# DIGITAL F.M. DEMODULATION USING FREQUENCY COUNTING TECHNIQUES

BY

DANIEL ALBERT LEPIC, B.Sc.

PART A: McMASTER (ON-CAMPUS) PROJECT

A project *report submitted in partial fulfillment of the requirement for the degree of

Master of Engineering

Dept. of Engineering Physics

McMaster University

* One of two required project reports.

PART B: is an "Industrial" project report.

MASTER OF ENGINEERING (1972)
(Engineering Physics)

McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: Digital F.M. Demodulation Using Frequency Counting Techniques

AUTHOR: D. A. Lepic, B.Sc. (McMaster University)

SUPERVISOR: Dr. A. R. Elliott, Dept. of Electrical Engineering

NUMBER OF PAGES: iii, 40

SUMMARY: The demodulation of analogue F.M. signals using frequency counting techniques is examined and implemented through the use of modern high speed T.T.L. integrated circuit technology. The entire demodulation unit was derived from exclusively digital components particularily compatible to frequency counting methods. The device was tested with carrier frequencies up to 2MHz and signal frequencies over the entire audio range with varying degrees of modulation. The main limitations appear to lay not in the hardware but in the actual counting technique itself which required quite large frequency deviations to resolve the higher audio frequency signals employed.

CONTENTS

ACKNOWLEDGEMENT

I would like to express my gratitude to Dr. A. R. Elliott for his knowledgeable supervision, guidance and encouragement throughout the course of the work for this report.

## ABSTRACT

The demodulation of analogue F.M. signals using frequency counting techniques is examined and implemented through the use of modern high speed T.T.L. integrated circuit technology. The entire demodulation unit was derived from exclusively digital components particularily compatible to frequency counting methods. The device was tested with carrier frequencies up to 2MHz and signal frequencies over the entire audio range with varying degrees of modulation. The main limitations appear to lay not in the hardware but in the actual counting technique itself which required quite large frequency deviations to resolve the higher audio frequency signals employed.

# INTRODUCTION

In the demodulation of an analogue F.M. wave using digital techniques there appear to be two alternative approaches. One involves the use of frequency follower circuits, and the other is a simplified form of frequency counting. The former method carries with it a host of problems such as synchronization, inter-symbol interference, coincidence error, etc.; the latter method, being inherently simpler, relies solely on the fact that in an analogue F.M. wave all the information is contained in the zero crossings. In this report a frequency counting technique is developed, incorporated into a strictly digital circuit, and applied to the demodulation of analogue F.M. signals.

At the very beginnings of radio technology it was known from basic electromagnetic radiation theory that one could not efficiently transmit audio signals directly any significant distance with feasible size antennas. Elementary antenna theory dictates that to obtain any appreciable power output, the dimensions of the transmitting antenna must be of the same order of magnitude as the wavelength of the signal to be transmitted. As a consequence, to transmit signal frequencies in the audio range (20Hz to 20KHz) one would need a radiating antenna of over 100 miles in height - an insurmountable engineering problem. It was in solution to this problem that radio engineers first turned to techniques of signal modulation using high frequency "carrier" signals.

Basically, modulation is the impressing or encoding of audio signal information onto a single high frequency carrier signal to enable effective transmission from real-scale antennas. There are several ways to modulate a high frequency signal. The most common analogue techniques involve varying some characteristic of a high frequency sinusoid and include:

(I) A.M. (Amplitude Modulation) - the amplitude of a high frequency sinusoidal carrier is made to vary in direct proportion to the audio signal amplitude.

(II) F.M. (Frequency Modulation) - the frequency of a high frequency carrier is made to vary in direct proportion to the amplitude of the audio signal.

(III) P.M. (Phase Modulation) - the phase of a high frequency carrier is made to vary in direct proportion to the amplitude of the audio signal.

If one can represent the carrier signal by the expression below:

$$E_c(t) = E_o \sin (2\pi ft + \phi) \qquad (1)$$

where $E_o$ = amplitude of carrier signal,

$f$ = frequency of carrier signal,

$\phi$ = phase of carrier signal,

and if:

$E_s = E_s(t)$ is the audio signal,

then:

$E'_o = E_o + K_1 E_s(t)$ represents the instantaneous amplitude in A.M. systems

$f' = f + K_2 E_s(t)$ represents the instantaneous frequency in F.M. systems

$\phi' = \phi + K_3 E_s(t)$ represents the instantaneous phase in P.M. systems

With the advent of digital technology several strictly digital modulation techniques have evolved. These techniques have one basic feature in common - the amplitude of the audio signal is sampled at a finite number of places and then represented by some function of a high frequency (binary) carrier pulse train. In P.A.M. (pulse amplitude modulation) it is the amplitude of the pulse train which is varied, in P.P.M. (pulse position modulation) it is the position or phase of the carrier reference pulses which is varied and in P.C.M. (pulse code modulation) the sampled amplitudes are actually quantized and coded in binary format for transmission.

Regardless of the modulation technique employed one always has the problem of recovering or demodulating the audio signal from the carrier at the receiver. In both the analogue and digital areas demodulation techniques have been devised using analogue or analogue/digital circuitry combined. In this report the problem of the demodulation of analogue F.M. signals is examined and a technique and feasible circuit are developed which perform such a function using exclusively digital components.

## PRELIMINARY CONSIDERATIONS

### (a) F. M. Characteristics and Frequency Counting

In typical analogue frequency modulation, as mentioned earlier, the frequency of a carrier wave is modulated or varied in proportion to the amplitude of an impressed audio signal wave. Analogous to compressions and rarefactions induced by sound waves in media, the carrier wave is compressed and rarified in direct proportion to the increasing or decreasing amplitude of the audio signal. The relative periodicity of these frequency variations about the carrier is the same as the periodicity of the impressed signal. The relative degree of compression or rarefaction is characterized by a quantity called frequency deviation and for any point localized on the F.M. wave it is basically equal to the difference between the F.M. instantaneous frequency at that point and the carrier frequency. The maximum frequency deviation permitted is determined by the standards committee of the D.O.C. * and for conventional radio is of the order of 75KHz.

For a typical F.M. wave a signal voltage of zero volts would leave the instantaneous F.M. frequency at carrier value; maximum signal voltage would result in an instantaneous F.M. frequency equal to the sum of the carrier frequency and the maximum frequency deviation. Fig. 1a, b, c illustrate audio signal, carrier, and F.M. waves respectively. ** As can be seen, the periodicity of compression and rarefaction is the same as the

* D.O.C. (Dept. of Communications), Radio Standards Specification No. 153

** The reason for a square wave is explained later.

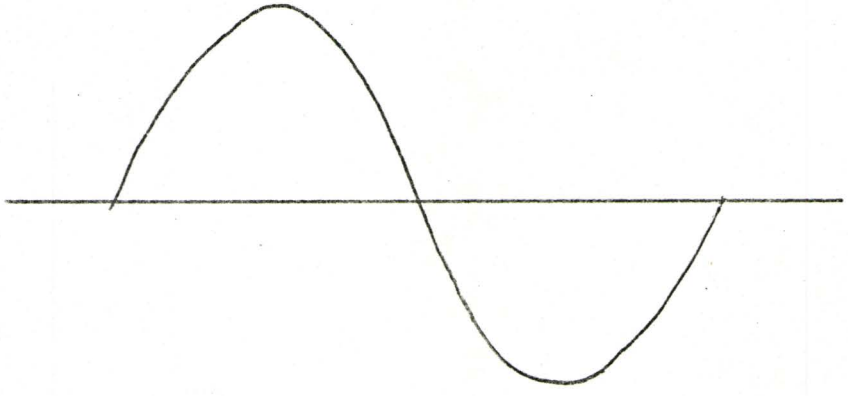


Fig. 1a
Audio Signal

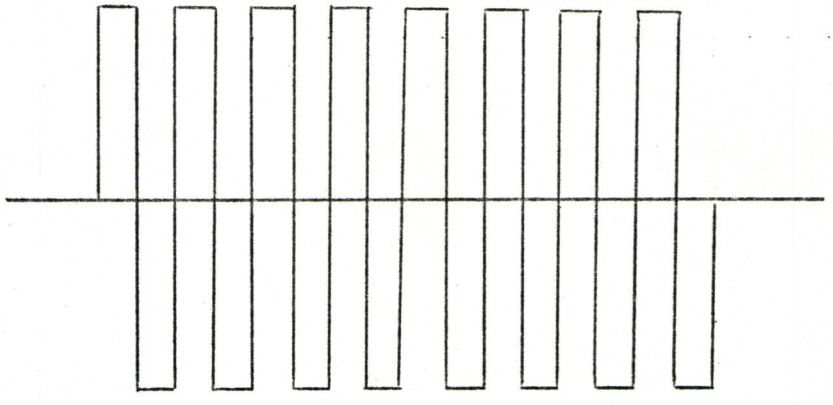


Fig. 1b
Carrier Signal

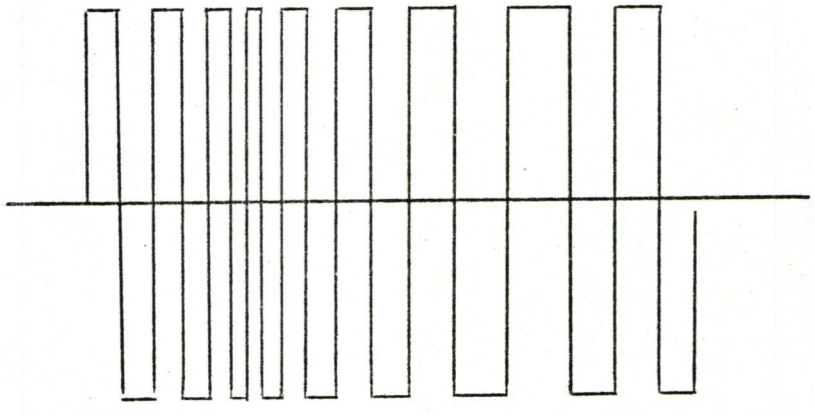


Fig. 1c
F.M. Signal

periodicity of the signal. The degree or intensity of compression or rarefaction is directly proportional to the impressed signal amplitude with the constant of proportionality determined by the imposed M.F.D. (maximum frequency deviation). Again, the main problem in F.M. receivers is to "pick up" an F.M. signal such as that shown in Fig. 1c and extract from it the information content given by the signal in Fig. 1a. The carrier is used solely for transmission.

In the INTRODUCTION it was mentioned that the instantaneous frequency of an F.M. wave can be described by the expression

$$f_i = f_c + K_2 E_s(t) \qquad (2)$$

where $f_i$ = instantaneous F.M. frequency,

$f_c$ = carrier frequency,

$E_s(t)$ = signal amplitude

In terms of this instantaneous frequency expression and assuming a sinusoidal carrier one can describe the F.M. wave by the expression below:

$$E_c(t) = E_o \sin 2\pi \left[ f_c t + K_2 \int E_s(t)\, dt \right] \qquad (3)$$

Though not particularly relevant to this discussion, the close association between phase modulation and frequency modulation in the above relation is clearly implied.

Actually the waveform in Fig. 1c represents one that has been strongly "limited". In this representation only the zero crossings are preserved and all amplitude content is lost. According to F.M. theory, however, all the information content is still contained in this "chopped" signal. To put the

signal into a more convenient form for processing one need only use a Schmitt trigger arrangement and constant area pulse generator (some form of multivibrator) to detect the zero crossings and generate a pulse to represent each crossing. One would now have the F.M. signal representation shown in Fig. 2. It is only after F.M. signals have been put into this form that they can be processed by the demodulation unit to be described in the next section.

Now from Fig. 2 one can see qualitatively that there is a greater density of pulses in the regions of signal wave maxima and a smaller density of pulses in the regions of signal wave minima. For regions in between, one has corresponding intermediate densities of pulses. It seems plausible, therefore, that if one were to count a number of these pulses in the F.M. wave for a given period of time and compare the number of carrier pulses occurring over the same period of time, one should obtain a quantity that can characterize the audio signal over the period of time measured. In actual practice the period of time over which the pulses are counted is known as the sampling interval and the technique of counting pulses is known as frequency counting.

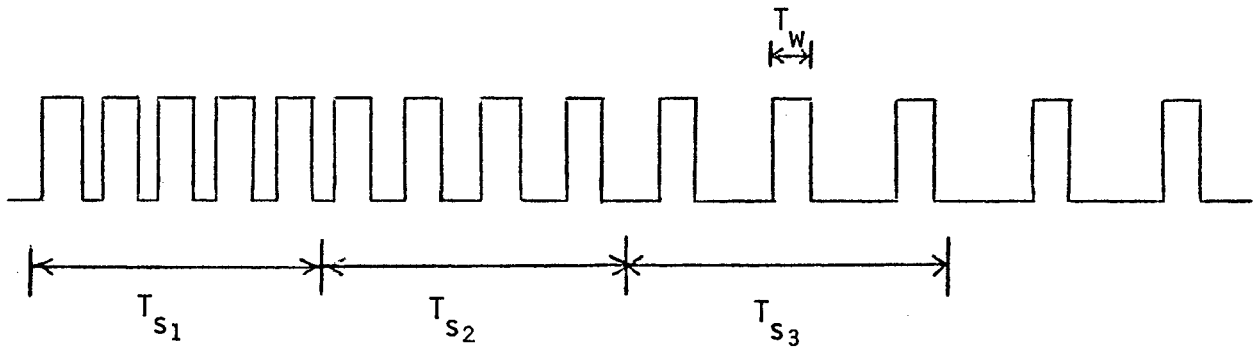
(b) Sampling Considerations

From equation (2) it was seen that the instantaneous frequency of an F.M. wave could be represented by:

$$f_i = f_c + K_2 E_s(t)$$

In this expression the quantity $K_2 E_s(t)$ actually represents the "frequency deviation" referred to earlier. Also, as mentioned earlier, in an effort to conserve frequency space an upper limit

MODIFIED F.M. WAVE




$T_{S_i}$ = Sampling Interval

$T_W$ = Constant Pulse Width

Fig. 2

or maximum frequency deviation has been imposed. One can represent this maximum frequency deviation by $\Delta f_m$. For all conventional F.M. systems, audio signal information must be characterized by frequency deviations within this range.

$$\text{i.e. } -\Delta f_m \leq K_2 E_s(t) \leq \Delta f_m$$

Accordingly, if one has a carrier of frequency "$f_c$" and a modulating signal $E_s(t)$ of frequency "$f_s$", then the instantaneous frequency of the resultant F.M. wave must vary between $f_c \pm \Delta f_m$ with a periodicity, "$f_s$". Now, if one converts this standard F.M. wave to a pulse form as mentioned above and counts the number of equal width pulses during a sampling interval $T_s$ seconds long, then during one sampling interval one would detect from $(f_c + \Delta f_m)T_s$ to $(f_c - \Delta f_m)T_s$ pulses depending upon the relative location of the sampling interval (S.I.) over the signal waveform. In the sampling interval a total of $f_c T_s$ carrier pulses would have been generated by an unmodulated carrier wave. If one could somehow obtain the difference between the number of "instantaneous" F.M. pulses and the corresponding number of carrier pulses occurring for each sampling interval $T_s$, then one would have a characterization of the signal wave for all sampling intervals and essentially reproduce the signal.

