AD A086995

(12) LEVEL II

NRL Memorandum Report 4233

Studies on the Reduction of Intermodulation Generation in Communications Systems

G. H. Stauss (Editor)

G. C. Bailey, C. D. Bond, C. A. Carosella, A. C. Ehrlich, C. S. Guenzer, G. N. Kamm, and C. E. Young

AD A086995

July 7, 1980

DTIC ELECTE JUL 23 1980 S D B

DDC FILE COPY

NAVAL RESEARCH LABORATORY
Washington, D.C.

Approved for public release; distribution unlimited.

80 7 11 018

(18) SBIE / (19) AD-E000 492

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: NRL Memorandum Report 4233	2. GOVT ACCESSION NO.: AD-A086995
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): (6) STUDIES ON THE REDUCTION OF INTERMODULATION GENERATION IN COMMUNICATIONS SYSTEMS	5. TYPE OF REPORT & PERIOD COVERED: (9) FINAL REPORT
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): (10) G. H. Stauss, G. C. Bailey, C. D. Bond, C. A. Carosella, A. C. Ehrlich, C. S. Guenzer, G. N. Kamm, and C. E. Young	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Naval Research Laboratory, Washington, D.C. 20375	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NRL Problem R08-73, Program Element 33109N, Project X-0731-CC
11. CONTROLLING OFFICE NAME AND ADDRESS: Naval Electronic Systems Command, Washington, D.C. 20360	(11) 12. REPORT DATE: July 1980; 13. NUMBER OF PAGES: 95
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
(14) NRL-MR-4233	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

(12) 94

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

(16) X0731CC

18. SUPPLEMENTARY NOTES

(17) X0731CC

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Satellite communications	Connector design
Intermodulation interference	Ferromagnetic materials
Nonlinear conduction	Nonlinear circuit analysis
Tunneling junctions	Multiplex systems

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Potential means of intermodulation generation in multiplex systems are reviewed and evaluated theoretically. The dominant mechanisms involve non-ohmic junctions and ferromagnetic components. Junctions through aluminum oxide films have been studied experimentally and a possible means of improvement is described. Experimental tests of ferromagnetic components show the need to exclude them entirely from multiplex and high performance systems. With care in the choice and assembly of components, it is concluded that intermodulation signals can be kept to the level of thermal noise in most configurations.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(Page ii Blank)

251950

CONTENTS

ACCESSION for
NTIS White Section ☑
DDC Buff Section ☐
UNANNOUNCED ☐
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
Dist. AVAIL. and/or SPECIAL
A

FOREWORD

The production of interfering signals by nonlinear responses in multichannel communications systems has in the past been epitomized by telephone crosstalk and the "rusty bolt" effect on shipboard. The problem, however, is not a thing of the past. Advances in technology have led to higher transmitter powers and greater receiver sensitivities in combination with the use of multiple closely spaced channels and a greater physical density of components. As a result, hitherto unimportant nonlinear behaviors of passive components are compromising what should be high performance systems. High sensitivity receivers are being subjected to significant products of intermodulation generation (IMG) from the mixing of multiple simultaneous transmitter signals either diplexed (using a common antenna) or radiating from nearby antennas. Receiver filtering can eliminate only those extraneous signals which fall outside the intended operating band of the receiver.

The Naval Research Laboratory has been concerned with experimental investigations and theoretical reviews of mechanisms for IMG and corrective measures in connection with the development of the Fleet Satellite Communications System (FLTSATCOM) by the Naval Electronic Systems Command. The work reported here was supported by that agency and directed at NRL by V. J. Folen of the Magnetism Branch, Electronics Technology Division. A special acknowledgement is due M. Frazier of the Illinois Institute of Technology for many valuable consultations throughout this study.

This report comprises six chapters, of which four have appeared previously in somewhat altered forms. Each chapter contains its own conclusions and recommendations. The general nature of the IMG problem is set forth in the introduction to the first chapter. Throughout this study the two principal sources of IMG have appeared to be poor mechanical junctions and the presence of ferromagnetic components. Junctions are discussed theoretically in Chapter I and experimentally in Chapters I and III. Experimental studies of magnetic components appear in Chapters II-IV and theoretical studies in Chapters IV and V. Chapter V is a theoretical survey of the inherent nonlinear properties of materials to identify the relative importance of various IMG mechanisms and the lower limits on system sensitivity. The first five chapters concentrate on the third order IM signal at $2\omega_1 - \omega_2$ of two closely tuned transmitters with frequencies ω_1 and ω_2. (The order of an IM product at $a\omega_1 + b\omega_2$ is defined as $|a| + |b|$.) This is often the signal of greatest strength and most practical significance, but as discussed in Chapter III other signals can be important. If more than two primary signals exist analysis becomes difficult. Chapter VI discusses a computer-based scheme for the prediction and possible diagnosis of the complete IM spectra from analytically difficult non-linearities or multiple primary signals.

Chapter I

GENERATION OF INTERMODULATION BY ELECTRON TUNNELING THROUGH ALUMINUM OXIDE FILMS*

C.D. Bond, C.S. Guenzer and C.S. Carosella

Radiation Effects Branch
Radiation Technology Division

INTRODUCTION

The Navy's Fleet Satellite Communication System (FLTSATCOM) consists basically of three synchronous-orbit satellites with UHF links to various earth-surface terminals. Each satellite and surface terminal simultaneously operates both transmitter and receiver units in close physical proximity. The Navy's transmit and receive bands are approximately 240 to 270 MHz and 290 to 320 MHz respectively. These two carrier bands are separated by only 20 MHz. For such a system to operate successfully, any undesirable frequency mixing caused by nonlinear components (intermodulation (IM) generation) must be minimized.

The severe magnitude of this problem can best be illustrated by calculating the acceptable IM power level at a receiver for typical operating conditions; for example, for a satellite transmitter power of $\sim$100W (+50 dBm) the space attenuation alone is $\sim$175 dB ($-$125 dBm, or down to 3×10^{-16} W) for either the uplink or downlink signals. At the satellite receiver the uplink signals should normally be $\sim$15 dB above any interference or noise level. This requires the interference level to be below $-$140 dBm, 10^{-17} W. Thus the acceptable IM level at the satellite receiver is required to be 190 dB below the local transmitted power level. This is a ratio of IM power to signal power of 10^{-19}. The same requirement must also be met at the earth-based terminals.

This problem is diagramed in Fig. 1 by an illustration from Young [1]. Here the IM products are presented in a simplified two-signal analysis. Consider two frequencies f_1 and f_2 in the transmit band at voltage levels of V_1 and V_2. The input voltage V_{in} across some network can be written

$$V_{\text{in}} = V_1 \cos \omega_1 t + V_2 \cos \omega_2 t. \tag{1}$$

For a linear network the output current will be

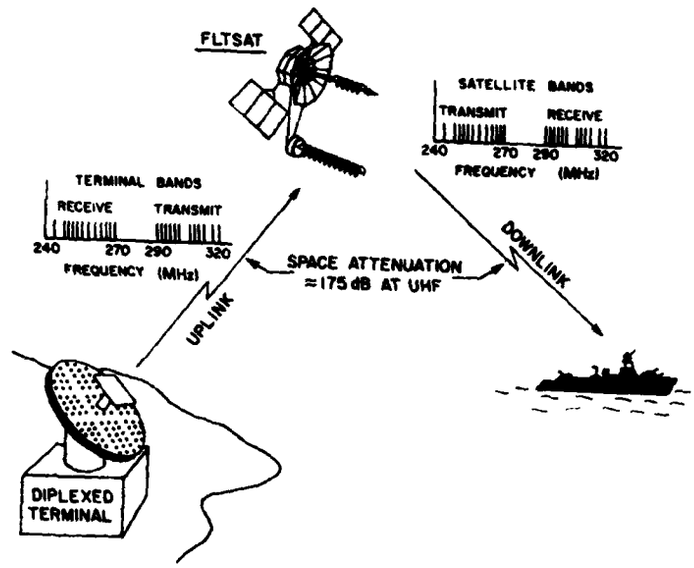
$$I = G_O V_{in}.$$

For a nonlinear network the output current can be expressed as the power series

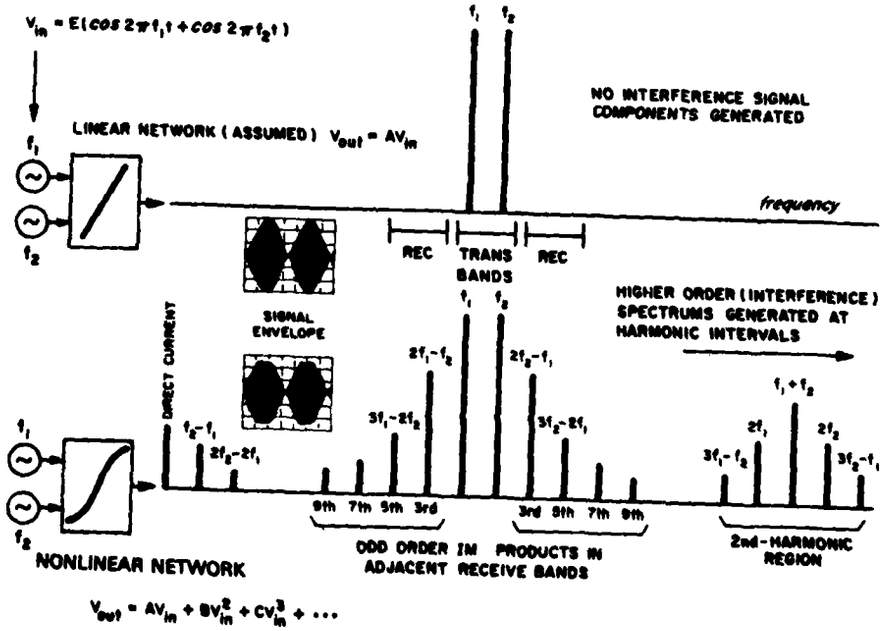
$$I = G_0 V_{\text{in}} + G_1 V_{\text{in}}^2 + G_2 V_{\text{in}}^3 + \dots . \tag{2}$$

The separate IM term in the nonlinear output can be seen by substituting the input voltage of Eq. (1) into Eq. (2) to give

*Previously published as NRL Report 8170.



(a) Basic satellite communication system



(b) Simplified two-signal analysis

Fig. 1 — Basic satellite communication system showing simplified two-signal analysis for a linear and a *nonlinear network*. The intermodulation-frequency terms result from a power series representation of the nonlinear signal-voltage output and constitute interference in the receive bands.

$$I = G_0(V_1 \cos \omega_1 t + V_2 \cos \omega_2 t) +$$

$$G_1 \left[\frac{V_1^2 + V_2^2}{2} + \frac{1}{2} V_1^2 \cos 2\omega_1 t + \frac{1}{2} V_2^2 \cos 2\omega_2 t \right.$$

$$\left. + V_1 V_2 \cos (\omega_1 - \omega_2)t + V_1 V_2 \cos (\omega_1 + \omega_2) t \right] +$$

$$G_2 \left[\frac{1}{4} V_1^3 \cos 3\omega_1 t + \frac{1}{4} V_2^3 \cos 3\omega_2 t + 3V_1 \left(\frac{V_2^2}{2} + \frac{V_1^2}{4} \right) \cos \omega_1 t \right.$$

$$+ 3V_2 \left(\frac{V_1^2}{2} + \frac{V_2^2}{4} \right) \cos \omega_2 t + \frac{3}{4} V_1 V_2^2 \cos (2\omega_2 - \omega_1)t$$

$$+ \frac{3}{4} V_1 V_2^2 \cos (2\omega_2 + \omega_1) t + \frac{3}{4} V_1^2 V_2 \cos (2\omega_1 - \omega_2) t$$

$$\left. + \frac{3}{4} V_1^2 V_2 \cos (2\omega_1 + \omega_2)t \right] + \ldots . \qquad (3)$$

The frequencies $f_1 = 250$ MHz and $f_2 = 270$ MHz have been chosen for most of the device testing at the NRL IM test facility.

At these two typical satellite transmitter frequencies, most of the IM terms shown in Eq. (3) include frequencies that lie outside the satellite receive band. However, two of the largest magnitude IM terms of frequencies $2f_2 - f_1$ and $3f_2 - 2f_1$ (290 MHz and 310 MHz) do fall within the local receive band. (The source of the latter IM frequency is the V^5 term and is not shown in Eq. (3).) This example illustrates the problem resulting from a narrow separation of the transmit-receive bands.

The NRL investigation is concerned with the generation of intermodulation due to various nonlinear mechanisms in normally passive hardware such as contacts, flanges, connectors, and other current-carrying metal structures attendant to the transmitter-receiver system. The investigation described in this report is specifically concerned with the generation of intermodulation due to the nonlinear conduction by electrons tunneling through thin oxide films, particularly Al_2O_3 on aluminum structures.

The detailed mechanisms of surface-to-surface contacts is extremely involved and complex and remains incompletely understood. There are many interacting variables, such as the density of microscopic contact points, number of contacts, contact pressure, oxide growth, oxide fracture and regrowth, metal-to-metal bonding, diffusion, and time and temperature effects.

Despite the lack of a complete microscopic theory, contact models such as those described in the Philco-Ford report [2] do allow predictions adequate for assessment and control of the intermodulation generated in real hardware contacts. Because of the numerous contact mechanisms which have been identified, an experimental study of tunneling in real contacts is not amenable to reproducible results and to correlation with theory [3]. Consequently the intent of this experimental investigation was to fabricate well-characterized tunneling junctions of $Al-Al_2O_3-Al$, to directly measure the intermodulation generated, and to attempt to correlate the intermodulation with the device circuit parameters and with tunneling theory. In addition some preliminary efforts were made to improve the conduction characteristics of such junctions by ion implantation.

TUNNELING THEORY

The theory of electron tunneling through an insulating layer dates back to the 1920's. However controversy still continues as to the correct form of the tunneling equations for thin fiims ($\lesssim$5 nm). Some of the principal difficulties include the following:

- Applicability of macroscopic parameters, such as the dielectric constant, to a few atomic layers,
- Complexity of integrals which can presently be solved only by approximations which affect the accuracy of the final results to an undetermined degree, and
- Accurate evaluation of contaminant effects on the electron scattering surfaces.

This last effect has been developed into a valuable qualitative method for investigating the vibrational modes of the contaminants, but quantitatively the effect is poorly understood. Thus comparison of the experimental results with the electron tunneling theory is only approximately quantitative.

The theory of electron tunneling is based on the fact that the wave function of a free electron ψ_e extends only a few nanometers into an insulator at any metal-insulator interface. In large-dimension insulators the wave function quickly decreases to a vanishingly small value such that the effect goes unnoticed; but if the insulator is only 3 nm (a typical aluminum oxide film), the wave function ψ_e remains finite throughout the insulator and yields a measurable probability for the electron to pass from one side of the insulator to the other. This effect is shown in Fig. 2, where the extended wave function ψ_e on the left is matched in magnitude and slope to the decaying exponential within the insulator, which before it vanishes, similarly matches with the extended wave function in the metal on the right.

Since temperature has a small effect, one needs only the zero-temperature expression for the tunneling current density J in one direction [4]:

$$J = \frac{me}{2\pi^2\hbar^3}\left[eV\int_0^{E_{F_1}-eV} P(E)\,dE + \int_{E_{F_1}-eV}^{E_{F_1}} (E_{F_1} - E)P(E)\,dE\right].$$

The parameters are shown in Fig. 2, where V is the voltage applied across the junction and E_{F_1} is the Fermi level of the electron distribution in the metal area on the left. $P(E)$ is the transmission probability by tunneling for electrons of transverse energy E.

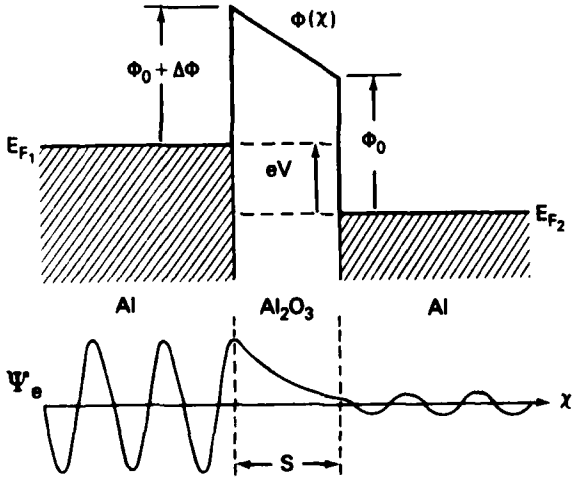


Fig. 2 — Electron tunneling through the potential barrier of a thin insulating film between two conductors. The conductors of the sandwich structure are maintained at a potential difference of V such that there is a finite electron probability function ψ_e everywhere to the right of the first interface. The resulting current is a nonlinear function of voltage.

The usual method for calculating the transmission probability involves the WKB approximation, which yields a transmission probability given by

$$P(E) = \exp\left[-2\int_S \sqrt{\frac{2m}{\hbar^2}\,[\phi(x) - e]}\,dx\right].$$

In this expression, $\phi(x)$ is the local barrier potential acting on the electron within the classically forbidden region of the aluminum oxide, and the integral is carried out over the width of the forbidden region S. For the simple trapezoidal barrier shown in Fig. 2, the transmission probability from left to right is given exactly by

$$P(E) = \exp\left\{-\frac{4}{3}\frac{2m}{\hbar}\left(\frac{S}{eV + \Delta\phi}\right)\left[(E + \phi_0 + eV + \Delta\phi)^{3/2} - (E + \phi_0)^{3/2}\right]\right\}.$$

This exponential expression is sufficiently complex so that the last integral in the initial equation for J cannot be carried out exactly. Forlani and Minnaja [5] have expanded the argument of the exponential about the average of the Fermi levels and obtained a low voltage expression for the tunneling current density J. As part of this investigation a Taylor expansion of the Forlani-Minnaja equation to the fourth order in voltage has been published [6]:

$$J = \frac{me}{2\pi^2\hbar^3}\, e^{-A\bar{\phi}^{1/2}}\left[\left(\frac{2\bar{\phi}^{1/2}}{A} + \frac{2}{A^2}\right)eV + \frac{A}{48\bar{\phi}^{1/2}}(eV)^3\right], \tag{4}$$

where the parameter A is defined as

$$A = \frac{4\pi S\sqrt{2m}}{\hbar}.$$

The dependence on the interface potential is through $\bar{\phi}$, the average of the two interface potentials. The lack of dependence on $\Delta\phi$, the difference in interface potentials, is reflected by the lack of a V^2 term. Apparently this lack of polarity is caused by the choice of expansion about an average Fermi level. Stratton [7] has chosen instead to expand about the Fermi level on one side. Brinkman et al. [8] give a low-voltage expansion for this case as

$$J = \frac{me}{2\pi^2\hbar^3}\, e^{-A\bar{\phi}^{1/2}}\left[\frac{2\bar{\phi}}{A}\,eV - \frac{\Delta\phi}{12\bar{\phi}}\,(eV)^2 + \frac{A}{16\bar{\phi}^{1/2}}\,(eV)^3\right],$$

where

$$\bar{\phi} = \phi_0 + \frac{1}{2}\,\Delta\phi.$$

Here, not only is the cubic term 3 times larger than in Eq. (4), but there is a significant quadratic term proportional to $\Delta\phi$. Unfortunately this expression is only approximate, and Brinkman *et al.* do not state the approximations involved. Thus the simplest type of theory is somewhat uncertain even for the second- and third-order terms.

JUNCTION FABRICATION

The laboratory fabrication of an electron tunneling junction is illustrated in Fig. 3. The procedure involves the vacuum evaporation of an aluminum strip ~150 nm thick onto a clean glass substrate. The strip is then allowed to oxidize at room temperature from 12 to 24 hours. Following this oxidation a second similar strip is deposited at right angles to the first strip. The resulting junction sandwich area consists of two parallel plates of thin aluminum separated by an aluminum oxide insulating film approximately 3 nm thick. Such a structure serves as a useful tool in the experimental study of the electron tunneling phenomena described in the previous section.

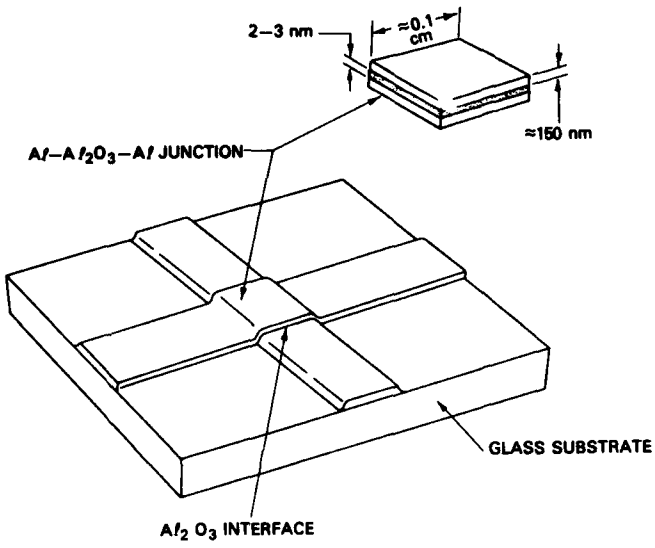


Fig. 3 — Laboratory technique used in fabricating tunneling-junction sandwich structures of Al-Al_2O_3-Al. The aluminum oxide film is grown at room temperature on a vacuum-evaporated aluminum strip and overlayed with a second aluminum strip as shown.

In practice, five to ten such junctions were usually fabricated simultaneously under the same conditions to establish controls and statistics. The vacuum-chamber arrangement for evaporation, masking, and monitoring is shown in Fig. 4. Prior to evaporation a Vac-Ion pump is used to evacuate the

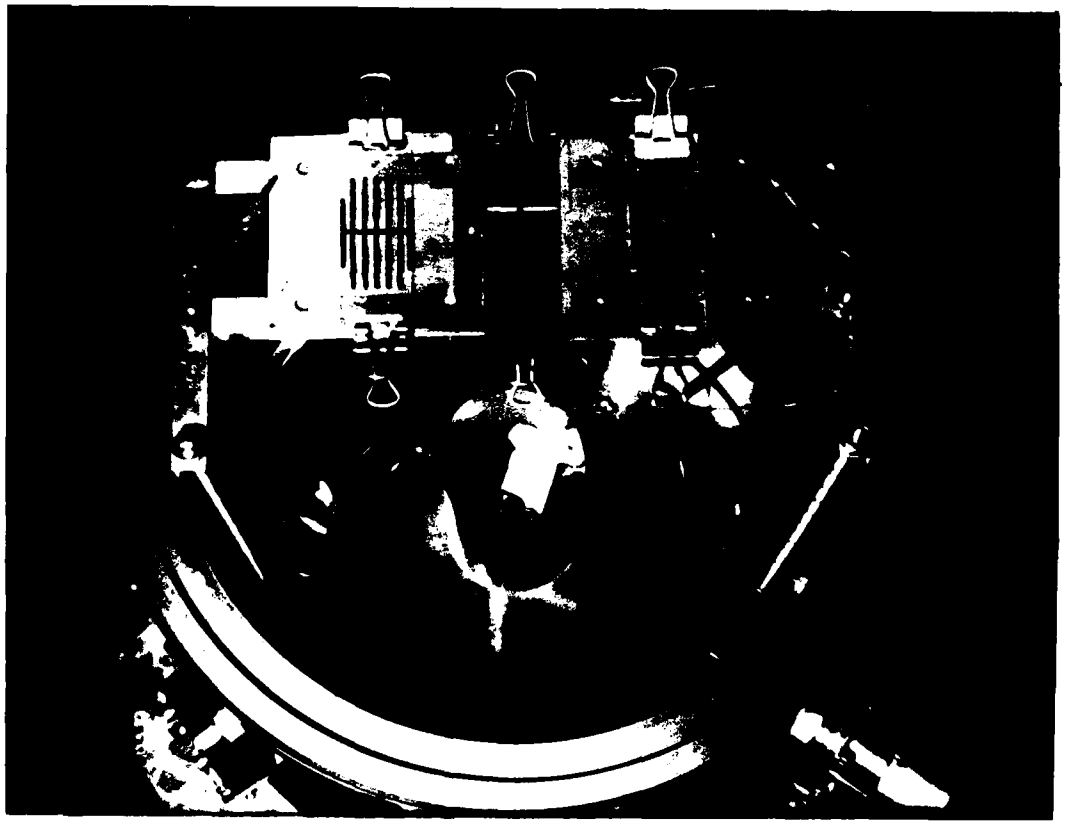


Fig. 4. — Vacuum-chamber arrangement for evaporation, masking, and monitoring. The crystal head of the Sloan thickness monitor is normally mounted in the plane of the substrate surfaces and is not shown in this photograph.

chamber to a pressure of $\lesssim 5\times10^{-7}$ torr. After degassing the tungsten basket filament and its aluminum contents (0.5 gof 99.999+% aluminum), the aluminum is evaporated rapidly in $\lesssim 1.0$ minute. During the evaporation the chamber pressure typically increases to $\lesssim 8\times10^{-6}$ torr. A shutter arrangement is used to control the starting and stopping of aluminum deposition. (The circular shutter is displaced to the left in Fig. 4.) The film thickness is monitored during evaporation both by measuring the film resistance and by a direct-reading digital monitor of the crystal oscillator type (Sloan 200). The programmable monitor is also used to follow directly the subsequent growth of the aluminum oxide. After the aluminum film evaporation, the monitor is reset to zero thickness, and the density of Al_2O_3 is dialed into the monitor. After the pressure in the chamber is increased to 1 atmosphere of pure oxygen, the Al_2O_3 thickness increases from zero to $\sim$2 nm in the first 5 minutes and then grows more slowly, reaching 2.2 to 2.7 nm after about 1 hour. Although no oxide growth was ever observed in this study after a few hours, the second evaporation of aluminum over the oxide growth was not carried out until 12 to 24 hours later.

Figure 5 shows a typical assembly of five junctions at three stages of fabrication. Initially thin strips of indium are "soldered" onto the glass microscope slide to form electrical contacts. (Molten indium readily adheres to a clean glass surface.) A mask is carefully aligned over the contacts, and a strip of aluminum is evaporated onto the glass and indium contacts. After the strip is oxidized, a set of five aluminum-film strips are similarly evaporated simultaneously at right angles to the first strip. Finally No. 36 copper magnet-wire leads are soldered onto the indium contacts.

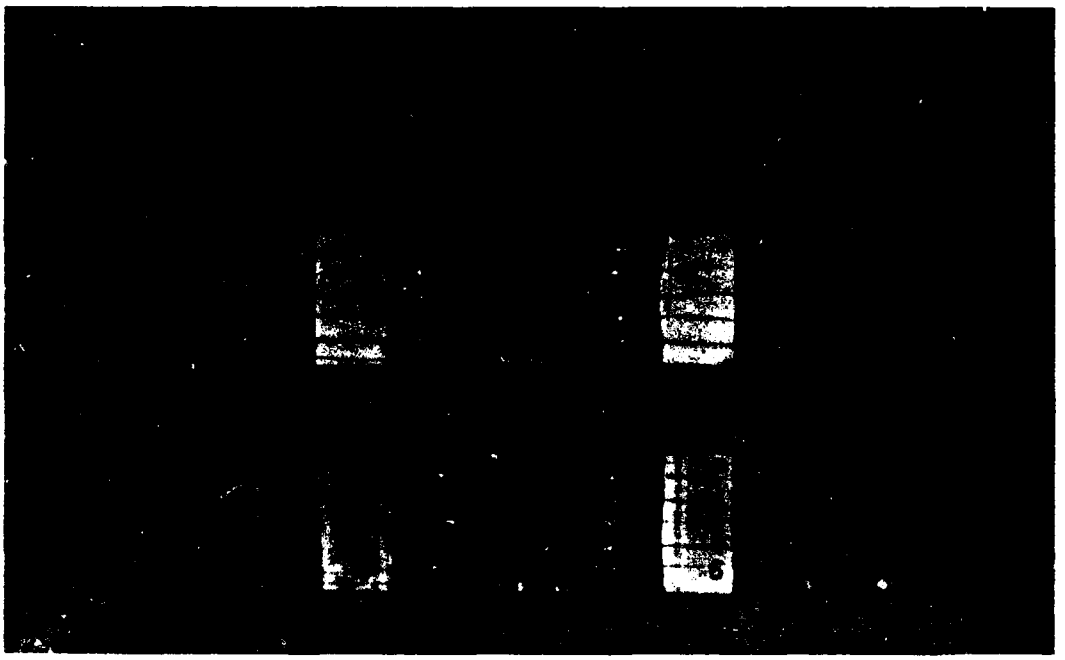


Fig. 5 — A typical assembly of five junctions shown at three stages of construction: (A) Pattern of indium contacts are attached to clean glass substrate. (B) Cross strip of aluminum is vacuum deposited and oxidized in pure oxygen at room temperature for $\approx$24 hours. (C) Finally, five parallel strips of aluminum are vacuum deposited at right angles to the cross strip to form five junctions.

CURRENT-VOLTAGE CHARACTERISTICS

The electrical properties of these junctions which are of primary interest in this application are their nonlinear resistance characteristics and their generation of RF intermodulation. The nonlinear DC current-voltage ($I-V$) characteristics of such junctions were measured with a standard four-point arrangement (Fig. 6). A programmable constant-current supply (Keithly 227) is used to drive from

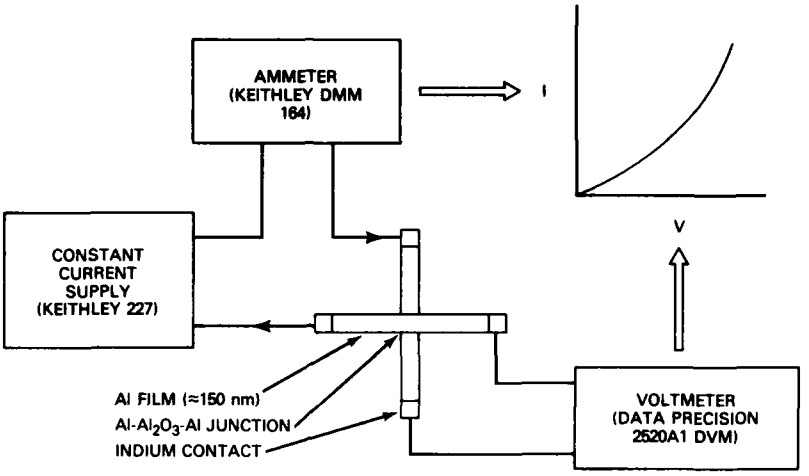


Fig. 6 — Standard four-point arrangement for measuring the current-voltage characteristics of the electron tunneling junctions

3 nA to 1 mA through two legs of the junction film leads. A 5-1/2-digit digital voltmeter is used to measure the voltage developed across the junction. Since the input impedance of the voltmeter is $>10^8$ ohms, negligible current from the current source flows through the voltmeter circuit, and the effects of lead resistance and contact resistance are essentially eliminated; that *s*, the only IR drop seen by the voltmeter is across the junction area. The current is measured by a digital multimeter (Keithley 164).

Figure 7 shows the current-voltage characteristic curve for one of the early 0.0168-cm^2 devices. On the scale shown the experimental curve becomes visibly nonlinear at about 80 mV and exhibits a behavior typical of all the junctions measured. On the basis of tunneling theory, one would expect the curve to be described by the cubic dependence of the form shown in the fourth-order expansion in Eq. (4). A computer program, FORLAN, was written to facilitate rapid calculation of the tunneling current as a function of the junction parameters. This program uses the complete Forlani-Minnaja expression together with the near-linear (low-voltage) experimental values of I, V, and G_O to determine the values of S for selected values of $\bar{\phi}$. The program can then generate full-range characteristic curves for different values of $\bar{\phi}$ so as to obtain a best fit to the experimental $I-V$ curve.

The quality of fit to the experimental data was found to vary, depending on the particular set of junctions. In some cases an excellent fit could be obtained, and in other cases the calculated curvature deviated substantially from the experimental data beyond 100 mV.

Using the best-fit values of S and $\bar{\phi}$, one can also calculate the tunneling current from the expansion in Eq. (4). The dashed line in Fig. 7 shows an example of such a calculation. It was found, however, that the current-voltage characteristics of all of the junctions could be fitted considerably better by using an empirical cubic equation of the form of Eq. (4): $I = G_O V + \alpha G_O V^3$. Here, when experimental values of G_O and I and V at 100 mV and 200 mV were used, an average value of α could be obtained to give a satisfactory fit to the data. The solid line in Fig. 7 shows an example of this procedure. The α's calculated from the tunneling theory in Eq. (4) were generally 20% to 50% lower than the empirically determined α's. Since α is a measure of the nonlinearity and will later occur explicitly in the IM expression, we will use the empirical value of α as determined for each junction.

It is also important to emphasize that the junction parameters $G_O, \alpha, \bar{\phi}$, and S do not remain constant in time. As indicated, the characteristics shown in Fig. 7 were measured at 90 hours after fabrication. The junction resistance as a function of applied voltage for this same junction is shown in Fig. 8

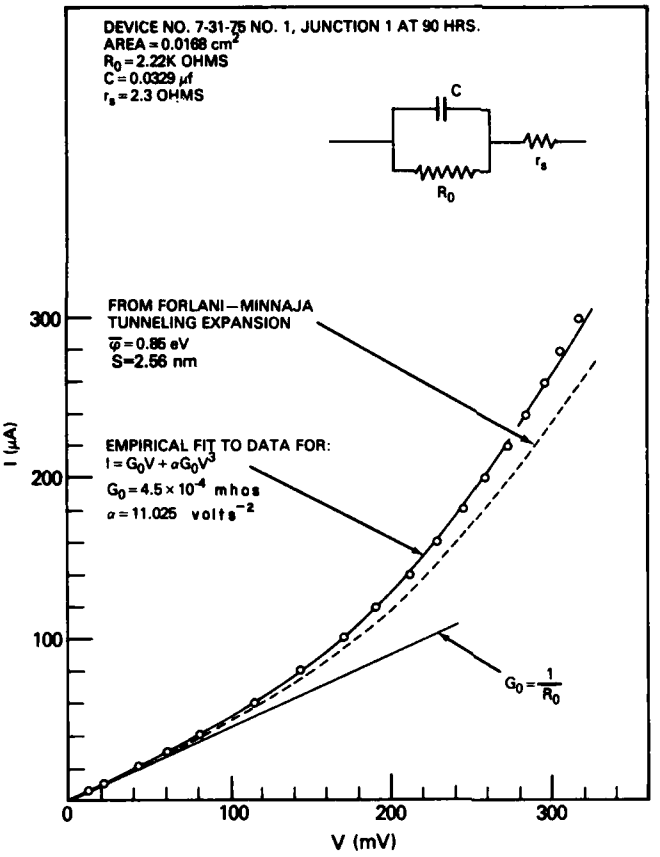


Fig. 7 — The current-voltage characteristic curve for a 0.0168-cm^2 junction measured 90 hours after fabrication. The equivalent circuit of the junction is shown in the upper right.

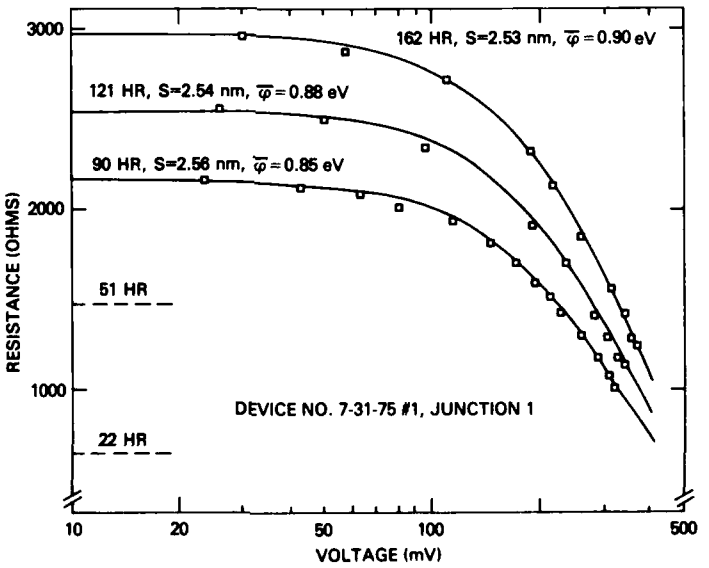


Fig. 8 — The behavior of junction resistance as a function of voltage and time. The solid curves are generated by the FORLAN program and represent a best fit to the experimental points for the indicated values of $\bar{\phi}$ and S.

with time as a parameter. Here the solid curves are a best fit to the experimental points using the FORLAN program. The curves show the sensitivity of the junction resistance to $\bar{\phi}$ and particularly to the insulator thickness S.

Although the marked change of resistance with time does not prevent the device from being well characterized, it does necessitate the additional inconvenience of keeping a time history of the individual junctions and of measuring the intermodulation within a few hours of any characteristic measurement. A time history of four typical junctions is shown in Fig. 9. These four junctions were fabricated simultaneously under the same conditions. The low-voltage resistance R_O of each junction was tracked for 2000 hours or more. The intermodulation was generally measured from 6 to 48 hours after completion of fabrication. This time lapse was found to be convenient both because of the slow rate of resistance change and because the initial resistance (300 Ω to 3 kΩ) was more suitable for IM measurements.

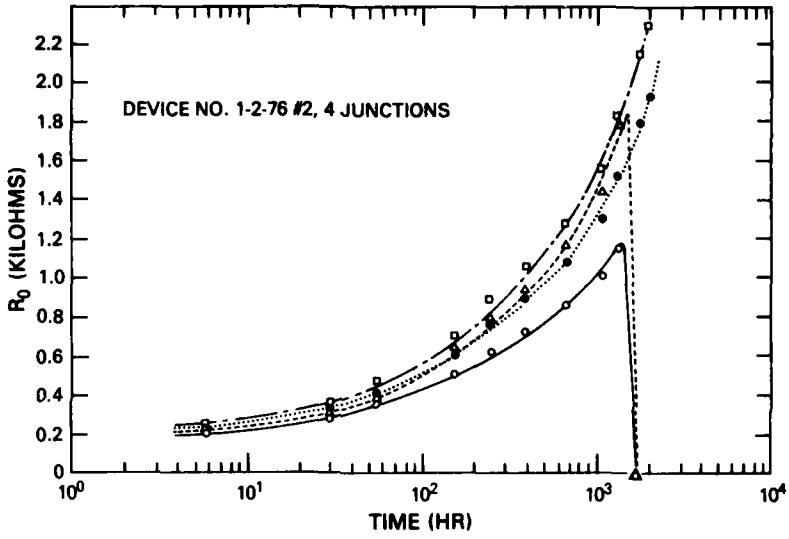


Fig. 9 — The time history of junction resistance R_0 for four typical junctions. The rate of resistance change increases continuously for several hundred to several thousand hours, followed usually by an abrupt self-shorting drop to fractions of an ohm.

