics complexity.

6.7 Comparison of Diversity Methods

From the foregoing discussion it should be clear that no single diversity method is "best" for all conceivable applications. The choice can, in general, be made only after other parameters are known. For instance, systems containing only a few ground stations will tend to favor those methods that can be designed into each ground station separately (such as power and path diversities). Systems containing many ground stations will tend to favor those methods that can take advantage of the fact that only a small fraction of stations suffer high attenuation simultaneously (such as low-frequency, bandwidth, and spacecraft antenna diversities). Systems in which an information delay of up to several hours is permissible should find temporary storage advantageous.

Also, we saw that it is difficult to meet 99.99% service availability objective by a single diversity method, except for low rain-rate regions and 11/14 GHz frequencies. Table 4 compares the service availability (listed as $[1 - \text{S.A.}] \times 10^{-4}$) that can be expected in practice for the various methods discussed, assuming a clear weather margin of 3 dB. Frequency and delay diversities have been excluded, since the level of

Table 4—Service Availability for Various Diversity Methods

Diversity Method	Attn. Range (dB)	$(1 - \text{S.A.}) \times 10^{-4}$; for Fig. 22 curves:									
		a	b	c	d	e	f	g	h	i	j
Nothing	3	13	24	37	75	100	160	230	330	330	670
Power	10	1	1	2	10	11	22	40	90	85	300
Bandwidth	6	1	2	6	17	18	34	60	120	110	360
Dual Path*	6–18	1	1	4	20	2	8	23	80	80	300
Satellite Antenna Gain	20	1	1	1	3	3	7	15	38	30	190

* Attenuation range varies from ~18 dB @ 11 GHz to ~6 dB @ 30 GHz for 30-km path separations (see Fig. 20).

service availability depends on other constraints in these cases (e.g., number of low-frequency channels or amount of temporary storage available). The "attenuation range" column lists the maximum signal level improvement considered feasible for each method, as previously discussed. We note that for the higher frequencies (18/30 GHz) in high rain-rate regions, service availabilities of only 99% or less (100 or higher in Table 4) may be feasible.

In some cases a combination of several diversity methods may prove advantageous. This is particularly true when one of the methods is path diversity, since it provides progressively larger improvements at the higher rain-attenuation levels (see Fig. 20). For instance, the low attenuation levels could be compensated by transmitter power diversity and the higher levels by path diversity. Table 5 shows the amount of additional transmitter power (above the 3-dB clear-weather margin) that would be required to reach (in conjunction with path diversity) a 99.99% service availability objective.

We note that for all cases of Fig. 22 except curve j (30 GHz, high rain-rate region, 50 km path), the combination of power and path diversity could readily provide the desired service availability.

Table 5—Peak Power Required in Power-Plus-Path-Diversity System

Fig. 22 curve	a	b	c	d	e	f	g	h	i	(j)
Power Diversity (dB)	0	0	1	5	0.5	2.5	5	13	10	(52)

7. Conclusions

The results of any study are no better than the assumptions that went into it. As long as our knowledge about rainfall remains fragmentary, all predictions about earth-space attenuation and its impact on future satellite communications systems are bound to contain large uncertainty factors. It is therefore important that a more generally valid theory of rain-attenuation be developed, particularly since cost and time constraints prevent us from waiting until a complete set of measured data becomes available for every potential earth-station site.

Ideally, a complete theory of rain-induced attenuation should include:

- Description of point statistics by a minimum number of parameters,
- Determination of overall distributions from extreme value observations,

- Interrelation between spatial and temporal statistics,
- Dependence of path statistics on path orientation,
- Characterization of rain-attenuation aloft from measurements on the ground.

Additional measurements are needed to check the validity of extrapolating results from one geographical location to others. There is a relative abundance of rainfall and attenuation data for the eastern U.S., but a lack of data for the low rain-rate regions in the western U.S. Also, most existing earth-space attenuation data were obtained at high elevation angles, whereas typical locations in Alaska involve low elevation angles, corresponding to path lengths through rainfall of up to 50 km.