As a consequence of the above the question immediately arises as to how many "samples" one must take to effectively characterize a signal. The sampling theorem of Shannon [8] states that any signal can be completely characterized by specifying

its value at a specific sampling time when the sampling rate is at least twice the highest frequency component of the signal. In other words, if a signal is band limited to an upper frequency of B Hz then one must sample at a rate > 2B Hz to obtain proper characterization of the signal. In terms of sampling intervals this means that one must have a sampling interval or window of at least half the period of the signal.

In consideration of the above it appears that to detect the complete audio frequency range from about 20Hz to 20KHz one must sample at a maximum sampling rate (interval) of 40KHz, or once every $2.5 \times 10^{-5}$ seconds. Slower rates will resolve lower frequency signals but blur out higher frequency signals. The system considered in this report is designed to count the number of pulses that occur during the time interval $T_s$ as shown in Fig. 2 and so the notion of sampling is not in the conventional sense (although sampling theory still holds). In this system for a given sampling time interval, signal resolution (or the number of sample intervals measured to specify a signal over one cycle) depends on the frequency being sampled. Fig. 3 is a plot of resolution (i.e. number of sample intervals measured) versus signal frequency for various sampling times. Also drawn in on this graph is the minimum resolution or minimum number of samples required by the sampling theory to characterize the signal. As is evident from the graph, for a given sampling time interval resolution is a strongly decreasing function of frequency.

Fig. 3

Resolution versus Frequency
for various
Sampling Interval Lengths

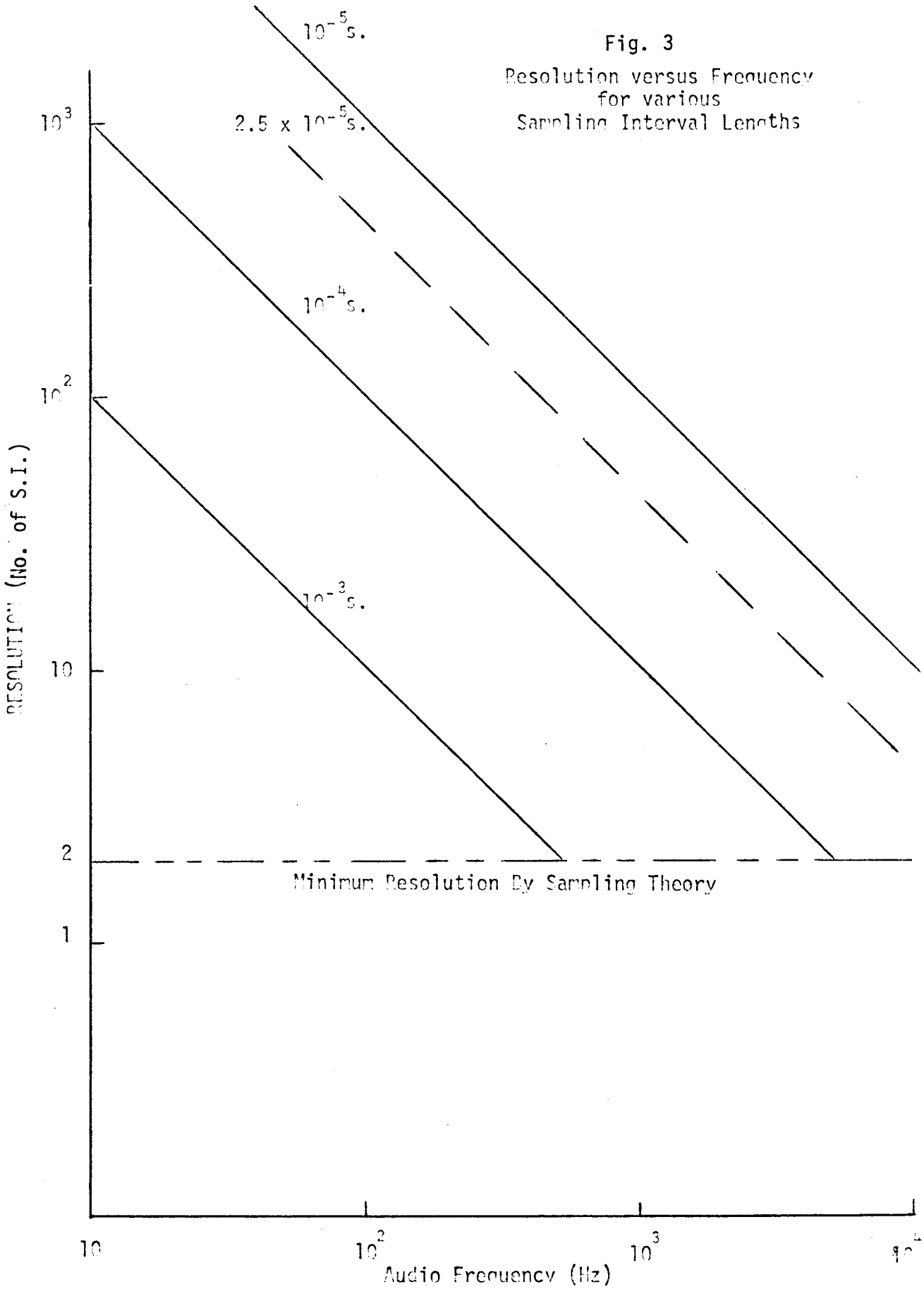


Fig. 4

NO. OF COUNTS
VERSUS
SAMPLING INTERVAL
(FOR VARIOUS M.F.D.)

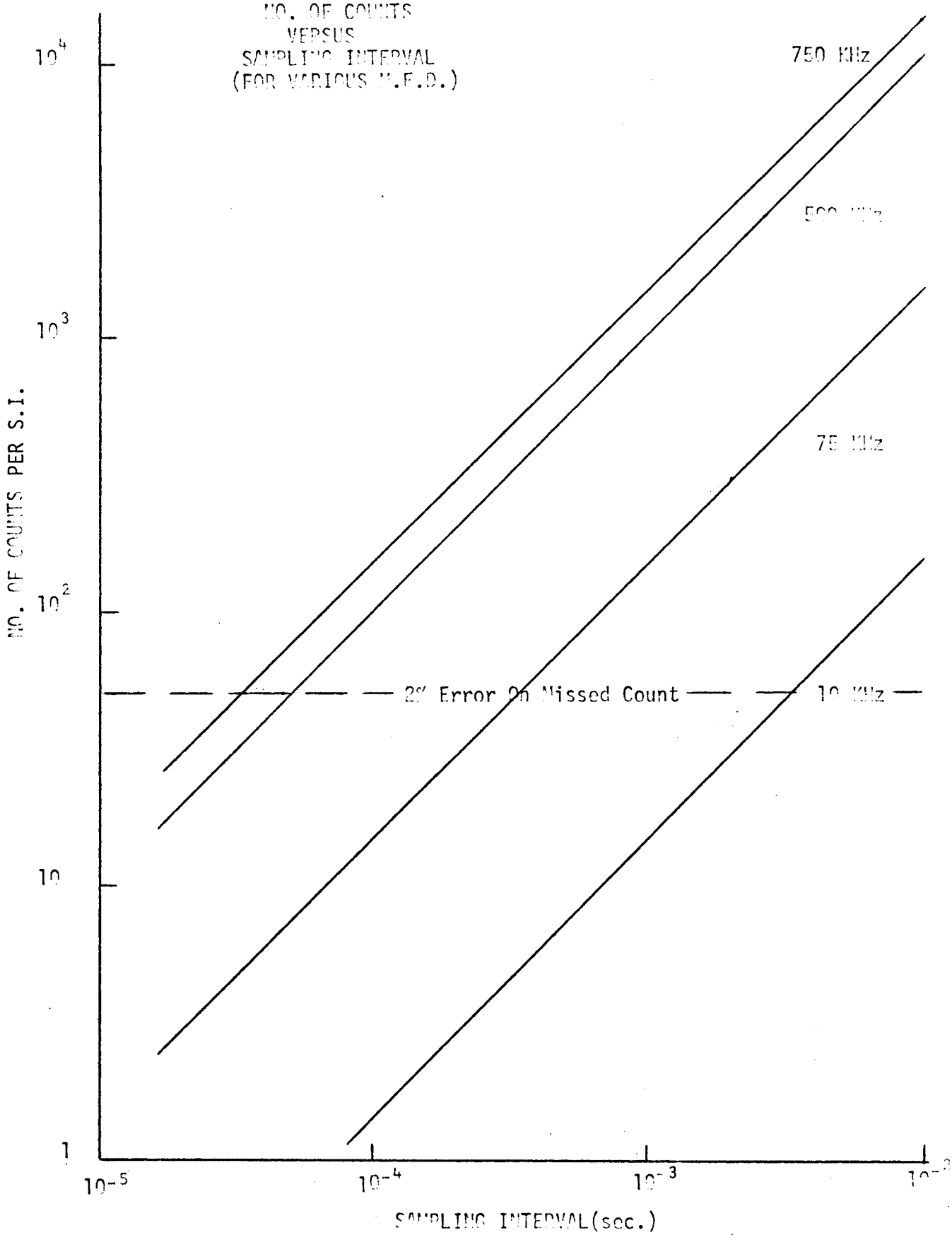


Fig. 4 is a graph of the number of counts that would be obtained per sampling interval versus sampling interval for various maximum frequency deviations (M.F.D.). Drawn in on this graph is the threshold for a 2% error limit on a missed pulse. Obviously one must have at least 50 pulses per sampling interval (S.I.) to insure only 2% error on a missed pulse. This 2% threshold was chosen to be a suitable tolerance level for proper signal perception. From the graph of Fig. 4 one can see that if the sampling time of $2.5 \times 10^{-5}$ seconds is chosen in order to resolve signals over the entire audible range, one would need an M.F.D. and bandwidth of greater than 750KHz to realize a practical 2% error tolerance in the signal amplitude. For conventional F.M. radio with its fixed M.F.D. of 75KHz one would have an error of 20% for every missed count. This value of 20% is obtained from Fig. 4 where one can see by inspection that a maximum frequency deviation of 75KHz and sampling interval of $2.5 \times 10^{-5}$ seconds yields only 5 counts to represent the signal in the sampling interval. If one misses a count, this represents an error of 20% and in any realizable frequency counting system this would entail almost complete distortion. Certainly then, one cannot apply this technique to conventional F.M. broadcast signals unless the M.F.D., inherent in the modulation process, is increased. Since increasing M.F.D. also entails an increasing bandwidth, frequency counting systems appear practical only in wide-band systems or systems where bandwidth conservation is not of prime importance.

In further illustration of this in-applicability to radio, consider a typical carrier of 100MHz frequency with the imposed M.F.D. of 75KHz. To resolve a signal of 15KHz modulating this carrier assume one samples every $10^{-5}$ sec. (i.e. about 7 samples over the signal). The maximum signal amplitude in this case, would be represented by 1000.75 counts over a sampling interval and the minimum represented by 999.25 counts over a sampling interval. The entire difference in signal amplitude is represented by only 1.50 counts or pulses. Since one is dealing with a frequency counting type detector and not a phase shift device, one would not be able to resolve this radio signal (discrete pulses are required). For conventional radio, as it stands, with a M.F.D. fixed at 75KHz one could only obtain effective reproduction for signals of frequency less than 900Hz using frequency counting techniques.

In this report, as will be seen carrier frequencies up to 2MHz and audio signals over the entire audio range were demodulated, but it should be emphasized that to effect this, M.F.D.'s much in excess of the conventional 75KHz were employed.

## DIGITAL HARDWARE REALIZATION

### (a) Elementary Design

In Fig. 5 is shown a sketch of the fundamental demodulator design. Ideally, the system would consist of a chain of cascaded up-counters connected by logic gates to a processor or difference unit into which could be toggled a binary number representing the carrier count over the given sampling time. The output from this processor would be a binary number representing a biased signal which when fed into a D/A converter should give a replica of the signal waveform.

Hypothetically, the F.M. signal, in the proper format (see INTRODUCTION) would start to fill up the counters during a sampling interval. Then, after a prescribed sampling period, the contents of the counters would be transferred through the logic gates by a clock pulse of narrow width and occurring at the sampling frequency. The trailing edge of this clock pulse would be used to trigger the clear inputs of the counters and thus reset them for another sampling interval.