An additional type of behavior that is characteristic of the junctions can be seen for two of the junctions, which self-shorted after about 1300 hours. Most of the junctions that were visibly free of fabrication defects lasted beyond a few hundred hours. In a few cases such shorts were observed to repair themselves (return to a resistance along the expected curve) following a voltage spike due to improper switching procedure. This behavior suggests the burnout of a localized filamentary short. The mechanism of the time-dependent behavior is not understood, but it appears to be due to a combination of diffusion at the interface, contaminant absorption, edge effects, and mechanical rupture [9]. However, all of the observed nonlinear and time-dependent behavior is assumed to realistically simulate the numerous microscopic tunneling points of real metal-to-metal contacts.

CAPACITANCE MEASUREMENTS

In addition to the nonlinear resistive characteristics of such junctions an important property which dominates the RF conduction is the junction capacitance. Although most of the junctions fabricated were of small area ($\lesssim 0.015$ cm^2), their capacitance is relatively large because of the small plate separation of ~ 3 nm. The capacitances of the junctions were measured at 1.0 MHz with a Boonton direct capacitance bridge (Model 75D).

Since the thin-film device is not purely capacitive, the bridge will see the equivalent circuit of the device (Fig. 7). At balance the bridge and device admittances Y_B and Y_D respectively, or reciprocally the impedances, will be matched, so that $Y_B = Y_D$ and

$$Y_B = G_B + j\omega C_B. \tag{5}$$

Here the conductance G_B and capacitance C_B are read directly from the bridge. For the typical junction parameters shown in Fig. 7 and for frequencies $f \geqslant 1$ MHz, $R_O >> X_C$, where $X_c = 1/2\pi fC$,

$$Y_D \approx \frac{r_s}{r_s^2 + \dfrac{1}{\omega^2 C^2}} + j \frac{\dfrac{1}{\omega C}}{r_s^2 + \dfrac{1}{\omega^2 C^2}} . \tag{6}$$

By equating the real and imaginary parts of Eq.s (5) and (6), one can obtain the device capacitance C and series resistance r_s in terms of the bridge values G_B and C_B.

In using the value of capacitance thus determined one is assuming that the value of C is not a function of frequency or applied voltage. The value of C showed no measurable change with applied bias voltage up to 300 mV. Cursory experiments at audio frequencies also showed no frequency dependence of C. The use of a fixed value of C for the junction implies that the dielectric constant of the Al_2O_3 is not a function of voltage and that the capacitive part of the junction will conduct UHF linearly. A measurement of the capacitance also permits a determination of the effective plate separation of the junction sandwich. For a parallel-plate capacitor the separation S in nanometers is simply

$$S = 0.8854 \frac{KA}{C} , \tag{7}$$

where K is the dielectric constant for Al_2O_3, A is the plate area in cm^2, and C is the capacitance in μF. For the junction values shown in Fig. 7, $A = 0.0168\ cm^2$ and $C = 0.0329\ \mu F$. For a value of $K = 10.44$ [10] ,Eq. (7) gives a plate separation of 4.7 nm. This separation is considerably larger than either the 2.56 nm resulting from a parameter fit to the $I-V$ characteristics (Forlani-Minnaja expansion in Fig. 7) or the 2.25 nm resulting from the digital thickness monitor measurements discussed in the previous section. This discrepancy has been observed by other investigators [11] and is believed to result from a different averaging process inherent in the different measuring techniques.

When IM effects are calculated, the bridge values of capacitance are used rather than the values calculated from tunneling or weight-gain measurements, since presumably the reactive RF current and voltages will behave in accordance with the effective value of C measured at 1.0 MHz. It will also be seen in the following section that the RF power developed by the junction is an extremely sensitive function of the capacitive reactance.

INTERMODULATION MEASUREMENTS

A block diagram of the basic IM test facility, together with the equivalent circuit of a tunneling junction, is shown in Fig. 10. Each transmitter power output can be varied independently from 0 to 100 W. The maximum third-order (290-MHz) intermodulation generated by the test facility alone is ≤ -140 dBm at 50 dBm (100 W) total power output from the diplexer. The tunneling junctions under test are placed between the diplexer output and 150 m of RG-214 coaxial cable. The cable approximates an infinite transmission line and ideal termination [1]. In the case of the tunneling junction devices, the power levels were adjusted so that $P_1 = P_2$, and the total power input P_I to the junction was usually set to 1.0 W. The total power input of any device was kept below about 4 W because of the limited ability of the thin-film conduction strips to dissipate power.

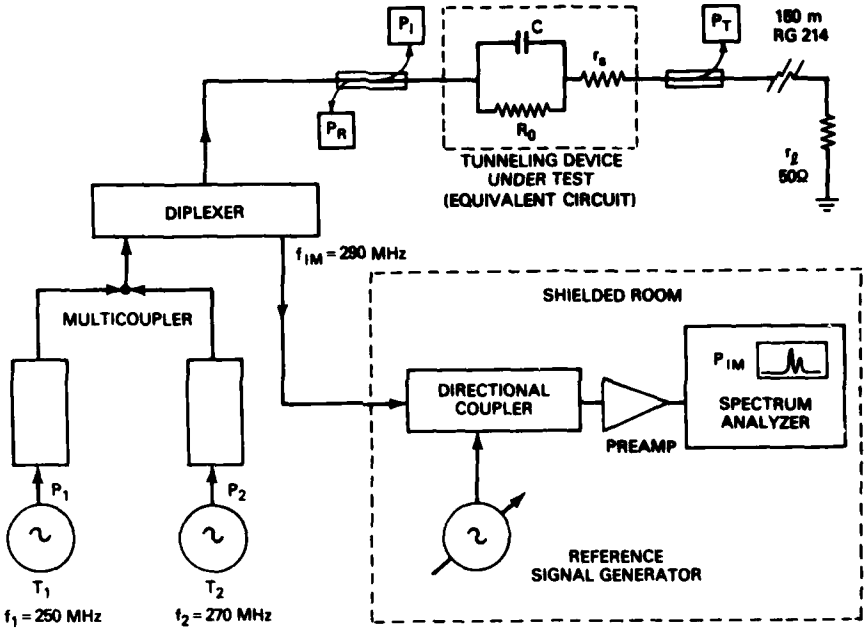


Fig. 10 — NRL intermodulation test facility (operated by the NRL Satellite Communications Branch). The equivalent circuit of a tunneling junction is shown in the normal test configuration.

As shown in Fig. 9, the junction resistance changes with time, making it necassary to measure the $I - V$ characteristics just prior to IM measurements. Capacitance values are also measured prior to such IM test runs, even though the capacitance remains essentially constant with time. The resistance change with time, however, affords the opportunity to measure the intermodulation as a function of junction resistance.

The assembly of junctions, usually in a set of five as shown in Fig. 5, is mounted in a special chassis box (Fig. 11). To minimize the generation of spurious intermodulation, the aluminum chassis box contains no ferromagnetic materials and is fitted with low-IM type-N UHF connectors [1]. The IM power level of the chassis box was tested using a No. 14 bare copper shorting bar soldered between the terminals. These tests consistently showed IM levels of ≤ -140 dBm at an input power of 1.0 W. One must also make sure that the thin-film structure associated with the junction assembly does not produce intermodulation. A measurement of this effect was made by using the film cross-strip on a set of five junctions as a shorting bar. At a 1.0-W input the IM level was measured at -140 dBm ± 5 dB. A thin-film strip, fabricated and mounted in the same way as the junction assemblies, also showed the same results. Therefore any intermodulation above the residual level of -140 dBm ± 5dB would be due to the junction and not caused by the chassis box, connectors, or device structure external to the junction.

Because of the impedance mismatch of the chassis box to the 50-ohm coaxial cable, 45% of the power was observed to be reflected when using a copper shorting bar. When an assembly of junction devices is mounted in the chassis box, the transmitted power P_T drops from 55% to as low as 30%, depending on the size of the junctions. In most cases the impedance of the junction and leads is less than 20Ω. The transmitted power P_T as experimentally measured is used in calculating the RF current passing through the junctions.

To calculate the IM power P_{IM} to be expected from a tunneling junction, we return to Eq. (2). For the tunneling junctions studied, it was found that the experimental DC current-voltage characteristics can be adequately described by

$$I = G_0 V + G_2 V^3. \tag{8}$$

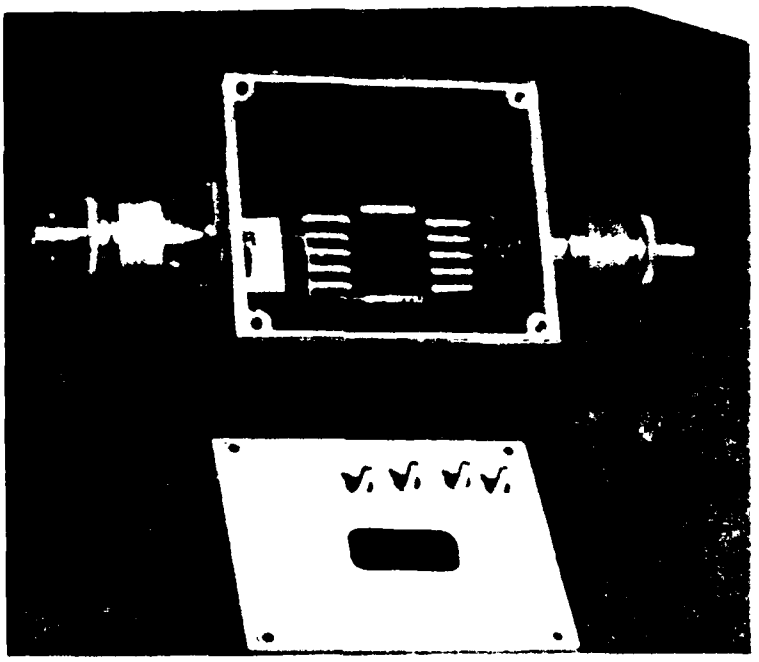


Fig. 11 — Junction assembly mounted in aluminum chassis box. Nonferromagnetic materials and low-IM type-N UHF connectors are employed to minimize spurious generation of intermodulation.

In this case there is no V^2 term, since the tunnelling current is essentially symmetric about the origin of the $I-V$ characteristics. Also, as can be seen from Eq. (3), any V^2 term would produce no IM frequencies falling within the receive band. Equation (8) can be written in a more convenient form by letting $\alpha = G_2/G_O$, so that

$$I = G_0 V + \alpha G_0 V^3, \tag{9}$$

where G_O is the usual linear conductance and α is a ratio that measures the departure from linearity. In the case of an RF input voltage of the form of Eq. (1), the current in Eq. (9) can be written as shown in Eq. (3), except that here only the experimentally observed angular frequencies ω_1, ω_2 and $2\omega_2 - \omega_1$ are retained:

$$\begin{aligned} I(t) &= G_0(V_1 \cos \omega_1 t + V_2 \cos \omega_2 t) + \frac{3}{4}\, \alpha G_0 V_1 V_2^2 \cos (2\omega_2 - \omega_1)t \\ &= I_0(t) + I_{\mathrm{IM}}(t). \end{aligned} \tag{10}$$

It is important to note that $I(t)$ is the RF current through the nonlinear resistance R_O, that $I_O(t)$ is the primary current at frequencies f_1 and f_2, and that $I_{IM}(t)$ is the intermodulation current at frequency $2f_2 - f_1$. The peak *IM* current is just

$$I_{\mathrm{IM(peak)}} = \frac{3}{4}\, \alpha G_0 V_1 V_2^2. \tag{11}$$

The value of V_1 and V_2 can be calculated from the total transmitted power P_T. The transmitted power P_T results from both power inputs P_1 and P_2 running simultaneously at frequencies f_1 and f_2. The circuit parameters of the equivalent circuits in Fig. 7 and 10 have the following ranges, depending on junction area and history:

$$300\ \Omega < R_0 < 100\ \text{k}\Omega,$$

$$0.026\ \Omega < X_c < 0.296\ \Omega \text{ at } 290\ \text{MHz},$$

$$2\ \Omega < r_s < 25\ \Omega,$$

$$r_l = 50\ \Omega.$$

For all cases $R_O >> X_c$, so that essentially all of the transmitted current I_T ($I_T(t) = I_1 \cos \omega_1 t + I_2 \cos \omega_2 t$) passes through the capacitance C and the total transmitted power P_T is dissipated in the series resistance r_s and the load resistance r_l. Since $r_s + r_l$ determines the current, the measured RF power P_T is

$$P_T = \langle I_T^2(t) \rangle (r_s + r_l), \tag{12}$$

where

$$\langle I_T^2(t) \rangle = \frac{1}{T} \int_0^T (I_1 \cos \omega_1 t + I_2 \cos \omega_2 t)^2\, dt.$$

For $P_1 = P_2$ and $I_1 = I_2$, Eq. (12) yields

$$I_{T(\text{peak})} = \left[\frac{P_T}{r_s + r_l}\right]^{1/2}.$$

Since essentially all of this current passes through the capacitance, the voltages V_1 and V_2 developed across the capacitance C and also across the nonlinear resistance R_O are approximately

$$V_1 = X_{C1}\left[\frac{P_T}{r_s + r_l}\right]^{1/2} \text{ and } V_2 = X_{C2}\left[\frac{P_T}{r_s + r_l}\right]^{1/2},$$

where X_{C1} and X_{C2} are the capacitative reactances at f_1 and f_2. With these values of V_1 and V_2, Eq. (11) becomes

$$I_{\text{IM(peak)}} = \frac{3}{4}\,\alpha G_0 \left[\frac{P_T}{r_s + r_l}\right]^{3/2} X_{C1} X_{C2}^2. \tag{13}$$

This IM current is viewed as being generated in the nonlinear element R_O, and as can be seen from Fig. 10, this current source is paralleled by the junction capacitance C and the external load $r_s + r_l$. Again $X_C << r_s + r_l$ at 290 MHz, so that most of this current is shunted through C, and the voltage developed across C and also across the external load $r_s + r_l$ is

$$V_{\text{IM(peak)}} = I_{\text{IM(peak)}} X_{C,\text{IM}} \tag{14}$$

where $X_{C,\text{IM}}$ is the capacitative reactance at 290 MHz. The IM power dissipated in $r_s + r_l$ is then

$$P_{\text{IM}} = \frac{\langle V_{\text{IM(peak)}}^2 \cos^2 (2\omega_2 - \omega_1)t \rangle}{r_s + r_l} = \frac{1}{2}\,\frac{V_{\text{IM(peak)}}^2}{r_s + r_l}. \tag{15}$$

Combining Eqs. (13), (14), and (15), we obtain finally

$$P_{\text{IM}} = \frac{9}{32}\,\alpha^2 P_T^3\,\frac{X_{C1}^2 X_{C2}^4 X_{C,\text{IM}}^2}{R_0^2 (r_s + r_l)^4}. \tag{16}$$

This equation expresses the IM power level in terms of readily measurable junction and test parameters. The RF test facility is set up to measure the IM power level in dBM units (decibels referenced to 1 mW), so that by definition Eq. (16) can be expressed as

$$(\text{dBm})_{\text{IM}} = 10 \log \frac{P_{\text{IM}} \text{ (watts)}}{10^{-3} \text{ (watts)}} \cdot \tag{17}$$

This dBm level of intermodulation is measured directly at the spectrum analyzer by comparison to a reference signal, as shown in Fig. 10.

The junctions initially measured in the IM test facility had relatively large areas ($A \geq 0.015$ cm^2) and were operated at an input power of $\sim$1.0 W. For a transmitter power $P_T = 0.5$ W these junctions showed intermodulation levels of -110 to -135 dBm, depending on the values of R_O and C of the particular junction. Figure 7 shows the $I-V$ characteristics and parameters typical of this type of junction. If one uses the values shown in Fig. 7 together with $P_T = 0.5$ W and $R_l = 50\ \Omega$, a calculation of the IM power from Eq. (16) gives a much lower IM level of -239 dBm.

In the preceding calculation it was assumed that the RF current is uniformly conducting through the film strip and junction area. At input frequencies of 250 and 270 MHz, however, one might expect UHF skin effects to significantly alter the IM level. An accurate calculation of the RF current distribution and resultant intermodulation is an extremely complex problem for the geometry of the thin-film device.

One simplifying approximation, which has been employed by Chapman et al. [2] regarding waveguide flanges, is to restrict the RF current across the tunneling area to the same depth as the skin depth in the adjacent conductor. Although the geometry of a thin-film device differs significantly from a comparitively large flange, a similar restriction arises through edge concentration of RF current in thin-ribbon conductors. (Evidence of such a current distribution was indicated in one set of junctions in which the RF power had exceeded 10 W. Small thermally produced blisters showed a marked concentration along the outer edge of the conducting strips.)

For this approximation the effective RF current is restricted to an average skin depth $\sim 5.16 \times 10^{-4}$ cm around the perimeter of the actual junction area. For the junction parameters shown in Fig. 7, the effective RF conduction area is 2.67×10^{-4} cm^2, or about 1/63 the apparent area. Thus the effective junction resistance and capacitance become $\sim 1.39 \times 10^5\ \Omega$ and $\sim 5.23 \times 10^{-10}$ F respectively. With these effective values of R_O and C, Eq. (16) gives an intermodulation level of -131 dBm, which is within the experimental range observed.

To raise the IM level and to examine any area effects, several sets of five-junction assemblies were fabricated in graded sizes ranging from 0.015 cm^2 down to 0.0015 cm^2. It was found that the size scaling effect on capacitance C and junction resistance R_O essentially followed the expected variation for such junction elements in parallel. As the junction area is decreased, one would expect, according to Eq. (16), the capacitive reactance product ($X_{C_1}^2 X_{C_2}^4 X_{C,IM}^2 \sim X_{av}^8$) to increase much faster than R_O^2 and to result in a rapid increase in P_{IM} with decreasing area. This increase is largely offset, however, by an increase in the ratio of effective RF conduction area to total juction area as the junction area decreases. When the present fabrication techniques are used, a reduction in junction area by more than an order of magnitude becomes impractical, and the power-handling ability becomes too low to conveniently make IM measurements.

Figure 12 shows typical results of IM measurements on a set of five junctions ranging in size from 0.0143 cm^2 to 0.0018 cm^2. The measured value of the junction capacitance is shown along the abscissa. The IM power level in dBm units and in watts is shown on the left and right ordinates respectively. The input power P_I is held constant at +30 dBm (1.0 W). Over the capacitance range examined the transmitted power P_T varies from 0.39 W to 0.51 W but remains relatively constant on the scale shown. The IM level for each junction is plotted as a square point along with the estimated error bars

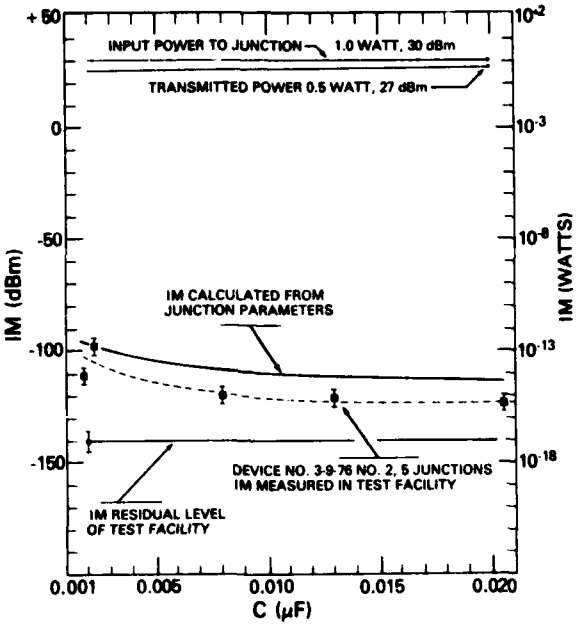


Fig. 12 — Experimental and calculated intermodulation power levels generated by electron tunneling as a function of junction capacitance. Here the variation in capacitance results from a gradation in junction areas ranging from 0.0018 cm^2 to 0.0143 cm^2. All junctions were fabricated simultaneously and have the same oxide thickness of ≈ 2.5 nm. The junction parameters and I-V characteristics for each of the five junctions are shown in Fig. 13.

of ± 3 dB. These experimental points are connected with a smooth dashed curve. The solid curve shows the intermodulation calculated from Eq. (16) using the experimentally measured junction parameters with the values of R_O and C corrected to the effective RF skin depths. The uncorrected junction parameters are shown in Fig. 13. Similar sets of junctions showed the same functional dependence of intermodulation on area and capacitance as illustrated in Fig. 12.

Although the intermodulation calculated in the preceding manner followed the experimental variation with junction parameters fairly well, the absolute magnitude of the calculated IM curve was found to be shifted from the measured intermodulation by ± 15 dB. This agreement is perhaps as good as can be expected in view of the crude RF skin-depth approximations used. Additional smaller effects which are not taken into account by Eq. (16) are distributed resistance and capacitance effects, nonlinear capacitive effects, and frequency-dependent effects.

The equivalent-circuit analysis leading to Eq. (16) tacitly assumes that the film resistance r_f over the junction area is zero. Actually this resistance is generally larger than the capacitive reactance at 270 MHz; for example, in junction 1 of Fig. 13, $X_C = 0.0286\ \Omega$ and $r_f = 0.185\ \Omega$. Thus, one should consider the effects of distributed resistance and capacitance along the junction area. With use of a two-dimensional model, both an exact continuous-distribution analysis and a numerical discrete analysis showed a voltage distribution $V(x)$ that varied significantly along the width of the junction. In addition the RF edge conduction along the film strip leading up to the junction area will increase the effective value of r_s. A calculation of these two combined effects for junction 1 of Fig. 13 using a discrete four-segment equivalent circuit yielded a value of P_{IM} which differed from the previously corrected value by only 1 dB. However, this small net correction is due to the fortuitous approximate cancellation of the two effects for the geometry chosen.

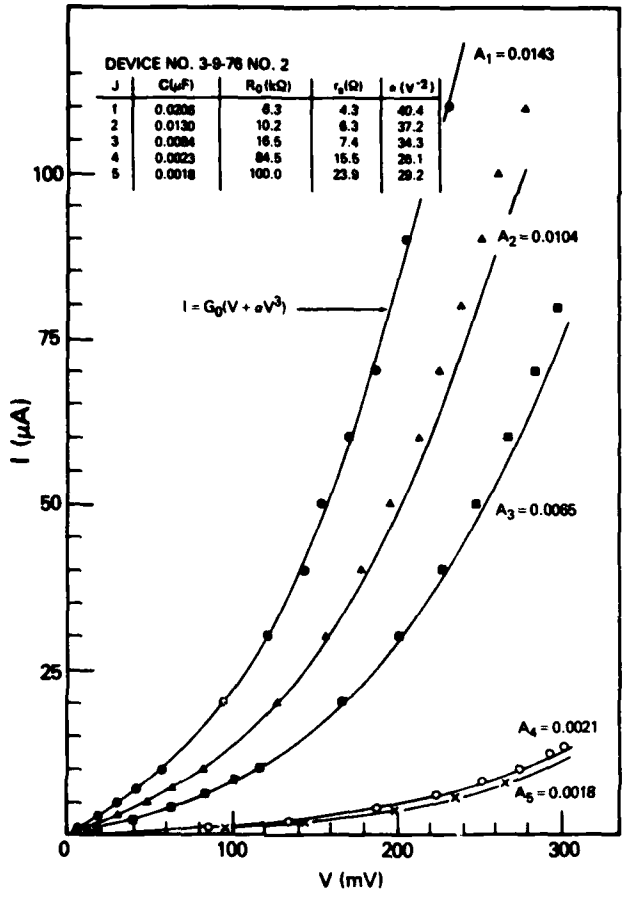


Fig. 13 — Current-voltage characteristics as a function of area and capacitance for the five junctions measured in Fig. 12. The corresponding junction parameters are shown in the upper-left table.

The α of Eq. (16) is derived from the nonlinear DC resistance characteristics of the junction. This nonlinearity is presumed to arise solely from electron tunneling, as described previously. However, known nonlinear capitance effects can arise through nonlinear dielectric and electrostriction effects [12]. Electrostriction can also change the tunneling resistance through a volume change. A rough calculation of the volume-change effect predicted an intermodulation well below the level of detectability.

Equation (16) is also applied to RF power measurements of P_{IM} and P_T in the UHF range (240 MHz to 290 MHz), whereas the parameters α, R_O, and r_s are determined from DC measurements and C is measured at 1 MHz. The assumption of frequency independence seems reasonable, since the resistivity and dielectric constant of Al_2O_3 appear to be relatively constant over a wide range of frequencies [13,14].

ION IMPLANTATION EFFECTS

To eliminate or reduce the IM effects of tunneling junctions in metal-to-metal contacts, one apparently must either eliminate the surface insulating films (oxides or otherwise) or basically modify the structure of the metal oxide surface. Fabrication of full-scale metal hardware that is essentially free of metal oxide films both at the surfaces and under metal platings such as gold would appear to offer a

reasonable solution. Alternatively, modification of the metal oxide surfaces should ideally make any contacts more conductive, linear, and free of time-dependent effects. We have explored implantation of metallic ions in the oxide surface as a possible way to achieve the desired results.

Several tunneling junctions were fabricated similarly as previously described, except Ag^+ ions of $\sim$2 keV energy were implanted into the Al_2O_3 surface prior to the evaporation of the second overlapping aluminum strip. Theoretical estimates indicate that 2-keV Ag^+ ions have a mean range in Al_2O_3 of $\sim$1.7 nm. Ideally these silver atoms will have a Gaussian distribution about the mean range with a standard deviation of $\sim$0.6 nm. A plot of Gaussian distribution shows range straggling extending from the surface to $\sim$3.2 nm, with about 12% of the silver atoms stopping in the aluminum beyond the Al_2O_3

The actual distribution of the silver atoms in the Al_2O_3 and Al may be quite different than just described because of three effects that have not been measured:

- Uncertainty in the low-energy range of Ag^+ ions in Al_2O_3 (existing data indicate that heavy-ion ranges in light substrates may be as much as twice the theoretical estimates),
- Removal of the Al_2O_3 surface by ion sputtering, and
- Oxide regrowth when the film is removed from the vacuum environment.

To evaluate the electrical characteristics of such implanted junctions, standard reference monitor junctions were also fabricated simultaneously on the same substrate and under the same conditions without ion implantation. All electrical measurements are then referenced to the monitor junctions. This procedure is necessary because different batches of junctions will generally have different values of R_O and because the electrical characteristics will be measured at different times.

Initially two sets of junctions (three reference junctions per set) were implanted. One set was implanted with a fluence of $\sim 10^{15}$ atoms/cm^2, and the other set was implanted with a higher fluence of $\sim 10^{16}$ atoms/cm^2. The junction implanted at the higher fluence showed a sharp drop in resistance from an average resistance of 36.3 kΩ to an average resistance of 13.5 Ω, or a ratio of implanted resistance R_O(Ag) to reference resistance R_O(Ref.) of $\sim 4\times10^{-4}$. The junctions implanted at lower fluence, however, unexpectedly showed an increase from 12.3 kΩ to 23 MΩ, or a ratio $R_O(Ag)/R_O$(Ref.) = 1.9×10^3. The different implant fluences produced opposite effects of change in resistance that differed by over six orders of magnitude!

In view of this surprising result a more elaborate set of junctions were fabricated to facilitate measurement of the junction resistance as a function of the fluence of silver atoms implanted. Figure 14 shows the results of these measurements. The ratio $R_O(Ag)/R_O$(Ref.) is plotted on the ordinate, and the fluence of silver atoms and of charge are plotted along the top and bottom abscissa respectively. The solid line connects points measured immediately after fabrication, and the dashed line connects points measured about 2 hours later. The resistances are changing rapidly during this initial period, but the ratio remains relatively constant in time. The two square points are from the original measurements and are seen to be consistent with the more detailed later results. The error bar on one square point is significantly higher than the errors for the other points and resulted from an equipment-produced uncertainty in the charge fluence at this one point. The increase in junction resistance for implants up to $\sim 4\times10^{15}$ atoms/cm^2 was completely unexpected, and the detailed physical mechanisms responsible for the functional behavior of $R_O(Ag)/R_O$(Ref.) with the silver implant fluence are not yet fully understood. Recent work [15] shows that ion implantation in the oxide surface of metals has the effect of increasing the oxide thickness. This effect would account for the initial increase in tunneling resistance. As the silver density is increased, the oxide conductivity apparently increases faster than the competing thickness-growth, resulting in a low-resistance junction. For those junctions implanted to fluences of $\geq 10^{16}$ atoms/cm^2 the resistance R_O was generally reduced to a few ohms or fractions of an

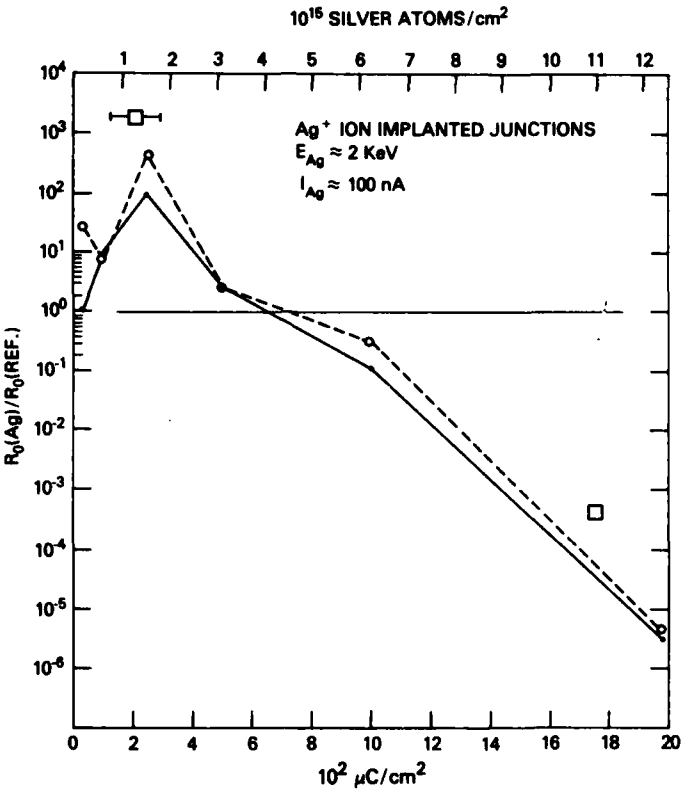


Fig. 14 — Effect on junction resistance of implantation of 2-keV silver ions in the Al_2O_3 surface as a function of fluence of silver atoms. The ratio of ion-implanted resistance R_0(Ag) to an identical unimplanted resistance R_0 (Ref.) is plotted on the ordinate.— initial; --- after 2 h.

ohm. The current-voltage characteristics of the implanted junctions showed better linear behavior than unimplanted junctions over the same current range. Although these junctions had greatly improved conduction characteristics, they continued to show the familiar time-dependent increase of R_O. (See Figure 14).

Preliminary test of these junctions at the IM test facility confirmed that the low resistance implanted junctions did indeed have a low intermodulation of ~ -135 dBm. For such an implantation technique to be useful on full scale hardware, however, one would have to solve the formidable problems of the time-dependent behavior and mechanically fragile surfaces.

CONCLUSIONS

From our experimental and theoretical investigation of the generation of intermodulation by electron tunneling in Al–Al_2O_3–Al junctions, we list the following conclusions:

- Tunneling junctions ranging in areas from ~ 0.015 cm^2 down to ~ 0.0015 cm^2 showed IM levels of ~ 150 dB to $\sim$ 110 dB below the total transmitted power of 0.5 W. This corresponds to a P_{IM}/P_{sig} ratio of 10^{-15} to 10^{-11} respectively. Since ideally the ratio P_{IM}/P_{sig} should be $\lesssim 10^{-19}$, the electron tunneling mechanism is seen to be a significant source of intermodulation.

- The functional dependence of the IM power on the effective RF junction parameters appears to be satisfactorily described by Eq. (16).

- Although the main features of the current-voltage characteristics of the junctions are described by the Forlani-Minnaja theory, an accurate fit to the experimental data was not always possible over the entire range of experimental values.

- No significant effects on hysteresis in the $I-V$ characteristics were observed. However, a small resistance asymmetry was always observed at the higher voltages. If the film strip oxidized first was operated at a positive polarity, the junction resistance was lower than the reverse polarity by 3% to 7%. Thus the junctions exhibit a small rectifying component that is thought to arise from the asymmetry of the Al_2O_3 density profile at the two interfaces.

- In well-characterized $Al-Al_2O_3-Al$ junction devices the time-dependent behavior is prominent, with the junction resistance increasing continuously for several hundred to several thousand hours, followed usually by an abrupt self-shorting drop to fractions of an ohm. During such resistance increase, the capacitance remains essentially constant. Such time-dependent behavior is assumed to take place randomly in the numerous microscopic tunneling points in real hardware contacts. This is a mechanism that could contribute to the erratic fluctuation of intermodulation observed in macroscopic contacts.

- Ion implantation of $\gtrsim 10^{16}$ silver atoms per cm^2 in the oxide surfaces produced low-resistance and low-IM junctions but failed to stop the characteristic increase of junction resistance with time. Our initial implantation results are sufficiently promising, however, to warrant further investigation of the possible beneficial effects of employing different metallic-ion species, greater implantation depths, and higher fluences.

ACKNOWLEDGMENTS

We are particularly indebted to B.J. Faraday, head of the Radiation Effects Branch, for his continuous guidance and assistance in addressing the various problems in all phases of the work.

We express special appreciation to Charles E. Young of the Satellite Communications Branch for his thorough and patient accommodation of the authors' numerous requests during device-testing at the intermodulation test facility.

We thank J. K. Hirvonen of the Materials Modification and Analysis Branch for his able consultation and assistance during the ion implantation experiments and also for the generous allocation of beam time on the newly installed ion implantation facility.

We thank Warren L. Bendel of the Radiation Effects Branch for a thorough review of the manuscript and for making important corrections to our analysis.

We also express an appreciation to Edwin P. Westbrook of the Radiation Effects Branch for his help on RF analysis problems during numerous consultations.

REFERENCES

1. C.E. Young, "The Danger of Intermodulation Generation by RF Connector Hardware Containing Ferromagnetic Materials," National Electronic Packaging Conference, 76-West Connector Symposium/Proceedings, Anaheim, Calif., Feb. 23-26, 1976. See also Chapters II and III, this report.

2. R.C. Chapman, J.C. Darlington, A. Savarin, R. Steinberg, A. Paul, and R. Moss, "Intermodulation Generation in Normally Passive Linear Components," Philco-Ford Corporation Report WDL-TR5242, Aug. 24, 1973.

3. W.H. Higa, *Proc.IEEE* **63**, 306 (1975).

4. C.B. Duke, *Tunneling in Solids*, Acedemic Press, New York, 1969, p. 60.

5. F. Forlani and N. Minnaja, *Il Nuovo Cimento* **31**, 1246 (1964).

6. C.S. Guenzer, *Proc. IEEE* **64**, 283 (1976).

7. R. Stratton, *J. Phys. Chem. Solids* **23**, 1177 (1962).

8. W.F. Brinkman, R.C. Dynes, and J.M. Rowell, *J. Appl. Phys.* **41**, 1915 (1970).

9. C.S. Guenzer, C.D. Bond, and C.A. Carosella, *Bull Am. Phys. Soc. II*, **21**, 273 (1976).

10. AIP Handbook, 2nd edition, McGraw-Hill, 1963, p. 9-93.

11. K.H. Gundlach and G. Heldmann, *Solid State Communications* **5**, 867 (1967).

12. G.H. Stauss, Chapter V, this report.

13. G.V. Samsonov, *The Oxide Handbook*, IFI/Plenum, 1973, p. 266 and 315.

14. H.P. Westman and J.E. Schlaikjer, *Reference Data* for *Radio Engineers*, ITT Corp., 1962, pp. 62 and 63.

15. V. Ashworth, D. Baxter, W.A. Grant, and R.P.M. Procter, *Corrosion Science* **16**, 775 (1976).

Chapter II

THE DANGER OF INTERMODULATION GENERATION BY RF CONNECTOR HARDWARE CONTAINING FERROMAGNETIC MATERIALS*

Charles E. Young

Satellite Communications Branch
Communications Sciences Division

INTRODUCTION

Not only satellite communications, but all systems which utilize multicarrier transmitters and readily achievable low noise receiving systems (NF $\leqslant$ 6 dB), are vulnerable to the IMG self-interference problem.

Figure 1 shows a portion of the diplexed system typically employed in earth terminals. This system consists of two filters, one for the receive band and one for the transmit band, coupled to a common output port to permit simultaneous operation with a single antenna.

Intermodulation generation by all elements in the system, including connectors, adapters, filters, etc., shown in Figure 1, plus transmission hardware to and including the antenna itself (not shown) should be at least 10 dB below the weakest desired carrier, nominally in the order of -130 dBm. Great care is required in the design of a diplexed system to ensure an IMG level $\leqslant -140$ dBm with transmitted power in the order of $+50$ dBm. One very important requirement is the selection of RF transmission connector hardware without ferromagnetic materials, the subject of this paper.

INTERMODULATION TEST SET

Figure 2 is a simplified block diagram of the intermodulation test set built to measure RF connector IMG levels. Transmitter signals f_1 and f_2, at 250 and 270 Megahertz (MHz) respectively, were coupled via the six cavity multicoupler to the transmit port, T_x, of the interdigital filter type diplexer. The output of each transmitter was controllable from 0 to 100 watts. IMG signals appearing at the receive port, R_x, of the diplexer were fed to a high gain, low noise, receiver/spectrum analyzer in the shielded room. By utilizing the best constructional techniques available, the 3rd order IMG component, $f_3 = 2f_2 - f_1 = 290$ MHz, of the test set alone was made less than -140 dBm with $+50$ dBm total power from the diplexer output (antenna) port. This represents a test set residual intermodulation conversion of -190 dB or less. Measurement of 3rd order IMG was made since this lowest order component is the largest interference generated in a normal system.