Finally, the various diversity methods must be evaluated for specific system configurations to determine their relative cost-effectiveness in achieving a desired service availability objective.

Appendix—Service Availability in Frequency Diversity Systems

Assume a satellite with M transponders of which N can operate at 4/6 GHz and the other $M - N$ only above 10 GHz. The satellite serves M earth-space links whose rain statistics are assumed identical but uncorrelated. The service availability (S.A.) of the system is then given by one minus the probability that all 4/6 GHz transponders are already in use when the rain-attenuation of a particular link exceeds the clear-weather margin, i.e.,

$$\text{S.A.} = 1 - pP, \tag{11}$$

where p is the probability of rain attenuation exceeding clear-weather margin for any one link and P is the probability of rain attenuation exceeding clear-weather margin for N or more links simultaneously in a system with $M - 1$ links.

The probability that no link out of $M - 1$ suffers excessive attenuation is

$$P_0 = (1 - p)^{M-1}. \tag{12}$$

The probability that exactly m links out of $M - 1$ suffer excessive attenuation simultaneously is

$$P_m = \frac{p^m(1 - p)^{M-m-1}(M - 1)!}{(M - m - 1)!m!} \tag{13}$$

Hence, the probability that N or more links out of $M - 1$ suffer excessive attenuation is

$$P = 1 - P_0 - P_1 - P_2 \ldots - P_{N-1}$$

$$= 1 - \sum_{m=0}^{N-1} \frac{p^m (1-p)^{M-m-1} (M-1)!}{(M-m-1)!\, m!} \qquad [14]$$

or, with $n = m + 1$,

$$P = 1 - \sum_{n=1}^{N} \frac{p^{n-1} (1-p)^{M-n} (M-1)!}{(M-n)!\, (n-1)!}.$$

Substitution into Eq. [11] gives the final result:

$$\text{S.A.} = (1-p) + \sum_{n=1}^{N} \frac{p^n (1-p)^{M-n} (M-1)!}{(M-n)!\, (n-1)!}. \qquad [15]$$

References:

[1] D. E. Kerr, "Propagation of Short Radio Waves," Lexington, Mass., Boston Tech. (1964), p. 679–681.

[2] G. Mie, Annalen der Physik (4), Vol. 25 (1908), p. 377 ff.

[3] P. S. Ray, "Broadband Complex Refractive Indices of Ice and Water," *Appl. Optics*, **11**, No. 8, (1972), pp. 1836–1844.

[4] D. E. Setzer, "Computed Transmission Through Rain at Microwave and Visible Frequencies," *Bell Syst. Tech. J.*, **49** (1970), pp. 1873–1892.

[5] J. A. Morrison and T. S. Chu, "Perturbation Calculations of Rain-Induced Differential Attenuation and Differential Phase Shift at Microwave Frequencies," *Bell Syst. Tech. J.*, **52**, No. 10 (1973), pp. 1907–1913.

[6] H. R. Pruppacher and R. L. Pitter, "A Semi-Empirical Determination of the Shape of Cloud and Rain Drops," *J. Atmos. Science*, **28** (1971), pp. 86–94.

[7] T. S. Chu, "Rain-Induced Cross Polarization at Centimeter and Millimeter Wavelengths," *Bell Syst. Tech. J.*, **53**, No. 8 (1974), pp. 1557–1579.

[8] M. J. Saunders, "Cross Polarization at 18 and 30 GHz due to Rain," *IEEE Trans. Ant. and Prop.*, **AP-19** (1971), pp. 273–277.

[9] L. T. Gusler and D. C. Hogg, "Some Calculations on Coupling Between Satellite Communications and Terrestrial Radio-Relay Systems," *Bell Syst. Tech. J.*, **49** (1970), pp. 1491–1512.

[10] U.S. Department of Commerce, NOAA, Environmental Data Service, National Climatic Center, Federal Building, Asheville, N.C. 28801.

[11] W. F. Bodtmann and C. L. Ruthroff, "Rain Attenuation on Short Radio Paths: Theory, Experiment, and Design," *Bell Syst. Tech. J.*, **53**, No. 7 (1974), pp. 1329–1349.