The contents of the counters, toggled across the logic gates into the processor represent in binary format a number indicative of the density of the F.M. wave in that particular sampling interval. The binary number is subtracted from a binary number representing the carrier in the processor. The binary number output, which indicates the pulse density difference, is converted to an analogue voltage by a D/A converter. This voltage level persists until the next clock pulse triggers another transfer.

Fundamental Demodulator Design

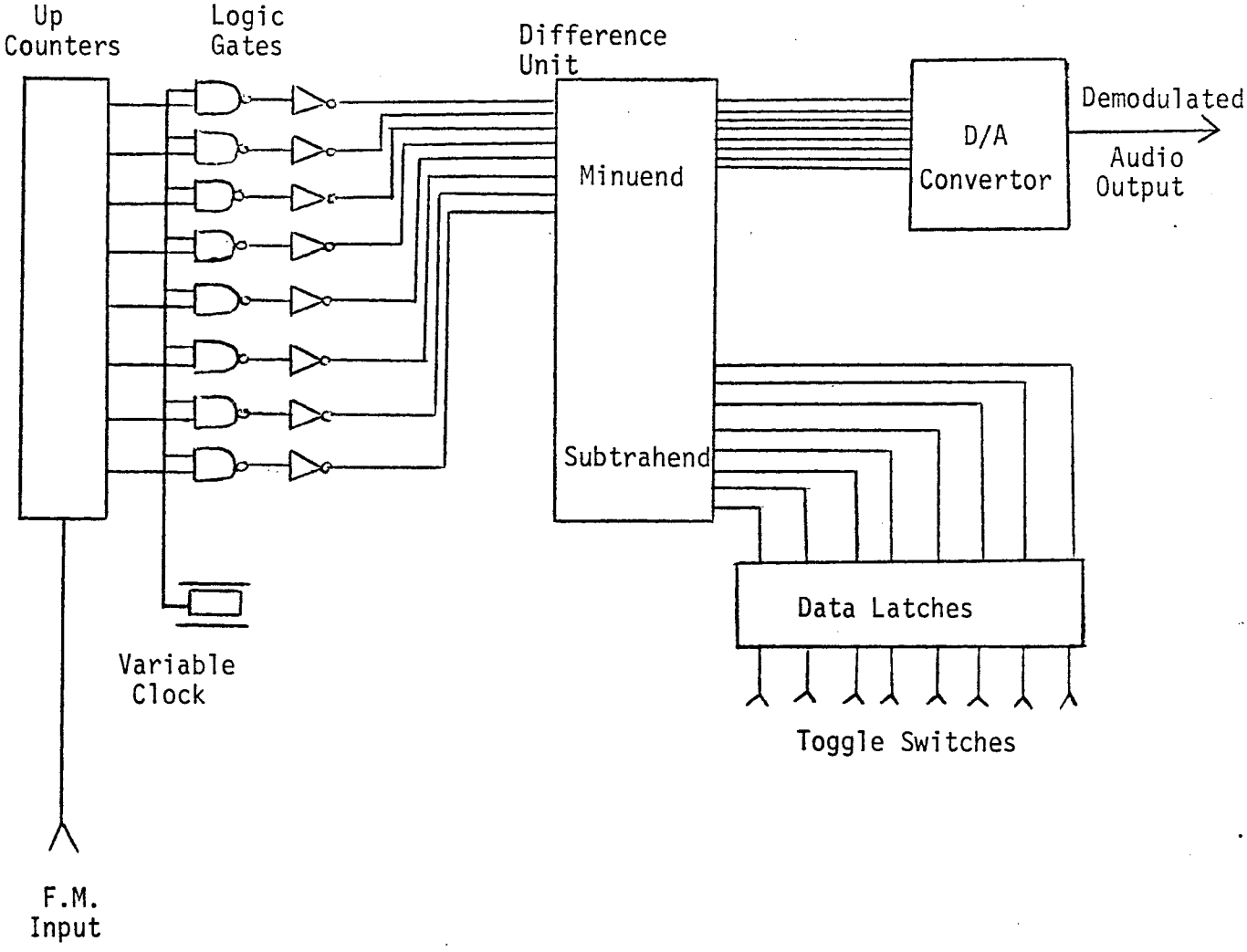



Fig. 5

(b) Practical Design Problems

In attempting to translate the initial elementary design into a feasible circuit a problem was immediately encountered. Because of the relatively high frequencies of the carrier signals with respect to the sampling rate one would need very long counters to contain all the counts that would be accumulated over a standard sampling interval. For a typical 100MHz carrier being sampled every $10^{-4}$ seconds one would need a binary counter with a capacity exceeding 1000 states (i.e. 10 bits). For higher carrier frequencies and for lower sampling rates one would need even larger counters. For every additional counter bit one would also need a corresponding logic element, an extra bit in the processor and an extra bit in the D/A converter. The amount of circuitry would become tedious and inconvenient after a certain point. In addition, for higher carrier frequencies, the digitized frequency deviations representing audio signals would occupy only a small fraction (the least significant bits) of the total number of bits entered into the D/A converter and thus would be represented by only a small fraction of the available output voltage. To overcome these difficulties a rather specialized technique was devised and termed "cycle filtering".

Cycle filtering is actually a low pass filtering technique that reduces the counter size and the number of counts utilized in processing without actually distorting the information content. For a given maximum frequency deviation and intermediate size counter one can actually neglect the higher order bits stored during a

sampling interval since the required audio signal information will be primarily reflected in the low order bits. With this in mind one can use a finite counter size and allow the counter to cycle itself through the sampling interval. The counter will fill up and dump an integral number of times in the sampling interval and retain only the residue that is in the counter when the sample clock toggles a transfer into the processor. In effect one has completely ignored the counts corresponding to the high frequency carrier and retained only the counts relevant to the much lower frequency audio signal. The only restriction is that the total frequency swing and counter size be correlated so that for all the sampling intervals representing sections of the signal, the counter undergoes the same number of "initial" cycles before the residue is transferred.

In this technique the proportionality inherent in the frequency deviation from the modulation procedure is retained and the total maximum frequency deviation range (i.e. the modulating amplitude) is reflected in a count range. Fig. 6a is an attempt to clarify this idea. From it one can see that all the counts or pulses below line AA have been neglected. All the counts above line AA represent the reduced or "cycle filtered" F.M. signal. The binary numbers that are transferred to the difference unit to characterize each sampling interval are actually these "reduced" count numbers.

In Fig. 6a the count range representing the maximum frequency swing is shown to be less than the actual counter

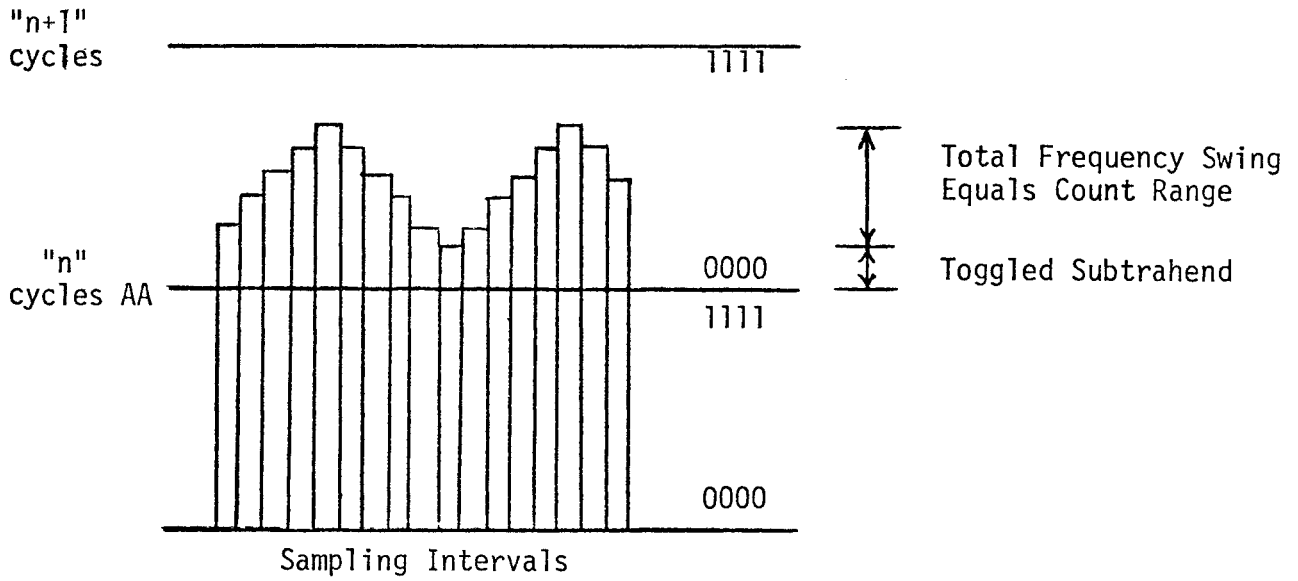



Fig. 6a

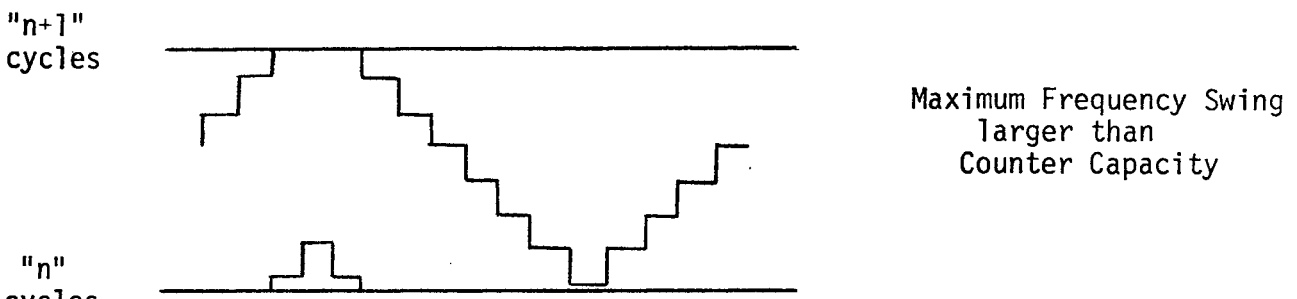



Fig. 6b

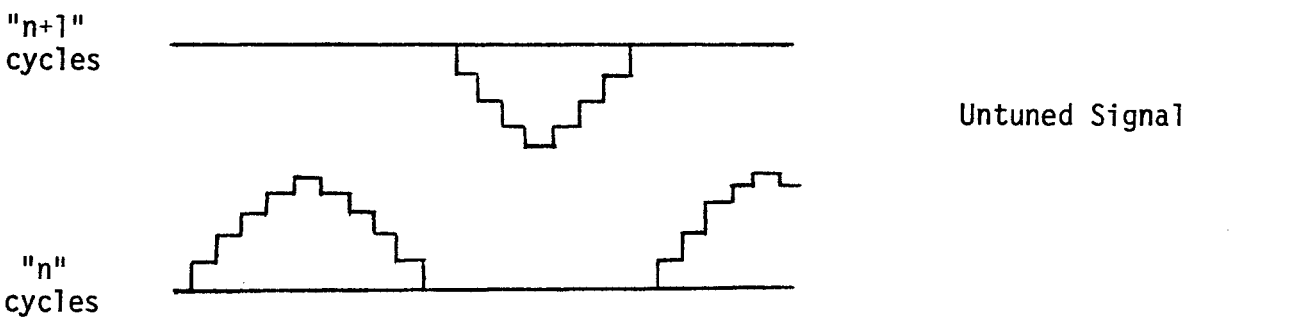



Fig. 6c

capacity. Ideally the counter capacity should be equal to the largest frequency swing permitted (i.e. $2\Delta f_m$) to make full use of the system resolution. Regardless, as long as the count number representing the total frequency swing is not greater than the counter capacity, one can "tune" the signal to minimum at zero count by simply toggling into the difference unit a pre-determined number representing the signal minimum. The difference unit (see Fig. 5 ) subtracts this number from incoming data after each sampling interval and outputs a series of numbers representing the audio signal with minimum amplitude set at binary zero. This pre-determined "tuning number" would ordinarily depend on carrier frequency, maximum frequency deviation, and sampling interval for a long counter system (i.e. $(f_c - \Delta f_m)T_s$); but, in attempting to reduce hardware requirements and increase resolution through cycle filtering one must take the counter size into consideration and employ suitably "reduced" tuning numbers. For a given system with fixed modulation parameters, the required reduced tuning numbers for a variety of carriers can be quite easily determined and are completely independent of audio signal parameters.

If the maximum "digitized" frequency swing is greater than the counter capacity or if the signal is not properly tuned, one will encounter a phenomenum which could be termed "cycle overflow". In this instance the signal would appear segmented and shifted because the residues transferred to the difference unit to represent different parts of the signal belong to different cycles. Proper

tuning or increasing counter size to accomodate the maximum frequency swing should eliminate this problem. Fig. 6b and c illustrate "overflow" conditions resulting from two different sources. In Fig. 6b although the signal is tuned, overflow occurs because the system capacity is less than the digitized maximum frequency swing. In Fig. 6c overflow occurs because of improper tuning.