The device under test (DUT), connector hardware in the subject tests, was placed between the diplexer output port and 500 feet of RG-214/U coaxial cable as indicated in Fig. 5. The cable approximates an infinite transmission line and ideal termination (return loss >30 dB). If the DUT being measured is more linear than the test set, no change in IMG level occurs. If the DUT is less linear than the test set, the IMG level increases.

*Presented at Ninth Annual Connector Symposium, Cherry Hill, N.J., 20-21 Oct. 1976.

PRECEDING PAGE BLANK-NOT FILMED

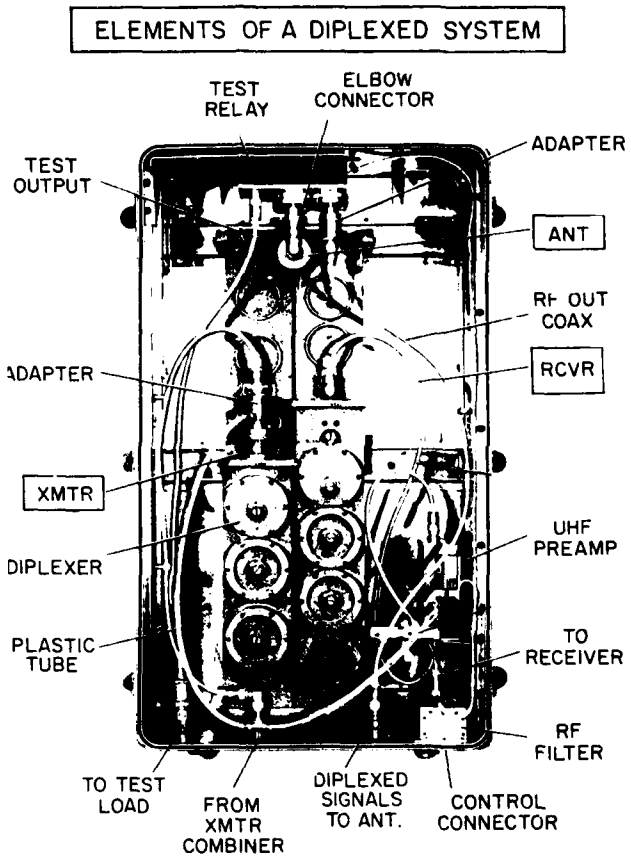
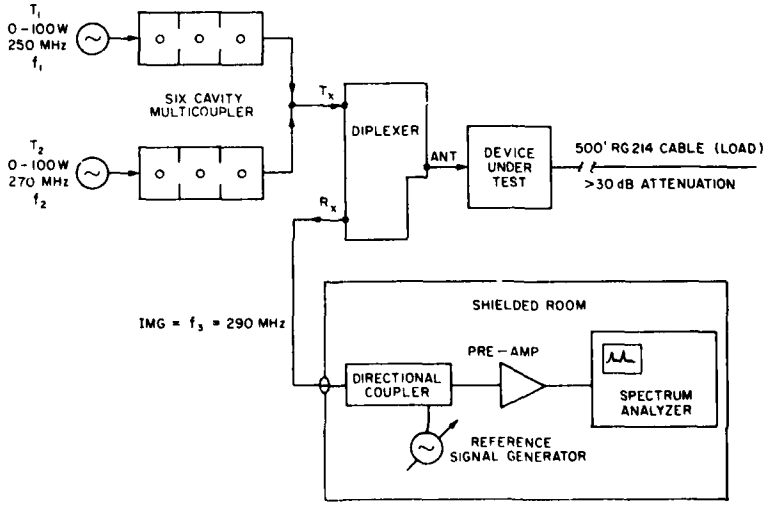


Figure 1



INTERMODULATION TEST SET

Figure 2

Figure 3 is a view of the transmitting equipment, multicoupler, diplexer, and cable load portion of the test set. Although two diplexers are shown, only one was required to measure the reflected IMG produced by connector samples placed between the diplexer output and cable load. The two electromagnets, in the foreground, were utilized as simple diagnostic tools. If IMG is predominantly due to ferromagnetic non-linearity, application of either an external axial or transverse field reduces the materials' permeability and hence the observed IMG level.* If IMG is not due to ferromagnetic materials, no control is obtained.

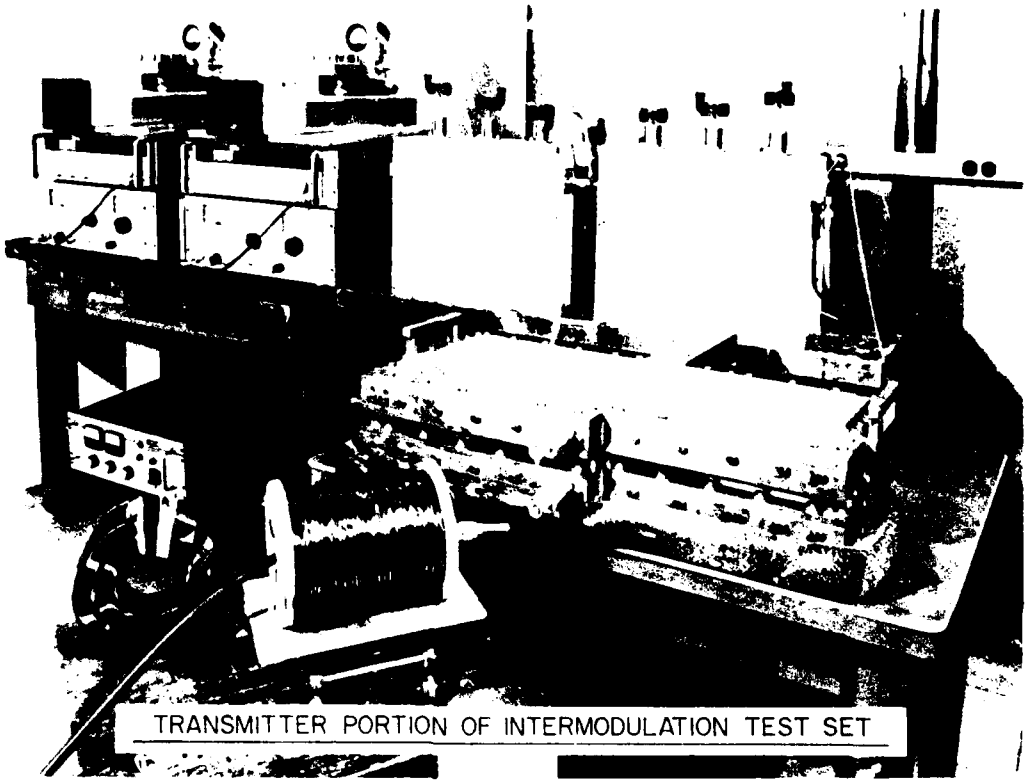


Figure 3

Figure 4 is a view of the test equipment in the shielded room. The basic receiving system utilizes a three cavity filter for IM component selection, a reference attenuator, a low-noise preamplifier and the spectrum analyzer. A calibrated signal generator is coupled into the line via a directional coupler to determine the intermodulation component level.

INTERMODULATION TESTS

In the search for improved RF hardware, very large intermodulation generation by certain types of connectors was discovered. Many of the newer high quality precision devices, such as the double jack N barrel adapter shown in Fig. 5A, utilize stainless steel outer conductors. These devices, even when gold plated, were found to produce much higher IMG levels than ordinary silver plated brass (non-ferromagnetic) connectors and adapters, such as the UG-29 and UG-57 shown in Fig. 5B.

*Conceptually, IMG in a ferromagnetic material is reduced because of decreased permeability while being subjected to a large steady magnetic field. Ideally, the material should be magnetically saturated to reduce the permeability to unity (air). This is not practical, however, with materials and insulation gaps involved in actual RF connector hardware.

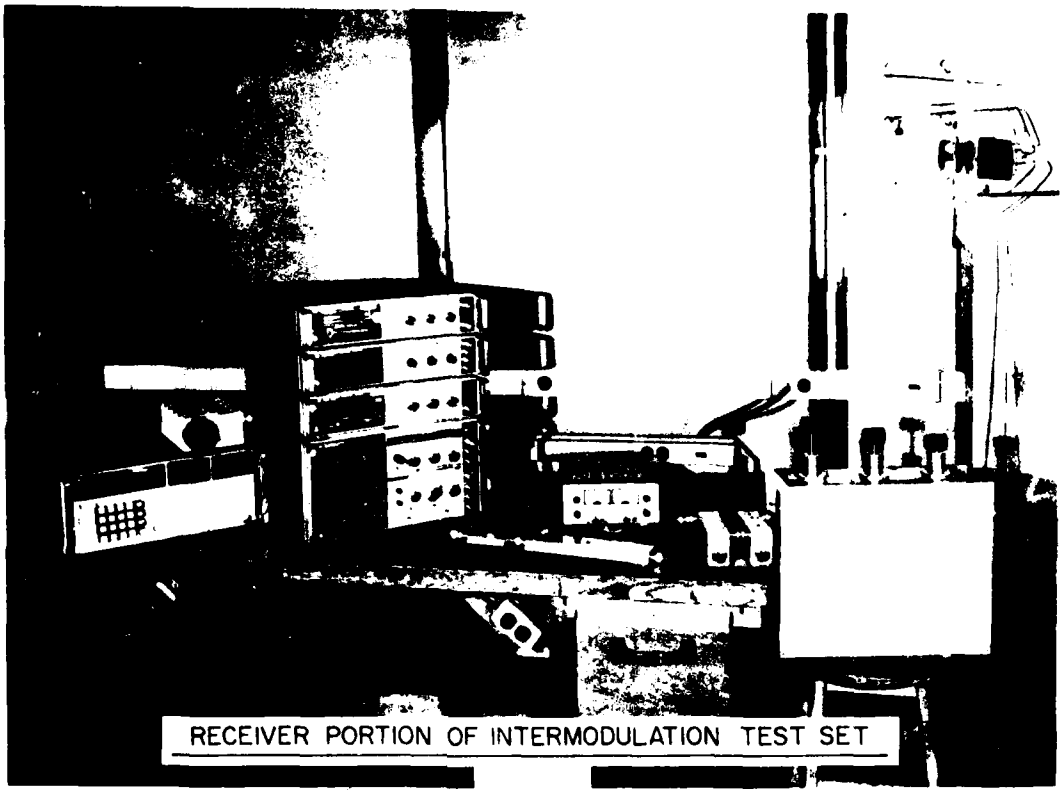
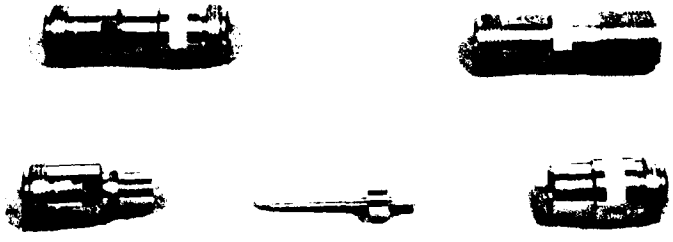
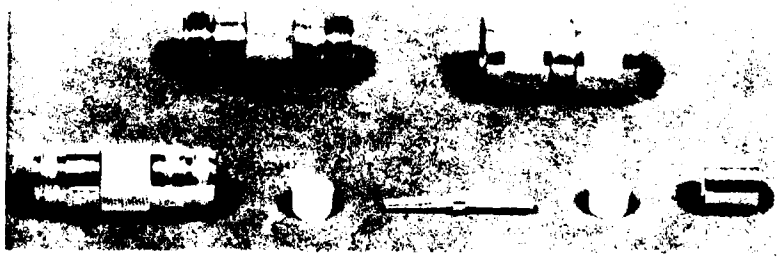


Figure 4



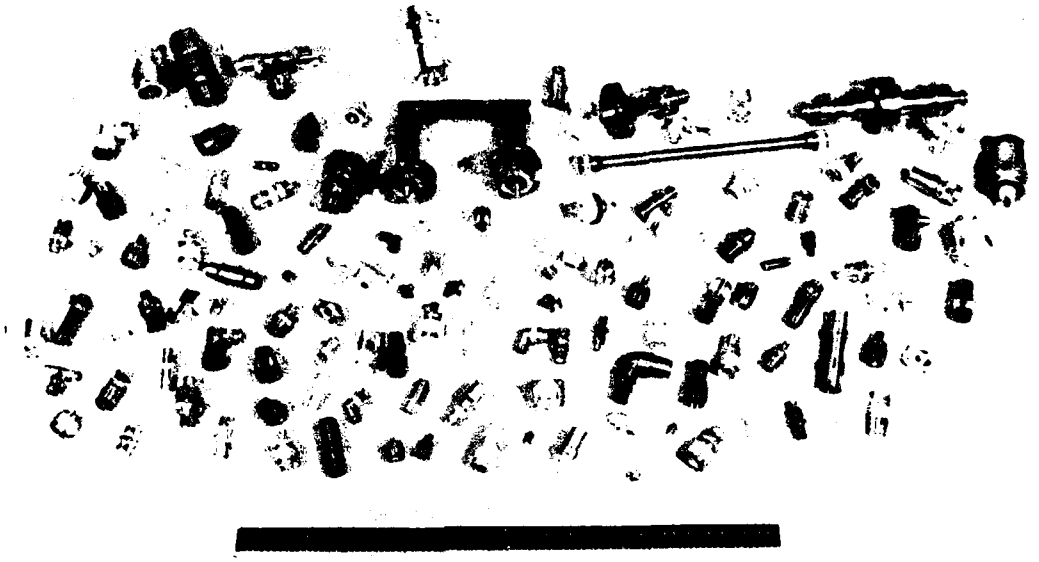
(a) Stainless steel precision adapters



(b) Silver plated brass adapters

Figure 5

The variety of available connectors and adapters is almost endless. A collection of common type coaxial connectors and adapters is illustrated in Fig. 6. A considerable number of connector types with and without ferromagnetic materials have been tested for IMG. However, only a few generic types will be reported here. The precision double jack N barrel adapter, previously shown in Fig. 5A, was selected for extensive IMG testing because its stainless steel body, which serves as the coaxial outer conductor, could be readily duplicated from other metals. Figure 7 is a close-up view of this device connected into the test bed. The axial magnet has been slid away for viewing.



Collection of coaxial connectors and adapters
Figure 6

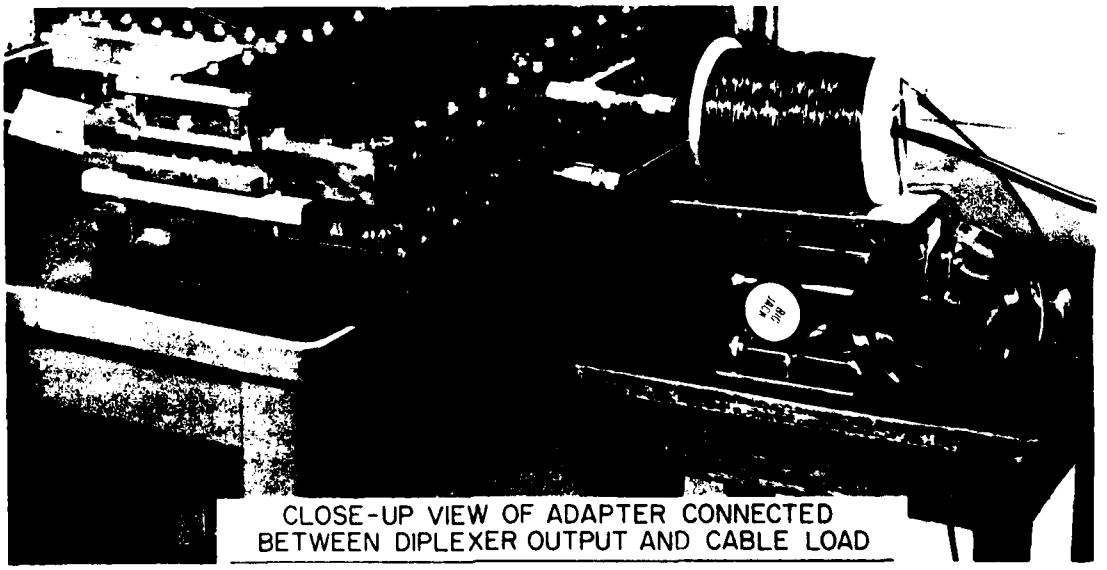
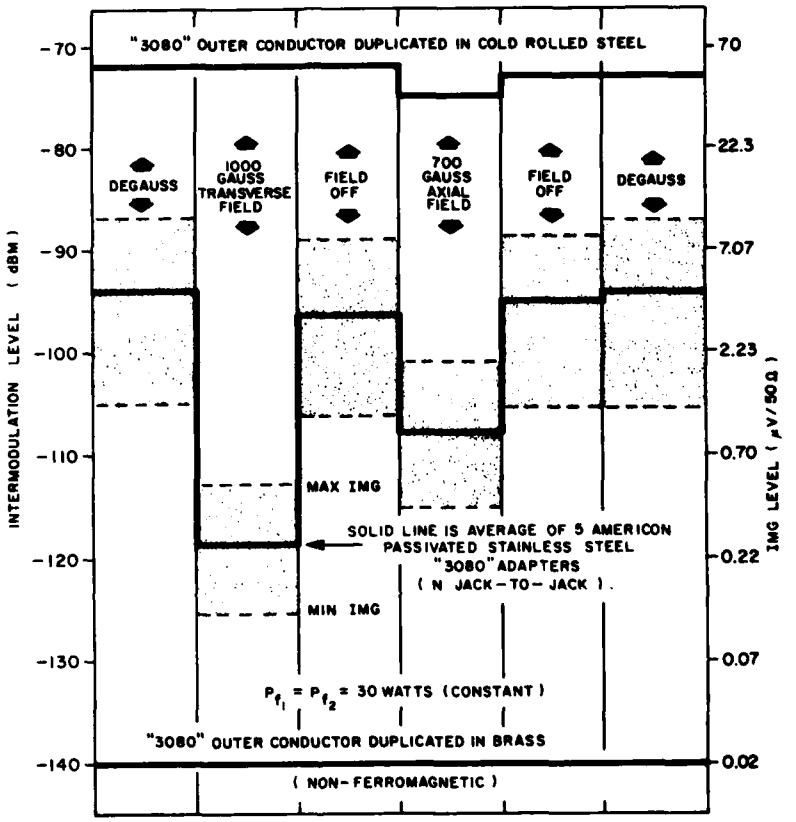


Figure 7

Stainless Steel Non-linear Effects

Third order IMG characteristics of 5 stainless steel N barrel adapters, using constant 60 watt total RF drive, are shown in Fig. 8. The shaded areas indicate the spread of IMG among the samples; the solid line represents the averaged values for the 5 samples. The columns indicate the external magnetic field conditions employed.



PRECISION ADAPTER COMPARISON

CONTROL OF IMG WITH EXTERNAL MAGNETIC FIELDS

Figure 8

Because of hysteresis effects and residual magnetism retained by ferromagnetic materials, an initial degaussing operation (using sinusoidal drive to the external electromagnets and described later) was employed to establish the "worse case" IMG level of each DUT. After degaussing, an average IMG level of −94 dBm was obtained as indicated in the first column. Application of the 1000 gauss transverse direct current (dc) magnetic field at the surface of the connector body, column 2, reduced the IMG level to −118 dBm, a drop of 24 dB. This is a definite indication of permeability reduction in the stainless steel outer conductor elements of the DUT. Removing the dc field, column 3, resulted in an IMG level of −96 dBm, a reduction of 2 dB from the initial degaussed condition, which indicates a change in residual magnetism. Applying the 700 gauss axial dc field, column 4, dropped the IMG level to −108 dBm or −14 dB relative to the initial degaussed level. Removing the field, column 5, gave −95 dBm and degaussing, column 6, again gave −94 dBm, the initial value measured.

To further verify that the non-linearity was in fact due to the stainless steel outer conductor elements, identical structures were machined from brass and reassembled with the original berryllium copper inner conductor. The adapter IMG then measured −140 dBm (the test bed residual) and no control was observable with external magnetic fields, as shown by the bottom curve in Fig. 8. The urgency to discontinue use of stainless steel in RF connectors is clear.

The top curve in Fig. 8 shows the effect of using cold-rolled steel, a high permeability ferromagnetic material, for the outer conductor. With this material, the IMG level measured −72 dBm, with only a 3 dB drop for the axial field. In later tests, using stronger fields to further reduce permeability, larger reductions of IMG were achieved. However, this material, because of its high permeability, terminates most of the magnetic field at its outer surface where minimum or no RF current flows. Hence, no control was observed with the transverse field. The axial field, however, was effective in reducing permeability at the inner surface of the cold-rolled steel outer coaxial conductor. Stainless steel, type 303, by contrast, is specified to have a permeability less than 2. Actually, permeability of the stainless steel devices measured were less than 1.1. This low permeability allows either field to penetrate to the inner surface of the material and decrease permeability where RF current is flowing.

Figure 9 shows oscillographic measurements of 3rd order IMG obtained from one of the precision adapters used in collecting data for Fig. 8. Frames A through D indicate IMG levels for the various

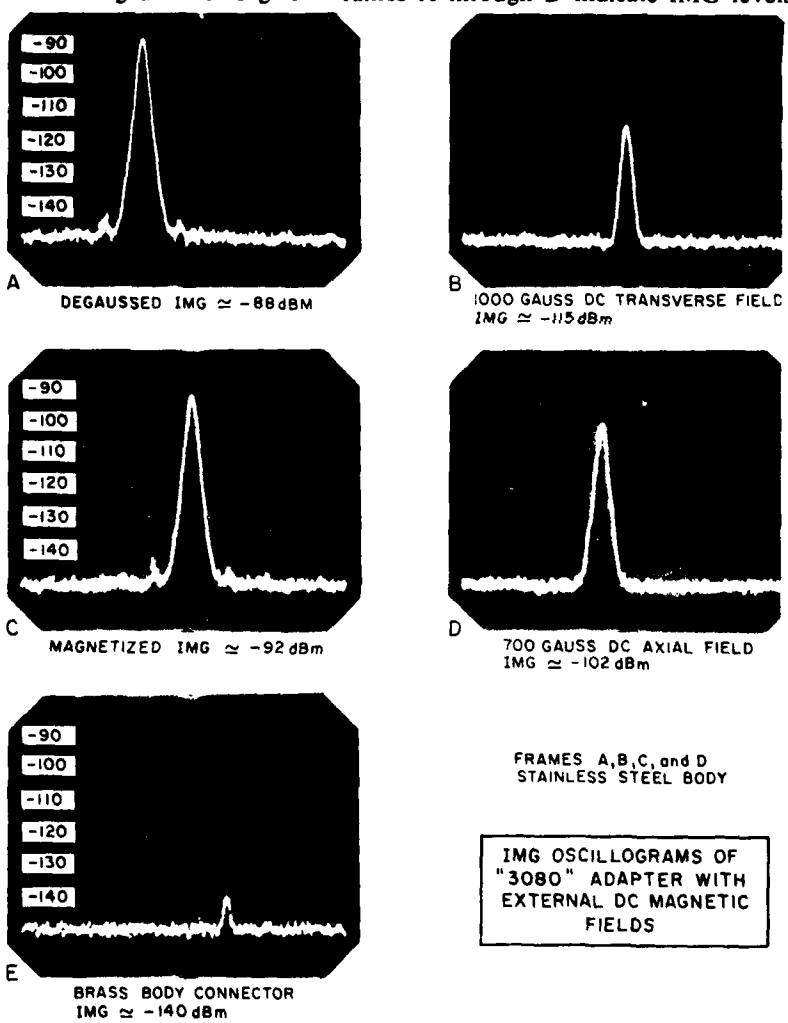


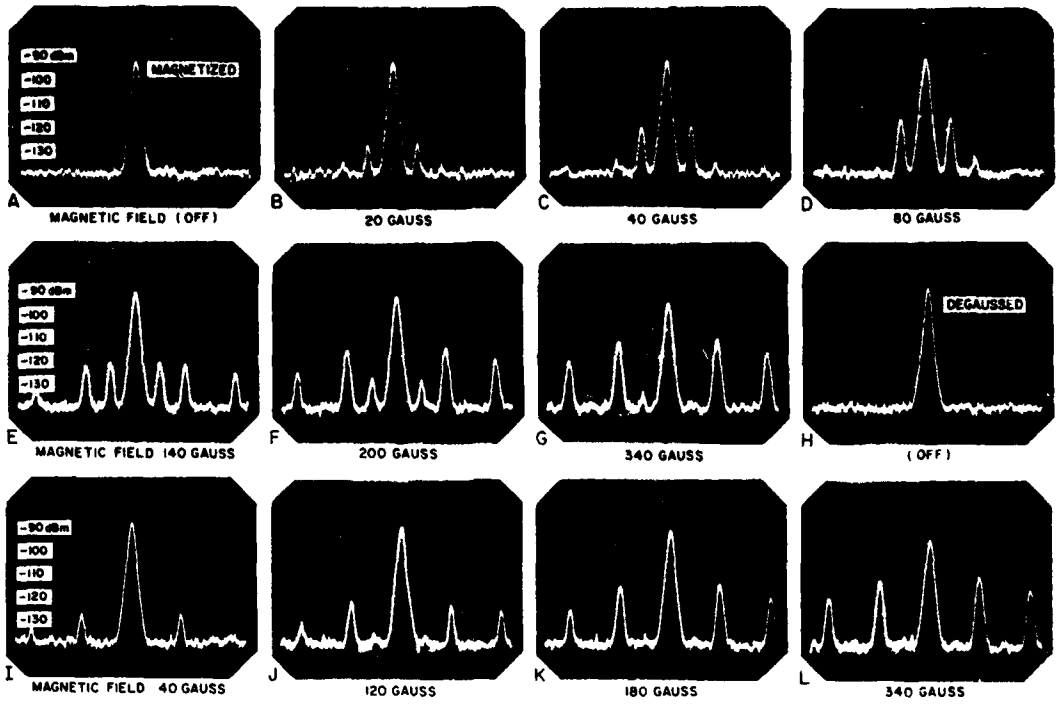
Figure 9

magnetic field conditions with the original stainless steel conductor elements. Frame A shows the degaussed IMG level, −88 dBm, after subjecting the adapter to large sinusoidal (transverse and axial) demagnetizing fields. Frame B shows the reduced IMG level, −115 dBm, with application of the 1000 gauss dc transverse field, a drop of 27 dB and indication of reduced permeability in the stainless steel body. Frame C shows an IMG level of −92 dBm after the field was removed. The 4 dB reduction of IMG from the initial degaussed condition indicates residual magnetism and reduced permeability. Frame D shows an IMG level of −102 dBm during application of the 700 gauss axial field, a drop of 14 dB from the initial degaussed condition. After degaussing, the DUT again produced an IMG level of −88 dBm, as in Frame A. The sample adapter IMG and dc magnetic field control characteristics are similar to the "averaged" device data shown in Fig. 8. Frame E, like the bottom curve in Fig. 8, shows the dramatic improvement obtained by replacing ferromagnetic (stainless steel) with nonferromagnetic (brass) conductor elements; IMG ~−140 dBm, an improvement of ~50 dB.

Several IMG spectrum characteristics are experimentally evident when a device containing low permeability ferromagnetic material is subjected to an external alternating current (ac) magnetic field. If residual magnetism is present, a relatively small ac field will in effect modulate the residual field of the DUT, forming RF sidebands spaced at the magnetic drive frequency from the IMG carrier. If the ac field is increased beyond some critical magnitude, the spectrum will suddenly "flip" to frequency doubled sidebands, indicating that demagnetization of the device has occurred. With no residual field, IMG depression occurs for either positive or negative magnetic field excursions, and hence at twice the magnetic field frequency. The DUT will remain in the demagnetized (degaussed) state at lower or higher ac drive levels until remagnetized by a large dc field. Experimental data of these phenomena are described next.

The oscillograms of Fig. 10 illustrate IMG characteristics obtained when the sample stainless steel adapter was subjected to a range of sinusoidal ac transverse magnetic fields, the largest of which was used for the degaussing operation. Frame A shows the "magnetized" IMG level, −92 dBm, due to a previously induced 1000 gauss dc transverse field and with no ac field applied. In Frame B, application of a small amplitude (±20 gauss) 2 kilohertz (kHz) transverse magnetic field in effect modulates the residual magnetic field of the DUT, forming RF sidebands spaced at the magnetic drive frequency from the IMG carrier. Increasing the field to ±40 gauss, Frame C, also doubled the apparent fundamental modulation sideband amplitudes, indicating a relatively linear modulation process. In Frame D the modulating field was increased another 6 dB (to ±80 gauss) but the sidebands only increased 4 dB, an indication that the peak ac field is exceeding the residual field of the device. Increasing the field to ±140 gauss, Frame E, decreased fundamental sideband amplitudes but increased second harmonic sidebands markedly. Note also the appearance of relatively strong 4th harmonic sidebands but essentially no 3rd harmonic components. A modulating field of ±200 gauss, Frame F, resulted in fundamental sidebands <2nd or 4th harmonic sidebands and IMG carrier depression ~2 dB. With a slightly larger field, the modulation spectrum flipped to even order sidebands only, the indication of residual field erasure or degaussed state of the DUT. (The precise flip point is affected by hysteresis and difficult to photograph.) To assure full erasure of residual magnetism, the ac field was raised well above the flip point, to ±340 gauss in Frame G. Removing the field, as in Frame H, produced maximum IMG (degaussed) response, −88 dBm, as also obtained in Frame A of Fig. 9. Frames I through L show only even order sidebands, verifying the degaussed condition of the DUT. Compare Frames C and I (same modulating field) before and after degaussing, respectively. Remagnetization with a dc field (or simultaneous use of ac and dc fields) will again produce the fundamental sideband/IMG carrier relationships previously obtained. Use of this modulation measurement technique provides a simple but effective means of determining the magnetic state of RF connector hardware.*

*The ferromagnetic modulation mechanism is analogous to the diaphragm motion in an earphone with sine wave excitation. If a polarizing magnet were not used, a sine wave of current sent through the earphone coil would attract the diaphragm during both the positive and negative alternations, producing two cycles of motion during each electrical cycle. When a permanent field is added, the diaphragm will be flexed even when no current is flowing. The diaphragm is flexed more or less by an ac signal; one cycle of current producing one cycle of mechanical motion.



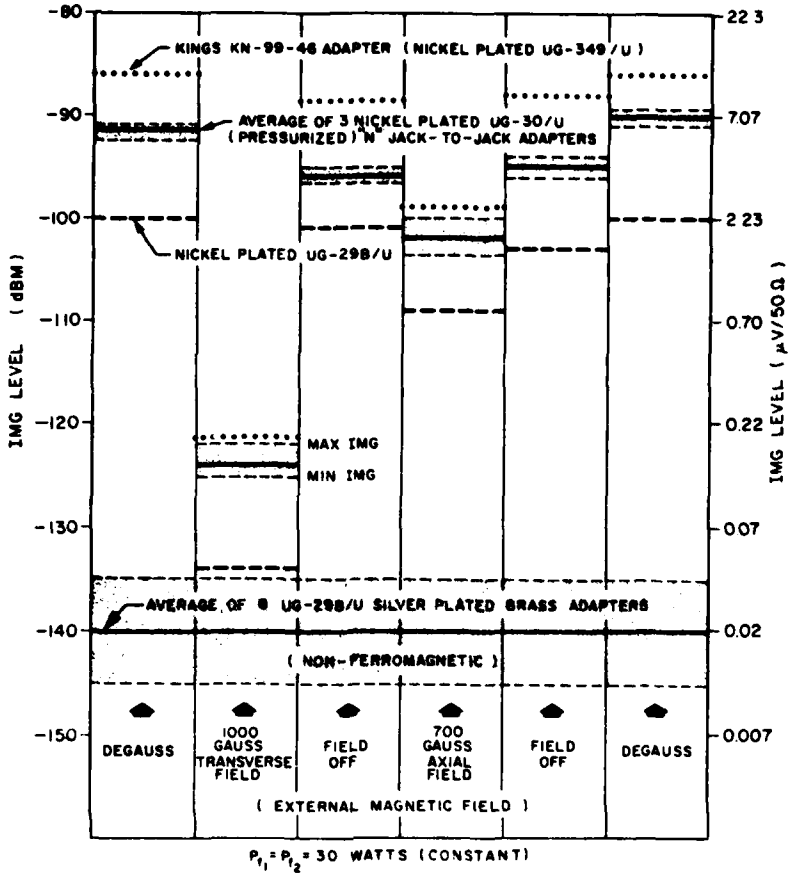
IMG OSCILLOGRAMS OF "3080" STAINLESS STEEL ADAPTER (SAMPLE A)
WITH EXTERNAL 2kHz AC MAGNETIC FIELD (PEAK)
Figure 10

Axial field modulation was also investigated and found to be similar in character to transverse field modulation. Degaussing was only obtainable with low permeability materials, such as stainless steel or nickel plated devices to be discussed. Kovar, a high permeability material used in hermetically sealed connectors, exhibited residual fields that did not yield to degaussing, using fields as large as 500 gauss (limit of the test set). IMG characteristics of Kovar devices will be covered later. Some devices appear to lose or gain residual magnetism during storage or handling. It should also be noted that exposure to large ac magnetic fields, as during the degaussing operation, for a significant period of time creates an appreciable temperature rise because of induction heating (eddy-current flow) in all metal elements of the connector or adapter. This temperature rise can induce mechanical stresses, create contact non-linearity type IMG, deform insulating material, and in some instances, permanently damage the device.

Nickel Plating Non-Linear Effects

Figure 11 shows 3rd order IMG levels of several connector types with nickel plating, another ferromagnetic material. Test set conditions were as described for the stainless steel devices. Nickel plating has come into wide use because of its "non-tarnish" characteristic and lower cost relative to silver or gold plating. Unfortunately, IMG is essentially as large as with stainless steel connectors. In fact, the "dotted" curve, obtained with a heavily nickel plated device, showed larger IMG than most stainless steel devices. Reduction of IMG with external magnetic fields was similar to that for stainless steel. A more direct comparison, as a function of field, is presented later.

A simple experiment utilizing a standard UG-29 adapter was conducted to observe IMG levels due to nickel plating. A silver plated adapter, as manufactured, with IMG initially at −140 dBm was



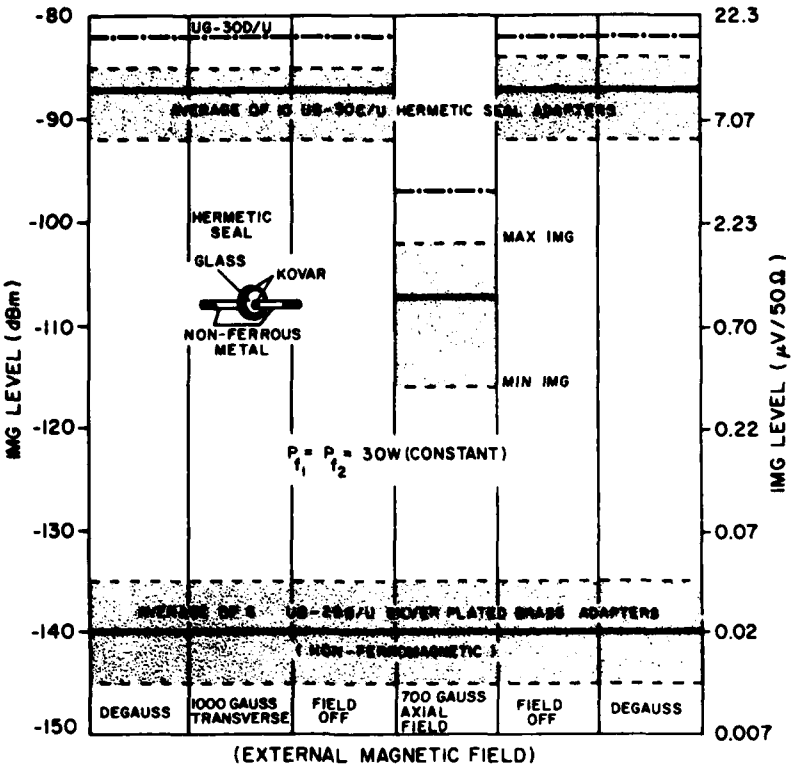
IMG COMPARISON OF NICKEL AND SILVER PLATED RF CONNECTOR / ADAPTERS

Figure 11

selected. The adapter was then disassembled, stripped of the original silver plating, replated with 500 microinches of nickel and reassembled. The "dashed" curve is the result, a 40 dB degradation over silver plating. Obviously, nickel plated connectors are also unacceptable in sensitive multicarrier communications systems. The shaded area bracketing the −140 dBm reference shows the spread in IMG levels for six UG-29 (non-ferromagnetic) adapters. The solid line represents the averaged values and also that which is readily obtainable through device selection.

Hermetic Seal Non-Linear Effects

Figure 12 shows 3rd order IMG data for hermetically sealed connectors. Test set conditions are the same as for Figs. 8 and 11. Hermetic seals are non-linear because the center conductor through the glass seal and the metallic rim around the glass seal are made of kovar or similar ferromagnetic material. Kovar is used because it is a nearly perfect match to the thermal coefficient of glass 7052. IMG averaged −87 dBm at 60 watts total drive for the ten UG-30C/U hermetic seal adapters tested. Interference is thus worse than from either the stainless steel or nickel plated connectors and only 15 dB less than from the cold rolled steel connector. Kovar contains 99.7% ferromagnetic materials (iron, nickel, and cobalt). The permeability of kovar is therefore much greater than the stainless steel or nickel plated devices. Tests conducted with and without the kovar elements in a given connector provided unmistakable proof of the intermodulation contribution and the need for their avoidance in IMG vulnerable systems.