[12] S. H. Lin, "Dependence of Rain-Rate Distribution on Rain-Gauge Integration Time," *Bell Syst. Tech. J.*, **55**, No. 1 (1976), pp. 135–141.

[13] S. H. Lin, "Statistical Behaviour of Rain Attenuation," *Bell Syst. Tech. J.*, **52**, No. 4 (1973), pp. 557–581.

[14] S. H. Lin, "A Method for Calculating Rain Attenuation Distributions on Microwave Paths," *Bell Syst. Tech. J.*, **54**, No. 6 (1975), pp. 1051–1087.

[15] S. H. Lin, "Rain-Rate Distributions and Extreme Value Statistics," *Bell Syst. Tech. J.*, **55**, No. 8 (1976), pp. 1111–1124.

[16] *Rainfall Frequency of the United States*, Weather Bureau Technical Paper No. 40, and *Precipitation Frequency Atlas of the Western U.S.*, NOAA Atlas No. 2, Vols. I to XI. Available from Superintendent of Documents, Government Printing Office, Wash., D.C. 20402.

[17] Ms. A. E. Freeny and J. D. Gabbe, "A Statistical Description of Intense Rainfall," *Bell Syst. Tech. J.*, **48** (1969), pp. 1789–1851.

[18] H. E. Bussey, "Microwave Attenuation Statistics Estimated from Rainfall and Water Vapor Statistics," *Proc. IRE*, **38** (1950), pp. 781–785.

[19] R. K. Crane, "Propagation Phenomena Affecting Satellite Communications Systems Operating in the Centimeter and Millimeter Wavelength Bands," *Proc. IEEE*, **59**, No. 2 (1971), pp. 173–188.

[20] W. Findeisen, *Meteor. Ztschr.*, **55** (1938), p. 121.

[21] D. Jarett and L. D. Spilman, "Impact of Rain Attenuation on 18/30 GHz Satellite Systems," paper 74-496, AIAA, 5th Communications Satellite Systems Conf., Los Angeles, Cal., April 22–24, 1974.

[22] C. L. Ruthroff, "FM Demodulators with Negative Feedback," *Bell Syst. Tech. J.*, **40** (1961), pp. 1149–1156.

[23] A. J. Giger and J. G. Chaffee, "The FM Demodulator with Negative Feedback," *Bell Syst. Tech. J.*, **42** (1963), pp. 1109–1135.

[24] D. C. Hogg anc T. S. Chu, "The Role of Rain in Satellite Communications," *Proc. IEEE*, **63**, No. 9 (1975), pp. 1308–1331.

[25] D. C. Hogg, "Statistics on Attenuation of Microwaves by Intense Rain," *Bell Syst. Tech. J.*, **48** (1969), pp. 2949–2962.

Correction Notice

In the paper *High-Power Low-Loss PIN Diodes for Phased-Array Radar*, by A. Rosen et al., which appears in the March 1979 issue of *RCA Review*, Eq. [2] on p. 24 is incorrect. It should read as follows:

$$Q = \frac{1}{\omega C_0[r_s + (\rho_C/A) + R_s]} + \left(1 + \frac{C_1}{C_0}\right)\omega\tau_D.$$