(c) Functional Demodulator Design and Operation

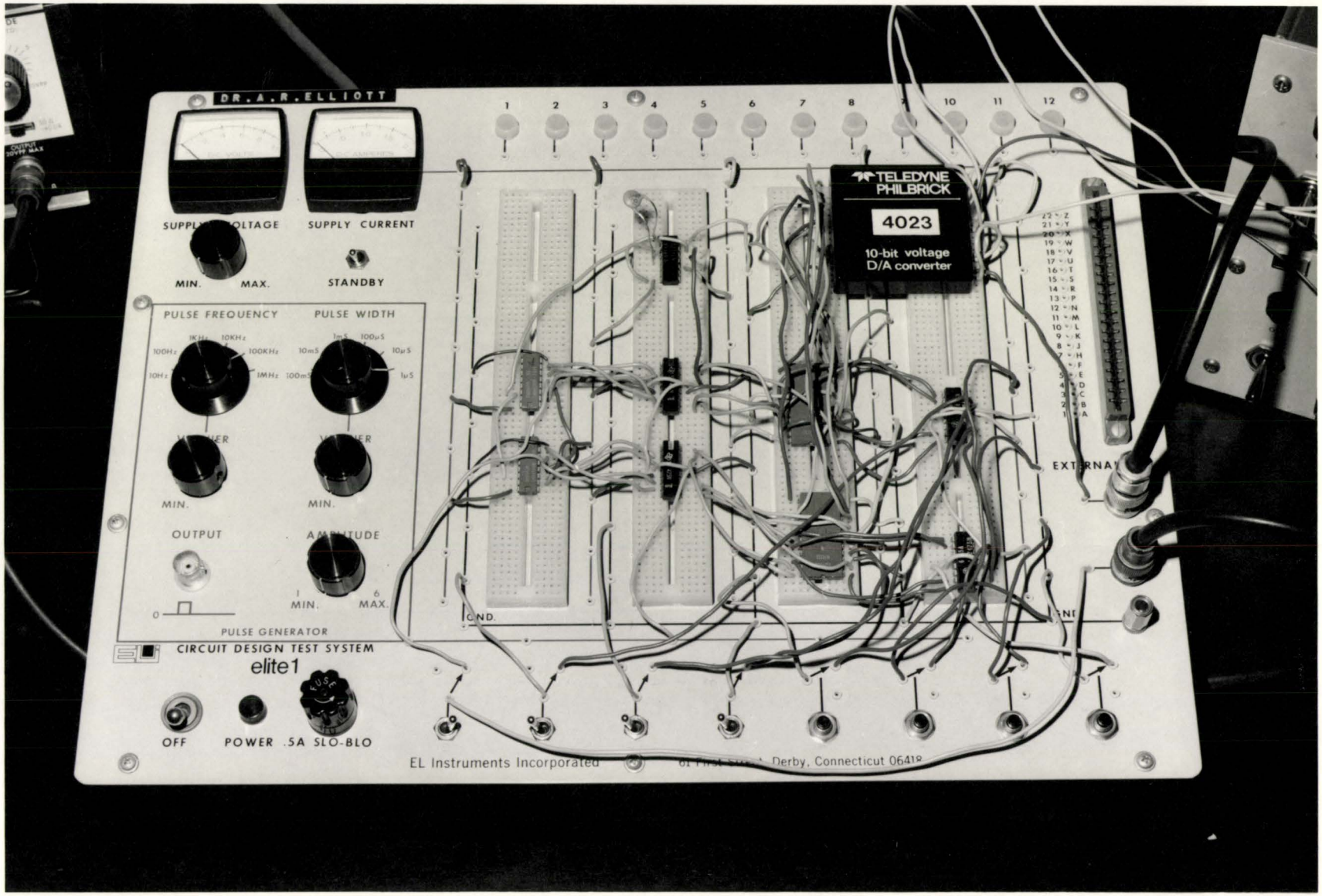
The final functional design of the demodulator was much like the elementary design initially considered. Two up-counters were cascaded to give 8-bit input capacity to the system. Instead of logic gates, data latches were employed for the transfer of the accumulated count into the difference unit after each sampling interval. The difference unit was composed of full adders that performed 1's - complement addition to implement binary subtraction. The "tuning" number was toggled in binary format across a set of data latches to the subtrahend entry of the difference unit. Since this number represented the signal minimum (i.e. $(f_c - \Delta f_m)T_s$ reduced) the possibility of having to contend with negative numbers and associated sign logic was eliminated. The output from the subtractor was entered into an 8-bit D/A converter to obtain the final demodulated analogue signal.

The clock used to transfer the contents of the counters accumulated during a sampling interval was a variable frequency, variable pulse width unit capable of output frequencies from 10Hz to 1MHz and pulse widths from 10 msec. to 0.1 μsec.

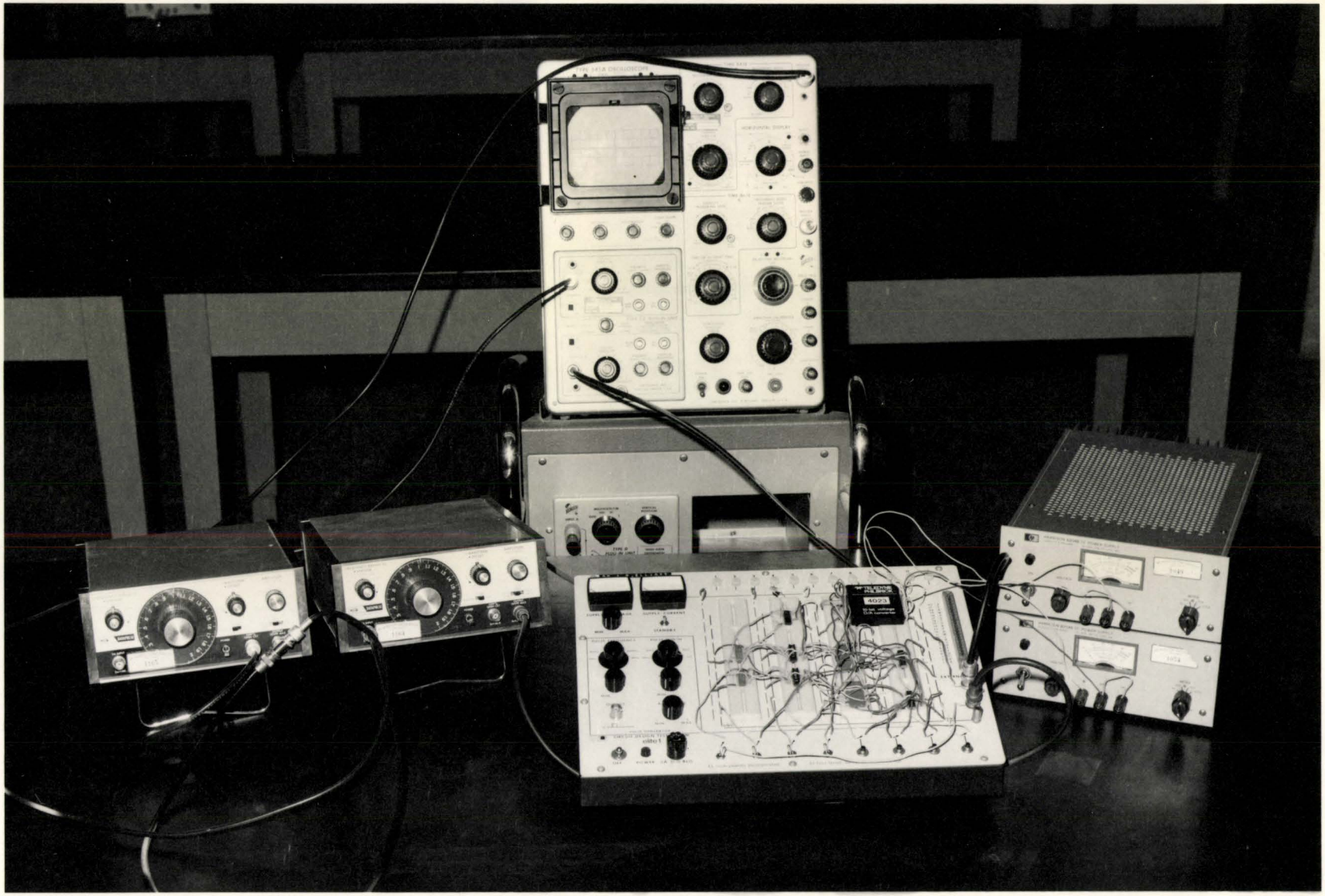
To reset the counters after each sampling interval a monostable (one-shot) multivibrator was used. The one-shot was triggered on the falling edge of the clock pulse to ensure that the accumulated data was successfully transferred before the counters were cleared. The output pulse width of the monostable was set at 100 nanoseconds and the width of the clock pulse was set at 1 μsec. to ensure that no pulses would be lost during transfer or resetting.

The F.M. input unit was capable of producing carrier signals up to 2MHz and audio signals over the entire audio range. The system also had variable modulation with a correspondingly variable maximum frequency swing. The system had to be modified slightly to put the output F.M. in a form compatible to entry into the demodulator unit. (See INTRODUCTION, APPENDIX B)

A detailed schematic of the circuit as well as a listing of the actual components used is given in Appendix A. Photographs of the demodulation unit and the experimental set-up used to test the unit out are shown in Fig. 7 and 8 respectively.




Demodulation Unit
Fig. 7



Experimental Set-up

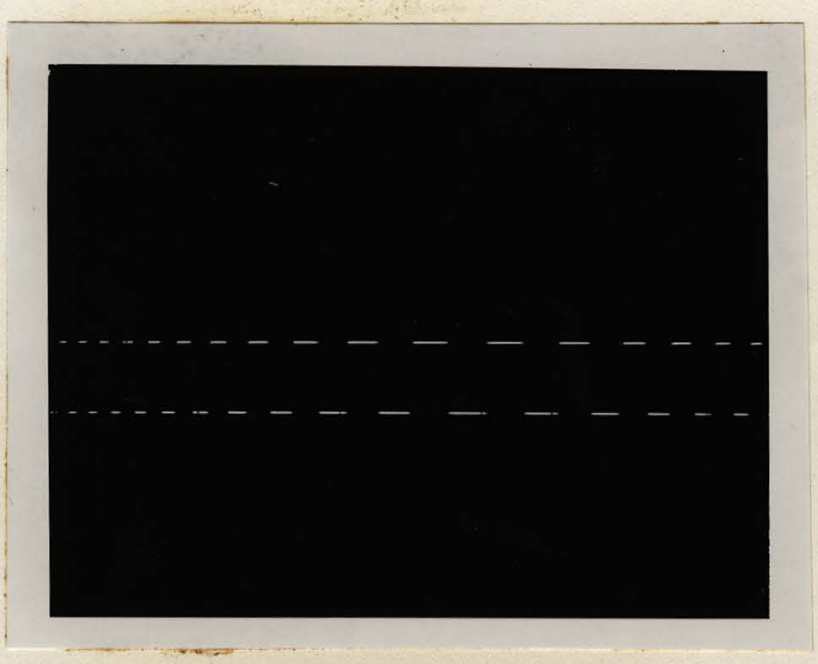
Fig. 8

## Demodulator Test Results

As mentioned earlier, conventional F.M. signals must be slightly modified (digitized) into a form compatible for entry into the digital demodulation unit. Fig. 9 is an oscilloscope photograph of part of an actual "modified" F.M. input used to test out the demodulator. In this instance a 1.4KHz sine wave was modulating a 1MHz carrier. Although the time scale had to be uncalibrated to facilitate the photography it appears that one is seeing only the minimum half of the audio cycle. Nonetheless, the photograph gives clear visual indication of the format required for processing by the digital demodulator.

In Fig. 10 is oscilloscope photograph of typical clock pulses and one-shots employed. To enable one to more easily discern the relative pulse widths involved, the clock frequency employed for the photograph was increased to 100KHz. For actual sampling purposes clock frequencies from 5KHz to 50KHz were employed. The sequence of events on the demodulator is indicated rather clearly in this figure. Counts accumulate in the counter until a clock pulse goes high. During the momentary high level the logic gates are enabled and the contents of the counters are transferred into the subtracter. When the clock pulse goes low it disables the logic gates and triggers a monostable multivibrator. The multivibrator applies a narrow pulse to the clear inputs of the counters to reset them for the next sampling interval.

From the account given in the first part of this report

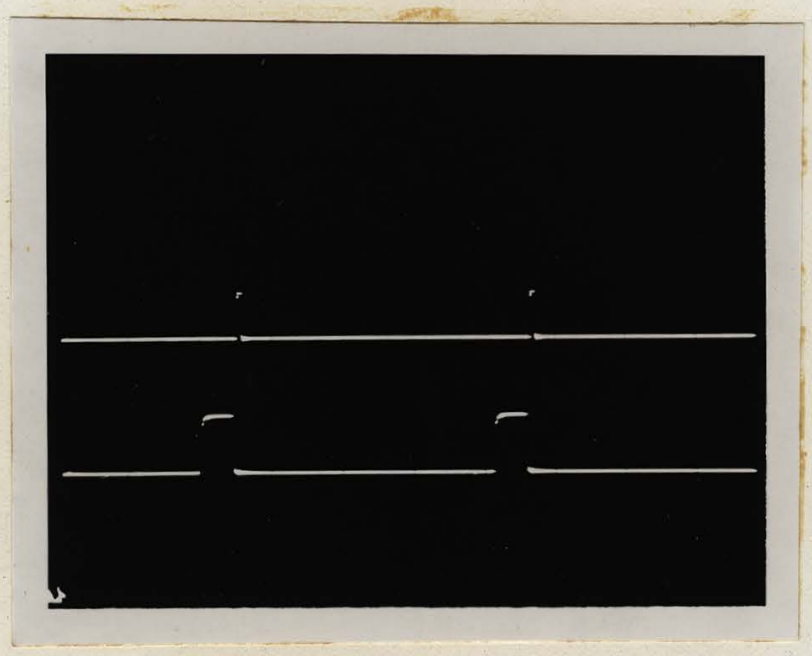


F.M. Modified Input

1.4 KHz signal
1 MHz carrier

Vertical scale = 5v/div.
Horizontal scale = 2 μsec./div. (uncal.)
1 div. = approx. 1 cm.

Fig. 9



Control Pulses

(A) One shot
(B) Clock pulses

Vertical scale = 5v/div.
Horizontal scale = 2 μsec./div.
1 div. = approx. 1 cm.