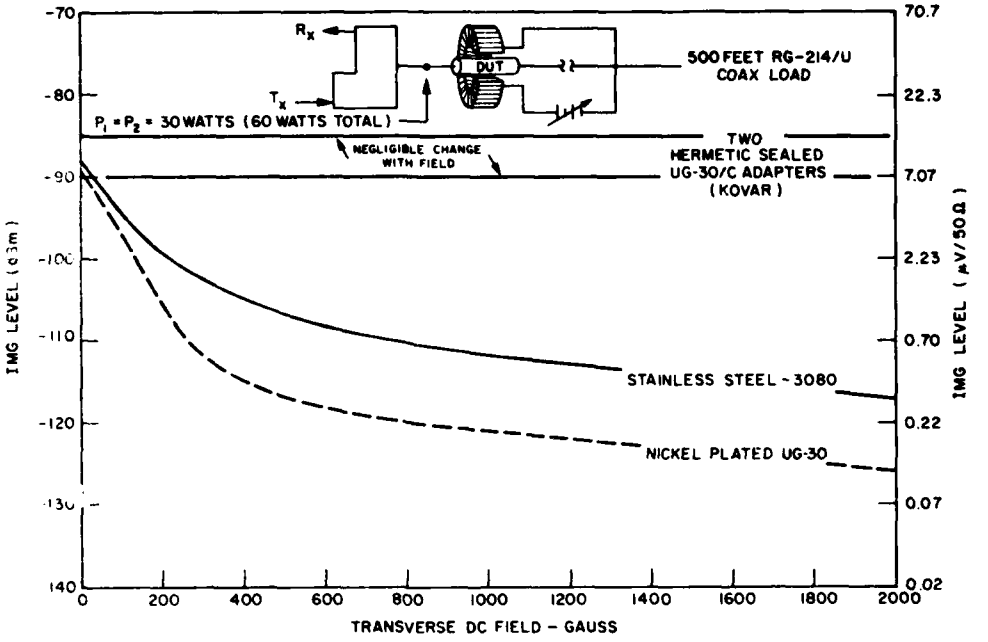
IMG COMPARISON OF HERMETICALLY SEALED AND SILVER PLATED JACK-JACK ADAPTERS

Figure 12

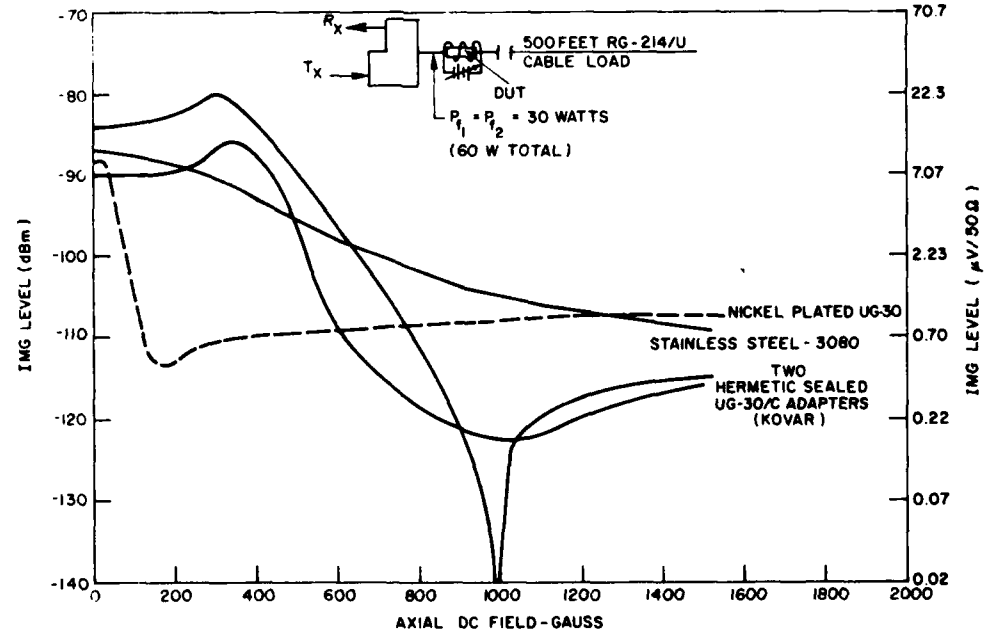
The hermetic-seal source of IMG should be avoided by using pressurized connectors wherever possible. The UG-30/U adapter employs a gasket seal construction instead of the kovar-glass seal and is said to be good for pressures in the order of 50 psi, which should be adequate for earth based systems. However, it should be stressed that such devices must not use stainless steel, nickel plating or other ferromagnetic materials. During the IMG study an order was placed for six UG-30/U silver plated adapters. When the order arrived the devices were nickel plated. In addition, only three of the devices could be used for the nickel plated UG-30/U data "solid curve" shown in Fig. 11; the remaining devices could not be used because of contact non-linearity and noise (when tapped) during operation in the test set. The tap test is a very effective method for revealing defective RF connector hardware, such as included metal chips or other contact type IMG problems. Unfortunately, an IMG test set is not normally available. However, such equipment is not difficult to construct and is strongly recommended for use by RF hardware manufacturers to avoid delivery of defective components.

IMG versus Magnetic Field Strength

Ferromagnetic connector IMG characteristics, as a function of external transverse and axial dc magnetic fields, are shown in Fig. 13A and B, respectively. The total RF power output (two equal carriers) was 60 watts to each DUT, as indicated. Progressive reduction in IMG with increasing transverse field was obtained for both the nickel plated and stainless steel devices as shown in Fig. 13A. The smaller reduction in IMG for a given field strength in the 3080 stainless steel adapter, relative to that obtained with the nickel plated device, is believed to be due to greater spread of magnetic flux throughout the stainless steel (ferromagnetic) body, resulting in less effective field at the inner surface of the conductor where RF current flows. In contrast, nickel plating on the UG-30 was 500 microinches or less (thickness) at either the inner or outer surface of the brass body. This represents



(A)



CONNECTOR IMG VARIATION WITH EXTERNAL MAGNETIC FIELD

(B)

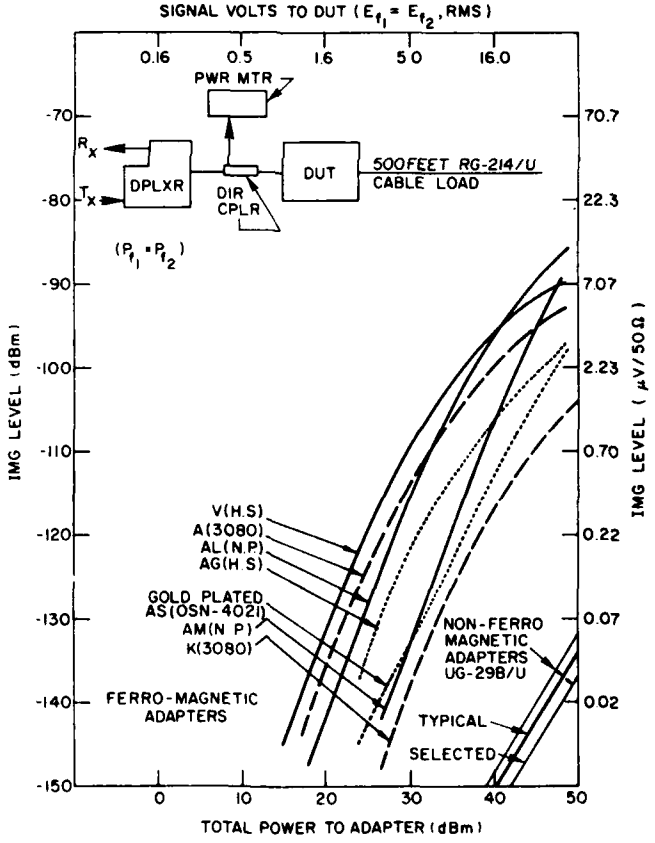
Figure 13

much less ferromagnetic material than provided by the stainless steel body for field spreading, resulting in a greater field concentration at the inner surface of the outer conductor. The "no-change" IMG response shown for the hermetic seal devices with the transverse field $\leqslant$2000 gauss is believed to be due to the termination of this field at the outer surface of the high permeability kovar ring (see Fig. 12). Hence, no magnetic field interaction with RF current flow along the inner surface of the ring or kovar center pin was possible.

Connector IMG response to axial fields is more complex, as shown in Fig. 13B. With the axial field, interaction with RF current surfaces occurs for all connector types. In the hermetic seal device, magnetic fields are interacting with RF current on both the kovar pin center conductor as well as the inner surface of the kovar ring. Two kovar connector samples were selected to reveal the range of IMG variations encountered. Note that both devices show an IMG maximum for an axial field of about 300 gauss. This characteristic was evident for either positive or negative field polarities and occurred to some degree with all kovar devices tested. The IMG increase may be the result of increased permeability, due to the external field acting as a bias. This response and the other features evident in Fig. 13B are discussed in Chapter IV.

IMG as a Function of RF Drive

Figure 14 shows IMG as a function of RF drive level for both ferromagnetic and nonferromagnetic connectors. As in the previously shown constant power measurements, the ferromagnetic units are always inferior. To identify each connector/adapter of the large number tested, letters were



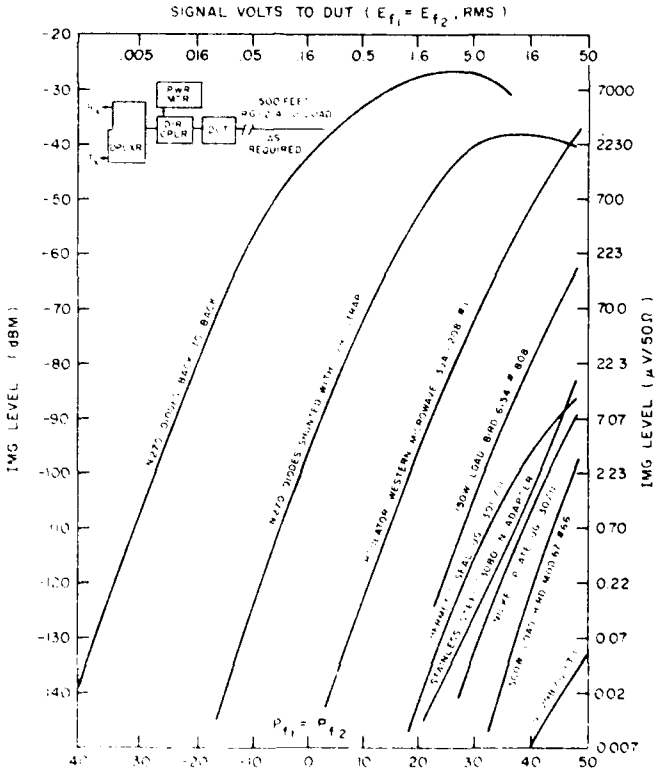
TYPICAL CONNECTOR/ADAPTER IMG VARIATION WITH INPUT POWER

Figure 14

assigned followed by a parenthesis indicating the (type of device). For example V(H.S) identifies sample V as a hermetic seal kovar adapter, A(3080) identifies sample A as the Americon 3080 stainless steel adapter, AL(N.P) identifies adapter AL as a nickel plated device, etc. In general, IMG levels from "kovar" devices exceed those by stainless steel, which in turn exceed those by most nickel plated devices. Gold plating of stainless steel connectors reduces IMG but does not provide the quiet levels attained with ordinary silver plated brass devices. For example, a "3080" stainless steel adapter was selected which measured about −90 dBm at +48 dBm total drive before plating. The device was then plated with 1000 micro-inches of gold, an amount which is not now economically feasible. This thickness is theoretically in excess of 5 skin depths at 250 MHz, the lowest transmitter frequency employed. The IMG characteristic was then essentially that of the minimum IMG device K(3080) without plating; a reduction of roughly 15 dB with plating at high power levels but still inferior to the silver plated brass connectors by about 30 dB.

IMG by Various Devices

Figure 15 compares 3rd order IMG characteristics of connectors and other non-linear devices such as resistive loads, a circulator and diodes back-to-back, as a function of input power. As to be expected, diodes are extremely non-linear, exceeding the worse connector IMG level by many orders of magnitude. Shunting the diodes with a very low impedance copper strap only partially removes the "junction" type non-linearity. Diodes, it should be noted, because of the hermetic seal construction also include the ferromagnetic non-linearity as well. The circulator, a ferromagnetic device, is extremely non-linear, and should not be directly used in intermodulation sensitive circuits. Resistive loads, and terminations, may exhibit both contact and ferromagnetic non-linearities because of the material and construction employed. These devices have been found to be greatly inferior to the coaxial cable load used for terminating the intermodulation test set samples indicated.



IMG COMPARISON OF VARIOUS NON-LINEAR DEVICES AS FUNCTION OF INPUT POWER

Figure 15

CONCLUSION

From the few experiments presented here, it should be clear that intermodulation generation by ferromagnetic materials (such as stainless steel, nickel plating and hermetic seals) in RF connector hardware is a very serious interference problem to satellite and other sensitive communication systems. Using ferromagnetic materials in RF connectors for military applications is particularly dangerous because of the possible misuse of such devices in IMG vulnerable systems. MIL-C-39012B and related specifications should be revised, prohibiting the use of ferromagnetic materials. Communication centers should be alerted to the potential interference problems of such materials. Their immediate removal is strongly recommended.

The necessity to exclude ferromagnetic materials in the fabrication of RF connector hardware, currently an industry-wide problem, cannot be overstressed as a very important step in linearizing RF systems for maximum communication capability.

Chapter III

CONNECTOR DESIGN TECHNIQUES TO AVOID RFI*

Charles E. Young

Satellite Communications Branch
Communications Sciences Division

BACKGROUND

An earlier paper[1], presented at the Ninth Annual Connector Symposium, gave detailed experimental evidence of the serious RFI levels produced by commonly available RF connectors which use ferromagnetic materials (iron, nickel, cobalt or their alloys) for electrical conductors. For example, the body structure of a wide variety of precision made coaxial connectors and adapters are currently manufactured only from stainless steel, type 303, a ferromagnetic alloy. To cut cost and provide a corrosion resistant finish, nickel plating, another ferromagnetic material, has been almost exclusively substituted for silver or gold, previously employed to plate brass stock connectors. That such materials could be considered for electrical conductor service is difficult to understand because of the known nonlinear effects of even minute quantities of ferromagnetic contaminants in RF systems.† Use of ferromagnetic materials, however, has become so widespread that silver or gold plated brass (nonferromagnetic) devices, which had been standard for many years, are no longer readily available as "off-the-shelf" items but must be specially ordered in quantities (500 or more) to obtain reasonable production cost.

As shown in reference (1), connectors fabricated from ferromagnetic materials typically produce IMG power levels 3 to 5 orders (1000 to 100,00 times) higher (worse) than without. Obviously such interference levels cannot be tolerated in todays highly sensitive communication systems. The ferromagnetic connector RFI problem came to light in 1975 during the Naval Research Laboratory's investigation of passive component nonlinearity and means for its reduction required by the Fleet Satellite Communications (FLTSATCOM) system, then under development.

To alert the communications community of the ferromagnetic connector nonlinearity problem, reference (2) was issued September 16, 1975 in advance of formal reporting and published under the government-Industry Data Exchange Program (GIDEP) as Alert No. Y1-A75-01, October 6, 1975. The experimental findngs of ferromagnetic connector nonlinearity are substantiated through theoretical analysis.[3] A literature search‡ and discussion with others in the communications field also confirms the NRL findings.

Although the ferromagnetic connector nonlinearity problem has been reported and widely discussed during the past 3 years, very little action has been taken to correct the problem. Many companies are continuing business as usual with a "take it" or leave it" attitude. It has been suggested that

*Presented at the Eleventh Annual Connector Symposium, Cherry Hill, N J , 25-26 Oct 1978

†The references listed at the end of this chapter are only a small sampling of the available literature

‡The ferromagnetic nonlinearity problem has been recognized, almost from the beginning of electrical communications. The history of resistance anisotropy in ferromagnetic metals goes back to 1857 (W. Thomson) and that of the anomalous Hall effect to 1893 (A. Kundt). See references (4), (5), (6), (7), (8), and (9) for further details

only the Navy is having a problem but this is not true. Cost reduction and corrosion resistance, if not accompanied with full operational capability, is not justification for deviating from sound metallurgical and mechanical design practices which have been established through basic research and development over many years.

This chapter was orginally intended to identify coaxial connector problems found in the NRL study other than the ferromagnetic nonlinearity reported.[1] However, because of the seriousness of the ferromagnetic problem, a further discussion of connector nonlinearity, as related to the basic concept of "skin depth" conductor current flow at RF (indicated but not developed in the earlier paper) will be presented. Both ferromagnetic and contact nonlinarity, the two major sources of RFI by connector hardware, will be discussed.

BASIC SYSTEM REQUIREMENTS

The need for extremely linear passive components, including RF connector hardware, is evident by considering antenna, receiver, and transmitter requirements in a typical communication center where simultaneous reception and transmissions are involved.

Antenna System Configurations

Figure 1 illustrates schematically the two basic systems employed in radio communication centers: (a) the single antenna diplexed receive/transmit system and (b) the use of separate antennas for reception and transmission. In terms of system size, weight, and cost, the single antenna diplexed (combined) receive/transmit system is of course preferred, as it avoids the need for two or more antenna structures and signal feeds for the required receivers and transmitters. However, the problem of self interference is potentially worse. Electrical nonlinearities within the multicoupler-diplexer filters or connecting RF hardware to and including the antenna stuctures will convert a fraction of the multiple signal transmit power into IMG product signals. The IMG signals that fall within the receive frequency bands and are not well below receiver thermal noise (amplitude) will interfere and degrade weak signal reception.

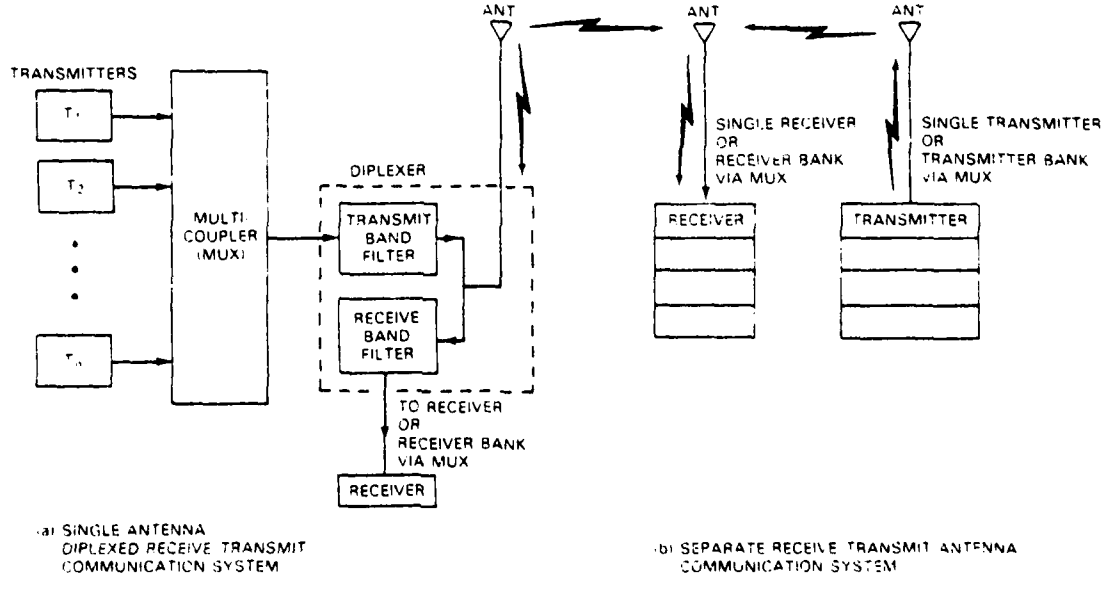


Fig. 1 — Radio communication center receiving/transmitting system arrangements

To avoid conducted RFI, as in the diplexed system, an attractive alternative (at sufficiently high frequencies) is the use of separate receive and transmit antennas spaced as far apart as is practicable. Free space attenuation between isotropic (nondirectional) antennas is:

$$\frac{P_r}{P_t} = \frac{\lambda^2}{(4\pi D)^2} \tag{1}$$

where

P_r = received power in watts

P_t = transmitter power in watts

λ = signal wavelength in meters or feet

D = separation in same units as λ

The attenuation, α, in decibels (dB) is:

$$\alpha = 10 \log \frac{P_r}{P_t} = 10 \log [\lambda^2/(4\pi D)^2] \tag{2}$$

At $D = \lambda$, the free space attenuation is 22 dB. For twice the separation, $D = 2\lambda$, attenuation is increased 6 dB, giving an isolation of 28 dB. An additional 6 dB is obtained for each doubling of the separation, as shown in Table 1.

Table 1 — Attenuation between Isotropic Antennas

Separation (ft)	λ	2λ	4λ	8λ	16λ	32λ
Attenuation (dB)	22	28	34	40	46	52

In practice, antennas are directional (not isotropic) and the isolation is increased by the directivity obtained with each antenna. At UHF, isolation between receive and transmit antennas can be in the order of 50 dB (as obtained on FLTSAT). This value of isolation not only reduces the transmit antenna IMG signal coupling to the receive antenna, but also reduces fundamental transmit signal levels impinging on the receive antenna, thereby minimizing nonlinear responses in the receive antenna system. Unfortunately, at HF it is difficult, if not impossible to obtain this order of isolation because of the longer wave length, λ, and reduced antenna directivity.

Receiver Sensitivity (Noise) Threshold-IMG Requirements

State of the art receiving systems can readily provide noise figures (NF)* in the order of 4 dB or less. To avoid sensitivity degradation, IMG products which fall in the receive frequency band must be less than receiver thermal noise by approximately 20 dB. Table II indicates the parameters used to to determine the maximum permissible IMG level in a 100 Hertz (Hz) signal bandwidth and the equivalent power in decibels relative to 1 milliwatt (dBm). The above receiver noise threshold-to-IMG margin of 20 dB represents a sensitivity degradation of only 0.04 dB and is imposed upon spacecraft and other critical system designs, not only to insure negligible RFI but to also provide some margin for system degradation with time and under environmental extremes. An IMG level equal to receiver thermal noise (−150 dBm for the 4 dB NF receiver) represents a 3 dB degradation in threshold sensitivity, the maximum acceptable RFI limit for most communication systems. Even this degree of performance is not attainable with ferromagnetic RF connector hardware, as shown.

*NF = $10 \log \left[1 + \frac{\text{Rec Noise Temp}}{290^\circ \text{K}}\right]$

Transmitter Requirements

Transmit signal levels vary but for discussion purposes may be assumed to be in the order of 100 watts, equal to +50 dBm. A receiver RFI limit of −170 dBm, or less, as indicated in Table II requires that transmit IMG products falling in the receive frequency band be at least 220 dB below +50 dBm, the desired transmit signal levels. With closely spaced receive and transmit frequency bands, less than 8 percent of band centers in the FLTSAT system, sufficient IMG attenuation has only been achieved with separate antennas and very linear RF hardware. The 3rd order IMG level of a diplexed system, carefully built to test RF connector hardware and described in reference (1), measured somewhat less than −140 dBm with +50 dBm total power from the diplexer output (antenna) port. This represents a diplexed system residual IM conversion of −190 dB or about 30 dB worse than desired.

Table II − Receiver Sensitivity/IMG Threshold

Parameter	Numerical Value	Equivalent
Boltzmann's Constant, k	1.38×10^{-23} Joule/ Kelvin	−198.6 dBm/Hz K
Ant. Noise Temp (~19°C)	290 K	
Rec. Noise Temp	440 K	
System Noise Temp	730 K	+28.6 dB K
Bandwidth	100 Hz	+20.0 dB
Rec. Noise Threshold	10^{-18} watts	−150.0 dBm
Margin (rec. Thresh/IMG)	100/1	−20.0 dB
Max IMG RFI level	10^{-20} watts	−170.0 dBm

Since IMG products appear to drop about 10 dB per order, a diplexed system becomes a viable technique, if a sufficiently wide unused band of frequencies (guardband) is allowed between the receive and transmit frequency bands to reject the 7th and lower order products. However, it should be noted that IMG products, or harmonics radiated by any system, whether diplexed or not, may fall into receive frequency bands of other nearby systems and there cause RFI and/or sensitivity degradation. Elimination of nonlinear passive components is therefore still required.

CONNECTOR IMG DATA

To again show the excessive RFI produced by commonly used ferromagnetic connectors relative to that permitted in the communication systems just described, characteristic connector IMG data will be presented. The NRL test set and extensive connector IMG data were described in detail in reference (1).

Ferromagnetic IMG as a Function of Product Order

Third order IMG, being the lowest order and largest interference which can occur in the receive frequency band of a multiplex system (with guardband width less than the transmit frequency band, as in FLTSAT) usually determines the degradation of threshold sensitivity. The extremely high level of 3rd order IMG by ferromagnetic connectors, however, implies the presence also of potentially degrading higher order IMG products. These characteristics were investigated, using connector samples V, A and U from the earlier study as the device under test (DUT) shown in Figure 2. To identify each connector/adapter of the large number tested, letters were assigned followed by a parenthesized indicator of each device. Sample "V" was a standard UG-30C/U kovar hermetic seal (HS) RF coaxial adapter. Sample "A" was a precision "N" double jack (female) adapter which uses a stainless steel body for the coaxial outer conductor. Sample "U" was a standard UG-29 nickel plated (NP) adapter.

Sample V was one of the ten UG-30C/U kovar hermetic seal coaxial adapters, originally tested. This commonly employed device is extremely nonlinear because the center conductor through the glass

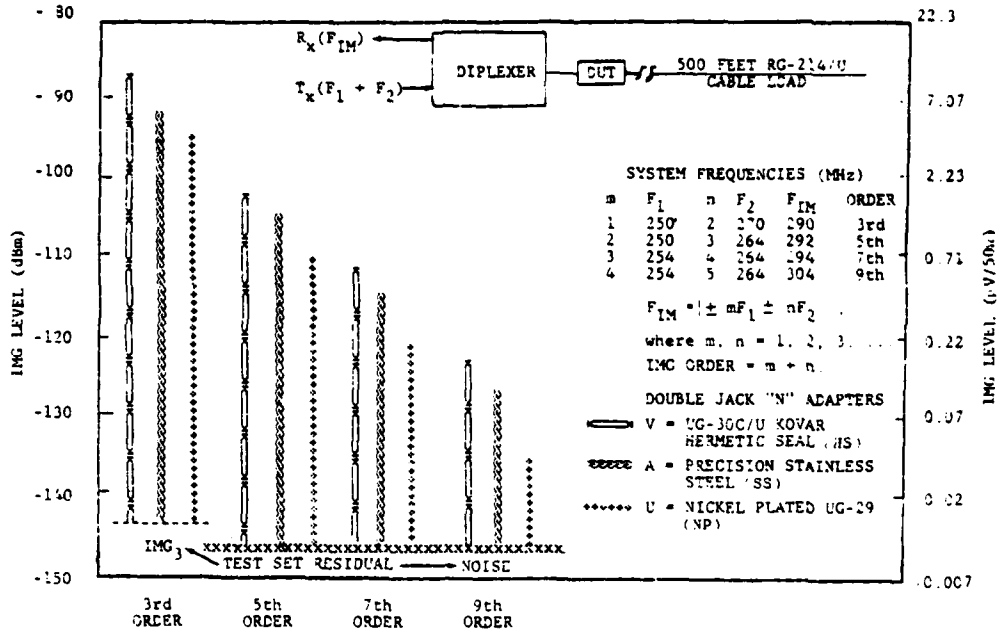


Fig. 2 — Ferromagnetic connector/adapter IMG as a function of product order

seal and the metallic rim around the glass seal is made of kovar or similar ferromagnetic material. Kovar contains 99.7% ferromagnetic materials (iron, nickel and cobalt) and is generally the most nonlinear of devices tested. Note its very large IMG level for all orders of nonlinearity shown in Figure 2.

For many years kovar was the primary material employed for hermetic seals of reported RFI problems, documented as early as 1966[11]. Fortunately, a new non-ferromagnetic hermetic seal has been announced recently. The Space and Communication Group at the Hughes AirCraft Company, Los Angeles California, have reported the successful development of a non-ferromagnetic hermetic seal because of ferromagnetic IMG RFI problems encountered in their MARISAT and related communication systems. The new seal is reported to have high RF power capability, no IMG problems and to withstand repeated thermal shocks from liquid nitrogen to boiling water with no detectable leakage. The radio communications community is in urgent need of such a device.

From Figure 2, it is evident that IMG from the stainless steel connector is only slightly less than that from the kovar device and that the nickel plated device is almost as nonlinear. The drop of about 10 dB for each progressively higher order of IMG is typical of known nonlinear devices. More importantly, the higher orders of IMG from all three ferromagnetic connectors far exceeded the 3rd order IMG residual of the diplexed test set (~ -144 dBm) and of course the test set noise threshold of ~ -147 dBm. (Note that 5th and higher order IMG from the test set alone was below noise threshold and therefore not detected.) Actually 11th order IMG from the nickel plated device, 13th from the stainless steel and 15th from kovar measured above the test set noise threshold.

Nonferromagnetic Connector IMG Comparison

Standard silver plated (non-ferromagnetic) UG-29 adapters, when operated as the DUT typically did not change the test set 3rd order IMG residual (~ -144 dBm), an indication that connector 3rd order IMG was below the test set residual by at least 10 dB*, or ~ -154 dBm. Higher orders of IMG

*Assuming simple power addition, a test set plus DUT IMG level 0.5 dB greater than that of the test set alone indicates that DUT IMG (alone) is theoretically 9.6 dB below the test set and hence that much more linear.

with nonferromagnetic connectors were also below the test set detection noise threshold, as might be expected. Except for obviously defective adapters, even a random selection of silver plated non-ferromagnetic devices usually gave IMG levels (DUT + test set) no more than 3 dB above that of the test set alone, indicating a maximum DUT IMG = test set IMG, or ~ -144 dBm. The vastly superior performance of the non-ferromagnetic connectors is not indicated in Figure 2, but was illustrated in numerous data comparisons in reference (1). The necessity to exclude ferromagnetic materials as electrical conductors in RF connector hardware is clear.

IMG MECHANISMS

Like thermal noise, nonlinearities are present to some degree in all elecrical networks. Many IMG mechanisms have been postulated[12, 13, 14, 3] but the two major sources of nonlinearity encountered in RF connector hardware are:

(1) Imperfect metal-to-metal electrical contacts, and

(2) Use of ferromagnetic materials for electrical conductors.

The problem of imperfect contacts is widely recognized, but nonlinearity due to ferromagnetic conductors is less well known. The nonlinearity of ferromagnetic conductors is related to the change in permeability experienced with current flow as described next.

Skin Depth*

It is well known that alternating current is not uniformly distributed over the cross section of a homogenous conductor (as with direct current) but is displaced more and more to the conductor surface as the frequency is increased. For very high frequencies, practically the entire current is concentrated in a very thin layer at the surface called the "skin depth", of the conductor. The skin depth, δ, at which the current density drops to $1/e(\sim 37\%)$ of its value at the conductor surface is given by

$$\delta = (\pi f \sigma \mu)^{-\frac{1}{2}} \tag{3}$$

where f is the frequency of operation, and σ and μ are the conductivity and permeability, respectively, of a given conductor. If the conductor is a nonferrous metal, such as silver, a linear relationship exists between the resulting magnetic flux, B, and the magnetic field intensity, H, with current flow; i.e., $B = \mu H$ where μ is a constant, very close to that of free space, $\mu_0 = 4\pi \times 10^{-7}$ henrys per meter (h/m). For this linear relationship, δ can be analytically determined. Silver, for example, which has the largest conductivity, $\sigma = 6.15(10^7)$ mhos/m, (least resistivity, $\rho = \frac{1}{\sigma}$, or loss of any metal at normal temperatures) has a skin depth of $\sim 2 \times 10^{-6}$ meters (m) at 1 GHz, (10^9 Hz). The minimum silver plating, in accordance with Federal Specification QQ-S-365a for nonferrous base metal conductors is 0.0005 inch ($\sim 12.7 \times 10^{-6}$ m) which provides >6 skin depths at 1 GHz, and thus conducts more than 99.8% of the total current.

On the other hand, if the current carrying conductor is a ferromagnetic material, μ is not constant but varies with H in a very nonlinear manner. B depends not only on H but also on previous values of H, the well known hysteresis effect. The skin depth equation is therefore nonlinear with "memory" and cannot be solved analytically. Of greater concern, however, is the variation of skin depth caused by permeability change; which is equivalent to a nonlinear circuit impedance change—being a function of instantaneous current amplitude. This effect is evident in the following numerical approximations of skin depth for nickel plating, based in part upon data from reference (9). The hysteresis loop

*An excellent analytical development of skin effect, as well as an interesting historical sketch of its discovery, starting with Maxwell in 1873, is given in reference 15.

(memory) of nickel at room temperature is shown in Figure 3, and the nonlinear change in relative permeability, μ_r, as a function of field strength, H, is shown in Figure 4. These parameters are comparable for other ferromagnetic materials and may therefore be used as a model.

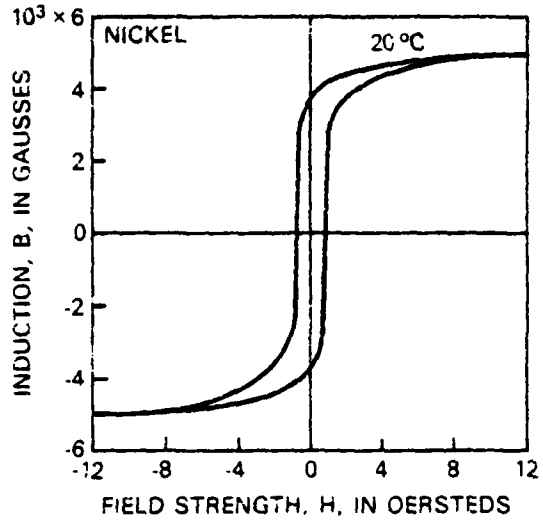


Fig. 3 — Ferromagnetic hysteresis nonlinearity

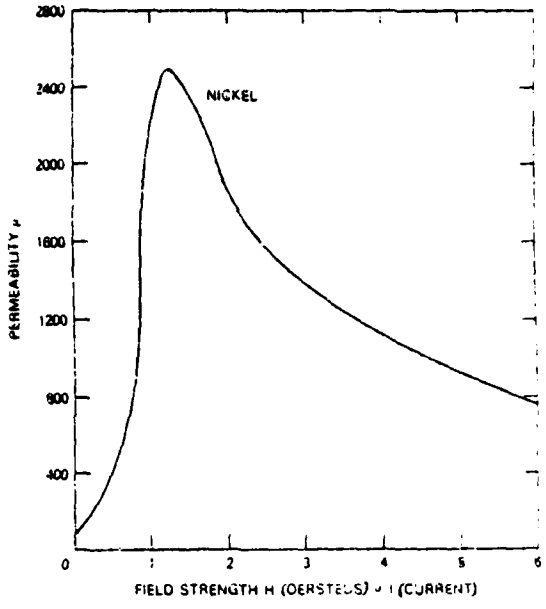


Fig. 4 — Ferromagnetic permeability nonlinearity

If one first assumes an extremely small signal current amplitude, the initial permeability of nickel, $\mu_i = 4\pi \times 10^{-5}$ h/m, may be used. The conductivity of nickel, σ_{Ni} is $\sim 1.3 \times 10^7$ mhos/m. These values result in an initial skin depth, δ_i of $\sim 4.4 \times 10^{-7}$m at 1 GHz., a value approximately 1/5 of that for silver. Since the thickness of nickel plating on connectors typically measures 10^{-5}m (> 20 skin depths) or more, the entire current is carried by the nickel plated surface. Besides the nonlinearity problem, a minute skin depth intensifies the effects of surface imperfections such as scratches, holes, oxide contaminants, etc. creating anomalous skin current paths and erratic metal-to-metal contact junctions.

As signal current increases, the permeability of nickel increases, reaching μ_{max}, 10 to 100 times μ_i. This further decreases skin depth by a factor of 3 to 10 times. Beyond some critical current, however, permeability decreases, finally reaching saturation, $\mu_{sat} = 4\pi \times 10^{-7}$ h/m (free space); resulting in a maximum skin depth for nickel of approximately twice that for silver, because of the poorer conductivity of nickel. Thus, a very large nonlinear change in skin depth, by at least an order of magnitude, can be visualized as a function of current flow. The additional nonlinear effects due to hysteresis are not possible to evaluate numerically.

There are undesired effects other than the nonlinear permeability change which disqualify ferromagnetic materials for use as electronic connectors. Besides the low conductivity, an effect exhibited by ferromagnetic metals is the anisotropy of electrical conductivity; different values for different current and/or field directions. Also, magnetoresistivity, the change in resistance associated with a change in magnetization, weakly found in all metals, is orders higher in ferromagnetic materials. Magnetostriction, the change in physical dimensions of a ferromagnetic material in a magnetic field is another undesired effect for an electrical connector. See the listed references for further information.

As noted earlier, non-ferromagnetic metals exhibit constant permeabilities, which differ only minutely from that of free space; being either paramagnetic (slightly larger than μ_0) or diamagnetic (slight less than μ_0). Skin depth is predictable and independent of current magnitude except for an

extremely small thermal modulation (discussed in Chapter V and in reference 13, for example). Figure 5 is a chart giving skin depth, δ, in parts of an inch for various metals over a wide range of frequencies (10 Hz to 1 MHz using the top and right hand scales, and 100 kHz to 10 GHz, using the bottom and left hand scales). The non-ferromagnetic metals, solid lines, exhibit increasing skin depths, as conductivity decreases, relative to that of silver. The ferromagnetic elements, nickel, Ni, and iron, Fe, shown dashed, indicate even less skin depth based upon their initial permeabilities, μ_i, as plotted. Recall, however, that ferromagnetic metal skin depth is dependent upon instantaneous current magnitude and is therefore modulated below and above these values with alternating current flow. For example, nickel, at sufficiently high current (permeability saturation, μ_{sat}) should approach the skin depth of brass because of their comparable conductivities.