Patents Issued to RCA Inventors First Quarter 1979

January

A. A. Ahmed Retrace Blanking Pulse Generator with Delayed Transition (4,134,046)
A. Anchutin Plural Panels Deployed Effectively as a Single Panel (4,133,502)
A. R. Balaban and S. A. Steckler Gating Signal Generator for Switched Pincushion Correction Circuit (4,132,927)
H. W. Becke Gate Turn-off Thyristor with Anode Rectifying Contact to Non-Regenerative Section (4,137,545)
A. Bloom and L. K. Hung Cholesteryl Carbonates and Carbamates of Azo Dyes (4,134,888)
T. V. Bolger Direct Digital Frequency Synthesizer (4,134,072)
C. J. Busanovich and R. M. Moore Polycrystalline Selenium Imaging Devices (4,132,918)
R. H. Chan, M. R. Mann, and F. M. McDonnell Dynamic Channel Allocation Buffer Matrix (4,136,399)
T. L. Credelle, W. J. Hannan, and F. W. Spong Broadening the Spatial Frequency Pass Band of a Thermoplastic Layer (4,137,077)
T. L. Credelle Color Flat Panel Television (4,137,478)
W. F. Dietz Horizontal Deflection System with Boosted B Plus (29,885)
A. G. Dingwall Insulated-Gate Field-Effect Transistor Having a Deep Channel Portion More Highly Doped than the Substrate (4,132,998)
R. E. Fernsler Ramp Generator for Harmonic Tuned Deflection (4,134,082)
R. E. Flory and C. B. Oakley Raster Registration System for a Television Camera (4,133,003)
S. O. Graham CCD Output Circuit Using Thin Film Transistor (4,132,903)
W. H. Groeneweg PAL Identification Circuit (4,133,002)
D. Hampel Multi-Function Logic Gate with One Gate Delay (4,133,040)
R. S. Hopkins, Jr. Digital Video Synchronizer (4,134,131)
L. R. Hulls and S. C. Hadden Identification of Engine Cylinder Having Fault (4,133,205)
R. N. Hurst Television Picture Size Altering Apparatus (4,134,128)
G. Kaganowicz, J. W. Robinson, and H. Yasuda VideoDisc with a Dielectric Layer Formed from Acetylene and Nitrogen (4,137,550)
W. R. Kelly and R. L. Barbin Method and Apparatus for Optimizing Color Purity in a Color Kinescope (4,137,548)
W. F. Kosonocky and D. J. Sauer Charge Transfer Circuits with Compensation for Transfer Losses (4,134,028)
L. H. Lin Resist Development Control System (4,136,940)
C. K. Mok Dual Mode Filter (4,135,133)
A. M. Morrell Cathode Ray Tube Having Improved Shadow Mask (4,136,300)
M. A. Polinsky Bipolar Transistor with High-Low Emitter Impurity Concentration (4,136,353)
J. A. Rajchman Cathode Addressing System (4,137,551)
M. D. Ross and J. K. Clemens Transcoding Apparatus (4,136,358)
J. J. Serafini Automatic Beam Current Limiter with Independently Determined Threshold Level and Dynamic Control Range (4,137,552)
R. L. Shanley, 2nd Brightness Control Circuit with Predictable Brightness Control Range (4,135,200)
J. M. Shaw and K. H. Zaininger Planar Silicon-on-Sapphire Composite (4,133,925)
L. A. Torrington Record Handling System for a VideoDisc Player (4,133,540)
J. A. VanRaalte Image Display Block Scanning Method (4,137,485)
J. K. Weimer Charge Injection Devices and Arrays and Systems Including Such Devices (4,134,031)

February

A. A. Ahmed Current Mirror Amplifier (29,910)
A. A. Ahmed Signal Translation Circuits (4,140,977)
J. R. Barkwith Idle-Busy Signalling Between Telephone System and Radiophone System (4,138,595)
A. Bloom Electro-Optic Device (4,141,627)