Fig. 10

one would expect that the output from the D/A converter of the demodulator when fed into an oscilloscope should yield a demodulated "histogram-type" signal representative of the input audio signal before modulation. Fig. 11a , b, and c illustrate typical input audio signals (before modulation) and the resultant demodulated audio signals after processing by the demodulator for three basic waveforms - sinusoidal, triangular, and square.

All three waveforms are 200Hz signals that modulated 1MHz carriers and were sampled at a frequency of 7KHz. There is good reproduction in all cases. The only actual difference in the two sets of signals is a slight difference in phase due to the fact that each sample was accumulated during one sampling interval but not displayed unitl the start of the next sampling interval. The apparent gross $90^{o}$ phase shift was due to inversion caused by the scope sweep triggering level. As well the apparent amplification of the demodulated signal is due to the fact that an 8-bit bipolar (±5V) D/A converter was used with the maximum frequency swing of the input F.M. adjusted so that the corresponding count range would equal the counter capacity.

For these audio frequencies with the sampling interval length of $1.4 \times 10^{-4}$ sec.(i.e. 7KHz) duration, there are 35 samples characterizing the signals. As mentioned earlier, however, for the higher audio frequencies and the same sampling interval rate there are correspondingly fewer samples to characterize the signal and consequently there is poorer resolution. This is explicity illustrated in Fig. 12a , and b. The demodulated audio waveforms

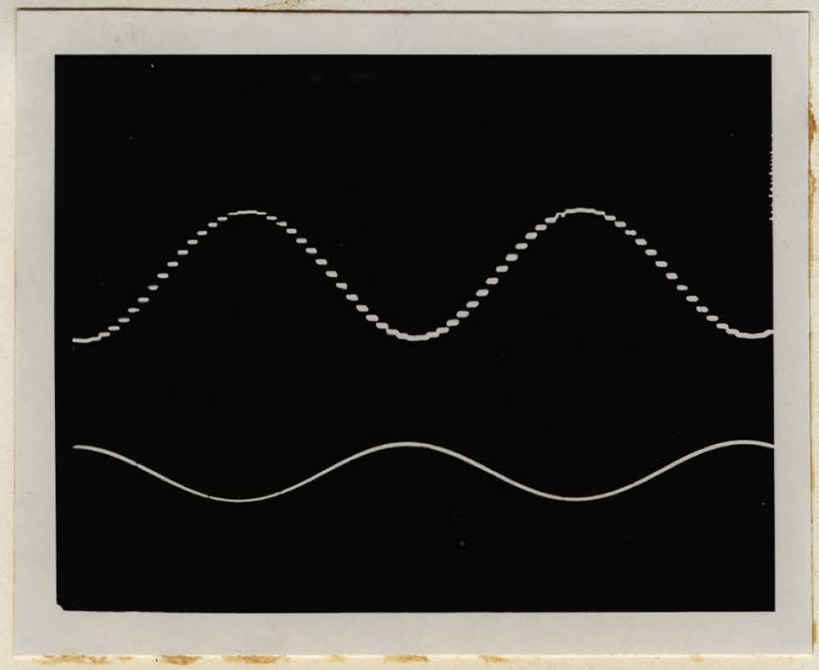


Fig. 11a

200 Hz Sine Wave

(A) Demodulated
(B) Original

Carrier = 1 MHz
Sampling Interval = 143 μsec.
(i.e. 7KHz)
Vert. Scale = 5v/div.
Horiz. Scale = 1 msec./div.
1 div. = approx. 1 cm.

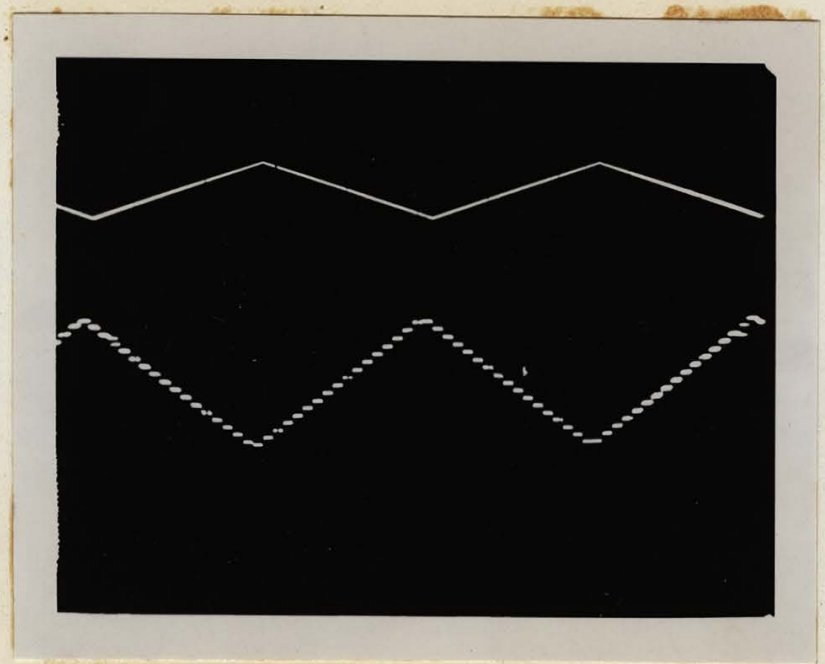


Fig. 11b

200 Hz Triangular Wave

(A) Demodulated
(B) Original

Carrier = 1 MHz
Sampling Interval = 143 μsec.
(i.e. 7KHz)

Vert. Scale = 5v/div.
Horiz. Scale = 1 msec./div.
1 div. = approx. 1 cm.

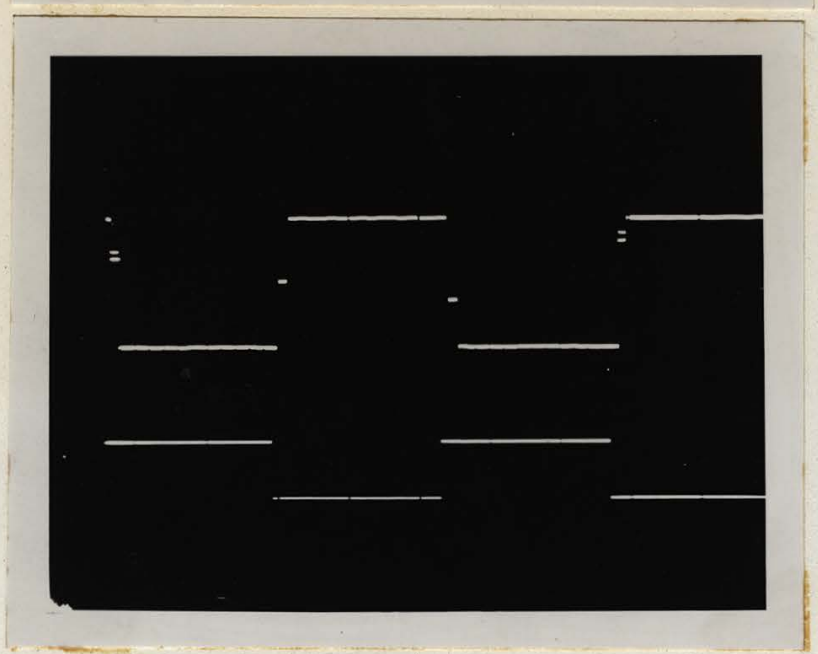


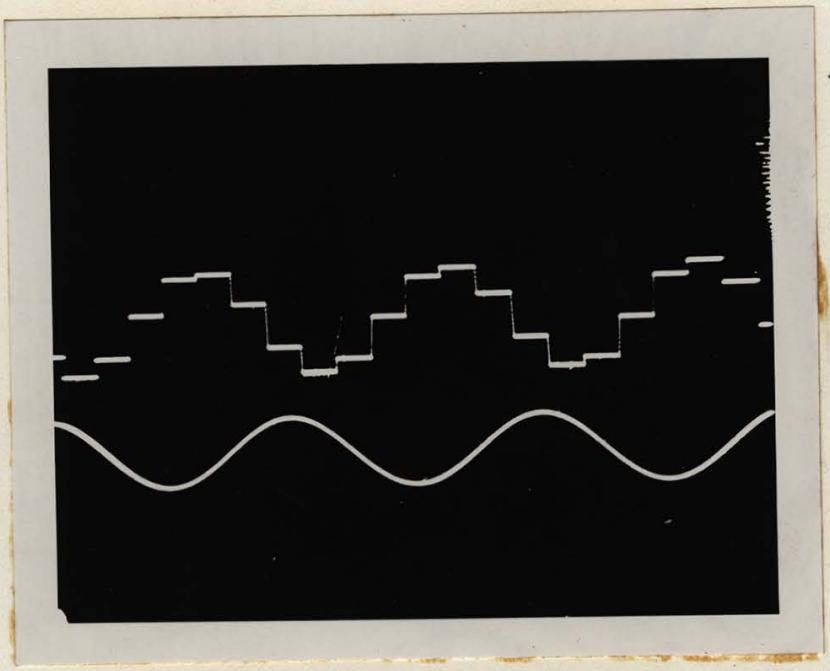
Fig. 11c

200 Hz Square Wave

(A) Demodulated
(B) Original

Carrier = 1 MHz
Sampling Interval = 143 μsec.
(i.e. 7KHz)

Vert. Scale = 5v/div.
Horiz. Scale = 1 msec./div.
1 div. = approx. 1 cm.

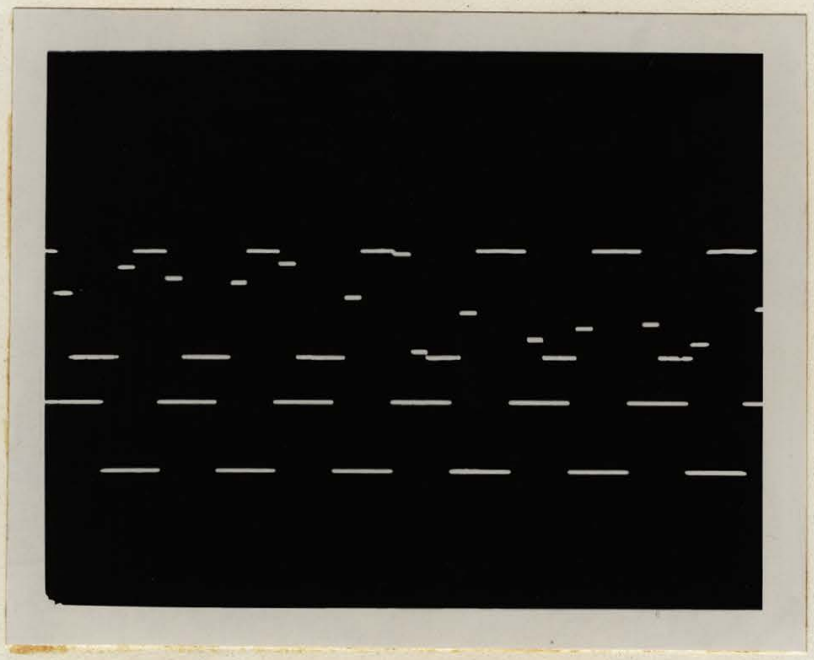


1.4 KHz Sine Wave

(A) Demodulated
(B) Original

Carrier = 1MHz
Sampling Interval = 100 μsec.
(i.e. 10KHz)
Horiz. Scale = 0.2 msec./div.
Vert. Scale = 5v/div.
1 div. = approx. 1 cm.

Fig. 12a



1.4 KHz Square Wave

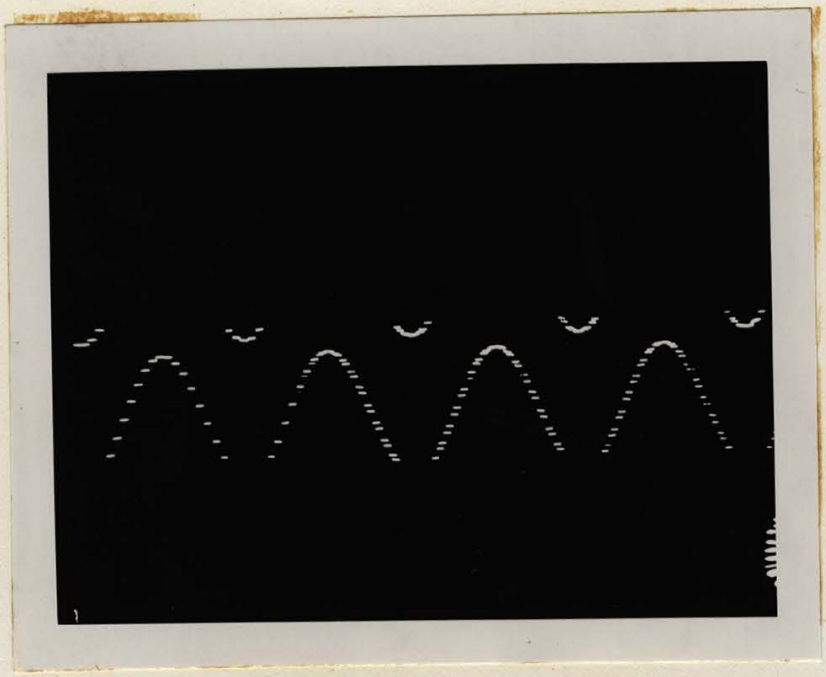
(A) Demodulated
(B) Original

Carrier = 1MHz
Sampling Interval = 100 μsec.
(i.e. 10KHz)
Horiz. Scale = 0.2 msec./div.
(uncal.)
Vert. Scale = 5v/div.
1 div. = approx. 1 cm.

Fig. 12b

in these oscilloscope photographs are 1.4KHz in frequency and the sampling interval rate is only slightly greater than before (i.e. 10KHz). At these higher frequencies, however, one now has only about 8 samples to characterize the signal and resolution is greatly reduced. Nonetheless, the signals are still discernable and should remain so until sampling theory is violated. To achieve the same resolution of these higher frequencies as for the 200Hz signals one need only increase the number of sampling intervals or add a low pass filter to the output. In this case, as well, the maximum frequency swing has been decreased on the F.M. input device and accordingly the number of counts representing this maximum frequency swing is less than total counter capacity (i.e. 8-bits, 10v p. to p.)