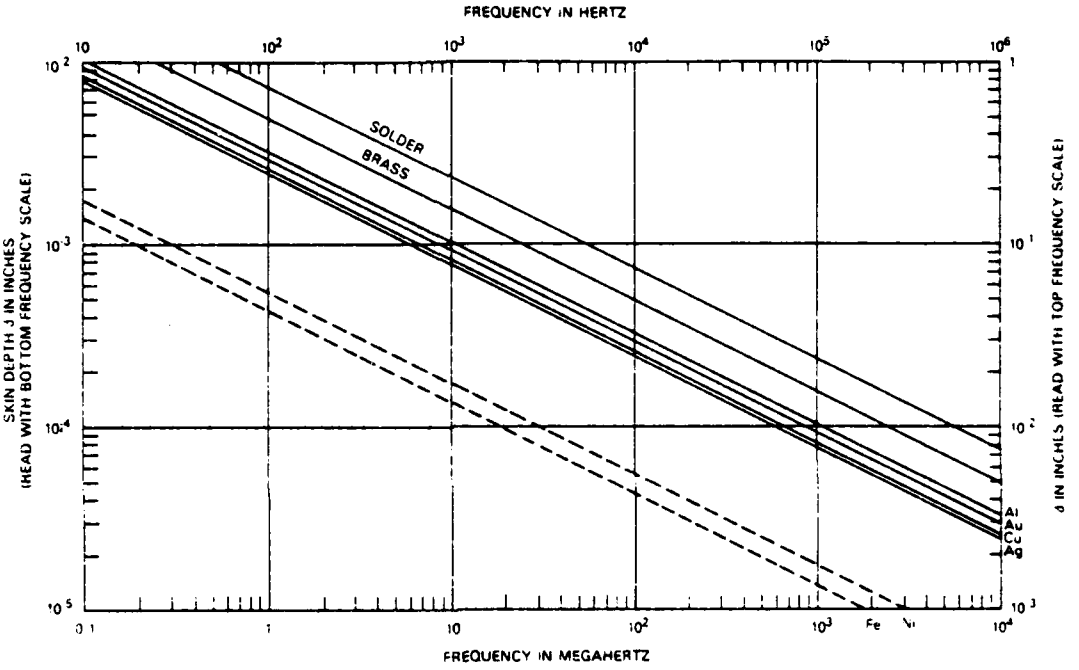


Fig. 5 — Conductor skin depth as a function operating frequency

The most important consequence of conductor skin depth is the greatly increased impedance (power loss) with alternating current (ac) flow, as compared to that with direct current (dc) flow. This effect is shown in Figure 6 for the same metals and frequency range presented in Figure 5. The surface resistivity, R_s,* defined as the resistance in ohms of a surface of equal length and width, becomes

$$R_s = \frac{\rho}{\delta} = (\pi f \mu \rho)^{\frac{1}{2}}. \tag{4}$$

Note, the superiority of silver (Ag) relative to all other metals as well as the maximum resistive (power) loss shown for the ferromagnetic metals, Ni and Fe.

Contact Imperfections

An illustration of a perfect contact between the end surfaces of a cylinder and a flattened sphere is shown in Figure 7. The shaded area represents the effective ac conductor surface while the total cross section represents the far greater dc conductor contact area. This drawing points up the fallacy of rating a coaxial connector in terms of some minimum dc resistance, as done by connector manufacturers. A much more meaningful measurement would be the RF resistance, at say, the upper frequency limit of the device.

*Schelkunoff, reference 16, p 550, defines R_s as the intrinsic resistance of the material.

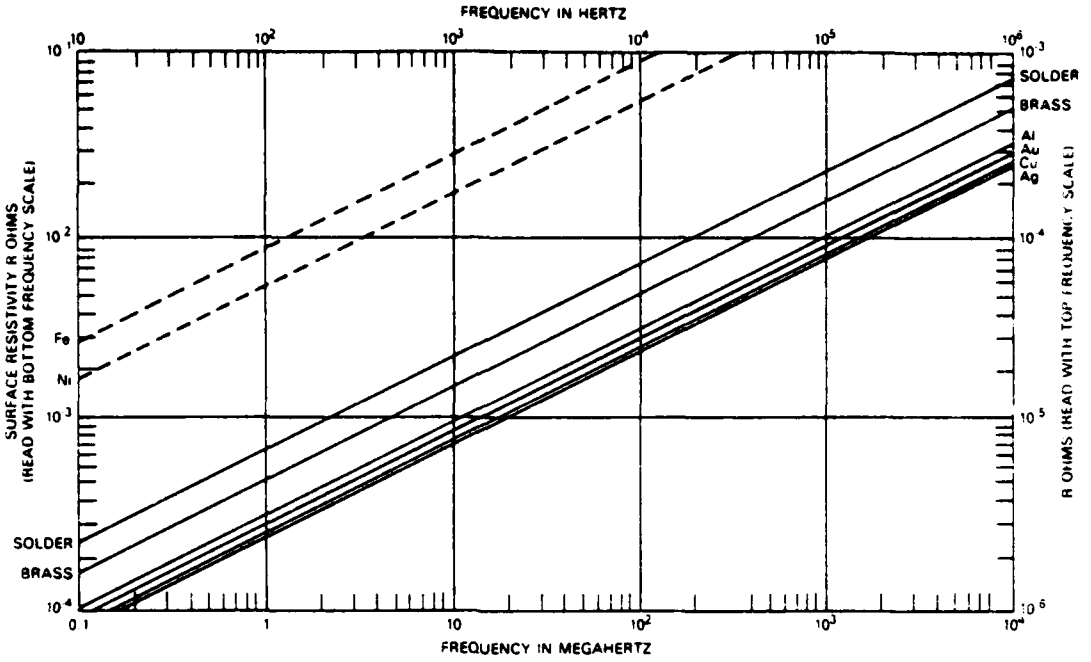


Fig. 6 — Conductor resistivity as a function of operating frequency

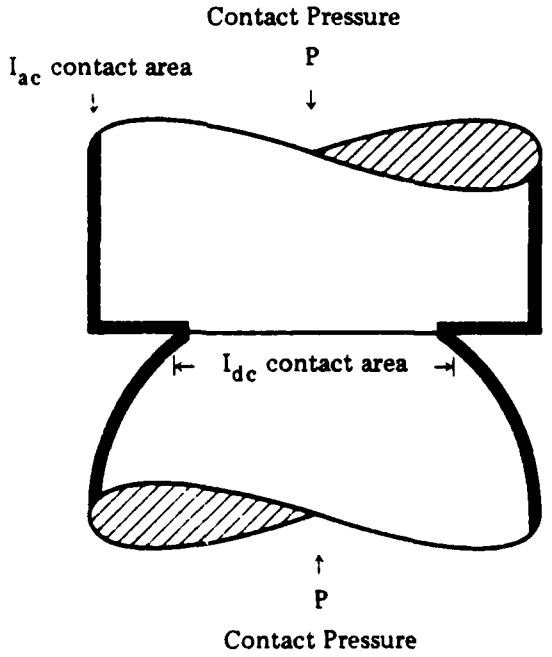


Fig. 7 — Ideal contact junction

Unfortunately, contacts are not as perfect as shown in Figure 7, but because of asperities, are more nearly as shown in Figure 8. This rough contact junction would appear considerably degraded at dc but may not be much different than Figure 7 for ac. Experimentally, it is often found that the IMG from a given connector will vary with contact pressure but not always be minimum at maximum pressure. One possible explanation is that with maximum pressure, the RF surface contact is warped or otherwise degraded at the expense of a better dc contact. Contaminants (oxides, sulfides, lubricants, etc.) tend to be pushed to the outside RF surface area, another possible contributor to poor performance. Gold plated contacts, in spite of the somewhat higher surface resistivity than silver for example, have been found to consistently give lower IMG levels than any other surface, apparently because of the relative freedom from corrosion products. Such contacts also show less criticalness to contact pressure, apparently because of the malleable characteristic of gold and the reduction of asperities. However, the use of a nickel undercoat to prevent base metal migration through gold plated surfaces must not be employed because of the ferromagnetic IMG interference caused.

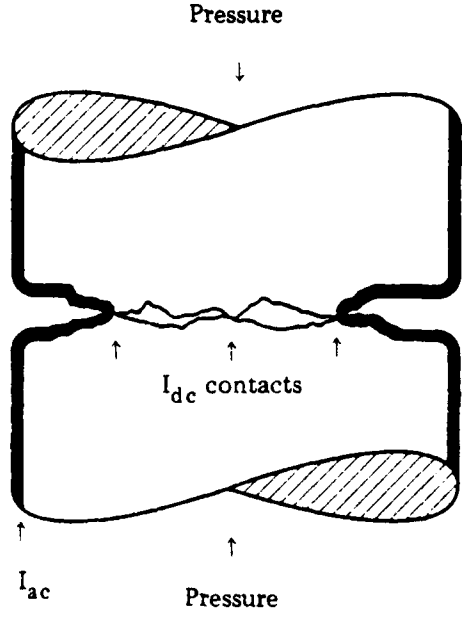


Fig. 8 — Imperfect contact junction

Figure 9 indicates the characteristic pin and socket arrangement usually employed for the center conductor contacts of coaxial connectors. The shaded surface area in the magnified view again illustrates the extremely small RF contact area relative to the dc contact area. Increased penetration depths of pins into sockets, although reducing dc contact resistance, have little effect on RF performance. As with the previous butt joints, IMG is often found to vary until "good seating" is secured. Again, the use of gold plating on both pin and socket appears essential for minimizing IMG.

Contact imperfections are also possible with the outer conductor elements of coaxial connectors. Construction practices today are, in some respects, inferior to what they were 30 years ago. For example, a potential source of IMG and contact failure has been the relatively insecure contact made between the outer conductor sleeve, and the main body of N male type connectors, accomplished by crimping. Mating, demating, shock, vibration, temperature, etc. eventually loosen this form of attachment, causing intermittent metal-to-metal contact, a large IMG source, or in other cases, complete failure, or breaking off of the contact ring. It should be noted that during and for a while after World War II, this type of connector problem did not exist because the outer conductor contact of the device

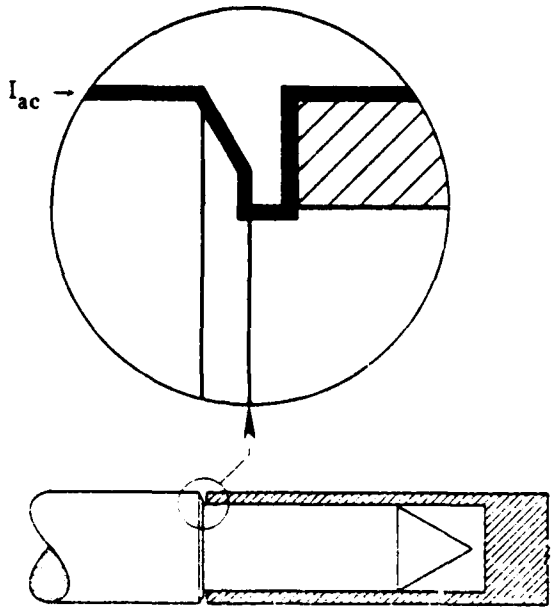


Fig. 9 — Typical pin-socket center conductor RF path

was machined from one piece of brass, which included the connector body. However, most manufacturers now crimp this body contact for reasons of economy. Although the original one piece construction is preferred, a full surface crimp, as used on certain connector types and by some manufacturers, is considered a minimum design. There are many similar instances too numerous to relate here. The need to use semi-rigid coaxial cables to obtain greater shielding has brought about many non-standard connector designs which have not yet been perfected. MIL-C-39012 only applies to flexible braided cable type connectors, except for the type SMA connector. A thorough re-evaluation of all cable-connector interfaces for both the center and outer contacts is urgently needed, based upon the NRL connector study.

CONCLUSIONS

It has been shown in this paper that the threshold sensitivity of many radio communication systems is currently limited, not by the associated low noise amplifiers (which were specially designed at high cost), but by IMG RFI which occur in improperly constructed RF connector hardware. The major contribution to this system performance degradation arises from the use of ferromagnetic materials (such as stainless steel or nickel plating) for electrical conducting elements in RF connector hardware, a practice which has been adopted by industry without sufficient research or user-consultation. This metallurgical problem can be corrected by returning to the use of non-ferromagnetic materials as previously employed. The long standing RFI problem associated with kovar or similar ferromagnetic hermetic seal type connectors can be also be elminated by use of recently developed non-ferromagnetic seals.

The contact nonlinearity problems have generally arisen from "short-cut" fabrication practices which have been adopted by manufacturers in recent years to reduce cost. These practices which differ with each connector type, should be permitted only if they do not detract from device performance and/or reliability.

REFERENCES

1. See Chapter II.

2. C. E. Young, "The Danger of Intermodulation Generation by RF Connector Hardware Containing Ferromagnetic Materials," NRL Tech Memo 5430-180A of 16 Sept 1975. Also issued under the Government-Industry Data Exchange Program (GIDEP) as Alert No. Y1-A75-01 of 6 October 1975.

3. See Chapter V.

4. A. R. Miedema and J. W. F. Dorleijn, "Electrical Conduction in Ferromagnetic Metals," Phillips Technical Review, Vol 35, 1975 No 2/3, pp 29-40.

5. E. M. Puch and N. Rostoker, "Hall Effect in Ferromagnetic Materials," Reviews of Modern Physics, Vol 25. No. 1, January 1953, pp 151-157.

6. E. Peterson, "Harmonic Production in Ferromagnetic Materials at Low Frequencies and Low Flux Densities," Bell Sys. Tech. Jour., Vol 7, 1928, pp 762-796.

   J. H. Van Vleck, "Fundamental Theory of Ferro-and Ferrimagnetic Magnetism," Proc. IRE, Vol. 44, Oct 1956, pp 1248-1259.

8. R. M. Bozorth, "Magnetoresistance and Domain Theory of Iron-Nickel Alloys," Physical Review, 70, 1946, pp 923-932.

9. R. M. Bozorth *Ferromagnetism*, D. Van Nostrand Co, 1951.

10. C. A. A. Wass, "Table of Intermodulation Products," J. IEEE (London), Vol 95, part III, Radio and Communication Engineering, January 1948, pp 31-39.

11. NELC LTR SER 3250-14 of 3 Mar 1966 (Information Bulletin) and Final Report for "Electrical Hull Interaction," Contract No. 123 (953) 55012A IIT Research Institute Project E6062 of 28 Feb 1967.

12. R. C. Chapman and J. C. Darlington, "Intermodulation Generation in Normally Passive Linear Components," Philco-Ford Corp., Final Study Report WDL 5242, Aug 24, 1973.

13. J. Z. Wilcox and P. Molmud, "Thermal Heating Contribution to Intermodulation Fields in Coaxial Waveguides," IEEE Trans. Vol. COM-24, No. 2, Feb 1976, pp 238-243.

14. W. H. Higa, "Spurious Signals Generated by Electron Tunneling on Large Reflector Antennas," Proc. IEEE, Vol. 63, No. 2, Feb 1975, pp 306-313.

15. H. A. Wheeler, "Formulas for the Skin Effect," Proc. IRE, Vol. 30, Sep 1942, pp 412-424.

16. S. A. Schelkunoff, "The electromagnetic theory of coaxial transmission lines and cylindrical shields," Bell Sys. Tech. Jour., Vol. 8, Oct 1934, pp 532-579.

Chapter IV

A STUDY OF RF NONLINEARITIES IN NICKEL*

G.C. Bailey and A.C. Ehrlich

Metal Physics Branch
Material Science and Technology Division

I. INTRODUCTION

The purpose of this chapter is to present a new technique for studying the behavior of nonlinear metallic systems subjected to radio-frequency magnetic fields. Intermodulation (intermod) generation (IMG), which takes place in a nonlinear system results in the excitation of a whole series of signals! The intermod frequencies which are excited depend on the type of nonlnearity of the system. The significance of this technique is its large relative sensitivity; that is, extremely small intermod signals (as low as the rf thermal noise in a system) can be observed in the presence of very large fundamental signals. For example, rf powers of the order of $-$ 145 dBm (i.e., $10^{-17.5}$ W) can be observed in the presence of fundamental freuqencies with powers of the order of watts or tens of watts. Thus, a relative sensitivity of $10^{-18.5}$–10^{-19} or better is possible. Thus, it may be possible to investigate at new levels of detail those physical phenomena accompanied by nonlinear electromagnetic responses.

Although the technique is quite general and applicable to the study of a large number of nonlinear metallic systems, the one we have chosen initially for illustrative purposes is that of a ferromagnetic material, namely, nickel that is electroplated onto copper. Electroplated nickel was chosen for three reasons: first, one would expect that nickel, being a ferromagnet and hence nonlinear, should exhibit appreciable intermodulation signals[2]; second, no information was available on the fundamental origins of intermod signals in nickel nor indeed in any ferromagnetic material up to the time the present study was begun; third, we wished to examine the feasibility of using IMG to study magnetic structure changes too small to examine sensitively with other techniques.

II. EXPERIMENTAL METHOD

We present here a rather complete description of the experimental method which involves considerations and difficulties which are conceptual and not simply of the nature of problems with experimental techniques. It is therefore important to give an account of a number of step-by-step improvements made in the experimental approach.

A schematic diagram of the apparatus is shown in Fig. 1. For the time being, let us consider all portions of the apparatus except those portions labeled A_1, A_2, and ϕ which are located above the two dashed lines. Most of the power of the fundamental signals, which passes through the sample, is absorbed in a load whereas the intermod signal, after passing through a set of sharply tuned high-Q filters, is detected by a sensitive spectrum analyzer. The transmitters consisted of two ICOM-21A amateur radio transceivers stabilized on internal crystals at the frequencies f_1 = 146.000 MHz and f_2 = 148.000 MHz. The transmitters were isolated from each other by two sets of three cavity-type resonant

*Published in Journal of Applied Physics 50, 453 (1979).

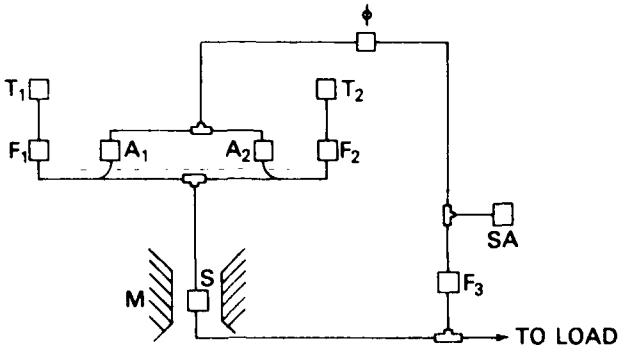


Fig. 1 — Schematic diagram of the experimental arrangement for studying intermodulation generation. The two transmitters (T_1 and T_2) operate at 146 and 148 MHz. Each set of filters (F_1, F_2, F_3) consists of three coaxial cavity resonant filters. Also shown are the dc electromagnet (M), the sample (S), and the spectrum analyzer (SA). The phase-canceling system includes the following components above the two dashed lines: A_1 and A_2 are two variable step attenuators and ϕ is a constant impedance variable-phase device.

filters represented by F_1 and F_2. The attenuation of each filter set was 55 dB at a bandwidth of ± 1 MHz with an insertion loss of 1.5 dB. The filters were tunable over the range 132-174 MHz. The maximum power from each transmitter was 10 W, approximately 5W of which was delivered to the sample as measured by a Byrd wattmeter. Most of the power loss occurred in the filters. The bulk of the exciting power at frequencies of f_1 and f_2 passed through the sample and was absorbed by a 500-ft roll of RG 223/U coaxial cable (the "Load" in Fig. 1) which had an approximate loss of 6 dB/100 ft for frequencies in the two-meter band. Double-shielded coaxial cable was used throughout the apparatus. The intermod signal ($2f_2-f_1$ = 150.000 MHz) which was generated in the sample passed through the set of three filters represented by F_3 (tuned to 150.000 MHz) and was detected by a Hewlett Packard spectrum analyzer (rf section model 8554B). A small fraction of the exciting power at f_1 and f_2 was also incident upon the spectrum analyzer and was used as a rough frequency calibration. As in any experiment at high frequencies, it would, of course, be desirable to know the incident rf magnetic field. Since this field was not precisely available and since its magnitude varied with depth into the sample as a result of the sample conductivity, all our measurements were on the power. Filters F_1 and F_2 were fine tuned to eliminate any reflections from the sample back into the transmitters. The standing-wave ratio at each transmitter output was close to unity.

Standard nickel electroplating was applied to copper wires (25 mm long × 1.62 mm in diameter) and to metallurgically polished oxygen-free high-purity copper plates (27 mm long × 6.3 mm wide × 0.74 mm thick). The samples were cleaned in NaOH, dipped in a solution of two parts sulfuric to one part nitric acid, rinsed in distilled water, copper plated with less than 0.001-in.-thick copper, and nickel plated in a solution of nickel choride, nickel sulfate, boric acid, and sulfuric acid with a pH of 4—5. The current density during plating was 27 mA/cm^2. A length of approximately 3 mm on each end of the samples was masked off from the nickel plating, and this area was soldered either directly to the ends of the studs of two UHF chassis connectors or to copper wires which were in turn soldered to the ends of the studs. This arrangement permitted studies to be made as a function of sample length, magnitude, and orientation of a dc magnetic field, temperature, input power, etc. In addition, two copper wires were soldered to the outside of the UHF connectors and acted as shields to prevent pickup into or leakage from the line.

In the experiment, an unwanted background signal always occurred that arose from a number of sources, some of which are known and some of which are unknown. With a copper-wire sample in place, one would expect this background signal to be quite low, and, in fact, the 150-MHz background signal was reduced, with special care, to — 145 dBm at 5 W per channel (i.e., at each frequency)

through the sample. This rather low background level was accomplished by removing all the visible tiny metal particles from the UHF-type connectors, by replacing the steel screws on the nine resonant-cavity filters in the circuit with brass screws, and most importantly by removing the silver oxide or sulfide from the connector plates on the filters.

III. INTERMOD POWER VERSUS APPLIED dc FIELDS

A. Experimental results

Samples of nickel-plated copper wire with various thicknesses of Ni were studied initially in order to simulate nickel-plated rf connectors and still have a more well-defined geometry than that of a connector. The data from one such sample with a 0.06-mm-thick nickel plating is shown in Fig. 2. The intermod signal at 150 MHz is shown as a function of applied dc magnetic field H_{dc}, with an rf magnetic field h_{rf} of constant amplitude. The magnetic field was applied perpendicular to the axis of the wire.

Several features of Fig. 2 are particularly noteworthy. It is seen that the intermod signal excited at zero applied field was a very large (a factor 10^4 in power) compared to the background level at high magnetic fields. The background level was taken as that intermod signal observed when an increase of the magnetic field no longer changed the intermod signal. Although it is not indicated in Fig. 2, there was very little hysteresis in the signal as the magnetic field was reversed in direction by 180°. The signal generally decreased with increasing field, but there was a sharp maximum at 600 Oe, a broad maximum at 3000 Oe, and a sharp minimum at 4900 Oe. The overall behavior of the intermod signal with applied field is strikingly similar to data obtained on commercial rf connectors.[3] Thus, the original goal of simulating a connector with a wire appears to have been met. The decrease of the intermod signal with increasing magnetic field and the disappearance of the signal when a field near that needed to magnetically saturate the sample was applied indicate that the intermod signal is related to the magnetic state of the sample and more specifically to domain wall displacement and domain rotation.

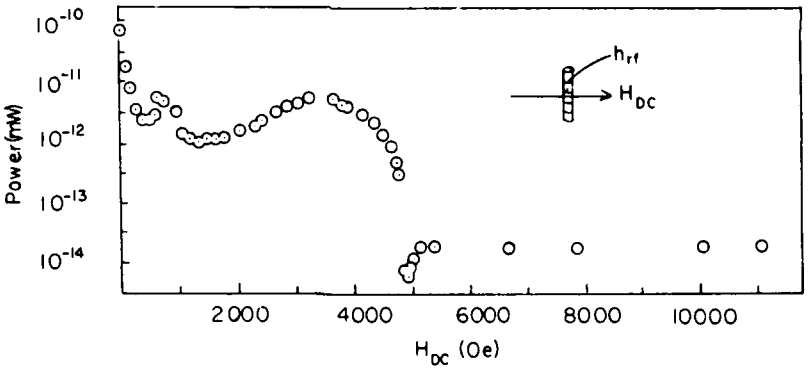


Fig. 2 — Intermodulation signal in a nickel-plated copper wire as a function of applied dc magnetic field

The interpretation of the dependence of the observed signal on the applied dc magnetic field is rather complex for the following reason. In a wire, the relative orientations of the rf magnetic field and the applied dc magnetic field are not unambigously defined. That is, with the dc field transverse to the wire, for example, there are regions in which the dc and rf fields are parallel, perpendicular, and at all angles in between parallel and perpendicular. If, however, a flat-plate orientation is used, the relative rf-dc magnetic field orientations and the demagnetizing factors can be well defined. Consequently, samples were prepared by electroplating nickel onto flat plates of highly polished pure copper to thicknesses of 0.025 and 0.25 mm. Two samples of each thickness were made with the nickel being plated on one side only as well as on both sides of the large flat faces of the copper plate. Thus, a number of well-defined dc-rf magnetic field relative orientations could be obtained.

The results did not depend significantly on whether the sample was plated on one or two sides. The plating thickness also did not affect the results significantly, as is reasonable since even the smaller thickness of 0.025 mm is still much larger than the skin depth. Therefore, typical data are shown in Fig. 3 for a sample with a plating thickness of 0.25 mm only one side of the copper plate. As is shown, the dc and rf fields were perpendicular to each other and in the plane of the sample. At low dc fields, the intermod signal behaved more or less as expected if it is associated with the number of domains in the sample. That is, as the dc field increased, the signal dropped rapidly as the sample became magnetically saturated at a few hundred oersteds. As the field was further increased, however, the signal showed a minimum at 400 Oe and then increased to a low constant value. It is interesting to note that most of the structure observed in the wire sample is now absent; that is, the peaks at 600 and 3000 Oe have been removed simply by choosing a flat-plate sample.

The origin of the minimum in Fig. 3 can be explained as follows. The intermod signal arising from the magnetic effects in the sample in this case are out of phase or nearly out of phase with the background signal that is always present in intermod experiments. At low fields, the magnetic signal dominates; at high fields the background signal, although small, dominates; at intermediate fields, where these two are comparable in magnitude but of different phase, a minimum will be observed.

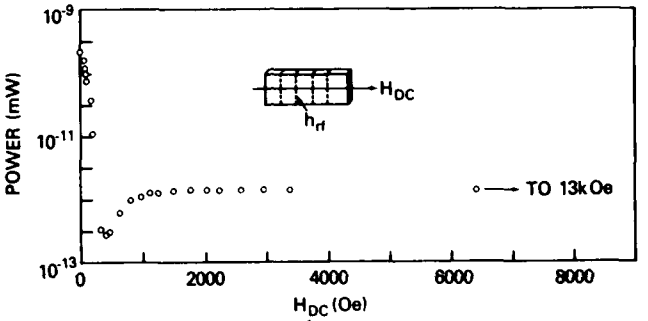
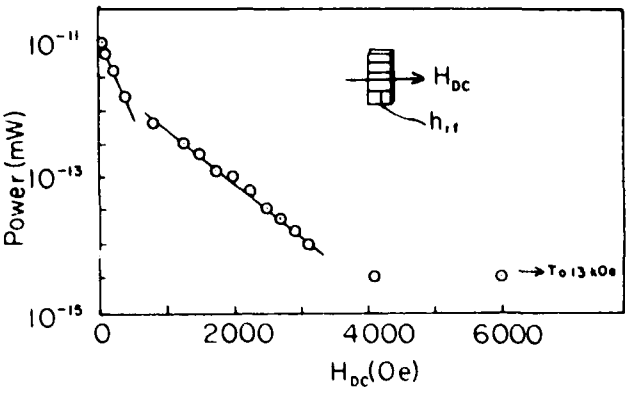


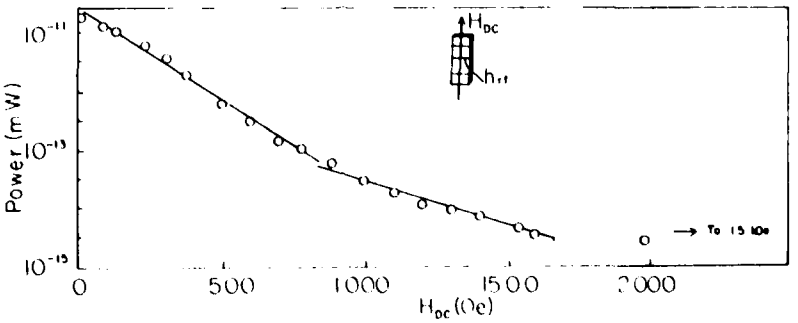
Fig. 3 — Intermodulation signal in a nickel-plated copper plate. The plating is only on one side of the copper and the rf and dc magnetic field directions are shown in the inset.

In order to test this hypothesis and, at the same time, to increase the sensitivity of the apparatus, the following additional components were added to the circuitry as shown by the components located above the two dashed lines in Fig. 1; two 20-dB couplers, two continuously variable attenuators, A_1 and A_2, and a trombone constant-impedance adjustable air line ϕ. With these modifications, an intermod signal whose phase and amplitude could be widely varied was added to the sample plus background signal just prior to the spectrum analyzer. The amplitude of the canceling signal alone was adjusted to give the same value as the level of the sample plus background signal alone at high magnetic fields. The two signals (canceling and sample plus background) were added at the Tee in front of the spectrum analyzer, and the phase of the canceling signal was varied until a minimum was obtained for the resultant signal at high magnetic fields.

The results for a sample, similar to the one whose data is shown in Fig. 3, are shown in Figs. 4 and 5, where the applied dc magnetic fields are in the sample plane and normal to the sample plane, respectively. It is seen that not only has the resultant new background signal (above magnetic saturation) been substantially reduced (from — 120 to — 145 dBm), but the minimum around 400 Oe has been removed. It is true that this would occur even if the minimum arose from an intrinsic magnetic effect vanishing and another reappearing at high fields. However, the fact that he observed signal does not change with applied dc field over a broad range from 4 to 13 kOe in Fig. 4 (a) implies that there is no magnetic contribution to the intermod signal beyond 4 kOe. It thus appears that this minimum, which had been observed before but had not been explained, is apparently a result of the addition of two signals of similar amplitude but different phase. Hence, the two modifications (a flat-plate sample



(a)



(b)

Fig. 4 — Intermodulation signal (sample plus background plus canceling signals) in a sample, similar to that whose data is shown in Fig. 3, as a function of a dc magnetic field applied in the plane of the sample (a) $H_{dc} \| h_{rf}$, (b) $H_{dc} \perp h_{rf}$

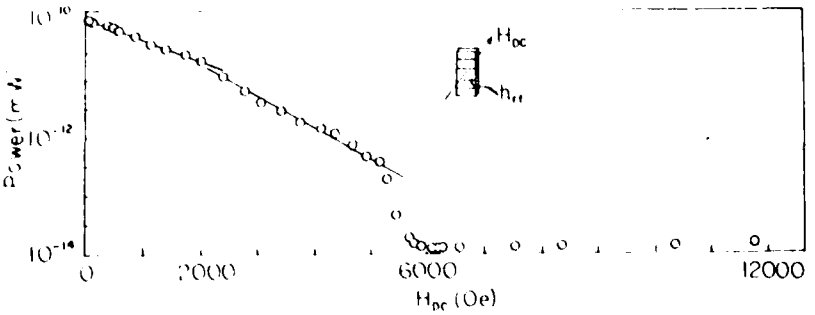


Fig. 5 — Same as Fig. 4, except that the applied dc field is normal to the sample plane.

and a background canceling signal) have eliminated all of the structure seen in the intermod signal of the wire sample (Fig. 2). Generalizing and considering the nature of Figs. 4 and 5, one can conclude that the kind of structure that is seen in Fig. 2, and that has also been seen in commercial connectors, is a consequence of interference between signals with various phases and various dependencies on dc magnetic fields that are generated at different positions around the circumference of the wire. Two different positions here are defined as two points which have different relative orientations of the dc and rf magnetic fields.

B. Discussion of Results

In addition to showing the removal of the background induced intermod effects on the resultant intermod power-field variation, Figs. 4 and 5 reveal a number of other interesting features which are qualitatively explained if we make the simple assumption that the dynamics of the domain walls are responsible for the generation of the intermod signals. There are two probable mechanisms that can account for this. The first is certain nonlinear intradomain-wall excitations which are most important for domain walls where the rf magnetic field is perpendicular to the magnetizations of the two domains separated by the domain wall and to the wall itself. A detailed treatment of this phenomena is being prepared for publication.[4] The second mechanism, to be treated in this paper in some detail, is the vibration of domain walls as a result of excitation by the two frequencies f_1 and f_2. In this second mechanism the intermod signal arises as a result of the absorption of the applied rf field energy by the oscillating wall and the subsequent reemission of energy at the same frequencies and at the intermod frequencies. Since this assumption is the basis of substantial subsequent discussion in this paper, we review some experimental evidence which supports it.

Reporting on a Kerr-effect experiment in a study of creep in a nickel-iron thin film, Kim *et al.*[5] observed the low frequency oscillations of a domain wall as a function of the frequency of the applied rf magnetic field which was parallel to the domain wall. A slowly varying (380 Hz) magnetic field was applied parallel to the wall, and the magnitude of this oscillation was observed as a function of frequency of an applied magnetic field h_{rf} from 1 kHz to 600 MHz. They observed a resonancelike displacement of the wall at 380 Hz at rf frequencies around 100 MHz, which they attributed to resonance absorption of energy from the rf field by the domain wall. They suggested that this problem is similar to the mechanical problem first investigated by Kapitza[6] in which a slowly varying force and a rapidly varying force are simultaneously applied to a particle in a potential well. It can be shown[7] that the low-frequency displacement of a particle will be very large when the high-frequency field is at its resonance. Thus, the wall ought to show large excursions in the vicinity of 100 MHz.

The results of the intermod power versus dc magnetic field H_{dc} must be explained on the basis of the interaction between the rf fields and the domain structure. The domain structure in turn depends on the magnitude and orientation of the dc field as well as the intrinsic magnetic characteristics of our electroplated Ni films. In particular, it is possible to correlate the various straight-line portions of the data shown in Figs. 4 and 5 with the evolution and disappearance of the different kinds of magnetic domains found in our sample. Among other things, what this correlation demonstrates is the ability of the IMG technique to detect very small changes in the details of the microscopic magnetic structure. This even occurs under conditions so close to magnetic saturation that other techniques known to us are insensitive to further changes in the magnetic domain structure. This high sensitivity to the detailed domain structure also can be seen in the differences in the zero dc field IMG powers seen in Figs. 4 and 5, which can be attributed to different states of magnetic domain structure.

To illustrate, in Fig. 4(a) a very large dc field (3.8 kOe) is necessary to completely suppress the IMG of magnetic origin, i.e., to completely saturate the sample as measured by IMG. Since the sample magnetization has reached 95% of its saturation value at a field of 1400 Oe according to Ref. 8, we see that the magnetization changes by only 5% while the IMG power changes by an accurately measurable factor of 30 for fields between 1400 and 3800 Oe. Thus, when more thoroughly understood, the present method should complement neutron diffraction and other techniques of investigating bulk domain structures.

1. dc Magnetic Field in Sample Plane

Figures 4(a) and 4(b) show the experimental data when the dc field is in the plane of the sample. Both Fig. 4(a), where the rf field is parallel to the dc field, and Fig. 4(b), where the rf field is perpendicular to the dc field, show two linear regions on these semilog plots of power versus field. In addition, the break in the curve in both cases occurs around 700-800 Oe. Features[8] of recent neutron

diffraction studies of electrodeposited nickel approximately 20 μm thick have been discussed in terms of a model that allows for two types of domains in the unmagnetized sample: bulk pillar-shaped domains with a magnetization normal to the plane of the sample and surface closure domains with a magnetization in the plane of the sample. As a dc magnetic field is increased in the plane of the sample, there is an abrupt break at 720 Oe in the curve of the neutron polarization parameter as a function of the dc magnetic field which is attributed to the vanishing of the pillar domains. Although the samples used in the present study are much thicker than 20 μm, the domain structure in both cases should be similar since the thickness of the sample in both cases is great enough for the samples to have bulk properties rather than thin-film properties. The initial decrease of the intermod signal with increasing dc field can therefore be ascribed to the pillar-domain walls being removed from the sample at a given rate. In fact, the linearity of the data indicates that the rate of change in the number of domain walls with field at a given field is proportional to the number of remaining walls. The mechanism by which the pillar domains give rise to IMG is an intradomain-wall excitation. As mentioned above, a quantitative mathematical treatment of this effect will be the subject of a future publication.

At 700-800 Oe, the removal of pillar domains is complete and the film consists primarily of domains having a magnetization parallel to the sample plane. At this value of the dc field, the slope of the signal-field curve changes because the evolution and disappearance of the remaining domain structure is quite different than at lower fields. The predominating domains are now those with a magnetization parallel to the plane of the sample, tending to be parallel to the dc field, and the rf field can now have a significant component parallel to the domain magnetization.

The important differences which exist for dc fields above 800 Oe between Figs. 4(a) and 4(b) can be understood using the model that assumes forced vibrations of the domain walls by the rf fields to be a major source of intermods. This mechanism will play an important role only for those domain walls separating domains whose magnetization has a component parallel to the rf field. If the rf field is perpendicular to the domain magnetizations then there is no reason, energetically, for any domain to grow at the expense of a neighboring domain, i.e., no reason for a domain wall to move.

Above 800 Oe the detailed domain structure is unknown, but what is certain is that there are far fewer domains than at zero field (and thus also domain walls) and that the domain magnetizations have a tendency to be parallel to the applied dc field rather than perpendicular to it. Thus, if domain-wall vibration is the major source of IMG at this point, the IMG power would be greater for H_{dc} parallel to h_{rf} than for H_{dc} perpendicular to h_{rf}. This we believe to be the reason for the much smaller value of the IMG power at 800 Oe in Fig. 4(b) than in Fig. 4(a).

The same reasoning accounts for the disappearance of the intermod signal at approximately 1500 Oe when H_{dc} is perpendicular to h_{rf}, while it persists up to values of H_{dc} of 4000 Oe when H_{dc} is parallel to h_{rf}. As the dc field is increased, the angle between the direction of the magnetization within a "typical" domain and the direction of the dc field becomes smaller and smaller. Also, the ratio of the number of walls aligned parallel or nearly parallel to the dc field to the number aligned at other larger angles to the dc field increases as the dc field increases. Hence, if the rf field is parallel to the dc field, as shown in Fig. 4(a), there are domain walls parallel to the rf-field direction until the sample is completely saturated magnetically when all walls disappear. On the other hand, as shown in Fig. 4(b), as soon as the overwhelming majority of the effective walls (from the intermod generation point of view) are nearly aligned with the dc field (and hence, perpendicular to the rf-field direction), the walls will not be affected by the rf field and the intermod signal becomes zero. For the field geometry depicted in Fig. 4(b), the near alignment of all the walls in a direction perpendicular to the rf-field direction occurs at a dc field far below that needed to saturate the sample. Thus, the intermod signal reaches a minimum value at a lower dc field for the conditions shown in Fig. 4(b) than for the conditions shown in Fig. 4(a).