D. E. Carlson, A. R. Triano, Jr., and C. R. Wronski Schottky Barrier Semiconductor Device and Method of Making Same (4,142,195)
S. H. Cohen and J. J. Fabula Process for Manufacturing a Radiation Hardened Oxide (4,139,658)
A. G. Dingwall Drain Extensions for Closed COS/MOS Logic Devices (4,142,197)
R. A. Dischert and R. E. Flory Video Signal Amplitude Registration System (4,141,040)
J. G. Endriz Image Display Device with Optical Feedback to Cathode (4,142,123)
W. J. Hannan Color Image Storage and Display Utilizing Holography (4,142,204)
L. F. Hart Apparatus and Method for Aligning Wafers (4,141,456)
L. V. Hedlund and A. C. Luther, Jr. Disc Eccentricity Compensating System (4,138,741)
L. V. Hedlund, R. P. Fink, and D. I. Wright Disc Track Servo System (4,142,209)
E. P. Herrmann Charge Transfer Output Circuits (4,140,923)
W. Hulstrunk Contact Clip (4,141,028)
H. P. Kleinknecht and J. Kane Optically Monitoring the Thickness of a Depositing Layer (4,141,780)
R. S. Mezrich Switchable Depth-of-Focus Pulse-Echo Ultrasonic-Imaging Display System (4,138,895)
M. Nowogrodzki Indicating Temperature Within Living Tissue (4,138,998)
J. I. Pankove Solar Cell with a Gallium Nitride Electrode (4,139,858)
W. T. Patton and N. R. Landry Short Radiating Horn with an S-Shaped Radiating Element (4,138,683)
W. Phillips Surface Acoustic Wave Device with Reduced Spurious Responses (4,142,163)
B. M. Pradal Crystal Overtone Oscillator Using Cascade Connected Transistors (4,139,826)
J. J. Risko Two Level Threshold Circuit (4,139,851)
A. Rose Photoconductor for Imaging Devices (4,139,796)
D. J. Sauer CCD Input Circuits (4,139,784)
J. L. Smith Magnetizing Method for Use with a Cathode Ray Tube (4,138,628)
F. R. Stave and L. A. Torrington VideoDisc Package (4,138,703)
L. R. West Ground Station Data Storage System (4,139,900)
C. M. Wine Memory Type Tuning System for Storing Information for a Limited Number of Preferred Tuning Positions (4,138,647)

March

R. W. Allen and A. Jackson, 3rd Frequency Activated Circuit (4,145,660)
F. Aschwanden SECAM Modulator (4,145,711)
W. H. Barkow Deflection Yoke with Permanent Magnet Raster Correction (4,143,345)
H. R. Beelitz Decoder Circuit (4,143,359)
C. A. Berard, Jr. Bilateral Energy Transfer Apparatus (4,143,282)
F. X. Conaty Video Record Package (4,145,726)
A. R. Dholakia and J. Alexander VideoDisc Pickup with Preplay Stylus (4,145,718)
J. A. Diciurcio Voltage Level Generator Using Digital Integration (4,145,743)
W. F. Dietz Regulated Deflection Circuit (4,146,823)
R. E. Fernsler Television Horizontal Oscillator Frequency Control Arrangement for Use with Tape Recorder (4,144,544)
R. E. Fernsler and M. L. Henley Television Horizontal Oscillator Frequency Control Arrangement for Use with a Tape Recorder (4,144,545)
M. R. Freeling and H. J. Wolkstein RF Burst Signal Recirculation Memory System Having a Diplexed Feedback Loop (4,145,691)
J. Goel Method of Making a Short Gate Field-Effect Transistor (4,145,459)
L. A. Harwood and E. J. Wittmann Automatic Brightness Control Circuit Employing a Closed Control Loop Stabilized Against Disruption by Large Amplitude Video Signals (4,143,398)
D. D. Holmes Comb Filter Apparatus (4,143,397)
R. J. Klensch Range Rate Measurement (4,146,890)
H. G. Lewis, Jr. Analog-to-Digital Converter (4,143,366)
M. J. Lurie Redundant Hologram Recording Method Employing Temporal Information Signal (4,142,772)
A. M. Morrell Cathode-Ray Tube with a Corrugated Mask Having a Corrugated Hinging Skirt (4,146,816)
D. B. Nichols, Jr. Multi-Dye Textile Dyeing Process (4,146,362)
E. J. Nossen and E. R. Starner Arithmetic Synthesizer Frequency Generation with Reduced Phase Jitter (4,144,579)

M. Packer Extrudable, Non-Flowing and Non-Aqueous Solvent Soluble Hold-Down Compound for Printed Wiring Board Assembly (4,143,005)
K. D. Peters and C. H. Anderson Modular Guided Beam Flat Display Device (4,145,633)
A. D. Robbi Signal Sampling Circuit (4,143,329)
T. O. Stanley Flat Panel Display Device (4,143,296)
E. R. Starner and E. J. Nossen Accurate Phase-Measuring System Using Arithmetic Synthesis (4,144,572)
L. C. Upadhyayula FET-TELD Combination with Capacitively Coupled Output Electrode Means (4,145,624)
D. H. Willis Television Receiver Protection Circuit (4,145,639)
J. E. Wojslawowicz GTO Bi-Directional Motor Control Circuit (4,146,826)