In both the above cases the demodulated signals were "tuned" until the signal minimum coincided to counter zero and the D/A output of -5.0v. When the signals are not tuned in, one has the phenomenon of "cycle overflow" mentioned earlier. In Fig. 13a and 13b oscilloscope photographs of signals that could be termed under-tuned and over-tuned respectively are shown. In the under-tuned case the carrier signal frequency is such that the number of counts representing the lower part of the audio signal occurs in a lower counter cycle than the counts representing the upper portion of the signal. As a result, samples around the signal minimum will be represented by a high count in cycle "i" and when transferred to the D/A converter would appear as high level. Samples over the rest of the signal waveform would be represented by low counts in cycle "i + 1" and would be outputed from the D/A converter as low

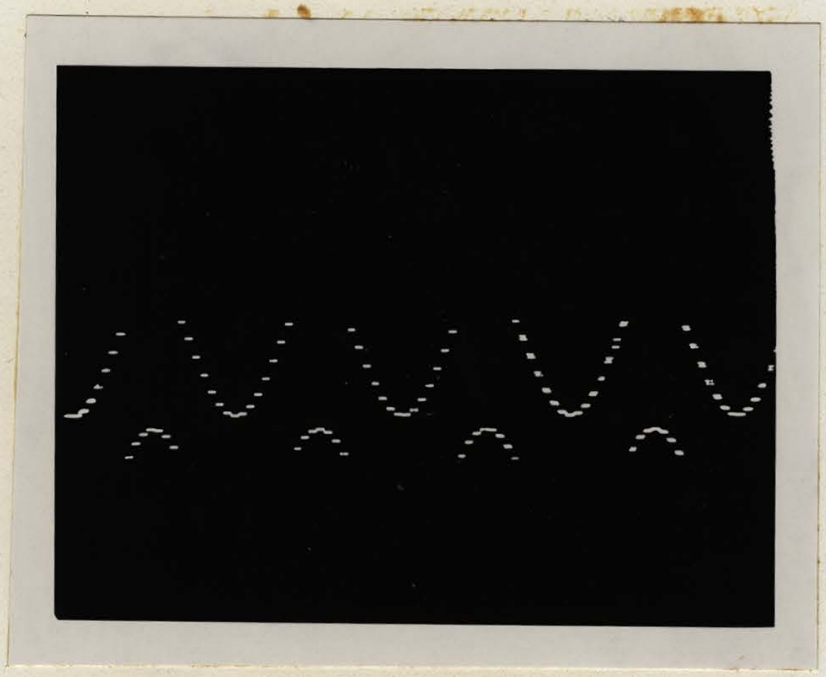


Under-tuned Signal

Audio = 400 Hz
Carrier = 1 MHz

Sampling Interval = 125 μsec.
(i.e. 8KHz)
Vert. Scale = 5v/div.
Horiz. Scale = 1 msec./div.
1 div. = approx. 1 cm.

Fig. 13a



Over-tuned Signal

Audio = 400 Hz
Carrier = 1 MHz

Sampling Interval = 125 μsec.
(i.e. 8KHz)
Vert. Scale = 5v/div.
Horiz. Scale = 1 msec./div.
1 div. = approx. 1 cm.

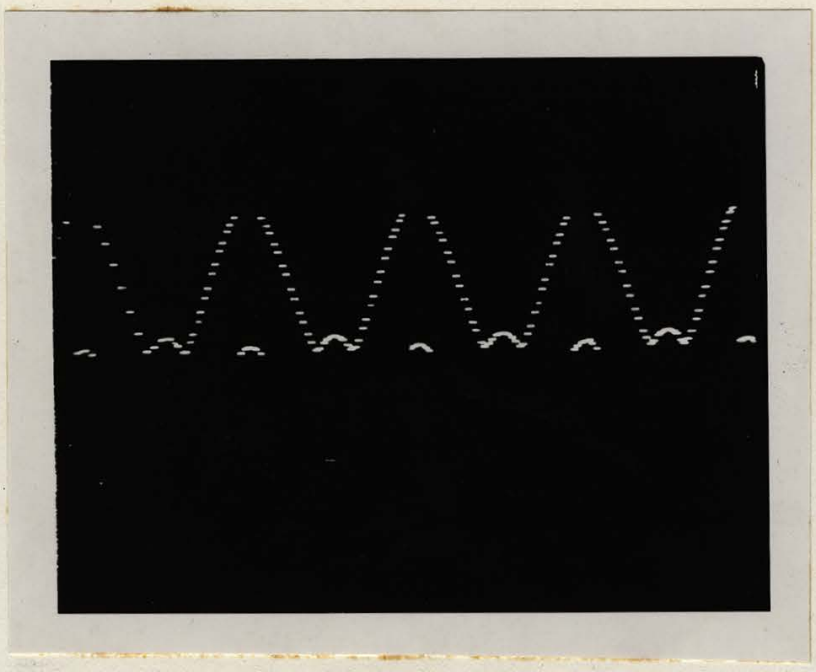
Fig. 13b

levels since the converter recognizes only the number fed into it after a sampling interval and has no indication of the number of initial cycles completed by the counter. To correct this problem one need only toggle in a suitable binary subtrahend into the subtracter. This would lower all of the signal into the same cycle and thus regroup the segmented signal. A similar treatment could be applied to the signal in Fig. 13b . It should be noted that double levels appear in the right hand side of these pictures. This is merely a manifestation of lack of proper correlation between oscilloscope sweeptime and camera exposure period.

In Fig. 14 another phenomenon is displayed in an oscilloscope photograph. For this particular sample the modulation has been increased to over 100%. As a result, one not only has over-flow segmentation but also a signal foldover effect due to over-modulation.

It should be mentioned at this point that to obtain proper resolution of most of the higher frequency audio signals, the maximum frequency swing had to be increased by increasing the degree of modulation. As mentioned earlier this was to be expected. Since there was an upper limit to the maximum frequency swing that would be obtained with the modulation unit, proper resolution of audio signals over 10KHz was unattainable. To obtain sufficient resolution for such signals one would need quite large frequency deviations (i.e. > 750KHz ). A typical example of the type of distortion that occurred is shown in Fig. 15 for a 5KHz square wave. The bluriness is only further illustration of lack of correlation between camera shutter speed and scope sweep speed; however, the

significant rounding of the signal edges is quite symptomatic of insufficient bandwidth.

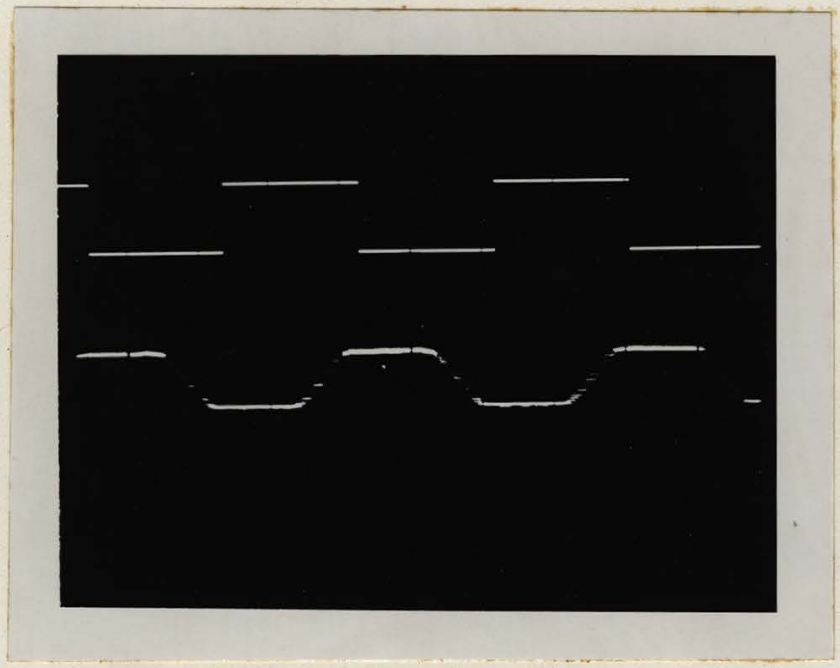


Over-modulated Signal

Audio = 400 Hz
Carrier = 1 MHz

Sampling Interval = 125 μsec.
(i.e. 8KHz)
Vert. Scale = 5v/div.
Horiz. Scale = 1 msec./div.
1 div. = approx. 1 cm.

Fig. 14



Bandwidth Distortion

(A) Demodulated Signal
Horiz. Sclae = 5 μsec./div.
Vert. Scale = 1v/div.

(B) Original Signal (5 KHz)
Horiz. Scale = 5 μsec./div.
Vert. Scale = 5v/div.

Carrier = 10 μsec.
(i.e. 100 KHz)
Sampling Interval = 10 μsec.
1 div. = approx. 1 cm.

Fig. 15

## CONCLUSIONS

The digital F.M. demodulator unit functioned as anticipated, illustrating both the merits and limitations expected of it. It would appear that F.M. demodulation using frequency counting techniques augmented by modern high speed digital hardware is achievable, but not particularly feasible in systems requiring bandwidth conservation. The main limitation here is not in the hardware, but in the input requirements imposed by the demodulation technique. The "digitizing" of analogue F.M. inputs into the proper format is not particularly difficult and can be easily realized through digital hardware. However, the fact that the frequency deviations must yield at least 100 integral pulses (50 cycles) at sampling intervals of less than $10^{-4}$ seconds to achieve high frequency resolution demands excessive input bandwidth. Frequency space for conventional F.M. radio would be quite overcrowded if the necessary frequency deviations of over 750KHz required by this system were imposed.

Nonetheless, if one considers only lower frequency audio signals (<5000 Hz) the demodulator functions quite well. It eliminates completely the need for A.F.C. (automatic frequency correction) hardware since the digital components make drift problems virtually non-existent. If one uses a variable sampling clock it also eliminates the need for much of the other conventional F.M. receiver components such as I.F. amplifiers and mixers. Alternatively one could use a fixed clock and retain an I.F. arrangement. Either way one would still require some kind of isolater or filter that

would allow only the frequency band of interest into the system (i.e. range of $f_c \pm \Delta f_m$). This would have to be tunable and would serve as a type of "rough" tuning complementing the finer tuning in the demodulator itself.

As to future recommendations, one would have to emphasize that pursuits should centre around the problem of making the system input bandwidth requirements more amenable to conventional inputs. The frequency counting technique could perhaps be modified. Alternately, a type of conversion network to quantize phase shifts into pulses or extend count ranges by multiplying pulses could be considered.

# BIBLIOGRAPHY

(1) Cash, A.B. and Liff, A.A.; Frequency Modulation Receivers, Prentice-Hall, New Jersey (1968)

(2) Bennett, W.R. and Davey, J.R.; Data Transmission, McGraw-Hill New York (1965)

(3) Malmstadt, H.V. and Enke, C.G.; Digital Electronics for Scientists, W.A. Benjamin Inc., New York (1969)

(4) McMullen, C.W.; Communication Theory Principles, MacMillan, New York (1968)

(5) Richards, R.K.; Electronic Digital Systems, Wiley, New York (1966)

(6) Schwartz, M.; Information Transmission, Modulation and Noise New York, Second Edition (1970)

(7) Seeley, S.; Radio Electronics, McGraw-Hill, New York (1956)

(8) Shannon, C.E.; Proc. Inst. Radio Eng. 37, 10-21 (1939)

(9) Stwiley, K.R.; Frequency Modulated Radio, Newnes Press, London (1958)

(10) Sunde, E.D.; Communications Systems Engineering Theory, Wiley, New York (1969)

(11) Tibbs, C.E. and Johnstone, C.G.; Frequency Modulation Engineering, Wiley, New York (1956)

# APPENDIX A

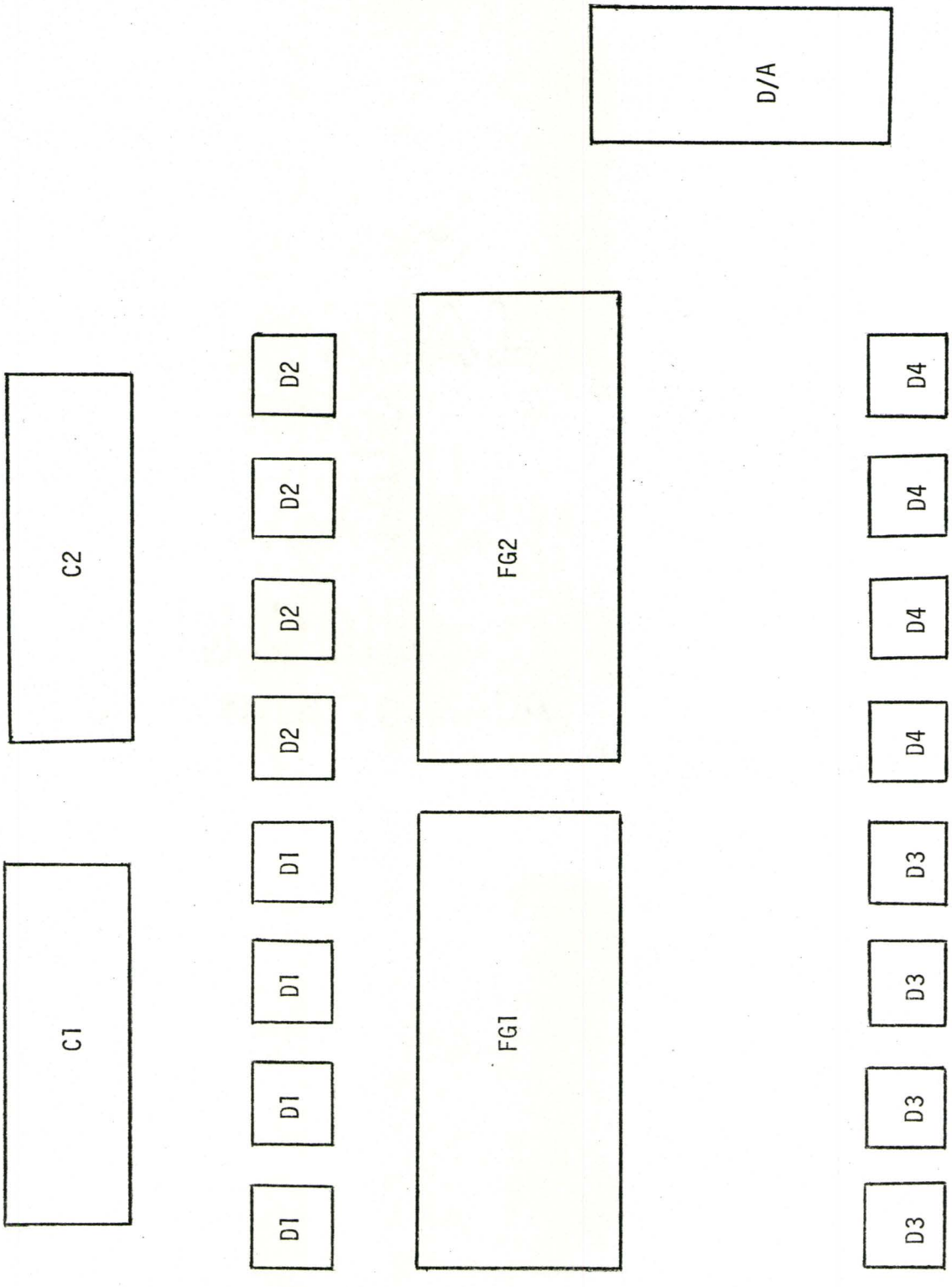



Fig. A1

* Outline refers to photograph on page 23.