It should be pointed out that at around 800 Oe, with H_{dc} perpendicular to h_{rf}, there is the possibility of having the relative orientations of the domain magnetizations, domain walls, and rf field for

some domains such as to permit IMG production by the intradomain-wall excitation mechanism. That there is little or no contribution of this sort suggests that the IMG production per domain wall is much greater for domain-wall vibration than for the intradomain-wall excitation mechanism. Further support for this conclusion comes from a comparison of the IMG power level in electro-deposited Ni with that from magnetically soft Ni at zero applied dc magnetic field. In the latter there is no reason to expect a strong uniaxial anisotropy perpendicular to the sample surface as is found in electrodeposited Ni. Thus, one does not expect pillar-shaped domains with magnetization normal to the sample surface, which are not susceptible to forced domain-wall vibration, to dominate the magnetic domain structure. Instead, a large fraction of the domains are expected to have magnetizations parallel to the rf field and thus have vibrating domain walls. We have measured the IMG power level in an annealed solid Ni wire at zero applied field and found it to be larger by a factor of approximately 10^4 than the zero-field power level shown in Fig. 2. Since the lower magnetic anisotropy expected in the solid Ni wire suggests larger domains and thus fewer domain walls, we attribute this larger IMG power to the presence of vibrating domain walls.

On the other hand, we have neglected domain-wall vibrations from the walls between the surface closure domains and the pillar domains that exist at fields below 800 Oe. Such vibrations require that spins be rotated against a very large uniaxial anisotropy, and thus such vibrations are expected to be very small. Once all spins are in the plane of the sample, however (perpendicular to the easy direction of magnetization), the relative anisotropy energy within this plane is much smaller and significant domain-wall vibration is expected.

2. dc Magnetic Field Perpendicular to the Sample Plane

Figure 5 shows the experimental data when the dc magnetic field is perpendicular to the plane of the sample. It consists of at least two and probably three linear regions. The break at 2000 Oe is probably real, but this is somewhat uncertain since it is just outside experimental scatter. Our interpretation of this data is that in the region between zero and 5000 Oe the bulk pillar domains are being swept out. We do not know of any information regarding the magnetic structure of electrodeposited Ni that suggests any change in the nature of this process in the vicinity of 2000 Oe. Between 5000 and 6000 Oe, as complete magnetic saturation occurs (the demagnetizing field of Ni, $4\pi M$, is approximately 6000 Oe), the last of the bulk domain structure disappears and, in addition and more important, the surface closure domains disappear.

The abrupt falloff (on a semilog plot) in IMG power around 5500 Oe we associate with the disappearance of the surface closure domains. Although sharply defined and large on the semilog plot, this decrease in IMG power represents something less than 0.5% of the zero-field IMG power. This is consistent with the negligible contribution attributed to the surface closure domains in discussing the case of H_{dc} parallel to the sample surface.

IV. INTERMOD SIGNAL VERSUS INPUT POWER

A study of the power of the intermod signal as a function of input power into the sample was carried out in order to understand more fully domain-wall motion effects in exciting the intermod signal.

The sample studied is the same as that whose intermod power-dc field variation is shown in Figs. 4 and 5. The dc magnetic field in the plane of the sample was maintained fixed as the total power through the sample was varied from 0.04 to 4W. The power at each frequency was first adjusted to be identical by observing the actual power transmitted through the sample using a Byrd wattmeter placed just prior to the infinite load. This reading was correlated with the spectrum analyzer reading of the power of the fundamental frequencies that leaked through the filters which were placed in the line (see F_3 of Fig. 1) in front of the spectrum analyzer. Thereafter, all power measurements, including that of the intermod signal at 150 MHz, were read from the spectrum analyzer. In this way, a variation of intermod power P' with the relative power P of the 148- or 146-MHz fundamental frequency could be

obtained. The powers of both fundamental frequencies were maintained equal throughout the experiment and were varied either by changing a step attenuatior (for the 0.04 – 0.4 W range) at the output of the transmitter or by varying the rf power control potentiometer on the transmitters themselves (for the 0.4 – 4 W range).

The results of the measurements of P' versus P for two different relative orientations of the dc and rf magnetic fields are shown in Figs. 6 and 7. In Fig. 6, data are shown for four different settings of the dc field (100, 200, 1500, and 2500 Oe), and lines each with a slope of 2 are drawn through the points for the three lower dc field cases. For a dc field of 2500 Oe, the line drawn through the data has a slope of 3. In Fig. 7, data are shown for two settings of the dc field (200 and 1000 Oe), and lines with slopes of 2 and 3 again are drawn through the points.

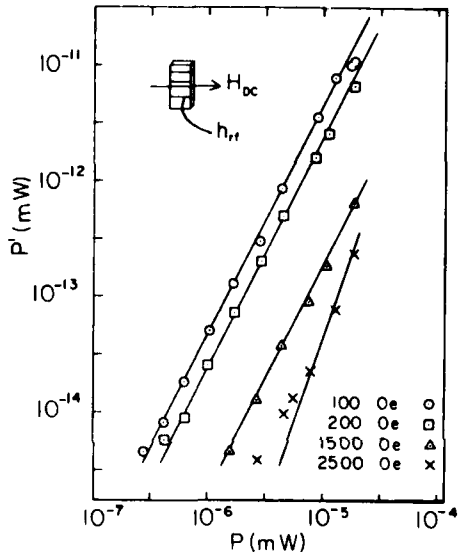


Fig. 6 — Intermodulation power P' as a function of the relative input power P at four different values of the dc magnetic field parallel to the rf magnetic field.

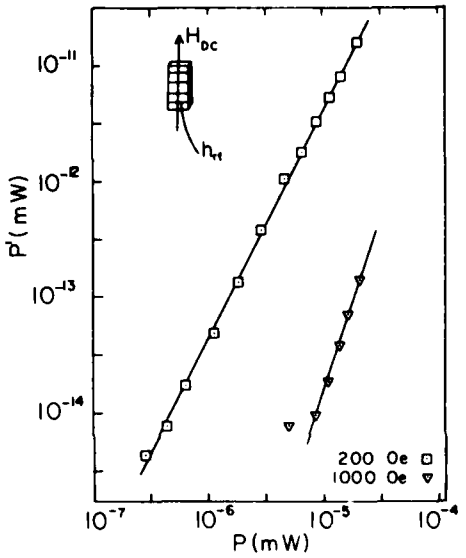


Fig. 7 — Same as Fig. 6 except for two different values of the dc magnetic field normal to the rf magnetic field.

The explanation of this data is not clear since the exact intermod-input-power dependence is not explainable in the absence of a detailed theory for the domain-wall dynamics. The change from an initial quadratic at low dc magnetic fields to a cubic dependence at fields near magnetic saturation of the sample may be due to a change in the type of nonlinearity. For example, when there is a large number of domains and domain walls (at low applied dc fields) there may be a large interaction between domain walls, as opposed to the high-field case when the wall density is small. Thus, the parameters affecting the excitations of the domain walls and hence their behavior as a function of applied power, may change drastically in the two cases.

It can be shown rigorously, however, that any physical phenomena that can be described by well-behaved mathematical functions (continuous functions with continuous derivatives) cannot give rise to third-order intermods whose power varies as the square of the input power. There are phenomena, however, which behave in a discontinuous way. For example, a pinned domain wall is immobile up to some critical field H_c . Thus, this pinned domain wall has a velocity versus applied field which is discontinuous at H_c. The intermod power produced by such a wall cannot be easily described mathematically and could conceivably vary in an unusual way with rf input power (i.e., a quadratic dependence of the third-order intermod power on input power). See Chapter VI.

V. THE VIBRATING DOMAIN WALL AS A NONLINEAR SYSTEM

Although a number of papers have been written in recent years on domain-wall vibration both in metals and in insulating ferromagnets, there has been very little mathematical treatement of the wall as a nonlinear system, particularly at the frequencies with which we are concerned here. Minakov and Fedosov[9] have recently discussed the energy loss of an rf field in a ferromagnetic metallic plate containing Bloch walls by solving the domain-wall equation of motion and Maxwell's equations self-consistently. In addition, Mitsek *et al.*,[10] in discussing the propagation of electromagnetic waves in magnetodielectrics, have assumed a solution of the same two equations in a form which involves the second harmonic of the fundamental wave. There still remains, however, a need for the solution of these two equations in a way which will yield not only the second harmonic but all higher harmonics and, in the case of two input frequencies, the intermodulation frequencies.

We therefore consider here in a very simple way how a vibrating domain wall may lead to intermodulation signals by writing the equation of motion for the wall as [11]

$$m\ddot{x} + \beta\dot{x} + \alpha x = 2MH, \tag{1}$$

where m is the mass per unit area, $\beta\dot{x}$ is a damping term, α is the restoring force constant, and $2MH$ is the applied force per unit area on the wall (M is the saturation magnetization and H is the applied field).

We extend Eq. (1) to our case by neglecting the damping for simplicity, by adding a nonlinear restoring force term, and by exciting the wall with two different rf frequencies, ω_1 and ω_2. Equation (1) then becomes

$$m\ddot{x} + \alpha x - \epsilon x^3 = 2M\,(H_1\cos\omega_1 t + H_2\cos\omega_2 t), \tag{2}$$

where H_1 and H_2 are the amplitudes of the rf magnetic fields in the sample. The parameter ϵ is chosen as positive, and thus the term $-\epsilon x^3$ makes the net restoring force weaker at any given value of x than it is without this term.

The solution of Eq. (2) can be obtained by a method of successive approximations[12] We assume a first approximate solution of the form

$$x = a\cos\omega_1 t + b\cos\omega_2 t, \tag{3}$$

where a and b are two parameters to be determined.

Substituting Eq. (3) into Eq. (2), we obtain a condition on parameters a and b by setting the resultant coefficients of the $\cos\omega_1 t$ and $\cos\omega_2 t$ terms equal to zero. These conditions are

$$-ma\omega_1^2 + \alpha a - 2MH_1 - \frac{3}{4}\epsilon a^3 - \frac{3}{2}\epsilon ab^2 = 0, \tag{4}$$

$$-mb\omega_2^2 + \alpha b - 2MH_2 - \frac{3}{4}\epsilon b^3 - \frac{3}{2}\epsilon a^2 b = 0. \tag{5}$$

We rewrite Eq. (2) as

$$m\ddot{x} = -\alpha x + \epsilon x^3 + 2M(H_1\cos\omega_1 t + H_2\cos\omega_2 t) \tag{6}$$

and substitute Eq. (3) for x on the right-hand side of Eq. (6). We also make use of Eqs. (4) and (5) to simplify the resulting equation. Doing this, integrating twice over time, and dividing by m, we obtain the second approximation for x as

$$
\begin{aligned}
x = {} & a\cos\omega_1 t + b\cos\omega_2 t \\
& - \frac{\epsilon}{m}\,\frac{a^3}{36\omega_1^2}\,\cos 3\omega_1 t - \frac{\epsilon}{m}\,\frac{b^3}{36\omega_2^2}\,\cos 3\omega_2 t \\
& - \frac{\epsilon}{m}\,\frac{3a^2 b}{4}\left[\frac{\cos(2\omega_1-\omega_2)t}{(2\omega_1-\omega_2)^2} + \frac{\cos(2\omega_1+\omega_2)t}{(2\omega_1+\omega_2)^2}\right] \\
& - \frac{\epsilon}{m}\,\frac{3ab^2}{4}\left[\frac{\cos(2\omega_2-\omega_1)t}{(2\omega_2-\omega_1)^2} + \frac{\cos(2\omega_2+\omega_1)t}{(2\omega_2-\omega_1)^2}\right] \\
& + \text{transient solution.}
\end{aligned}
\tag{7}
$$

We will consider only the long-time asymptotic solution and therefore disregard the transient solution. It is seen that there are only odd-order terms in the expression for the domain-wall amplitude.

Now, if we consider the power of the two frequency inputs as identical, as is the case experimentally, we can write Eq. (7) as

$$
\begin{aligned}
x = {} & a\cos\omega_1 t + a\cos\omega_2 t - \frac{\epsilon}{m}\,\frac{a^3}{36}\left[\frac{\cos 3\omega_1 t}{\omega_1^2} + \frac{\cos 3\omega_2 t}{\omega_2^2}\right] \\
& - \frac{\epsilon}{m}\,\frac{3a^3}{4}\left[\frac{\cos(2\omega_1-\omega_2)t}{(2\omega_1-\omega_2)^2} + \frac{\cos(2\omega_1+\omega_2)t}{(2\omega_1+\omega_2)^2} + \frac{\cos(2\omega_2-\omega_1)t}{(2\omega_2-\omega_1)^2} + \frac{\cos(2\omega_2+\omega_1)t}{(2\omega_2+\omega_1)^2}\right]
\end{aligned}
\tag{8}
$$

and Eqs. (4) and (5) as

$$
-ma\,\omega_1^2 + \alpha a - 2MH_1 - \frac{9}{4}\,\epsilon a^3 = 0. \tag{9}
$$

If we solve Eq. (9) for a in terms of m, ω_1, α, M, H_1, and ϵ, we can find the dependence of the power of the third-order intermod, i.e., the square of the coefficient of the $\cos(2\omega_2-\omega_1)$ term, on the input power. From Eq. (8) and Parseval's theorem,

$$
P_{2\omega_2-\omega_1} \sim \left[\frac{\epsilon}{m}\,\frac{3a^3}{4}\,\frac{1}{(2\omega_2-\omega_1)^2}\right]^2; \tag{10}
$$

but the input power goes as

$$
P_{\omega_1} \sim H_1^2.
$$

Hence, we must solve Eq. (9) for a as a function of H_1 and determine how the third-order intermod power varies with H_1^2.

In order to solve Eq. (9) for a, which we have done numerically, we shall insert typical or reasonable values of the parameters for nickel and vary ϵ. The parameters chosen here are typical and their exact values do not affect the functional variation obtained for $P'(H)$. The values used are $M = 590$ Oe, $m = 1.4\times10^{-10}$ g/cm^2, $\beta = 160$ g/cm^2 sec, $\alpha = \omega_0^2 m = 2.21\times10^8$ g/cm^2 sec^2, $\omega_1 = 2\pi f_1 = 9.173\times10^8$/sec, and $\omega_2 = 2\pi f_2 = 9.300\times10^8$/sec (Ref. 13).

The results of the solution of Eq. (9) for various values of ϵ and H_1 are given in Table 1. When there is no nonlinear term (i.e., $\epsilon = 0$) only one root exists which we call a_2 ; this root is proportional to H_1. For $0 < \epsilon \leqslant 10^{14}$ there are three real roots for a, each of which is a valid solution to Eq. (9). The particular root which is applicable depends on the way in which the systen is brought to its physical state,[14] i.e., to a given frequency and amplitude. We identify the second root, a_2, with the root

Table 1 — Solution of Eq. (9) for various values of the parameters ϵ and the rf magnetic field (in Oe). a's in units of 10^{-4}.

ϵ	H_1	a_1	a_2	a_3
0	5		0.577	
	10		1.15	
	20		2.31	
10^2	5	6.74×10^4	0.577	-6.74×10^4
	10	6.74×10^4	1.15	-6.74×10^4
	20	6.74×10^4	2.31	-6.74×10^4
10^{14}	5	6.43	0.581	−7.01
	10	6.06	1.19	−7.25
	20	4.90	2.78	−7.69
3×10^{14}	5	3.56	0.591	−4.15
	10	3.07	1.30	−4.37
	20	− 4.74	$2.37 + i\,1.32$	$2.37 - i\,1.32$
6×10^{14}	5	2.40	0.607	− 3.00
	10	− 3.21	$1.60 + i\,0.388$	$1.6 - i\,0.388$
	20	− 3.54	$1.77 + i\,1.35$	$1.77 - i\,1.35$
3×10^{15}	5	− 1.45	$0.727 + i\,0.269$	$0.727 - i\,0.269$
	10	− 1.61	$0.806 + i\,0.66$	$0.806 - i\,0.66$
	20	− 1.85	$0.92 + i\,1.02$	$0.92 - i\,1.02$

corresponding most closely to the root when $\epsilon = 0$. For any value of ϵ up to approximately 1×10^{14}, the root a_2 is proportional to H_1. From Eq. (10) , we see that $P_{2\omega_2-\omega_1}$ is proportional to a^6 or H_1^6; and since the input power is proportional to H_1^2, the third-order intermod power is proportional to the cube of the input power. This cubic dependence agrees with the experimental findings for high values of the dc field. When $\epsilon = 3\times10^{14}$, the roots start becoming complex with increasing H_1. When $\epsilon = 3 \times 10^{15}$ there are always two complex roots and one real root. The complex roots of a for the higher values of ϵ do not correspond to a real physical situation. The dependence of a on H_1 at higher values of ϵ is very weak. This is probably not significant since at these values of ϵ we have begun to violate the basic assumption of the calculation which is that the nonlinearity is a small perturbation of a linear system.

Although the intermod power variation with the cube of the input power obtained from this model of a vibrating domain wall agrees with the cubic behavior of P' versus P in Figs. 7 and 8, where the applied dc magnetic field is high, the quadratic power dependence shown in Figs. 7 and 8 at low dc fields still cannot be so explained. Terms proportional to x^2 or any other power of x, alone or in combination with an x^3 term, will lead to the generation of different intermods but will not affect the power dependence of the third-order intermods with input power.

VI. SUMMARY AND CONCLUSIONS

We have described a method to investigate nonlinear effects of rf fields in metals and have applied this method to a study of the third-order intermodulation generation signal in nickel as a function of an externally applied dc magnetic field and as a function of the applied rf power. The following comments can be made as a result of this investigation.

(1) The intermod signals excited in a nickel-plated rf connector, a nickel-plated wire, and flat nickel-plated copper samples all have the same origin.

(2) The intermod signal decreases monotonically with the magnetic domain-wall density in the sample.

(3) The sharp minimum in the intermod-signal vs. applied dc field is due to a phase cancellation of the sample signal and the background signal.

(4) The shape of the intermod-signal magnetic field variation can be qualitatively explained on the basis of the variation of the domain structure with field. This signal-domain-structure correlation suggests one use of the intermod technique to the study of domain density and distribution in bulk magnetic materials. Thus, the technique can complement neutron diffraction methods of domain structure studies.

(5) The intermod power input power variation is quadratic at low dc magnetic fields and cubic at high fields. The high field dependence is well explained by a model which ascribes the intermod signal to domain-wall oscillations that are driven by the applied rf field.

ACKNOWLEDGMENTS

We thank C.E. Young and E.L. Bruns for discussions involving the experimental apparatus and B.N. Kamm for suggesting the idea of canceling the background signal in order to improve the sensitivity of the apparatus. We are also grateful to V.J. Folen for many interesting and informative discussions.

REFERENCES

1. If f_1 and f_2 are the frequencies of two input signals, then the intermod frequencies may be any linear combination of f_1 and f_2. A signal with frequency $nf_1 \pm mf_2$ is called an $(n + m)$th-order intermod frequency. For example, if $n = 2$ and $m = 1$, we have $2f_1 - f_2$ or a third-order intermod signal.

2. C.E. Young first reported on intermod signals from nickel-plated connectors, 1976. See Chapter II.

3. See Chapter III.

4. D.I. Paul, G.C. Bailey, and A.C. Ehrlich, J. Appl. Phys., to be published.

5. P.D. Kim, D.M. Rodichev, and I.A. Safanov, Proc. Intern. Conf. on Magnetism Moscow, 1973, Vol. 4, p. 187 (in Russian).

6. See L.D. Landau and E.M. Lifshitz, *Mechanics*, 2nd ed. (Pergamon, New York, 1969), p. 94.

7. D.L. Mills (private communication).

8. W.H. Kraan and M. Th. Rekveldt, J. Magn. Mag. Mater,5, 247 (1977).

9. V.I. Minakov and V.N. Fedosov, Izv. Vyssch. Uchebn. Zaved. Fiz. No. 8, 95 (1977) [Sov. Phys. J. **20**, 1059 (1977)].

10. A.I. Mitsek, D.I. Sirota, and N.P. Kolmakova, Fiz. Nizk. Temp. **2**, 1031 (1976) [Sov. J. Low Temp. Phys. **2**, 507 (1976)].

11. W. Döring, Z. Naturforsch. **3a**, 373 (1948).

12. G. Joos, *Theoretical Physics*, 2nd ed. (Hafner, New York, 1950), p.99.

13. S. Chikazumi, *Physics of Magnetism* (Wiley, New York, 1964).

14. A.W. Marris and C.E. Stoneking, *Advanced Dynamics* (McGraw-Hill, New York, 1967), pp. 49-56.

Chapter V

INTRINSIC SOURCES OF IM GENERATION

George H. Stauss

Magnetism Branch
Electronics Technology Division

In addition to avoidable sources of intermodulation signals introduced in manufacture or assembly of a multiplex system, there will be some sources inherent to the materials of which the system is constructed. At some point these sources must provide a lower bound to possible system sensitivity. Such a limitation will only be significant if it is reached before thermal noise becomes dominant. The thermal noise power is $P_N = kT\Delta\nu$ W for a frequency bandwidth $\Delta\nu$. At 20°C and a bandwidth of 2500 Hz, the noise power becomes −140 dBm. It is potentially possible to lower either the temperature or the bandwidth by a factor of about 100, so an ultimate sensitivity limit of about −180 dBm is perhaps significant. Above these levels, a number of intrinsic IM mechanisms can be identified. This chapter constitutes a review of significant intrinsic IM sources.

1. RESISTIVE HEATING IN NON-MAGNETIC CONDUCTORS

This problem has been discussed in several places including Philco[1] and TRW[2] studies. This section is aimed at unifying the discussions and eliminating apparent discrepancies in the conclusions. It is concluded that this effect can be significant at high power levels at all frequencies.

The calculation which follows is done classically (no quantum effects; cf. Chapter I) and to first order. It is similar to that of the Philco study, but with fewer approximations. The TRW calculations, done differently and specifically for a waveguide configuration, reach substantially the same conclusions. Two primary frequencies ω_1 and ω_2 ($\omega_1 > \omega_2$) are assumed present with electric fields of amplitudes E_1 and $E_2 = \beta E_1$ parallel to the metallic surface, and the normally strongest 3rd order IM signal at $(2\omega_1 - \omega_2)$ is sought. The surface is taken as part of an infinite plane and $\omega_1, \omega_2, (2\omega_1 - \omega_2)$ are assumed close enough in frequency that a single skin depth $\delta_0 \equiv \left(\frac{2\rho_0}{\omega\mu}\right)^{1/2}$ in terms of symbols in the appended list can be used for all.

To lowest order, the density of primary power being dissipated at a depth z below the surface is then

$$p_{IN}(z) = \frac{E_1^2}{2\rho_0} e^{-\frac{2z}{\delta_0}} \left\{(1+\beta^2) + \cos\left[2\omega_1 t - \frac{2z}{\delta_0}\right] + \beta^2 \cos\left[2\omega_2 t - \frac{2z}{\delta_0}\right] + 2\beta \cos\left[(\omega_1+\omega_2)t - \frac{2z}{\delta_0}\right] + 2\beta \cos(\omega_1 - \omega_2)t\right\}, \quad (1)$$

where ρ_0 is the material resistivity at ambient temperature. [For copper $\rho_0 = 1.72 \times 10^{-8}$ Ω m at room temperature]. This power dissipation produces resistive heating and changes the local value of the resistivity. Heating can in principle also lead to local dimensional changes which could produce intermodulation. However, with temperature changes expected to be $\leq 10^{-6}$K even at the surface, a linear

thermal coefficient of 10^{-3}/K, weak dependence of the propagated signal on small dimensional changes, and considerable bulk thermal inertia, this effect is considered negligible. From (1), one finds that the total power absorbed at all depths per unit surface area, averaged in time, is

$$p_{IN} = \frac{\delta_0 E_1^2 (1 + \beta^2)}{4\rho_0}. \tag{2}$$

To find the change in local resistivity, one uses the fact that ρ is essentially proportional to absolute temperature and then looks for the temperature distribution. This is obtained from the diffusion equation

$$-G \frac{\partial^2 T(z)}{\partial z^2} + C_h \frac{\partial T(z)}{\partial t} = p_{IN}(z). \tag{3}$$

Here G is the thermal conductivity and C_h the heat capacity. [For copper at room temperature, $G = \sim 2 \times 10^2$ W/Km, $C_h = 3.44 \times 10^6$J/Km^3]. Since Eq. (3) is linear in $T(z)$, the solution will be a superposition of solutions for the various terms in $p(z)$ of Eq. (1). We neglect the steady state term; for the others, boundary conditions are taken to be $\frac{\partial T(z)}{\partial z} = 0$ at $z = 0$ (no heat flow into air space) and $T(z)_{z=\infty} = 0$ (the thickness of ordinary components is sufficient to be considered infinite).

In solving for $T(z)$, one will obtain contributions corresponding to all the different frequency dependent power terms. When these are converted to changes in resistivity and combined with the input signals, they will yield currents at several frequencies. However, the desired $(2\omega_1 - \omega_2)$ current only arises either from modulation of the ω_1 input by an $(\omega_1 - \omega_2)$ resistivity term, or from modulation of the ω_2 input by a $2\omega_1$ term. Furthermore, evaluation of the current amplitudes shows that modulation by the $(\omega_1 - \omega_2)$ resistivity term is dominant. Hence, only this part will be discussed in what follows. From the solution for $T(z)$ we then obtain

$$\rho(z) = \rho_0 \left[1 + \frac{T(z)}{T_0}\right] = \rho_0 + \frac{\rho_0}{T_0} \left(\frac{\beta E_1^2}{2\rho_0 G}\right) \left(\frac{4}{\delta_0^4 + W^4}\right)^{1/2} \left\{\sqrt{2}\, e^{-\frac{Wz}{\delta_0 W}} \cos\left[(\omega_1 - \omega_2)t - Wz + \phi - \frac{\pi}{4}\right] - e^{-\frac{2z}{\delta_0}} \cos\left[(\omega_1 - \omega_2)t + \phi\right]\right\}. \tag{4}$$

Here $W^2 = \frac{(\omega_1 - \omega_2) C_h}{2G}$ and $\tan\phi = \frac{W^2 \delta_0^2}{2}$.

This resistivity will modulate the local currents produced by the primary input at ω_1. Writing the skin effect in differential form, the current density at z, $J(z)$, is found from

$$\frac{1}{J(z)} \frac{dJ(z)}{dz} = -\frac{(1 + i)}{\delta(z)} = -\frac{(1 + i)}{\delta_0 \left(\frac{\rho}{\rho_0}\right)^{1/2}} \approx -\frac{(1 + i)}{\delta_0} \left[1 - \frac{T(z)}{2T_0}\right]. \tag{5}$$

For an input at ω_1 one obtains

$$J(z) = \frac{E_1}{\rho_0} e^{-\frac{z}{\delta_0}} \left\{\left[1 - \frac{T(0)}{T_0}\right] \cos\left[\omega_1 t - \frac{z}{\delta_0}\right] + \frac{\int_0^z T(z)\,dz}{\sqrt{2}\,\delta_0 T_0} \cos\left[\omega_1 t - \frac{z}{\delta_0} + \frac{\pi}{4}\right]\right\}. \tag{6}$$

When the quantities $T(0)$ and $\int_0^z T(z)dz$ are evaluated and substituted into (6), the IM current density with frequency $2\omega_1 - \omega_2$ is found to be

$$J_{IM}(z) = \frac{\beta E_1^2}{\rho_0}\,\frac{E_1\delta_0^2}{\rho_0}\,\frac{1}{2GT_0}\,(4+x^4)^{-1/2}\,e^{-\frac{z}{\delta_0}} \times$$

$$\left\{\left[\left(\frac{3}{8}-\frac{1}{2x}+\frac{1}{4x^2}\right)\cos\left[(2\omega_1-\omega_2)t-\frac{z}{\delta_0}+\phi\right]+\left(\frac{1}{8}-\frac{1}{2x}+\frac{1}{4x^2}\right)\times\right.\right.$$

$$\left.\sin\left[(2\omega_1-\omega_2)t-\frac{z}{\delta_0}+\phi\right]\right.$$

$$-\frac{e^{-Wz}}{2\sqrt{2}\,x^2}\sin\left[(2\omega_1-\omega_2)t-\frac{z}{\delta_0}+\phi+\frac{\pi}{4}-Wz\right]+\frac{e^{\frac{2z}{\delta_0}}}{4\sqrt{2}}\times \tag{7}$$

$$\left.\left.\cos\left[(2\omega_1-\omega_2)t-\frac{z}{\delta_0}+\phi+\frac{\pi}{4}\right]\right\}.$$

The dimensionless parameter $x \equiv \delta_0 W$ is of order of magnitude unity here, but is material dependent.

In treating an unperturbed input signal, one finds that $\bar{p}_{IN}$, the power absorbed per unit surface area, is related to the total current density in the same element, $J \equiv \int_0^\infty J(z)dz$, by way of the surface resistance $R_s \equiv \frac{\rho_0}{\delta_0}$; i.e.,

$$\bar{p}_{IN} = \overline{J^2}R_s \tag{8}$$

We will apply this same expression to find the power lost by $J_{IM} \equiv \int_0^\infty J_{IM}(z)dz$ in the same volume. Integrating Eq. (7), we obtain

$$J_{IM} = \frac{\frac{\beta E_1^2}{\rho_0}\left[\frac{E_1\delta_0^2}{\rho_0}\right]\delta_0}{2GT_0(4+x^4)^{1/2}}\left\{\frac{7}{40}\cos\left[(2\omega_1-\omega_2)t+\phi\right]+\left(\frac{9}{40}-\frac{1+2x}{4x(1+x)}\right)\times\right. \tag{9}$$

$$\left.\sin\left[(2\omega_1-\omega_2)t+\phi\right]\right\}.$$

Hence from (8) and (9) the total IM power absorbed per unit area becomes

$$\bar{p}_{IM} = \frac{\rho_0}{\delta_0}\,\overline{J_{IM}^2} = \frac{\beta^2\left[\frac{\delta_0 E_1^2}{4\rho_0}\right]^3\delta_0^2}{G^2T_0^2(4+x^4)}\left[\frac{13}{20}-\frac{(9x^2-x-5)(2x+1)}{10x^2(x+1)^2}\right], \tag{10}$$

or, using Eq. (2),

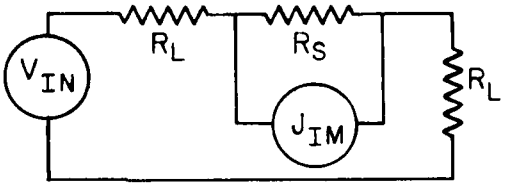
$$\bar{p}_{IM} = \frac{\beta^2}{(1+\beta^2)^3}\,\frac{\bar{p}_{IN}^3\delta_0^2}{G^2T_0^2}\,f(x), \tag{11}$$

where $$f(x) \equiv \frac{1}{(4+x^4)}\left[\frac{13}{20}-\frac{(9x^2-x-5)(2x+1)}{10x^2(x+1)^2}\right].$$

Reasonable values for x in our approximation that $(\omega_1 - \omega_2) << \omega_1$ are in the range $x = 1$ to 5 for copper. In order to satisfy the assumption of first order perturbation, $\frac{T(0)}{T_0}$ and $\int_0^z \frac{T(z)dz}{\delta_0 T_0}$ in Eq. (6) must be much less than 1. This requires that $x > \left[\frac{\delta_0 \bar{p}_0}{GT_0}\right]$, which here roughly means $x > 10^{-3}$. Hence, the infinity in $f(x)$ at $x = 0$ is meaningless. For reasonable x values, $f(x)$ is controlled mainly by the $\frac{1}{4 + x^4}$ factor. Explicit values are

x	=	1	2	3	4	5
$f(x)$	=	.085	.012	.0036	.0013	.0006

The reason for obtaining Eq. (11) is that the IM power received at the output load can be expressed in terms of this quantity. We assume that we are dealing with a well designed circuit with matched input and load. Then we can represent a unit IM source included in a linear system by the equivalent circuit illustrated.



From this one IM source the power received at the load $\bar{p}_{IML} = \overline{J_{IM}^2} R_S \left[\frac{R_L R_S}{(2R_L + R_S)^2}\right]$. But the input power lost in the unit IM source is $\frac{V_{IN}^2 R_S}{2(R_L + R_S)^2}$ while that available to the load is $\frac{\overline{V_{IN}^2}}{4R_L}$; the ratio of these gives $\frac{4R_S R_L}{(2R_L + R_S)^2}$. Hence we can write, using (10) or (8) and (2)

$$\bar{p}_{IML} = \frac{\bar{p}_{IN}\bar{p}_{IM}}{4P_{IN}} \tag{12}$$

for the delivered IM power from a unit IM source element. If the assumption is made that all IM signals arrive in phase at the load, the total $P_{IML} = (\text{Area})^2 \bar{p}_{IML}$. This then gives the result that

$$P_{IML} = \frac{\bar{p}_{IM}(P_{IN\,\text{lost}})^2}{4\bar{p}_{IN}P_{IN}}. \tag{13}$$

Using Eq. (11), we obtain finally

$$P_{IML} = \frac{\beta^2}{(1 + \beta^2)^3}\left[\frac{\delta_0}{2GT_0}\right]^2 \bar{p}_{IN}^2 \frac{(P_{IN\,\text{lost}})^2 f(x)}{P_{IN}} \tag{14}$$

As numerical examples, we consider two cases. (i) A high-Q UHF structure with high power density in a very restricted area with $\nu_1 = 270$ MHz, $\nu_2 = 245$ MHz, $\beta = 1$, $\bar{p}_{IN} = 4.5 \times 10^3$ W/m^2, $P_{IN} = 100$ W, $P_{IN\,\text{lost}} = 20$ W, $x = 3.2$, $f(x) = .0028$. This leads to $P_{IML} = .71 \times 10^{-17}$ W or -142 dBm. (ii) The x-band waveguide case considered in the Philco report, with $\nu_1 = 8050$ MHz, $\nu_2 = 7900$ MHz, $\beta = 0.1$, $\bar{p}_{IN} = 145$ W/m^2, $P_{IN} = 1000$ W, $P_{IN\,\text{lost}} = 45$ W, $x = 1.45$, $f(x) = .031$. This gives $P_{IML} = 1.1 \times 10^{-22}$ W or -190 dBm. Correction of a numerical error brings the Philco value to -232 dBm; the remaining difference is associated with the assumption in that report that modulation of the

input at ω_2 by the thermal effects at $2\omega_1$ is dominant, whereas we find modulation of the input at ω_1 by $(\omega_1 - \omega_2)$ thermal terms produces an effect about 40 dB stronger. Both of these examples indicate that this mechanism is capable, under certain conditions, of producing significant IM signals.

A particularly important non-intrinsic IM source which has been identified is the pressure-closed junction. Because of oxidation and surface roughness this junction contains insulating, semiconducting, and metallic regions in dimensions small enough to permit tunneling. However, consideration of the restricted area of metallic contact within the skin layer and across a gap of perhaps 100 Å suggests that the junction would be an important IM source simply through resistive heating. For example, in the waveguide discussed above, if a flange junction were introduced in which 1 W is dissipated or $\bar{p}_{IN} = 2 \times 10^9$ W/m^2, then P_{IML} would be increased 110 dB to a significant level of -122 dBm. It is to be expected that increasing contact pressure would decrease this contribution to junction IM production.

To see the important dependences more clearly, we will write $P_{IN\,\text{lost}} = \bar{p}_{IN} A$ where A is the effective conductor area; $\frac{1}{T_0}$ is generalized to ψ in the expression $\rho = \rho_0(1 + \psi T)$; $\bar{p}_{IN} \propto QP_{IN}R_s = QP_{IN}\rho_0/\delta_0$, where Q is descriptive of the element in question and R_s is surface resistivity of the conductor; $f(x) \approx x^{-3}$ quite well in the range $1 < x < 5$ and $x = \left[\frac{\delta_0^2(\omega_1 - \omega_2)C_h}{2G}\right]^{1/2}$. Thus Eq. (14) lead to

$$P_{IML} \propto \frac{\beta^2}{(1+\beta^2)^3}\left(\frac{A\psi}{2G}\right)^2\left[\frac{2G}{\delta_0^2(\omega_1-\omega_2)C_h}\right]^{3/2}\delta_0^2 P_{IN}^3\left(\frac{\rho_0}{\delta_0}\right)^4 Q^4$$

$$\propto \frac{\beta^2}{(1+\beta^2)^3}\,\frac{A^2\psi^2Q^4}{G^{1/2}C_h^{3/2}}P_{IN}^3\rho_0^4\delta_0^{-5}(\omega_1-\omega_2)^{-3/2} \tag{15}$$

$$\propto \frac{\beta^2}{(1+\beta^2)^3}\,\frac{A^2\psi^2Q^4}{G^{1/2}C_h^{3/2}}P_{IN}^3\rho_0^{3/2}\,\frac{\omega_1^{5/2}}{(\omega_1-\omega_2)^{3/2}}$$

Thus IM production due to resistive heating is enhanced by:

(a) Large power inputs,

(b) β values near 1 (equal power in both carriers),

(c) Large-area elements,

(d) High-Q elements,

(e) High carrier frequencies,

(f) Small carrier separations,

(g) Conductors with high resistivity, low heat capacity, low thermal conductivity, and high-thermal coefficient of resitivity ψ.

(It might also be noted that Eq. (14) is, to the level of approximations made, in functional and quantitative agreement with the calculations by TRW, after correction of a numerical error. The TRW result, obtained in the manner of Section VI (dielectrics), develops on the fact that for long signal paths and sufficient attenuation the value of P_{IM} will saturate and then decline for longer paths. For most applications the critical length, where the primary signals are attenuated 4.8 dB, will be quite long.)

2. MAGNETORESISTIVE GENERATION IN NON-MAGNETIC CONDUCTORS

Magnetic fields applied to a conductor have the effect of altering its resistivity. Hence those fields associated with currents in the material can potentially create IM signals. An approach similar to that used in Part 1. can be taken. We will evaluate the variations produced in the local resistivity by the primary currents and then find J_{IM}. This effect is found to be small.