Arnold R. Moore received the B.S. in Chemistry from the Polytechnic Institute of New York in 1942, with thesis work in Chemical Kinetics. Thereafter, he worked on gas discharge and photoelectric surface problems with RCA manufacturing divisions in Harrison, NJ, and Lancaster, Pa. In 1945 he returned to graduate school at Cornell University, where in 1949 he received the Ph.D. degree in Experimental Physics, with a minor in Theoretical Physics and Mathematics, under a National Research Council Fellowship. His work at Cornell consisted chiefly of experimental research into the motion of electrons in alkali and silver halides, along with the growth of the required single crystals. Since starting work at RCA Laboratories in 1949, he has been active in the fields of transistor physics, optical absorption in semiconductors and semiconductor alloys, magnetic susceptibility, acoustoelectric effect in cadmium sulfide and gallium arsenide, and most recently photovoltaic solar converters. In 1970–71, as a visiting professor at Brown University, he taught a graduate course in physics of solid-state devices and participated in research on photovoltaic cells.

Since his return to RCA Laboratories in 1971, Dr. Moore has been working on insulators and photo devices. He has received three RCA Achievement Awards and is a Fellow of the American Physical Society and a member of Sigma Xi.

David R. Patterson after completing a course in electronic technology at Trenton Technical Institute, Trenton, N.J., joined RCA Laboratories in 1962 as a research technician in the Conversion Devices Laboratory. His duties included system design, mechanical packaging, and construction of infrared television test equipment and closed circuit camera systems. He completed a program leading to a B.S. degree in Electrical Engineering awarded with high honors from Drexel University in 1972. Additional graduate work was undertaken at the Moore School of E.E. of the University of Pennsylvania. Mr. Patterson was a Research Associate in the Semiconductor Devices Research group of the Materials Research Laboratory with responsibilities including the design of electronic circuits for various semiconductor device research activities including power semiconductors, solar cells, light emitting diodes, and laser diodes. He is now employed at Optical Information Systems, Exxon Enterprises, Inc., Elmsford, N.Y.

Marvin Rozansky received the B.E.E. from Rensselaer Polytechnic Institute in 1964, the M.S.E.E. from Columbia University in 1965 and the Ph.D. from the University of Pennsylvania as a David Sarnoff Fellow in 1978. He has been with Missile and Surface Radar in Moorestown, since June 1965. His contributions there have been many and varied, being within the areas of design and development, circuit and servo design, radar system and subsystem analysis, radar computer programming, and evaluation and test. Major programs he has worked on include the FPS-95 and AEGIS. He is currently coordinating the system definition and analysis effort on the P-50M radar program.

Mr. Rozansky is a member of the Institute of Electrical and Electronics Engineers and Tau Beta Pi and Eta Kappa Nu honorary societies.

William L. Sked majored in Physics for three years at Stevens Institute of Technology, Hoboken, N.J. He joined RCA Solid State Division, Somerville, N.J. in 1969 as a Technician for the RF/Microwave Power Devices group, where he did work on bipolar process development. In 1971 he assumed the position of Foreman for the wafer processing area of the RF/Microwave Pilot Production Facility. He worked briefly for the COSMOS engineering group in the analysis of phosphosilicate glass composition. In 1975, he joined the Microwave Technology Center as a Research Technician and has worked on thin film technology processing of microwave hybrid circuitry and passive components. In 1978, he became a Technical Associate on the staff at RCA Laboratories.

Frank J. Wozniak attended RCA Institute from 1964 to 1966. Upon graduation, he worked at N.Y.U. where he took several technical courses. He joined RCA in 1969, working in the Defense Microelectronics Group in Sommerville, N.J. In 1971, he transferred to the Microwave Technology Center at RCA Laboratories, Princeton, N.J., where he is engaged in the fabrication of components for microwave integrated circuits.