## APPENDIX A (cont'd)

Demodulator Components

| Components | Description | Manufacturer | Type Number |
|---|---|---|---|
| D1,D2,D3,D4 | 4-Bit Bistable Latches | Texas Instruments | SN7475 |
| FG1,FG2 | Arithmetic Logic Unit | Texas Instruments | SN74181 |
| C1,C2 | Binary 4-Bit Up/Down Counter | Texas Instruments | SN74193 |
| D/A | 10-Bit Digital to Analogue Converter | Teledyne Philbrick | 4023 |

# APPENDIX B

Accessory Equipment Used:

| Unit | Manufacturer | Type Number |
|---|---|---|
| Power Supplies (D.C.) | Hewlett-Packard | 6204 B |
| Function Generators | Datapulse | 401 |
| Oscilloscope | Tektronix | 545 A |
| 'Scioe' Camera | Tektronix | C-12 |

"Modified" F.M. Synthesis: To synthesize the required F.M. input for the demodulator unit two Datapulse Type 401 function generators were employed. These function generators were capable of producing sinusoidal, triangular, or square wave outputs at frequencies from 1Hz to 2MHz. The units also had F.M. inputs whereby one unit could be cascaded to modulate the output from another similar unit. In modifying this F.M. into a form suitable for entry into the demodulator, offsets were employed. A square carrier wave was used with a peak to peak amplitude of +5v and positive offset of 2½v which simulated a pulse train. This was modulated bt audio signals without offsets and yielded the required format.

# RESISTIVITY - TEMPERATURE BEHAVIOUR OF $SnO_2$:B:Sb RESISTOR SPECIES

BY

DANIEL ALBERT LEPIC, B.Sc.

PART B: McMASTER (INDUSTRIAL) PROJECT

A project* report submitted in partial fulfillment of the requirements for the degree of

Master of Engineering

Dept. of Engineering Physics

McMaster University

* One of two required project reports.

PART A: is an "on-campus" project report.

MASTER OF ENGINEERING (1972)
(Engineering Physics)

McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: Resistivity - Temperature Behaviour of $SnO_2$:B:Sb Resistor Species

AUTHOR: D. A. Lepic, B. Sc. (McMaster University)

SUPERVISOR: Dr. J. P. Marton

NUMBER OF PAGES: IV, 38

CONTENTS

## LIST OF TABLES



## LIST OF ILLUSTRATIONS

ACKNOWLEDGEMENT

I would like to express my gratitude to Dr. J. P. Marton for his knowledgeable supervision, guidance and encouragement throughout the course of this report. As well, I am thankful to Mr. J. R. Lemon for providing technical advice and assistance.

ABSTRACT

Investigation of $SnO_2$:B:Sb semiconductor species over the temperature range - 60°C to + 175°C reveals that electrical resistivity in this region is determined by the complex superposition of stable thin film scattering phenomena. Transient effects due to lattice imperfections inherent in the fabrication process start to "anneal" out at temperatures greater than 50°C and can be characterized by an activation energy of the order of .013 eV. Uncompensated samples doped heavily with boron illustrate a trend toward ionized impurity scattering at lower temperatures but mainly the species exhibits a complicated interplay of acoustical and optical phonon scattering modulated by doping level in such a manner as to lower T.C.R. An empirical expression relating resistivity-temperature behaviour to doping is developed.

# INTRODUCTION

The study of tin oxide conduction mechanisms with boron and antimony doping is complicated in a twofold manner. Firstly, $SnO_2$ is a compound semiconductor which in the natural, undoped state exhibits extrinsic n-type behaviour. This behaviour arises from a stoichiometric deficiency of oxygen cations in the $SnO_2$ lattice. The resultant unbalanced distribution of charge gives rise to an effective donor concentration. Any attempt to characterize behaviour of this species by conventional elemental semiconductor laws must certainly include modifications.

Secondly, the tin oxide samples studied here were in thin film ($\simeq$ 1μm) polycrystalline form. On a microscopic scale the crystallite domains have internal order and periodicity; however, on a macroscopic level one has only pseudo-periodicity and must be very cautious in attempting to explain resistivity behaviour without taking this into account. Because of the dependence of crystallite structure and size on film thickness, one expects thin film phenomena to assume a prominent role in the determination of resistivity behaviour of the samples studied.

In general, when examining the resistivity-temperature behaviour of doped, thin film semiconductors one must consider:

(I) excess charge carrier effects

(II) scattering effects

(III) film thickness effects

and the relative temperature behaviour of these. Since the intrinsic energy gap of undoped $SnO_2$ is of the order of 3.4 eV (cf. Arai [1], Kohnke [2]) one would not expect intrinsic behaviour until much higher temperatures (>300°C), certainly not in the range considered in this report. As well, boron and antimony form fairly shallow doping levels in $SnO_2$. One is justified in assuming that these dopants are either completely ionized in the temperature range of interest or at least compensated to the extent that they contribute little net effect to the "excess" charge concentration. In fact, with antimony doping alone in $SnO_2$ one has complete ionization of its donor level at 77°K, (cf. Imai [3]). It shall be assumed, therefore, that for the temperature range of interest, - 60°C to +175°C, because of compensation or impurity ionization, excess charge carrier effects will be negligible.

Apparently then, for the temperature region in this report, the resistivity properties of $SnO_2$:B:Sb species are determined primarily by scattering and thin film phenomena and their effect on mobility. According to Matthiessen's rule, the relevant thin film phenomena include scattering from film boundaries and scattering from imperfections. Boundary scattering is a function of film thickness and usually assumes importance when the mean free path of the charge carriers is of the same order of magnitude, as the film thickness. Imperfection scattering depends directly on the imperfection density which again, in turn, is a sensitive function of film size.

At the onset, then, it appears that some or all of the following factors determine the resistivity-temperature behaviour of $SnO_2$:B:Sb species:

(I) ionized impurity scattering

(II) acoustical phonon lattice scattering

(III) optical phonon lattice scattering

(IV) scattering from imperfections (i.e. vacancies, grain boundaries, dislocations, crystal agglomerates)

(V) scattering from film boundaries

It is the purpose of this report to examine resistivity behaviour of $SnO_2$:B:Sb as a function of temperature, doping and film thickness in the light of the above factors, to deduce which ones assume precedence in determining T.C.R. behaviour over the temperature range $-60^{o}C$ to $+175^{o}C$. As a consequence of this investigation, it is anticipated that conditions will be established to lower the present accepted factory standard T.C. rating of ±200 ppm. to a more ideal rating of approximately ±50 ppm.

## APPARATUS AND SAMPLE FABRICATION

### (a) Sample Description

The samples employed for analysis were in the form of thin tin oxide films on cylindrical ceramic substrates. The cylindrical ceramic substrates were employed for basically two reasons - they are chemically inert with $SnO_2$ and the cylindrical shape is the one commonly used in production of $SnO_2$ resistors for industrial consumption. The cylinders were approximately 1" long by 1/4" in diameter. Samples were classed into four different resistance groups (grades) of approximate resistances of 14, 20, 28, 40 ohms (each group differing by a factor of $2^{1/2}$. The choice was somewhat arbitrary - industrial resistors, before being spiralled to precision, come out of the preparation furnaces with "pre-value" resistances of 10 to 100Ω and 10 to 50Ω is the most common range. The four different resistor groups were chosen to reflect any film thickness effects that might have appeared, since for similar dopings different resistance grades are produced by varying film thickness.

Throughout this report, doping is referred to in terms of mole percent of $SnO_2$. A 4% boron and .6% antimony doping would mean that the firing solution contained .04 moles of boron and .006 moles of antimony for every mole of $SnO_2$ assumed formed. In terms of number density 4%B, .6% Sb would mean, assuming complete incorporation of boron and antimony into the $SnO_2$ lattice, about $10^{20}$ atoms/cm$^3$ of dopant.

The $SnO_2$ polycrystalline film was prepared from the chemical hydrolysis of $SnCl_4$ in the presence of HCl at high temperatures. The relevant reaction is:-

$$SnCl_4 + 2H_2O \rightleftharpoons SnO_2 + 4HCl \quad \text{(endothermic)}$$

This is a reversible reaction and varying parameters during "firing" (i.e. pressure, temperature) will undoubtedly result in different "initial" sample structure and composition. As it is primarily the purpose of this report to examine $SnO_2$:B:Sb resistivity phenomena, physical firing parameters were maintained constant for all samples. The only chemical quantity that was altered was the amount of $H_3BO_3$ or $SbCl_5$ added to the firing solution for the boron and antimony doping respectively. A standard firing solution recipe is shown in Table 1.

(b) Sample Preparation

The samples were prepared in an experimental furnace modelled after the standard factory furnaces. Fig. 1a is a sketch of the apparatus used in preparation of the samples. Basically the furnace consisted of coil-type heating elements wrapped in helical fashion around a cylindrical oven core. Insulating material made of ceramic wool enclosed the oven within a rectangular asbestos fibreboard box of dimensions 1' by 1 1/2' by 6'.

Quartz glass rods were inserted into the cylindrical oven chamber and connected to a rotary motor.

As seen in Fig. 1a the rods were discontinuous in the firing chamber but were positioned such that the cylindrical ceramic substrates could pass from one into the other without falling out or getting stuck.

The ceramic substrates were propelled along the quartz-glass tube by a miniature conveyor system at the entrance and after entry into the oven they would be pushed forward by the successive entry of other substrates. The combined rotation of the tube and linear progression of the conveyor belt resulted in a helical progression of the resistor substrates towards the firing chamber. From the rotation speed of the quartz glass tube and speed of the conveyor belt, it was estimated that the substrates underwent about 10 revolutions per inch and moved forward at a speed of 6 inches per minute. These settings seemed to give good uniform films.

The furnace voltage was set to give 1050°C temperatures; however, the actual temperature of the substrates upon reaching the firing chamber appeared to be 850°C ± 10°C as measured by an optical pyrometer calibrated with a thermo-couple.

In brief then, the ceramic substrates would enter the furnace by a conveyor belt and be helically propelled to the firing chamber where they would be sprayed with the firing vapour. At the substrate surface temperature of 850°C the solution would chemically react upon contact with it and coat the substrate through several rotations with a uniform $SnO_2$ film. The resistors would then move through a cooler part of the furnace to an exit point.

Fig. 1b is a sketch of the firing nozzle. Basically it consisted of a heating coil surrounding a 1" diameter copper pipe to vaporize the firing solution and a 5/16" nozzle exit. Before the solution entered the nozzle to be vaporized, it was isolated from the heat of the vaporizing coils by a coolant system. A long thin "drip pin" was inserted into the firing nozzle chamber to prevent solution from precipitating or clogging by allowing a smoother flow of the solution into the nozzle heating chamber. The solution would flow into the nozzle chamber, along the "drip pin", be vaporized by the heating coils ($\approx 350^{\circ}C$) and then blown out through the nozzle exit onto the hot ceramic substrate. The rate of flow of the firing solution into the nozzle chamber was controlled by a fine screw adjustment, and the flow rate of the firing vapour out of the nozzle exit was controlled by the air pressure blown into the nozzle chamber. A small exhaust positioned at the far side of the furnace opposite the nozzle exit controlled any vapours that did not coat the substrate. Flow rates of 1 to 5 cc./min. were used for P.V. resistances of 14-40Ω (i.e. film thicknesses of 1.00 to 2.50 μm).

## OBSERVATIONS AND INTERPRETATION OF RESULTS

Throughout this report it will be assumed that resistance is proportional to resistivity and, since all substrates were of the same geometry, it will be further assumed that the constants of proportionality for all samples are approximately the same. Thus one can assume that resistivity changes with temperature, doping, or film thickness are reflected in a proportionate manner through resistance changes.