First consider a simple numerical result. With $\delta_0 = 10^{-6}$m typical of copper, a dissipation $\bar{p}_0 = 10^3$ W/m^2 using (8) leads to $J \approx 240$ A rms as the current through a 1 m width of skin layer. But in MKS units we can write $\oint \vec{B} \cdot d\vec{l} = \mu_0 I$ where I is the current through the loop around which $\vec{B}$ is integrated. Taking the loop as a rectangle transverse to the current, 1 m wide and several δ deep, with its upper side at the surface of the conductor, the contributions of the short ends to the integral cancel, and B on the lower long side is essentially zero. Hence B on the upper surface becomes roughly $B = 240\mu_0$ Tesla, which is equivalent to about 3 Gauss. But for the transverse magnetoresistive effect (B perpendicular to current flow as it will be for self-produced fields) we know[3] for copper that $\frac{\rho - \rho_0}{\rho_0} = 3.98 \times 10^{-17} H^2$ in MKS units. Thus resistance changes of a part in 10^{12} are possible in our example.

Given $$\rho(z) = \rho_0(1 + \sigma H^2(z)), \tag{16}$$

which is quite general since a linear contribution does not produce a 3rd order IM of interest anyway, we need to know $H(z)$. Using the same path integral formulation used in the example above, if we take $H(z)_{z\to\infty} = 0$,

$$H(z) = \int_z^\infty J_{IN}(z)\,dz. \tag{17}$$

For inputs at two frequencies, we can write the primary current, allowing for skin effect,

$$J_{IN}(z) = \mathrm{Re}\left\{(e^{i\omega_1 t} + \beta e^{i\omega_2 t})\ \frac{E_1}{\rho_0}\ e^{\frac{-(1+i)z}{\delta_0}}\right\}. \tag{18}$$

This then leads to $H(z)$ and to $\rho(z)$ through (16),

$$\rho(z) = \rho_0\left[1 + \frac{\sigma\delta_0^2 E_1^2}{4\rho_0^2}\ e^{-\frac{2z}{\delta_0}}\left\{(1+\beta^2) + \sin\left[2\omega_1 t - \frac{2z}{\delta_0}\right] + 2\beta\cos(\omega_1 - \omega_2)t\right.\right.$$
$$\left.\left. + 2\beta\sin\left[[\omega_1 + \omega_2]t - \frac{2z}{\delta_0}\right] + \beta^2\sin\left[2\omega_2 t - \frac{2z}{\delta_0}\right]\right\}\right]. \tag{19}$$

If Eq. (19) is combined with initial input signals, various frequency terms again result. The only parts of $\rho(z)$ which lead to IM products at $(2\omega_1 - \omega_2)$ are

$$\rho(z) = \frac{\sigma E_1^2 \delta_0^2}{4\rho_0}\ e^{-\frac{2z}{\delta_0}}\left[\sin\left[\omega_1 t - \frac{2z}{\delta_0}\right] + 2\beta\cos(\omega_1 - \omega_2)t\right]. \tag{20}$$

Unlike the resistive heating case, here both terms in the brackets remain significant. When (20) is combined with the input signals, we obtain a form similar to Eq. (6)

$$J(z) = \mathrm{Re}\left\{\frac{E_1}{\rho_0}(e^{i\omega_1 t} + \beta e^{i\omega_2 t})\left[1 - \frac{\rho(0) - \rho_0}{\rho_0}\right]\left[1 + \frac{(1+i)}{2\delta_0}\int_0^z \frac{\rho(z) - \rho_0}{\rho_0}\,dz\right]e^{-\frac{(1+i)z}{\delta_0}}\right\}. \tag{21}$$

The results analogous to Eqs. (7), (9), (10) and (11) then become

$$J_{IM}(z) = -\frac{\beta\sigma E_1^3\delta_0^2 e^{-\frac{z}{\delta_0}}}{16\rho_0^3}\left\{\left[\left(\frac{7}{2}+\frac{1}{2}e^{-\frac{2z}{\delta_0}}\right)\cos\frac{z}{\delta_0}+\left(1+e^{-\frac{2z}{\delta_0}}\right)\sin\frac{z}{\delta_0}\right]\cos(2\omega_1-\omega_2)t\right.$$

$$\left.+\left[\left(\frac{5}{2}+\frac{1}{2}e^{-\frac{2z}{\delta_0}}\right)\sin\frac{z}{\delta_0}+\left(3-e^{-\frac{2z}{\delta_0}}\right)\cos\frac{z}{\delta_0}\right]\sin(2\omega_1-\omega_2)t\right\}, \tag{22}$$

$$J_{IM} = \frac{-5\beta\sigma E_1^3\delta_0^3}{32\rho_0^3}\left[\cos(2\omega_1-\omega_2)t+\sin(2\omega_1-\omega_2)t\right], \tag{23}$$

and thus

$$\overline{J_{IM}^2}R_S = \frac{25}{16}\beta^2\left(\frac{\sigma^3\delta_0^2}{\rho_0^2}\right)\left(\frac{\delta_0E_1^2}{4\rho_0}\right)^3 = \frac{25}{16}\frac{\beta^2}{(1+\beta^2)^3}\left(\frac{\sigma^2\delta_0^2}{\rho_0^2}\right)\bar{p}_{IN}^3 = \bar{p}_{IM}. \tag{24}$$

Comparing this with Eq. (11) and (13) we find that for equal input power and input losses

$$\frac{P_{IML}\text{ (magnetoresistive)}}{P_{IML}\text{ (thermoresistive)}} = \frac{\bar{p}_{IM}\text{ (magnetoresistive)}}{\bar{p}_{IM}\text{ (thermoresistive)}} = \frac{25}{16}\left(\frac{\sigma^2}{\rho_0^2}\right)\left(\frac{G^2T_0^2}{f(x)}\right). \tag{25}$$

Thus the same functional behavior exists as for Eq. (14), except that the result is not sensitive to reasonable changes in the sideband position, signified by the absence of the x dependence in (24). Numerically for copper the ratio (25) takes the value $1.34\times10^{-7}/f(x)$ and for the two cases considered in Part 1. with $f(x) = .0028$, or $f(x) = .031$, the magnetoresistive IM power is found to be down from the thermoresistive values by 43 dB and 54 dB, respectively. For non-magnetic materials consideration of reasonable values of $f(x)$ indicates that the magnetoresistive effects are insignificant.

3. DIRECT VARIATION IN RESISTIVITY WITH CURRENT

If $\rho(z) = \rho_0(1+\xi J_{IN}(z)^2)$ where $J_{IN}(z)$ represents the primary current density, then the appropriate part of $J_{IN}(z)^2$ leading to the 3rd order intermodulation signal can be extracted from (1) as

$$J_{IN}(z)^2 = \frac{E_1^2}{2\rho_0^2}e^{-\frac{2z}{\delta_0}}\left[\cos\left(2\omega_1t-\frac{2z}{\delta_0}\right)+2\beta\cos(\omega_1-\omega_2)t\right]. \tag{26}$$

Then writing

$$\frac{1}{J(z)}\frac{dJ(z)}{dz} = -\frac{(1+i)}{\delta_0(\rho/\rho_0)^{1/2}} = -\frac{(1+i)}{\delta_0(1+\xi J_{IN}(z)^2)^{1/2}} \approx -\frac{(1+i)}{\delta_0}(1-\xi J_{IN}(z)^2/2), \tag{27}$$

we obtain

$$J(z) = \mathrm{Re}\left\{J(0)\exp\left[-\frac{(1+i)z}{\delta_0}\right]\left[1+\frac{\xi(1+i)}{2\delta_0}\int_0^z J_{IN}(z)^2dz\right]\right\}. \tag{28}$$

With $J(0) = \frac{E_1}{\rho_0}(e^{i\omega_1t}+\beta e^{i\omega_2t})(1-\xi J_{IN}(0)^2)$, to lowest order one gets

$$J_{IM}(z) = \frac{\xi\beta E_1^3 e^{-\frac{z}{\delta_0}}}{2\rho_0^3}\left\{\left[-\left(1+\frac{1}{2}e^{-\frac{2z}{\delta_0}}\right)\cos\frac{z}{\delta_0}+\left(\frac{5}{8}-\frac{3}{8}e^{-\frac{2z}{\delta_0}}\right)\sin\frac{z}{\delta_0}\right]\cos(2\omega_1-\omega_2)t\right.$$

$$\left.-\left[\left(\frac{3}{8}-\frac{3}{8}e^{-\frac{2z}{\delta_0}}\right)\cos\frac{z}{\delta_0}+\frac{1}{2}e^{\frac{-2z}{\delta_0}}\sin\frac{z}{\delta_0}\right]\sin(2\omega_1-\omega_2)t\right\}. \tag{29}$$

Integrated over z, then this leads to

$$J_{IM} = -\frac{\beta\xi E_1^3\delta_0}{16\rho_0^3}\left[3\cos(2\omega_1-\omega_2)t+\sin(2\omega_1-\omega_2)t\right]. \tag{30}$$

Thus

$$\overline{J_{IM}^2}R_S = \frac{\rho_0}{\delta_0}\left(\frac{\beta E_1^3}{\rho_0^3}\right)^2\frac{5\xi^2\delta_0^2}{256} = \frac{5}{4}\frac{\beta^2\xi^2}{\delta_0^2\rho_0^2}\left(\frac{\delta_0 E_1^2}{4\rho_0}\right)^3 = \frac{5}{4}\frac{\beta^2}{(1+\beta^2)^3}\left(\frac{\xi^2}{\delta_0^2\rho_0^2}\right)\bar{p}_{IN}^3 \tag{31}$$

which can be compared with (11) or (24).

This effect is hard to evaluate numerically. Any experimental determination of the coefficient ξ would have to be separated from the magnetoresistive effect dissussed in Part 2. Although the influence of skin depth is different in (23) and (31), one might anticipate IM contributions of similar magnitudes in the two cases.

4. RESISTIVE HEATING AND MAGNETORESISTIVE IM IN FERROMAGNETIC COMPONENTS

When components are composed of ferromagnetic materials, the kinds of components chosen in the previous numerical cases become so lossy as to be impossible for consideration. For a more suitable component we chose a 10 cm long coaxial line similar to RG-19 with $Z_0 = 50\Omega$, ID = .435 cm, OD = 1 cm. For such a line, standard expressions give attenuation = $8.68\alpha_c$ dB/m, where $\alpha_c = 1.05 R_s$ (cf. Eq. (8)). The frequencies will be chosen as $\nu_1 = 270$ MHz, $\nu_2 = 245$ MHz. Thus with copper elements in this line there will be a loss of .00376 dB or 9×10^{-4} of input power. If nickel components instead are used, the corresponding loss will change with skin depth depending on the proper value of the permeability μ. This will be taken as the average small signal value of $\frac{dB}{dH}$ and will be presumed constant here (but see Part 5.). If μ is assumed to be 100 μ_0 or 500 μ_0 the fractional loss of input power becomes .019 or .043, respectively. Further assuming $\beta = 1$ and a total input power of 60W, we can then evaluate the IM products for copper or for nickel elements. In this part and in Part 5., the permeabilities used are those suitable to low frequencies. They are still roughly appropriate in UHF systems, but may decrease by an order of magnitude on going to microwave frequencies.[4]

The coaxial component will be treated in an average fashion, with equal power densities on both conductors. One can alternatively distinguish the two, assuming equal total currents on both. In this latter case, the current densities and the total powers absorbed vary inversely as the radii. This implies that the inner conductor is by far the more significant IM source. But since the ratios are known, one can relate the separated conductor results to the average case with the same total absorption. We find for $q = \frac{r_2}{r_1}$ that $\bar{p}_{IM}$ from the outer conductor alone is $1/q^2(1+q)^2$ times the average calculated, while that from the inner conductor is $q^4/(1+q)^2$ times the average. With the coax dimensions chosen, this means that the resultant P_{IM} from the inner conductor is 4 dB above that found from the average, while the outer conductor gives a result which is down 18 dB and negligible in comparison.

If we evaluate P_{IML} from Eq. (14), averaged as above, we find for copper, since $\bar{p}_{IN} = 12$ W/m^2 and $P_{IN\,lost} = 0.054$ W, that $P_{IML} = .62 \times 10^{-27}$W or only -242 dBm due to resistive heating. Magnetoresistive effects are still 43 dB below this, since the frequencies and thus $f(x)$ were chosen as for the first example in Part 1.

But going now to nickel elements we must alter the material parameters: $\rho_0(Ni) = 4.5\rho_0(Cu)$, $G(Ni) = \frac{1}{6.5}G(Cu)$, $C_h(Ni) = 1.15C_h(Cu)$, and we will look at both $\mu = 500\mu_0$ and $\mu = 100\mu_0$. We then obtain

$\mu = 500\mu_0$	$\mu = 100\mu_0$
$P_{IN\,lost} = 2.58$ W	$P_{IN\,lost} = 1.14$ W
$R_s = 47.5\ R_s(Cu)$	$R_s = 21\ R_s(Cu)$
$\delta_0 = \frac{1}{10.5}\,\delta_0(Cu)$	$\delta_0 = \frac{1}{4.7}\,\delta_0(Cu)$
$x = 0.26x(Cu)$	$x = 0.58x(Cu)$
$f(x) = 0.136$	$f(x) = 0.0153$
$P_{IML} = 0.59 \times 10^{-19}$ W *or* -162 dBm	$P_{IML} = 0.25 \times 10^{-21}$ W *or* -186 dBm

The IM levels here are for resistive heating effects. Although they are still at low levels, they are considerably stronger than the corresponding copper value of -242 dBm.

To evaluate the magnetoresistive contribution, we need a value for the coefficient σ in Eq. (16). Material is available in a discussion[5] by Jan which shows that below about half of saturation magnetization, $\left(\frac{M_s}{2}\right)$, both transverse and longitudinal fields give $\frac{\Delta\rho}{\rho_0}$ proportional to M^2. Numerically, Jan's curves for nickel in transverse fields show $\frac{\Delta\rho}{\rho_0} = -4 \times 10^{-2}\left(\frac{M}{M_s}\right)^2$. Generalizing this somewhat, if a strong transverse field H_0 is applied, an expansion about H_0 gives

$$\frac{\Delta\rho(H) - \Delta\rho(H_0)}{\rho_0} = b\left\{2M(H_0)\frac{dM}{dH}(H_0)\,(H - H_0) + \left[\left(\frac{dM}{dH}(H_0)\right)^2 + M(H_0)\frac{d^2M}{dH^2}(H_0)\right](H - H_0)^2\right\} \tag{32}$$

where $b = -\frac{4 \times 10^{-2}}{M_s^2}$. Only the quadratic term will contribute to the 3rd order IM. (For M near M_s, $\frac{\Delta\rho}{\rho_0}$ is experimentally quite linear in H, so saturation must drastically reduce IM production.) Within the coefficient of the quadratic term, the first part is positive; the second will be positive and augment the first at the lower end of the magnetization curve, but must eventually become negative and weaken the first. To simplify, we will take the first coefficient alone, giving $\sigma = b\left(\frac{dM}{dH}\,(H_0)\right)^2 = b\left(\frac{\mu}{\mu_0} - 1\right)^2$. Since $M_s = 4.85 \times 10^5$A/m we obtain

$\mu = 500\mu_0$	$\mu = 100\mu_0$
$\sigma = -4.24 \times 10^{-8}$ (A/m)$^{-2}$	$= -1.69 \times 10^{-9}$ (A/m)$^{-2}$
$\frac{P_{IML}\ \text{(magnetoresistive)}}{P_{IML}\ \text{(thermoresistive)}} = 1.3 \times 10^9$ or $+91$ dB	$= 1.8 \times 10^7$ or $+72$ dB
Magnetoresistive $P_{IML} = -71$ dBm	$= -114$ dBm

Thus for magnetic materials, magnetoresistive effects are extremely strong sources of IM signals. Despite this, however, we find that an even more important source exists in the direct variation of the permeability with current. This is discussed in the next part.

It should be emphasized that these results are only approximate. The value of μ is a continuously varying function of field strength and frequency and the skin depth is thus itself a function of depth. Also, especially with external fields applied, there are directional variations in effective permeability since the ability of a small field to change the magnitude of magnetization is much greater for fields parallel rather than perpendicular to existing magnetization. For example, a solenoidal field parallel to the coaxial component is parallel to the currents in the conductors and normal to the fields which they produce, and so should decrease the effective permeability and consequent IM production more effectively than a transverse applied field. Furthermore, shape effects on the demagnetizing field make the solenoidal field much more effective in saturating the magnetic components, requiring only about $10^{-2}T$ as opposed to several tenths Tesla for a transverse applied field.

5. IM PRODUCTION DUE TO VARIATIONS IN PERMEABILITY IN FERROMAGNETIC COMPONENTS

To illustrate the significance of this effect, ignoring other contributions including imaginary μ'', we allow the permeability to vary in the form

$$\mu(z) = \mu_u (1 + DH(z)^2) \tag{33}$$

since linear variations will not lead to 3rd order IM's. We are interested here in fields produced by the currents in the conductors. From the primary currents, the parts of $H^2(z)$ which contribute to the 3rd order IM become (cf Eqs. (16) and (20))

$$H^2(z) = \frac{E_1^2\sigma_0^2}{4\rho_0^2} e^{-\frac{2z}{\delta_0}} \left[\sin\left(2\omega_1 t - \frac{2z}{\delta_0}\right) + 2\beta \cos(\omega_1 - \omega_2)t\right]. \tag{34}$$

This is to be incorporated into an expression analogous to Eq. (5)

$$\frac{1}{J(z)}\frac{dJ(z)}{dz} = \frac{(1+i)}{\delta_0(\mu_u/\mu(z))^{1/2}} - \frac{(1+i)}{\delta_0}\left[1 + \frac{DH^2(z)}{2}\right] \tag{35}$$

for which the solution, similar to (21), becomes

$$J(z) = \mathrm{Re}\left\{\frac{E_1}{\rho_0}(e^{i\omega_1 t} + \beta e^{i\omega_2 t})\left[1 - \frac{(1+i)D}{2\delta_0}\int_0^z H^2(z)\,dz\right] e^{-\frac{(1+i)z}{\delta_0}}\right\}. \tag{36}$$

From this we obtain

$$J_{IM}(z) = \frac{\beta D\delta_0^2 E_1^3}{32\rho_0^3} e^{-\frac{z}{\delta_0}}\left\{\left(e^{\frac{-2z}{\delta_0}} - 1\right)\left(\cos\frac{z}{\delta_0} + 2\sin\frac{z}{\delta_0}\right)\cos(2\omega_1 - \omega_2)t \right. \tag{37}$$

$$\left. + \left[(e^{-\frac{2z}{\delta_0}} - 1)\left(\sin\frac{z}{\delta_0} - 2\cos\frac{z}{\delta_0}\right) - 2\sin\frac{z}{\delta_0}\right]\sin(2\omega_1 - \omega_2)t\right\}.$$

The integrated form of (37) gives

$$J_{IM} = -\frac{\beta DE_1^3\delta_0^2}{32\rho_0^3}\left[\cos(2\omega_1 - \omega_2)t + \sin(2\omega_1 - \omega_2)t\right]. \tag{38}$$

in close correspondence with Eq. (23), as we would expect from the similarities of (16) and (33). (One point of difference arises from the fact that at the surface ($z = 0$) the present model introduces no IM current whereas magnetoresistivity does.) From Eqs. (37) and (23) we can thus write

$$\frac{J_{IM}\ (\text{variable } \mu)}{J_{IM}\ (\text{magnetoresistive})} = \frac{D}{5\sigma}, \qquad \frac{P_{IML}\ (\text{variable } \mu)}{P_{IML}\ (\text{magnetoresistive})} = \left(\frac{D}{5\sigma}\right)^2. \tag{39}$$

In this formulation, Eq. (33) shows that $D = \dfrac{d^2\mu}{\mu_u dH^2}$ where μ_u represents the average value of permeability in the average static field present. Typically we expect μ to be an increasing function of H for small fields and to be decreasing in fields above a few hundredths Tesla. This suggests that as applied fields increase μ will pass through regions of constant slope, which should produce minima in IM power levels. The ratio in (39), using (32), becomes

$$\frac{D}{5\alpha} = \frac{d^2\mu}{dH^2} \Bigg/ 5b\mu_u\left[\left(\frac{dM}{dH}\right)^2 + M\frac{d^2M}{dH^2}\right]. \tag{40}$$

Using $M = \left(\dfrac{\mu}{\mu_0} - 1\right)H$, (40) can be written as

$$\frac{D}{5\alpha} = \frac{1}{5b\mu_u}\frac{d^2\mu}{dH^2}\left[\left(\frac{\mu}{\mu_0} - 1\right)^2 + 4\left(\frac{\mu}{\mu_0} - 1\right)\frac{H^2}{\mu_0^2}\left(\frac{d\mu}{dH}\right)^2 + \left(\frac{\mu}{\mu_0} - 1\right)\frac{H^2}{\mu_0}\frac{d^2\mu}{dH^2}\right]^{-1}. \tag{41}$$

If we are concerned mainly with small external applied fields, curves for the magnetization of iron[6] can be used to estimate these values. (Nickel and iron values are similar enough to give comparable results.) As in Part 4., we use low frequency permeability data. At microwave frequencies we can anticipate an order of magnitude decrease μ.[4] We also assume that the b value of $-.04/\text{M}^2$ is approximately correct. For iron, at $H = 80$ A/m (1 Oe), $\mu_u \approx 3000\,\mu_0$, $\dfrac{d\mu}{dH} \approx 125\,\mu_0$, $\dfrac{d^2\mu}{dH^2} \approx 11\,\mu_0$, and $M_s \approx 1.6 \times 10^6$A/m. Hence $\dfrac{D}{5\sigma}$ becomes about 106 which then implies a power ratio of about 40 dB. This factor is difficult to establish accurately, and is very subject to material and environmental influences. Nonetheless, this contribution to IM signals is obviously so extremely significant that it reinforces the argument that ferromagnetic materials must be entirely excluded from high sensitivity multiplex circuits.

6. INTERMODULATION DUE TO NON-LINEAR DIELECTRICS

If dielectrics are present in a system, any variation in dielectric properties with applied field will serve to modulate incoming signal voltages in a manner analogous to the modulation of currents in Part 1. This effect has been addressed in both the Philco[1] and TRW[2] studies cited, and again here the aim is to rectify some apparent discrepancies. The approach below is similar in general outline to that of the TRW study. A solution has also been obtained in the manner of Part 1.; it yields results substantially equivalent to those below when the largely reactive nature of the IM source impedance is taken into account.

Variations in dielectric properties may occur in direct response to electric fields, through heating, or by electrostriction. The general form of the solution is applicable to a variety of dielectric media, such as molecular absorbers or ionizing gases. In each case, of course, the nature of the response to applied fields must be determined.

We start by writing equations for a generalized transmission line in the form

$$\frac{\partial V}{\partial y} = -L\frac{\partial I}{\partial t} - RI \text{ and } \frac{\partial I}{\partial y} = -C_0\epsilon\frac{\partial V}{\partial t} \tag{42}$$

where L, R, and C_0 (vacuum value) are the inductance, resistance, and capacity per unit length of line. The conductance of the dielectric will be included in a complex relative dielectric constant ϵ. We now take

$$\epsilon = \epsilon_1(1 + \alpha V^2) - i\epsilon_2(1 + \eta V^2) \tag{43}$$

both because quadratic terms are needed for IM production and because the dielectric response is expected to be insensitive to the sign of the applied field. The relationship between the factors α and η and corresponding factors suitable for multiplying E^2 will be discussed later. Substitution of (43) into Eq. (42) leads to

$$\frac{\partial^2 V}{\partial y^2} = -L\frac{\partial^2 I}{\partial t \partial y} - R\frac{\partial I}{\partial y} = LC_0\frac{\partial^2 V}{\partial t^2}(\epsilon_1 - i\epsilon_2) + LC_0(\alpha\epsilon_1 - i\eta\epsilon_2)\frac{\partial}{\partial t}\left[V^2\frac{\partial V}{\partial t}\right] \tag{44}$$

$$+ RC_0(\epsilon_1 - i\epsilon_2)\frac{\partial V}{\partial t} + RC_0(\alpha\epsilon_1 - i\eta\epsilon_2)V^2\frac{\partial V}{\partial t}.$$

We assume that we can write $V = V_0 + V_1 + \ldots$ in descending order of magnitude based on the smallness of α and η. Equations for V_0 and V_1 are then given by

$$\frac{\partial^2 V_0}{\partial y^2} - LC_0(\epsilon_1 - i\epsilon_2)\frac{\partial^2 V_0}{\partial t^2} - RC_0(\epsilon_1 - i\epsilon_2)\frac{\partial V_0}{\partial t} = 0 \tag{45}$$

and

$$\frac{\partial^2 V_1}{\partial y^2} - LC_0(\epsilon_1 - i\epsilon_2)\frac{\partial^2 V_1}{\partial t^2} - RC_0(\epsilon_1 - i\epsilon_2)\frac{\partial V_1}{\partial t} = LC_0(\alpha\epsilon_1 - i\eta\epsilon_2)\frac{\partial}{\partial t}\left[V_0^2\frac{\partial V_0}{\partial t}\right]$$

$$+ RC_0(\alpha\epsilon_1 - i\eta\epsilon_2)V_0^2\left[\frac{\partial V_0}{\partial t}\right]. \tag{46}$$

The lowest order solution for one signal propagating to positive y can be written in the form

$$V_0 = U\ \mathrm{Re}\ e^{i\omega t}e^{-i\tau y} = U\ \mathrm{Re}\ e^{i(\omega t - \kappa y)}e^{-\gamma y}. \tag{47}$$

Here $\tau^2 \equiv \omega C_0(\omega L - iR)(\epsilon_1 - i\epsilon_2) \equiv (\kappa - i\gamma)^2$.

We now assume that there are two primary input signals, so that

$$V_0 = U\ \mathrm{Re}\ [e^{i(\omega_1 t - \kappa_1 y)}e^{-\gamma_1 y} + \beta e^{i(\omega_2 t - \kappa_2 y)}e^{-\gamma_2 y}]. \tag{48}$$

Then Eq. (46) will contain on the right side terms at several different frequencies. We are interested in the IM at $\omega_{IM} = 2\omega_1 - \omega_2$. After evaluating $V_0^2\dfrac{\partial V_0}{\partial t}$ we find that we can rewrite Eq. (46) as

$$\frac{\partial^2 V_1}{\partial y^2} - LC_0(\epsilon_1 - i\epsilon_2)\frac{\partial^2 V_1}{\partial t^2} - RC_0(\epsilon_1 - i\epsilon_2)\frac{\partial V_1}{\partial t} = \tag{49}$$

$$-\frac{1}{4}\beta U^3(LC_0\omega^2 - iRC_0\omega)(\alpha\epsilon_1 - i\eta\epsilon_2)e^{i(\omega t - \kappa y)}e^{-\gamma y}$$

where we want to keep the real part of V_1. Here $\omega \equiv \omega_{IM} \equiv 2\omega_1 - \omega_2$, $\kappa \equiv 2\kappa_1 - \kappa_2$, and $\gamma \equiv 2\gamma_1 + \gamma_2$. Note that κ and γ are not in general the propagation constants for a signal at ω_{IM}, although to lowest order κ will be. These quantities will be discussed again later.

The solution of Eq. (49) which satisfies the conditions that $V_1 = 0$ both at $y = 0$ and as $y \to \infty$ is, to first order in α and η,

$$V_1 = \frac{\beta U^3\omega C_0}{4}\ \mathrm{Re}\ e^{i\omega t}\frac{\{(\omega L - iR)(\alpha\epsilon_1 - i\eta\epsilon_2)\}\{e^{-[\omega C_0(\omega L - iR)(\epsilon_1 - i\epsilon_2)]^{1/2}y} - e^{-i(\kappa - i\gamma)y}\}}{\{(\gamma + i\kappa)^2 + \omega C_0(\omega L - iR)(\epsilon_1 - i\epsilon_2)\}} \tag{50}$$

If the expression for τ^2 (Eq. 47) when $\omega = \omega_{IM}$ is used to define new quantities κ_{IM} and γ_{IM} we can then write

$$V_{IM} = \frac{\beta U^3}{4} \operatorname{Re} e^{i\omega_{IM}t} \frac{\{\tau_{IM}^2(\alpha\epsilon_1 - i\eta\epsilon_2)\}\{e^{-i\kappa_{IM}y} e^{-\gamma_{IM}y} - e^{-i(2\kappa_1 - \kappa_2)y} e^{-(2\gamma_1 + \gamma_2)y}\}}{\{\tau_{IM}^2 - (2\kappa_1 - \kappa_2 - 2i\gamma_1 - i\gamma_2)^2\}(\epsilon_1 - i\epsilon_2)} \tag{51}$$

Evaluated at the full length Y of the line, this then lead to

$$P_{IML} = \overline{V_{IM}^2(Y)}/Z_0. \tag{52}$$

The characteristic impedance Z_0 will also be frequency dependent, so that ideal matching cannot occur at all frequencies if $\frac{R}{\omega L}$ is significant. It is to be expected, however, that both $\frac{R}{\omega L}$ and $\frac{\epsilon_2}{\epsilon_1}$ will be small for real cases of interest in which case Z_0 can be taken to sufficient accuracy as $\left(\frac{L}{C_0\epsilon_1}\right)^{1/2}$.

Two cases of special interest are the general case when $\omega_{IM} \approx \omega_1 \approx \omega_2$, and the case where $\frac{\epsilon_2}{\epsilon_1}$ and $\frac{R}{\omega L}$ are both small, but frequencies are general.

When the IM frequency is very close to both carriers, we can approximate $\omega_{IM} = \omega_1 = \omega_2$, $\kappa_{IM} = \kappa_1 = \kappa_2$, and $\gamma_{IM} = \gamma_1 = \gamma_2$. From Eq. (51) we then find

$$V_{IM} = \frac{\beta U^3}{16} \operatorname{Re} e^{i(\omega_1 t - \kappa_1 y)} \frac{\{(\kappa_1 - i\gamma_1)^2(\alpha\epsilon_1 - i\eta\epsilon_2)\}}{\{(2\gamma_1^2 + i\kappa_1\gamma_1)(\epsilon_1 - i\epsilon_2)\}}(e^{-\gamma_1 y} - e^{-3\gamma_1 y}) \tag{53}$$

The y dependence of the amplitude, all contained in the last parentheses, leads to a saturation effect as described by TRW, with a maximum at $e^{-2\gamma_1 y} = \frac{1}{3}$, i.e., where P_{IM} is down 4.8 dB. This will generally involve transmission lines significantly longer than 3 m so that the effect would usually not be observed. A similar saturation effect will occur also in other IM processes.

In the case where $\frac{\epsilon_2}{\epsilon_1}$ and $\frac{R}{\omega L}$ are both small, the quantities κ and γ can be written

$$\kappa \approx \omega (LC_0\epsilon_1)^{\frac{1}{2}} \text{ and } \gamma \approx \frac{1}{2}\kappa \left(\frac{\epsilon_2}{\epsilon_1} + \frac{R}{\omega L}\right). \tag{54}$$

Hence we can take $\kappa_{IM} = 2\kappa_1 - \kappa_2$ and $\gamma_{IM} = 2\gamma_1 - \gamma_2$. Taken to lowest order, Eq. (51) leads to

$$V_{IM} \approx \frac{\beta U^3}{8} \kappa_{IM} y \left\{\alpha \sin(\omega_{IM} t - \kappa_{IM} y) + \left(\frac{\epsilon_2}{\epsilon_1}\right)(\alpha - \eta)\cos(\omega_{IM} t - \kappa_{IM} y)\right\}. \tag{55}$$

Note that to this approximation, the exponential factors have vanished from the result. If, in fact, $\epsilon_2 = R = 0$, so that only the real part of ϵ is changing, Eq. (46) can be solved directly to obtain the first term here. It can be seen that generally α will dominate η as an IM contributor. Only if α is essentially zero, perhaps as in an absorbing gas, will the η contribution be significant.

For the two special cases above, we can readily write the IM power at the load, taking $Z_0 = \left(\frac{L}{C_0\epsilon_1}\right)^{\frac{1}{2}}$. From Eq. (53) we obtain

$$P_{IML} = \frac{\beta^2 U^6}{512 Z_0} \frac{(\kappa_1^2 + \gamma_1^2)^2}{\gamma_1^2(\kappa_1^2 + 4\gamma_1^2)} \frac{(\alpha^2\epsilon_1^2 + \eta^2\epsilon_2^2)}{(\epsilon_1^2 + \epsilon_2^2)} e^{-2\gamma_1 y}(1 - e^{-2\gamma_1 y})^2. \tag{56}$$

From Eq. (55), the result for small ϵ_2 and R becomes

$$P_{IM} = \frac{\beta^2 U^4}{128 Z_0} \kappa_{IM}^2 \alpha^2 Y^2 . \tag{57}$$

In the absence of α, α^2 in Eq. (57) or (58) below is to be replaced by $\eta^2 \left[\frac{\epsilon_2}{\epsilon_1}\right]^2$. In terms of the input powers $P_1 = \frac{U^2}{2Z_0}$ and $P_2 = \frac{\beta^2 U^2}{2Z_0}$, the quantity $(\beta U^3)^2 = 8 Z_0^3 P_1^2 P_2$, while the propagation constant κ_{IM} in Eq. (57) can be taken as $2\pi/\lambda_{IM}$. Hence Eq. (57) can also be written

$$P_{IM} = \frac{1}{4} \pi^2 Z_0^3 P_1^2 P_2 \alpha^2 \left[\frac{Y}{\lambda_{IM}}\right]^2 . \tag{58}$$

Here λ_{IM} is the IM wavelength in the transmission line. If Y^2 is large enough it should be replaced in the next order of approximation by $e^{-2\gamma_1 Y} (1 - e^{-2\gamma_1 Y})^2 / 4\gamma_1^2$. Then the maximum P_{IM}, when $e^{-2\gamma_1 Y} = \frac{1}{3}$, is given by Eq. (58) if Y^2 is replaced by $1/27\gamma_1^2$. On the other hand, a discrete dielectric element can also be treated using Eq. (58) if Y is taken to be the length of that element.

In order to connect the parameter α with the more significant coefficient α' which expresses dependence on electric field, we can treat the transmission line structure as a capacitor with ϵ dependent on E^2 and then find to lowest order the corresponding connection between capacity and V^2. For example, if the volume occupied by the dielectric does not change, a structure with plane parallel plates results simply in $\alpha = \frac{\alpha'}{d^2}$, where d is the spacing of the plates. In the case of coaxial cylindrical plates, the relationship becomes $\alpha = \frac{1}{2} \alpha' (r_1^2 - r_2^2) \left[\ln \frac{r_2}{r_1}\right]^2$. The same relationships, if needed, will connect η and η'.

As one example, we now apply Eq. (58) to a case considered in the Philco study: molecular absorption by 1% water vapor in air, all contained in an x-band waveguide. We use the same conditions given in the Philco study, $P_1 = 6$ kW, $P_2 = 60$ W, $Z_0 = 448\ \Omega$, $Y = 10$ m, $\lambda_{IM} = 0.0445$ m, $d = .01$ m, $\epsilon_1 = 1$, $\alpha \approx 0$, dielectric attenuation 2.9×10^{-6} dB/m, $\eta' E^2 = 2 \times 10^{-4}$ for $E = 1.4 \times 10^5$ V/m. Treating the waveguide as a parallel plate capacitor leads to $\eta' = 10^{-14}$, $\epsilon_2 = 4.7 \times 10^{-9}$, and to $P_{IM} = 1.2 \times 10^{-17}$ W or -139 dBm. This value is below the -129 dBm estimated in the Philco study, but is not negligible. With greater power on either input, higher humidity, a longer waveguide, or the 15 to 20 dB increase estimated by Philco for more refined analysis of the molecular absorption process, water vapor becomes a significant source of IM signals. (In comparison, the same waveguide, as treated in Part 1, with 33 dB added for increased power and length contributes only -157 dBm through resistive heating.) This water vapor effect is a result of the molecular rotational absorptions at 22.2 GHz and will become more pronounced as that frequency is approached. Below x-band frequencies, this is the only absorptions process of interest in normal air, but at higher frequencies both water vapor and molecular oxygen absorption (at 60 GHz) will result in strong IM generation as well as significant signal propagation losses.

When a solid dielectric is considered, the mass rather than volume of dielectric is conserved. Allowance for the change dielectric volume modifies the relationships between α and α' above. These will then be given by $\alpha = \alpha' \left[\frac{\epsilon - 1}{d^2}\right]$ or $\alpha = \frac{1}{2} \alpha' \left[r_1^2 - r_2^2 - \frac{2\epsilon}{\epsilon - 1} \ln \frac{r_2}{r_1}\right] \left[\ln \frac{r_2}{r_1}\right]^2$, respectively, on the assumption that $\frac{\Delta \chi}{\chi} = \frac{\Delta m}{m}$ (χ, susceptibility, m, density), as in electrostriction.

Electrostriction is expected to be the principal source of non-linearity in good non-polar dielectrics[1]. The TRW study considers the effect of electrostriction on teflon in a coaxial line. Using the

conditions of the TRW example for a 300 MHz IM signal, with $\alpha' \approx (0.8 \text{ to } 25) \times 10^{-21}\,(\text{V/m})^{-2}$, $r_2 = 2.4 \times 10^{-3}$ m $= 3r_1$, $\lambda_{IM} = 0.7$ m, $Z_0 = 47\Omega$, $P_1 = P_2 = 30$ W, $\epsilon_1 = 2$, the above relationship for α combined with Eq. (58) leads to $P_{IMI} = (.01 \text{ to } 10) \times 10^{-21} Y^2$ W. A line 1 m long will thus yield P_{IMI} of -170 to -200 dBm, probably below detectability.

The value of α' is obtained from the expression

$$\Delta\epsilon = \frac{\epsilon_0 E^2}{K}\,(\epsilon - 1)^2 \tag{59}$$

where K is the bulk modulus of compressibility. Equation (59) is given by Böttcher[7] based on thermodynamic arguments. The spread in values reflects the range of K values quoted by the TRW report.

Since K is about as low as possible for teflon, the effect of electrostriction is unlikely to be significant with other well chosen dielectrics. If materials with resonance absorption losses, polar properties (permament dipole moments) or anisotropic polarizabilities were used, however, the IM products could be significantly increased, so such materials should be avoided.

Functionally, Eq. (58) and the relationships between α and α' show that P_{IMI} will increase quadratically with the number of wavelengths in the transmission line length, and except as Z_0 is altered by dimensional changes will vary inversely with the fourth power of the transverse structural dimensions. Since structure size and power levels will generally be correlated, the conclusion is that this kind of dielectric IM production will be insignificant in all but extreme cases. (For instance, if in the example above, P_1 and P_2 are increased to 1000 W while r_1 and r_2 are trebled, P_{IMI} will be increased 27 dB.) The one configuration in which IM production might be significant with good dielectrics is a high-Q resonant structure where the power densities are correspondingly enhanced. Thus it is desirable to exclude all dielectric materials from such elements.

7. SUMMARY

The intermodulation signals contributed by the intrinsic non-linear properties of materials (other than semiconductors) used in multiplex circuits have been calculated. Functional dependences, in terms of parameters accessible externally, have been obtained for the normally dominant 3rd-order IM signals.

Non-magnetic conductors are considered in which resistivity changes either by resistive heating, by associated mangetic fields, or by current density directly. Resistive heating is the most important of these effects.

Ferromagnetic metallic components are treated separately. These materials are entirely unsuited to small signal applications. In addition to causing very high losses, they can generate IM signals 100 dB above thermal noise in realistic circumstances. The danger in the use of such components is illustrated in Chapters II-IV.

Dielectric elements, distributed or not, in which the dielectric properties are functions of electric field, are treated in a generalized fashion. It is found that water vapor in air produces non-negligible IM signals, particularly at microwave frequencies. Dielectric media with no polar properties or resonance losses are generally insignificant as IM sources, although in resonant structures they might lead to observable IM signals.

The results of this chapter lead to the following recommendations concerning materials in multiplex systems:

i.) Ferromagnetic materials (and also semiconductors) should be totally excluded from any part of a system in which multiple signals exist. Leakage effects may require shielding of such materials.

ii.) Water vapor should be excluded from all transmission lines and cavities, particularly for microwave frequencies. Attenuation in the atmosphere is probably the main deterrent to operating above x-band frequencies, but if higher frequencies are used, oxygen also should be removed from the system.

iii.) High power densities should be avoided; since these are inevitable for cavity filters, materials used there (even more than in other parts of the system) must have low resistivity, high heat capacity, high thermal conductivity and low thermal coefficients of resistivity. Surfaces must be clean and free of corrosion products.

iv.) Dielectric materials should be excluded from resonant structures. If used elsewhere they should be non-polar and with minimum loss characteristics.

v.) Paths in which multiple signals pass should be kept to minimum lengths, preferably a meter or less. Mechanical junctions and closures should be kept to a minimum, placed to minimize currents across them, and made to best insure metallic contact.

If the above conditions are satisfied, IM production will probably be dominated either by the junction effects or by resistive heating in resonant elements. In the latter case the example given in Part I suggests that for an input power level of 1 kW, 3rd order IM power can be held below about -140 dBm for UHF signals. Because of the functional dependences illustrated in Eq. (15), this level will increase as input power rises or system temperature falls, and will increase as the separation of primary signal frequencies shrinks or as the primary frequencies rise.

SYMBOLS USED

Quantities used only where defined not listed. MKS units throughout. The subscript IM on a quantity means the same quantity as without, but restricted to the IM signal (3rd order, $2\omega_1 - \omega_2$ only).

A	effective conductor surface area
B	magnetic induction
b	defined by Eq. (32)
C_0	capacitance per unit length of transmission line with vacuum dielectric
C_h	heat capacity of conductor
D	coefficient of quadratic magnetic field dependence of permeability
d	spacing of parallel plate capacitor
E	electric field strength
$f(x)$	defined by Eq. (11)
G	thermal conductivity of conductor
H	magnetic field strength; $H(z)$ instantaneous local value
I	instantaneous current in transmission line
$J(z)$	instantaneous local current density; $J_{IN}(z)$ lowest order value due to input signals
J	instantaneous total current at all depths in conductor per unit width of surface
J	rms value of J
k	Boltzmann constant
L	transmission line inductance per unit length
M	magnetization; M_s saturation magnetization
$p_{IN}(z)$	instantaneous local density of power dissipated, lowest order due to input signals
p_{IN}	rms power dissipated at all depths per unit surface area, lowest order due to input signals

$\bar{p}_{IML}$	rms IM power dissipated at load due to unit area of IM source
P_{IN}	total rms input power available to load
$P_{IN\,\text{lost}}$	total rms input power lost in transmission to load
P_{IML}	total rms IM power delivered to load
Q	circuit Q factor
R	resistance of transmission line per unit length
R_s	surface resistivity, cf. Eq. (8)
r_1, r_2	inner, outer radii of coaxial transmission line
T	temperature; T_0 ambient; $T(z)$ local instantaneous value above ambient
U	amplitude of rf voltage across transmission line
V	instantaneous voltage across transmission line; V_0, V_1 lowest, next lowest order values for changes in dielectric constants.
W	defined by Eq. (4)
x	defined by Eq. (7)
Y	length of transmission line
y	distance along direction of propagation
Z_0	characteristic impedance or matching load impedance
z	depth below surface of conductor
β	ratio of signal amplitudes for two inputs, signal 2/signal 1.
δ_0	skin depth for small signal levels; $\delta(z)$ instantaneous local value
$\epsilon = (\epsilon_1 - i\epsilon_2)$	*relative* dielectric constant for small signal levels
ξ	coefficient of quadratic current density dependent change in resistivity
λ	wavelength in transmission line
μ	magnetic permeability; μ_u permeability for very weak signals; $\mu(z)$ instantaneous local value
ν	circular frequency
ω	radian frequency
ρ_0	resistivity for small signals; $\rho(z)$ instantaneous local value
σ	coefficient of quadratic magnetic field dependent change in resistivity
ϕ	defined by Eq. (4)
ψ	generalization of T_0^{-1}, coefficient of linear thermal dependent change in resistivity
α', η'	coefficients of quadratic electric field dependent changes in real and imaginary dielectric constants
α, η	α', η' modified by transmission line geometry to give changes in dielectric constants with voltage across transmission line
$\tau = (\kappa - i\gamma)$	complex propagation constant

REFERENCES

1. Philco Study Report WDL-TR 5242 (1973).

2. TRW FLTSATCOM Document 24200-540-006-01 (1974). J.Z. Wilcox and P. Molmud, IEEE *COM-24*, 238 (1976).

3. J. DeLaunay, R.L. Dolecek, and R.T. Webber, J. Phys. Chem. Solids **11**, 37 (1959).

4. R.M. Bozorth, *Ferromagnetism*, (Van Nostrand, 1951) p 798ff.

5. J.-P.Jan in *Solid State Physics*, F. Seitz and D. Turnbull eds. (Academic Press, 1957), Vol. 5, pp 70-71.

6. AIP Handbook (McGraw Hill, 1957), p 5-214.

7. C.J.F. Böttcher, *Theory of Electric Polarization* (Elsevier, 1973), 2nd Ed., Ch. VII.

Chapter VI

INTERMODULATION GENERATION DIAGNOSIS BY ANALYTICAL AND COMPUTER TECHNIQUES

A.C. Ehrlich, G.N. Kamm and G.C. Bailey

Metal Physics Branch

Material Science and Technology Division

INTRODUCTION

Recent advances in communications technology suggest that more powerful systematic methods for the reduction of IMG must be developed to keep pace with this technology. In this chapter analytical approaches to IMG diagnosis are reviewed and discussed. It is found that a great deal of information about the nature of the IMG-producing nonlinearity can be deduced from the intermodulation (intermod) spectrum and this can provide important guidelines in identifying the source of the nonlinearity. On the other hand it is not possible to identify with certainty a unique physical source of an intermod from the intermod spectrum alone although the reverse is possible. The strengths and limitations of the analytical approach are delineated and the need for the application of high speed computers for Fourier transform analysis of signals containing intermods is demonstrated. Several model computer calculations are carried out and the results are discussed and compared when appropriate to what might be expected from analytical considerations. Insight is provided into such effects as the power of the intermod relative to primary signal power, the surprising decrease in power of certain intermods when additional primary signals are turned on, the relative power of different intermods, etc. The goal is to provide a mathematical and computational basis for predicting the intermodulation spectrum from particular physical models of IMG sources as well as to develop procedures for diagnosing sources of IMG.

ANALYTICAL CONSIDERATIONS

Intermodulation generation occurs when an electronic system contains one or more elements which do not exhibit purely linear current-voltage (i-e) relationships. It is mathematically well-known and obvious how non-linearities give rise to IMG.

Consider the functional relationship

$$e = Ai + Bi^2 \tag{1}$$

where e is the voltage and i is the current across an element in the system.

Then, for a two-carrier signal with frequencies ω_1 and ω_2 and zero relative phase at $t = 0$

$$i = i_1 + i_2 = I_1 \cos\omega_1 t + I_2 \cos\omega_2 t$$

and thus

$$\begin{aligned} e &= A[I_1 \cos\omega_1 t + I_2 \cos\omega_2 t] + B[I_1 \cos\omega_1 t + I_2 \cos\omega_2 t]^2 \\ &= A[I_1 \cos\omega_1 t + I_2 \cos\omega_2 t] + B[I_1^2 \cos^2\omega_1 t + I_2^2 \cos^2\omega_2 t \\ &\quad + 2I_1 I_2 \cos\omega_1 t \cos\omega_2 t] \end{aligned}$$

Recalling that

$$\cos^2 x = 1/2(\cos 2x + 1)$$

$$\cos x \cos y = 1/2[\cos(x + y) + \cos(x - y)] \tag{3}$$

we find

$$e = A I_1 \cos\omega_1 t + I_2 \cos\omega_2 t + \frac{1}{2} B\{I_1^2(\cos 2\omega_1 t + 1) + I_2^2(\cos 2\omega_2 t + 1) + 2I_1 I_2[\cos(\omega_1 + \omega_2)t + \cos(\omega_1 - \omega_2)t]\} \tag{4}$$

and the i^2 term is found to give rise to second order intermodulation signals (intermods).

Analogously, if a third order term were also in Eq. (1), Ci^3 say, then additional terms would appear in Eq. (2), to wit

$$\frac{1}{4} C\{I_1^3[\cos 3\omega_1 t + 3\cos\omega_1 t] + I_2^3[\cos 3\omega_2 t + 3\cos\omega_2 t] + 3I_1 I_2^2[2\cos\omega_1 t + \cos(\omega_1 + 2\omega_2)t + \cos(\omega_1 - 2\omega_2)t + \cos(\omega_1 - 2\omega_2)t] + 3I_1^2 I_2[2\cos\omega_2 t + \cos(2\omega_1 + \omega_2)t + \cos(2\omega_1 - \omega_2)t]\}$$

In principle, this same kind of procedure can be used to carry out an evaluation of the intermodulation signals for an arbitrary number of carriers and for any order term or terms in the current-voltage relationship.

What this implies is that regardless of the functional relationship between voltage and current, if voltage can be expressed as a power series in the current, i.e. a Taylor series, then one could apply the technique used in obtaining Eq. (4) to predict the frequencies and amplitudes of the intermods.

A Taylor series expansion would treat the current associated with each primary frequency as an independent variable so that the expansion would appear as

$$e = f(i) = f(0) + \left[\frac{\partial f(0)}{\partial i_1}\right] i_1 + \left[\frac{\partial f(0)}{\partial i_2}\right] i_2 + \ldots + \frac{1}{2!}\left[\frac{\partial^2 f(0)}{\partial i_1^2}\right] i_1^2 + \ldots$$

$$= \left[\sum_{k=1}^{n} i_k \frac{\partial}{\partial i_k}\right] f(0) + \frac{1}{2!}\left[\sum_{k=1}^{n} i_k \frac{\partial}{\partial i_k}\right]^2 f(0) + \frac{1}{3!}\left[\sum_{k=1}^{n} i_k \frac{\partial}{\partial i_k}\right]^3 f(0) + \ldots \tag{5}$$

where $f(0)$ implies evaluation of $f(i)$ where each of the i_k's is zero and n is the number of primary carrier frequencies. Each term in the Taylor series could be treated in the manner leading to Eq. (4) to determine the resultant intermods. Although this procedure would be very long and tedious, in principle it could be carried out just one time for each term in the series and for various numbers of primary carriers and the results tabulated. We have carried out a number of calculations along these lines which have proved to be very important for learning about the relationships between the nonlinearities and the intermods they cause as well as the relationships among different intermods. On the other hand the compilation of the results proves to be rather elaborate and could not include the effects of the relative phases of the primary signals at some particular time. (Consideration of phase will be discussed below.)

More important, however, is the fact that some simple i-e characteristics often found in real systems cannot be expressed in a Taylor series. For example, the simplest rectifier, whose i-e functional dependence is

$$e = Ai \qquad i > 0 \tag{6}$$

$$e = 0 \qquad i < 0$$

where A is some constant, does not fulfill the mathematical requirements for a Taylor series expansion. It can, however, be easily treated using computer techniques and therein lies one major advantage of the computer analysis as compared to the analytical approach.

A number of useful facts for an analytical diagnosis procedure are already obvious from these discussions. *First*, all terms in the *i-e* functional relationship contribute additively to the intermod spectrum. *Second*, the highest order intermod that occurs arises from the highest order (non-zero) term that appears in the Taylor series expansion; i.e. the term $\frac{1}{m!}\left[\sum_{k=1}^{n} i_k \frac{\partial}{\partial i_k}\right]^m f(0)$ will give rise to m^{th} order intermods and in general, all odd or even order intermods lower than m according to whether m is itself odd or even.

[As an aside, we note that if the trigonometric manipulations analogous to those leading up to Eq. 4 were carried out for m^{th} order terms in the Taylor series using exponential rather than conventional forms of the cosine function, one would conclude that only m^{th} order intermods arise from the m^{th} order term in the Taylor series. This would be incorrect. Use of the exponential functions with the taking of the real part at the completion of the calculation is not valid for non-linear problems. The difficulty arises from the difference between the power of the real part of a complex number and the real part of the power of the same number. E.g., $\text{Re}\{(a + ib)^2\} \neq \{\text{Re}(a + ib)\}^2$.]

Third, it is trivial to demonstrate that any particular intermod, say $\cos(2\omega_1 + \omega_2)t$, has the same phase regardless of whether it arises from the third, fifth, seventh, etc. order term in Taylor series. (We are assuming, of course, that the non-linear element(s) giving rise to the intermods are found at a single point in space.) To do this, imagine the quantity $\omega_1 t$ replaced by $(\omega_1 t + \phi)$, and similarly for $\omega_2 t$, $\omega_3 t$, etc. and the result is obvious. On the other hand, given a multicarrier signal, it is likely that two or more different intermods will have the same frequency. For example, if $\omega_1 = 10$, $\omega_2 = 11$ and $\omega_3 = 13$, (in arbitrary units) then $2\omega_2 - \omega_1 = \omega_1 + \omega_3 - \omega_2 = 12$. In this situation there is no reason to expect these two different intermods to have the same or nearly the same phase. This provides an explanation for an often seen phenomenon that is not widely understood; to wit, a given intermod is reduced in amplitude when an additional carrier frequency is turned on. This could occur for example with the signal whose frequency is 12 in the example above when ω_3 is turned on if the $\omega_1 + \omega_3 - \omega_2$ signal is out of phase or nearly out of phase with the $2\omega_2 - \omega_1$ signal.

Fourth, the amplitude or power (power is proportional to the square of the amplitude) dependence of an intermod on the primary signal input amplitude (or power) can be seen from Eq. (5). If an m^{th} intermod arising from the m^{th} order term in the Taylor series expansion has a frequency $(p_1\omega_1 + p_2\omega_2 + p_3\omega_3 + p_4\omega_4 + \ldots) = \sum_i p_i\omega_i$ where the p's are positive or negative integers and $\sum_i |p_i| = m$, then the intermod amplitude will vary as $I_1^{p_1} I_2^{p_2} I_3^{p_3} I_4^{p_4} \ldots = \prod_i I_i^{p_i}$. This dependence of intermod power on primary signal power can be a useful tool in determining the i-e functional relationship. It must be remembered, however, that an m^{th} order term in the Taylor series expansion will also generate, in general, intermods of order $m - 2$, $m - 4$, ... etc. and all of these will also be of m^{th} order in the I's. For these contributions, specific functional dependence of the amplitude of the intermod on the amplitudes of the various primary signals cannot be simply specified. All that can be said without tedious calculation for a specific situation is that if an intermod frequency is $\sum_i p_i\omega_i$, then the amplitude will consist of a sum of terms each of which varies as $\prod_i I_i^{\,p_i+2n_i}$ where $\sum_i |p_i| = m$ is the intermod order and n_i is an integer equal to or greater than 0. Thus, for example, a three carrier signal composed of frequencies ω_1, ω_2 and ω_3 passing through a nonlinear device may well produce components such as $\cos(2\omega_1 + \omega_2)t$ arising from, say a 5th or 7th order term in the current voltage relationship. Although this intermod does not "contain" an ω_3, it will have an amplitude that does in general depend on I_3. In fact, certain experimental observations, such as the variation of the amplitude of a term like

$\cos(2\omega_1 + \omega_2)t$ with I_3, suggests the presence of 5th or higher order terms. In this regard one should, however, be cautious. This could appear to occur if a different third-order term involving ω_3 happened, by accident, to have a frequency numerically equal to $2\omega_1 + \omega_2$.

Fifth, the relative magnitudes of the intermods as a function of the number of frequencies represented in the intermod, the number of carrier signals, and the order of the nonlinear term are straightforward if tedious to deduce. Suppose the equation describing the nonlinearity is

$$e = Ai + Bi^3 + Ci^5 \tag{7}$$

If one compares a three-carrier signal to a five carrier signal then the Bi^3 term will give rise to the same kind and magnitude of third order terms for both cases. For five carriers, there are simply more third order terms arising from the larger number of combinations of frequencies than for three carriers. On the other hand, the Bi^3 term also gives rise to first order terms and these nonlinear generated first order terms will be larger when there are five carriers than when there are only three.

Analogously, the Ci^5 term will generate more fifth order intermods when five signals are considered than when three are, but their magnitudes will be the same in both cases for "similar" intermods, i.e. $\cos(3\omega_1 - 2\omega_2)t$ is "similar" to $\cos(3\omega_3 + 2\omega_2)t$, but not to either $\cos 5\omega_2 t$ or to $\cos(2\omega_1 + 2\omega_2 + \omega_3)t$, etc. On the other hand, the third-order intermod frequencies that arise from the Ci^5 term do have a magnitude which depends on the number of carrier frequencies.

If the Ci^5 term is the highest order term in the i-e relationship, then the largest fifth order terms will be the so called five carrier fifths, e.g. $\cos(\omega_1 \pm \omega_2 \pm \omega_3 \pm \omega_4 \pm \omega_5)t$. In order of decreasing magnitude will come $\cos(2\omega_1 \pm \omega_2 \pm \omega_3 \pm \omega_4)t$, $\cos(2\omega_1 \pm 2\omega_2 \pm \omega_3)t$, $\cos(3\omega_1 \pm \omega_3)t$, $\cos(3\omega_1 \pm \omega_2)t$, $\cos(4\omega_1 \pm \omega_2)t$, $\cos 5\omega_1 t$. The trend is clear; the more frequencies and the more equally represented in the intermod, the larger that particular intermod will be.

In general, higher order intermods tend to have smaller amplitudes. For example, the Ci^5 term of Eq. (7) will give rise to third order as well as fifth order intermods, and in fact the three carrier third, i.e. $\cos(\omega_1 \pm \omega_2 \pm \omega_3)t$ will have a magnitude somewhat larger than the five carrier fifth mentioned above. It is by no means true, however, that all or even most third order intermods are larger than all the fifth order intermods.

COMPUTER DIAGNOSIS OF NONLINEAR SYSTEMS

Using Equations (2) through (4) is basically a simpler way of obtaining a Fourier transform than the usual analytical procedure. On the other hand, computer based Fourier transform analysis is now a well developed technique. This fact and the widespread availability of high speed computers strongly suggest the application of computers to diagnosing nonlinear systems from the system's intermodulation generation. The advantages of using computers are more than just speed. Unlike the mathematical requirements on the current-voltage relationships that are necessary in the analytical approach discussed above, the computer approach requires no particular restrictive mathematical criteria for the current-voltage dependence. For example, there is no special difficulty in Fourier analyzing a multicarrier signal imposed on a circuit element with circuit characteristics given by Eq. (6). Thus, given an i-e relationship, analytical or otherwise, one can predict the magnitude and relative phase of the intermods.

There are three characteristics of computer based Fourier transforms which should be understood if results are to be properly interpreted. *First*, a computer based Fourier transform (FT) differs from an analytical transform in that, with the former, one works with a discrete (rather than continuous) set of "data" (values of the function being transformed) summed over a finite (rather than infinite) range of the argument of the function. As a consequence of this, the relative phases of the input signals at some fixed time can influence, to some extent, the magnitudes of the resultant intermods. This influence, however, is quite small.

Second, another and much more important consideration associated with signal phase is that all computer results are numerical and there is no way to distinguish between two or more contributing intermods at a given frequency. Further and as mentioned earlier, there is no reason for different intermods of the same frequency to have the same or any other particular phase relative to each other. Thus certain characteristics of particular intermods cannot always be easily sorted out. This corresponds to the actual experimental situation where various intermods can overlap.

Third, a possible major contribution to the overlap of intermods is the phenomenon of aliasing. When N discrete data points are separated from each other by a time interval T, then there is a maximum possible frequency, ω_m, that can be resolved which is approximately $\frac{1}{2T}$ for large N. Frequencies less than ω_m will be correctly given by the computer based Fourier transform. For those frequencies greater than ω_m by $\Delta\omega_m$, the associated computer generated spectrum line will appear at a frequency $(\omega_m - \Delta\omega_m)$ which is, of course, within the range of allowable frequencies. This is the phenomenon known as aliasing. Thus, the probability of signal overlap in the range of allowable frequencies is increased simply because of the apparent increased density of intermod signals. This phenomenon will not usually present too great a problem for two reasons. First, the intermods with frequencies greater than ω_m are usually of higher order and thus can be expected to have very low amplitudes. Second, the value of the intermod frequency can be made to indicate which spectrum lines have appeared by aliasing. For example, if all primary frequencies are whole numbers but ω_{max} is, say, a whole number plus 0.3, then any spectrum line centered on an ω that is not a whole number can be assumed to have arisen from aliasing.

In order to illustrate the nature of the intermod response expected from an i-e relationship not amenable to analytical treatment, a number of computer experiments have been carried out. For this purpose we have worked with linear, quadratic and cubic rectification functions and various combinations thereof. Thus, the i-e relationship is

$$\begin{aligned} e &= ai + bi^2 + ci^3 \quad && i > 0 \\ &= 0 && i < 0 \end{aligned} \tag{8}$$

where any one or two of the coefficients a, b, or c may be zero. For example if $a = b = 0$, $c \neq 0$ then we refer to this as cubic rectification. If only $a \neq 0$, it is linear rectification while if only $b \neq 0$ it is quadratic rectification. Calculations using up to three input frequencies have been carried out with

$$i = I_1 \cos(\omega_1 t + \phi_1) + I_2 \cos(\omega_2 t + \phi_2) + I_3 \cos(\omega_3 t + \phi_3)$$

as the input signal, where $\omega_1 = 10$, $\omega_2 = 11$, and $\omega_3 = 13$ and the phases ϕ_1, ϕ_2 and ϕ_3 could be varied using a random number generator. Amplitudes of signals are in arbitrary units since only ratios of amplitudes are significant.

In the first computer experiment, two of the three primary signals and two second order intermods were examined for several combinations of the coefficients a, b, and c in Eq. (8). The parameters used and results obtained are summarized in Table I. Values of a, b, and c used are shown above the column corresponding to the results. If one or more of these coefficients are not indicated above the columns, the implication is that they are zero. The results themselves are averages of four separate runs using random phases of the input signals with each run. The intent was to minimize the effects of signal phase discussed above.

Table I — Signal amplitudes at various frequencies for the kinds of rectification indicated.

Intermod and Frequency	c=1	b=1	b=c=1	a=1	a=b=c=1
$\omega_1 = 10$	34.571	17.047	51.618	9.265	60.883
$\omega_3 - \omega_1 = 3$	22.835	9.153	31.988	3.299	35.287
$\omega_2 + \omega_3 = 24$	22.697	9.117	31.813	3.036	34.848
$\omega_3 = 13$	34.534	16.695	51.229	9.242	60.470

The most obvious feature of the data in Table I is the equality between the sum of the amplitudes for quadratic ($b = 1$) and cubic ($c = 1$) rectification calculated separately and the result for the sum of quadratic and cubic rectification, column three. The summability is also true for linear rectification. This result is not surprising since it merely implies the obvious summability of spectrum lines from different nonlinearities. A second interesting regularity is that for each type of rectification, "similar" intermods have very nearly the same amplitudes. The extent to which the amplitudes are not identical results from more than one intermod contributing at a given frequency.

The variability of signal amplitude because of multiple intermod contributions at a single frequency, and the manner in which this amplitude can vary with primary signal phase are illustrated in Table II. Here the results of several individual runs with random phases are presented for intermod frequencies of 7 and 12. The reproducibility of the signal whose frequency is 7 is much better than the signal whose frequency is 12 because the two major contributors to the former signal are third order and fifth order intermods. Since the third order intermod should be perfectly reproducible and is much larger than the fifth, whether the fifth adds or subtracts from the third (depending on their relative phase) the amplitude of the net signal is changed only to a limited extent. In contrast, the major contributions to the signal whose frequency is 12 come from two third order intermods whose magnitudes can be expected to be much more nearly alike. Thus, addition or subtraction (constructive or destructive interference) of the two results in large variations in the magnitude of the net signal.

Table II — Signal amplitudes for the $\omega = 7$ and $\omega = 12$ signals for the kinds of rectification indicated.

Intermod and Frequency	run	c=1	b=1	b=c=1	a=1	a=b=c=1
$2\omega_1 - \omega_3 = 7$	1	2.007	.786	2.793	.133	2.876
$3\omega_2 - 2\omega_3 = 7$	2	2.030	.7739	2.803	.137	2.838
	3	2.0412	.7955	2.837	---	2.839
$\omega_1 - \omega_2 + \omega_3 = 12$	1	3.145	1.422	4.563	---	4.590
$2\omega_2 - \omega_1 = 12$	2	4.182	1.813	5.994	.0418	6.030
	3	2.024	1.023	3.047	---	3.022

The surprising thing about Table II is that the sum of the $c = 1$ and $b = 1$ columns equals the $b = c = 1$ column for an intermod frequency of 12. Unlike the results of Table I, the $\omega = 12$ signal shown in Table II is a sum of two different intermods of comparable magnitude. It is easy to show that under these conditions, the phase of the net signal for a particular nonlinearity will not in general be equal to the phase of a signal of the same frequency arising from a second nonlinearity. (The phases would be the same if the same single intermod was the only contribution for both of the nonlinearities as is the case in Table I). Thus the sum of columns $c = 1$ and $b = 1$ would not be expected to equal column $c = b = 1$. Suppose, however, that A_c is the coefficient of the $\omega_1 - \omega_2 + \omega_3$ intermod resulting from $c = 1$ and M_c is the coefficient of the $2\omega_2 - \omega_1$ intermod resulting from $c = 1$ and A_b and M_b are the analogous coefficients for $b = 1$. Then, if $M_c/A_c = M_b/A_b$ it can be shown that the net signal from $c = 1$ and $b = 1$ will have the same phase, and the additivity of the $\omega = 12$ signals in Table II are to be expected.

In Table III a large number of intermod amplitudes are given. It is obvious that the amplitudes of the intermods fall into groupings as shown. Furthermore the relative magnitudes correspond quite well with what we would expect from the analytical discussions earlier in this paper, including the facts that the higher the intermod order, the lower is its amplitude and that similar intermods have similar amplitudes. Thus, it is possible to check the assumption of the previous paragraph that $M_c/A_c = M_b/A_b$. By examining the magnitudes of the intermods $\omega_1 + \omega_2 + \omega_3$ or $\omega_2 + \omega_3 - \omega_1$ and $2\omega_1 + \omega_1$ or $2\omega_2 + \omega_1$, which have frequencies at which no other intermod of comparable order contributes, one can estimate

Table III — Signal amplitudes of the first, second and third order intermods for the kinds of rectification indicated. Intermods at higher order are indicated (in parenthesis) when they overlap a lower order intermod.

Intermod Order	Intermod & frequency	c=1	b=1	a=1
1	$\omega_1 = 10$	34.595	16.984	9.209
1	$\omega_2 = 11$	33.789	16.704	9.007
1	$\omega_3 = 13$	34.589	16.728	9.206
2	$\omega_2 - \omega_1 = 1$	20.044	9.044	3.883
2	$\omega_3 - \omega_2 = 2$	20.999	8.966	3.404
2	$\omega_3 - \omega_1 = 3$	22.789	9.236	3.317
2	$\omega_1 + \omega_2 = 21$	22.350	9.028	3.103
2	$\omega_1 + \omega_3 = 23$	21.961	9.200	3.399
2	$\omega_2 + \omega_3 = 24$	22.579	9.031	3.107
2	$2\omega_1 = 20$			
(4)	$3\omega_2 - \omega_3 = 20$	11.764	4.604	1.745
2	$2\omega_2 = 22$	9.843	4.483	2.219
2	$2\omega_3 = 26$			
(10)	$6\omega_4 - 4\omega_1 = 26$	12.240	4.641	1.647
3	$\omega_1 + \omega_2 - \omega_3 = 8$			
(5)	$3\omega_1 - 2\omega_2 = 8$	13.235	3.899	.193
3	$\omega_1 - \omega_2 + \omega_3 = 12$			
3	$2\omega_2 - \omega_1 = 12$	9.566	2.964	.074
3	$\omega_2 + \omega_3 - \omega_1 = 14$	13.679	4.328	.111
3	$\omega_1 + \omega_2 + \omega_3 = 34$	13.601	3.912	.02
3	$3\omega_1 = 30$	2.422	.789	.189
3	$3\omega_2 = 33$			
3	$2\omega_1 + \omega_3 = 33$	5.481	1.416	----
3	$3\omega_3 = 39$	2.208	.516	.146
3	$2\omega_1 - \omega_2 = 7$			
(5)	$3\omega_2 - 2\omega_3 = 7$	6.923	1.767	.089
3	$2\omega_1 - \omega_2 = 9$			
3	$2\omega_2 - \omega_3 = 9$	4.350	1.545	.109
3	$2\omega_2 - \omega_1 = 12$			
3	$\omega_1 + \omega_3 - \omega_2 = 12$	9.566	2.964	.054
3	$2\omega_3 - \omega_3 = 15$	6.550	1.996	.154
3	$2\omega_3 - \omega_1 = 16$	6.776	1.665	.136
3	$2\omega_1 + \omega_2 = 31$	6.635	1.688	.198
3	$2\omega_2 + \omega_1 = 32$	6.822	1.833	.178
3	$2\omega_1 + \omega_3 = 33$	5.481	1.416	----
3	$2\omega_2 + \omega_3 = 35$	6.649	1.999	.152

closely the magnitude of M_c, A_c, M_b and A_b. It can be seen that the proportion $M_c/A_c = M_b/A_b$ is valid to the expected accuracy. It appears in fact to be valid for all three third order harmonics, but the relationship would not be expected to hold in general.

In Table IV a similar grouping of intermods using only two (different) fundamental frequencies is shown. These data represent averages of only two separate runs but nevertheless result in better reproducibility of intermod amplitude than Table III because of fewer possible intermods having the same frequency.

Table IV — Signal amplitudes of a variety of intermod orders for the kinds of rectification indicated.

Intermod order	Intermod & frequency	c=1	b=1	a=1
1	$\omega_1 = 7$	20.680	13.244	9.191
1	$\omega_2 = 17$	20.687	13.243	9.191
2	$\omega_2 - \omega_1 = 10$	15.950	9.224	4.985
2	$\omega_2 + \omega_1 = 24$	15.702	9.081	4.907
2	$2\omega_1 = 14$	9.457	4.557	1.640
2	$2\omega_2 = 34$	9.406	4.533	1.632
3	$\omega_2 - 2\omega_1 = 3$	6.911	2.656	---
3	$2\omega_2 - \omega_1 = 27$	6.875	2.642	---
3	$\omega_2 + 2\omega_1 = 31$	6.715	2.580	---
3	$3\omega_1 = 21$	2.246	.518	---
4	$2\omega_2 - 2\omega_1 = 20$	2.280	---	.995
4	$3\omega_1 - \omega_2 = 4$	1.353	---	.330
4	$\omega_2 + 3\omega_1 = 38$	1.307	---	.319
5	$2\omega_2 - 3\omega_1 = 13$	---	.380	---
5	$3\omega_2 - 2\omega_1 = 37$	---	.377	---
5	$4\omega_1 - \omega_2 = 11$	---	.074	---
6	$3\omega_2 - 3\omega_1 = 30$	.252	---	.428

A number of other regularities are noticed in Table IV where results for higher order intermods are shown. A cubic rectification ($c = 1$) does not show fifth order intermods, but does generate 3rd order intermods while a quadratic law does not show fourth and sixth order intermods. Linear rectification appears to give only even order intermods (except for the fundamentals) which implies that the "third" order intermods in the last column of Table III probably arise, in fact, from various high order even intermods.

The effect of primary signal amplitude on the amplitude of the intermods for the three kinds of rectification discussed above has also been investigated. To minimize multiple intermod contributions to a single frequency we have used just the two frequencies 7 and 17. To simplify the interpretation, these two frequencies are always taken with equal amplitudes and varied by factors of two.

The amplitude relationships are very obvious (see Table V). For linear rectification, if the imput signals are increased by a factor n, then each intermod is increased by the same factor. For quadratic or cubic rectification, if the imput signals are increased by a factor n, then the intermods are increased by a factor n^2 or n^3 respectively. It is interesting that this is precisely the kind of behavior that is obtained for analytical linear, quadratic and cubic relationships although the particular intermods obtained are not. For example, a cubic i-e relationship (no rectification) would not produce any intermods greater than order three and no even order intermods whatever.

The accuracy of the relationships of the amplitudes discussed in the previous paragraph is extremely high as is the agreement of the amplitudes between different similar intermods (calculated, but not included in Table V). This is a result of only one significant intermod contributing to each frequency. In Table VI the same kind of experiment is carried out except that three frequencies are used and, incidentally, the input signal amplitudes are varied somewhat differently. In spite of the numbers in Table VI being the result of averaging several runs with random phases the variation from the amplitudes expected is somewhat greater than in Table V where a single run is shown.

Table V — Signal amplitudes for a number of intermods for the kinds of rectification indicated. The values of the primary signal amplitudes used are above the columns corresponding to the results.

Intermod order	Intermod & Frequency	1/4	1/2	1	2
	Linear Rectification (a=1)				
1	$\omega_1 = 7$	2.2910	4.5820	9.1637	18.327
2	$\omega_2 + \omega_1 = 24$	1.2157	2.4295	4.8594	9.7181
2	$2\omega_1 = 14$	.4057	.8113	1.6213	3.2433
3	$2\omega_2 - \omega_1 = 27$	----	----	----	----
3	$3\omega_1 = 21$	----	----	----	----
4	$2\omega_2 - 2\omega_1 = 20$	.2510	.4995	.9998	1.9994
4	$3\omega_1 - \omega_2 = 4$	.0807	.1630	.3256	.6514
	Quadratic Rectification (b=1)				
1	$\omega_1 = 7$	.8253	3.3011	13.2044	52.818
2	$\omega_2 + \omega_1 = 24$	.5619	2.2475	8.9899	35.9594
2	$2\omega_1 = 14$	.2817	1.1266	4.5064	18.026
3	$2\omega_2 - \omega_1 = 27$	.1639	.6555	2.6223	10.4893
3	$3\omega_1 = 21$	.0320	.1279	.5120	2.0477
4	$2\omega_2 - 2\omega_1 = 20$	---	----	----	----
4	$3\omega_1 - \omega_2 = 4$	----	----	----	----
	Cubic Rectification (c=1)				
1	$\omega_1 = 7$	.32215	2.5772	20.6175	164.940
2	$\omega_2 + \omega_1 = 24$	.2429	1.9432	15.5454	124.363
2	$2\omega_1 = 14$	.1461	1.1689	9.3512	74.810
3	$2\omega_2 - \omega_1 = 27$	.1066	.8530	6.8239	54.590
3	$3\omega_1 = 21$	.0347	.2775	2.21974	17.758
4	$2\omega_2 - 2\omega_1 = 20$	.0357	.2858	2.28647	18.292
4	$3\omega_1 - \omega_2 = 4$	.0210	.1677	1.3415	10.733

Table VI — Amplitudes as in Table V for three inputs.

Intermod order	Intermod Type	1/2	2/3	1
	Linear Rectification			
1	ω_1 etc	4.59	6.1	9.26
2	$\omega_1 + \omega_2$	1.6	2.25	3.07
3	$\omega_1 + \omega_2 + \omega_3$	.01	.05	.06
3	$2\omega_3 - \omega_1$	.1	.1	.2
	Quadratic Rectification			
1	ω_1	4.20	7.45	16.9
2	$\omega_1 + \omega_2$	2.25	4.05	9.0
3	$\omega_1 + \omega_2 + \omega_3$	.42	.75	1.65
	Cubic Rectification			
1	ω_1	4.32	10.2	34.6
2	$\omega_1 + \omega_2$	2.75	6.65	22.2
3	$\omega_1 + \omega_2 + \omega_3$	1.70	4.05	13.6
3	$2\omega_3 - \omega_1$	.83	1.95	6.85

An additional experiment was carried out to determine the influence of the coefficients, a, b and c on the intermod amplitude. One coefficient was individually varied by a factor of two, while the others were held equal to zero. It was expected that the resulting intermods would vary directly with the magnitude of the coefficient and this is exactly what was found.

CONCLUSION

We have shown here two approaches, the analytical and numerical, for obtaining the intermod spectrum given a specific nonlinearity and given the number of primary frequencies involved. Although one can do much with the analytical approach, we have shown that the computer method is the only way to handle the problem for certain rather common nonlinear systems. Nevertheless it is clear that the IMG diagnosis, that is the deducing of the i-e relationship from the characteristics of the intermod spectrum, cannot be carried out on a "prescription" basis. No routine procedure will quickly and unambiguously provide the i-e functional relationship and thus information leading to identification of the physical origin of the nonlinearity. However, the results of the analytical considerations and the computer method set forth above provide guidance for procedures that can be carried out, on a case by case basis, to obtain the i-e functional form.