### (a) Transient Effects

The resistance versus temperature characteristics of typical low doped $SnO_2$:B:Sb resistor samples for different resistor grades (i.e. different film thicknesses) are shown in Fig. 2. For the 14Ω and 20Ω groups the behaviour appears similar to that of a typical elemental semiconductor - the resistance increases in a parabolic fashion with temperature similar to the acoustical scattering $T^{3/2}$ law. For the higher grade, 40Ω group, the resistance behaviour appears rather anomalous, reflecting some internal activation that causes a sharp decrease in resistance for temperatures greater than 50°C. In Fig. 3 to Fig.5, R versus T curves are shown for the four resistance groups and various dopings up to 9% B, 1.35% Sb. In each case one appears to have an anomalous resistance decrease around 50°C for the higher grade (thinner film) samples and furthermore, the effect seems to extend to lower grades (resistances) as one employs heavier doping.

In an attempt to isolate the mechanism responsible for this anomalous resistance decrease, activation graphs of

$\log_e(1/R)$ versus $1/T$ were drawn to see if one could assign a relevant activation energy to the process. For the lower doped, lower resistance samples, a pure activation energy operable over the significant range could not be extracted; however, for the seemingly more extreme cases (heavily doped, high (28, 40Ω) resistances) a linear activation region was definitely attained. Fig. 6 is a typical plot of what was obtained and in Table 2 are listed results for several other samples that were measured. From the Table, it appears that an activation energy of .013 eV could be assigned to describe the anomalous resistance decrease. Since this is of the same order as the activation energy for grain boundary annealing, it is suspected that the observed resistance decrease actually represented the removal of these imperfections from the film, either by annealing, lattice re-arrangement, chemi-sorption or some other thermally dependent mechanism.

To check the reversibility of this anomalous resistance decrease, thermal hysteresis was examined. Several selected samples of various resistance groups and dopings were measured from -60°C to +175°C, allowed to cool to room temperature, then remeasured again over the temperature range, a few days later. Each sequence, from room temperature, through the -60 C°to +175°C temperature range was termed a cycle. Fig. 7 to Fig. 10 illustrate the results obtained for 14, 20, 28 and 40Ω resistance groups respectively, of the same doping (9% B, 1.35% Sb) and for two successive cycles. Fig. 9 shows the results for the 28Ω resistance group and three cycles of measurement.

If one assumes that the parabolic R-T shape represents the stable state of the resistor - the state with the minimum of imperfections - and that the observed anomalous behaviour is merely a manifestation of a sequence of transient R-T states caused by the successive removal of imperfections from the film, then one can explain this hysteresis phenomenon very satisfactorily. Apparently, as was expected, the number of imperfections depends quite strongly on film thickness. For thicker film samples, (low resistance groups) such as the 14Ω sample in Fig. 7, the imperfection concentration is low enough, so that its removal is masked by stable resistivity phenomena. For thinner films the imperfection concentration becomes large enough so that its effect is no longer masked by stable resistivity mechanisms and one observes transient R-T states. With successive heat treatment, the samples tend towards more stable behaviour as is evidenced by the hysteresis graphs. A hypothetical representation of these transient states is sketched in Fig. 11.

At this point one should note that the second cycle for the 28Ω sample (Fig. 9) seems to exemplify a more stable form than that observed after the third cycle. This does not necessarily violate the laws of irreversibility but more likely is a manifestation of the different cooling procedures employed in bringing the samples back to room temperature. It would appear that imperfections can be re-introduced by the cooling rate and/or method. The only difference in the cooling procedures for cycle two and cycle three was that before cycle three the samples were cooled

quickly (removed from T-C test oven, left at room temperature), before cycle two the samples were cooled much more slowly (left to cool down in oven). This phenomenon was further verified using several other samples and, in all cases, rapid cooling or quenching left the samples in a less stable state than slow cooling.

In summary of this imperfection-induced transient phenomenon it appears from the hysteresis curves, particularly Fig. 9, that for low temperatures the effect of imperfections on resistivity (resistance) is linearly superimposed on the stable resistivity processes. Because the R-T curve appears to be undistorted, for temperatures less than 50°C, by the imperfection effect (merely translated to higher resistance values) it appears that the effect is temperature independent in this region. At approximately 50°C thermal energy becomes sufficient to initiate the removal of imperfections by annealing or lattice rearrangement. This exponential removal continues at a rate characterized by activation energy .013 eV until the effect is once again masked by stable resistivity processes. If the proper cooling technique is employed at this point, then one should obtain the final stable form of the resistor with no further detectable hysteresis in the operating temperature range. If the samples are improperly cooled, some imperfections will again be frozen into the lattice and the sample will still be in only a pseudo-stable state.

The observation that imperfection dominated behaviour occurs "more readily" for heavily doped and/or higher resistance (thinner film) samples implies that both doping and film

thickness affect the imperfection concentration. As regards impurity dependence, one can easily rationalize that the greater the density of impurity atoms forced interstitially or substitutionally into the regular lattice structure, the more the regular lattice potential is distorted. A distorted lattice potential is certainly more amenable to imperfections than a regular one.

(b) Structural (Thin Film) Effects

In order to examine structural effects due to the thin film nature of the resistor samples, volume resistivity ($\rho_v$) values were calculated. These resistivity values were computed by an elementary technique using the weight of the film, determined by precise measurements on a micro-balance, to estimate film volume. The technique basically involved measuring the resistance of the samples, weighing them, removing the film by chemical means, and re-weighing the resultant "barren" substrates. The difference in weight was taken to be equal to the weight of the film and proportionate to its volume. Using an empirical value for the density of $SnO_2$ as proportionally constant, the volume of the film was calculated and $\rho_v$ determined. As well, using the calculated surface area of the resistor substrate, a value for the film thickness (d) was obtained. These thickness measurements were checked and verified to be within the desired accuracy by ellipsometric techniques.

The main source of error in the above technique would apparently lie in the assumption of:-

- uniform film density
- total film removal
- no chemical interaction with substrate

However, these assumptions were independently examined and found valid within the accuracy range desired.

Using $\rho_v$ measurements obtained in the above manner, a graph of $\rho_v$ versus $^1/_d$ with doping as parameter was plotted and is shown in Fig. 12. As mentioned in the INTRODUCTION, the $SnO_2$ samples studied were in polycrystalline, thin film form. According, to Matthiessen's rule and subsequent works of Fuchs [4] such thin films involve diffuse boundary layer scattering and thus should exhibit a functional dependence of $\rho_v$ on film thickness (since thickness is the only dimension of the same order of magnitude as the carrier mean free path). The curves in Fig. 12 indeed illustrate a functional dependence, with $\rho_v$ decreasing in an apparently linear fashion with film thickness. In other words, as anticipated, the boundary layer scattering effect on mobility causes the resistivity to increase as the film becomes thinner.

Fuchs has developed an expression relating conductivity and electron mean free path to film thickness. The general form of the expression is rather complex, but it can be simplified in limiting cases. Assuming that one is dealing with a relatively thick "thin film", Fuchs cites the following expression as an approximation to the resistivity behaviour:

(1) $$^{\rho_v}/_{\rho_{vo}} = 1 + {}^3/_{8\eta} \qquad \eta \geq 1$$

where $\eta = d/\ell$, "$\ell$" is the mean free path of an electron in the bulk, "d" film thickness, and $\rho_{vo}$ is bulk resistivity. One can re-write (1) in the form:

$$(2) \quad \rho_v/\rho_{vo} = 1 + \frac{3\ell}{8}\left\{1/d\right\}$$

This is basically similar behaviour to that observed in Fig. 12 where $\rho_{vo}$ values appear to be about 100 μΩ-cm and the electron mean free path by comparison with the above expression is of the order of a few microns. Since the film thicknesses of the resistor samples are also of the order of a few microns, the assumption of boundary layer scattering appears consistent. Only order of magnitude quantities are cited in the above since it appears one is on the verge of thick film effects and linear extrapolation to obtain $\rho_{vo}$ is not entirely justified. In fact, the above quantities were obtained for the 4%B, .4%Sb case, since one could not obtain positive $\rho_{vo}$ values through linear extrapolation in the other cases.

(c) Effect of Doping

The main purpose of doping the $SnO_2$ species with two different types of dopants and to such an extent as to make the semiconductor partially degenerate, was in seeking a means to lower the temperature co-efficient of resistivity over the operating range of the commercial resistor products.

The resistivity behaviour of undoped or lightly doped $SnO_2$ over these temperature ranges seems to involve both acoustical and optical phonon lattice scattering [5] and it was believed that very heavy doping could effectively

dampen or distort these thermally induced oscillations to such an extent as to be reflected in a lower T.C. Two dopant types were employed so that compensation effects would essentially cancel any gross "excess" charge carrier temperature behaviour. Doping to degeneracy actually served a two-fold purpose - it made the mobility even less sensitive to excess charge carrier temperature effects [6] and the larger impurity concentration would more effectively burden lattice motion. In respect to optical phonon scattering, which would seem to be the more important of the lattice scattering moles, since $SnO_2$ is a polar semiconductor, the heavy doping concentration would also effectively shield the ion cores or distort the polar homogeneity in such a way as to lower the optical coupling constant. This would certainly weaken the functional dependence of this scattering mode on temperature.

Fig. 13 is a graph of resistance versus temperature for the 20Ω resistor group with doping as parameter. A decrease in T.C. appears observable and apparently dependent on doping level. In Fig. 14 selected samples from Fig. 13 have been replotted in $\log_e ({}^1/_R)$ versus $\log_e (T)$ format in an attempt to isolate any power law relation. Four samples are shown plotted in this format with dopings ranging from 4% B, .27% Sb to 9% B, .9% Sb. In each case one can isolate a linear region over the temperature range 25°C to 125°C, indicating a resistivity-temperature relation of the form:-

(3) $$\rho_v = \rho_o T^x$$

where the exponent appears to be some function of doping.

In Table 3 values of the exponent 'x' for these samples and several other 20Ω resistance group samples are listed. Also listed are values for several 14Ω group samples. In Fig. 15 these temperature exponent values have been plotted against mole doping product for both the 14Ω and 20Ω resistance groups.

From Fig. 15 one can extract basically two things. Foremost, one has quantitative verification of the temperature co-efficient dependence on doping since the temperature exponent actually reflects T.C. behaviour. Secondly, the functional relation between temperature exponent and doping, though not altered in form by film thickness, is altered in degree. Apparently, as one would expect, thin film phenomena play an active underlying role in net resistivity temperature behaviour.

It appears from Fig. 15 that the doping effect depends only on the mole product of the two dopant species and is basically independent of any discrete weighting factors. In fact, samples of 6% B, .4% Sb and 9% B, .4% Sb have the same temperature exponent and hence temperature behaviour over this 100$^{o}$C range as the 4% B, .6% Sb and 6% B, .6% Sb samples, respectively.

For light doping the curves in Fig. 15 increase asymptotically, probably to some value of the exponent characterizing acoustical ($x \simeq 1.50$) scattering, optical scattering ($x \simeq 2.50$) or some non-linear combination of these, that assumes importance in the undoped species. As the doping increases, the temperature exponent in Fig. 15, decreases rapidly for both resistance grades to a minimum

value at a mole doping product of about 6.00%, at which point it becomes independent of doping.

If one plots temperature exponent against the inverse of the mole doping product, one obtains the linear relation shown in Fig. 16. From this graph one can determine the limiting minimum values of the temperature exponent that can be obtained by doping. For the 20Ω resistance group it is .038 indicating an optimum behaviour of $T^{.015}$ and $T.^{038}$ respectively, for the two sample groups. Since the slope of the two curves is the same ($\simeq$ .118 mole percent $^{-1}$) one can characterize the resistance behaviour over the temperature range 25°C to 125°C by the expression below:-

(4) $$R = R_o T^{G(d)} \; T^{F(B,Sb)}$$

where $G(d) = .015$ for 20Ω group

$= .038$ for 14Ω group

and $$F(B,Sb) = \frac{.118}{[B]\;[Sb]}$$

## CONCLUSION

It appears that the resistivity-temperature behaviour of the stable $SnO_2$:B:Sb species is greatly affected by doping. As was anticipated, one can effectively lower the temperature co-efficient down to a saturation value determined by film thickness effects. One cannot decompose this obtained functional dependence of resistivity on temperature into terms characterized by acoustical or optical phonon scattering, since although qualitatively suggestive of these phenomena, quantitatively the behaviour is quite different. It appears that over the largest part of the desired temperature range one is basically observing an effective damping of the lattice vibrational modes by the heavy incorporation of impurity atoms in interstitial and substitutional positions in the lattice.

From Fig. 14 it seems also that as one approaches lower temperatures ($<25^{\circ}C$) the effect of damping is gradually being overshadowed by some factor tending towards a resistance decrease with lower temperatures. Although one cannot isolate this process, it appears reasonable that it is a manifestation of ionized impurity scattering which, for these lower temperatures, assumes precedence over damping.

As one approaches higher temperatures ($>125^{\circ}C$) it would appear that resistance increases much more rapidly with temperature than that characterized by the damping expression. This is probably indicative of the onset of more typical acoustical and/or optical phonon scattering

behaviour as thermal energy becomes great enough to minimize the effect of lattice damping.

Underlying these scattering modes and the effect of lattice damping, it should be emphasized that structural effects, such as boundary layer scattering also assume much importance in resistivity determination. These effects, however, appear to make more independent contributions to the mobility and seem less temperature dependent than other scattering modes. In the determination of resistivity levels such structural effects are of extreme importance, but as regards temperature variation of the resistivity or T.C.R. they can be effectively overlooked, at least for the range of temperatures and thicknesses considered in this report. Certainly for temperatures high enough to alter the crystal structure they cannot be neglected.

In regard to transient phenomena, it would appear that this behaviour is indeed due to removeable imperfections and very probably grain boundary annealing, since the activation energy of .013eV is of the required order of magnitude. These imperfections are probably incorporated into the lattice by non-uniformities in the cooling process during preparation or perhaps by a differential layered deposition of oxide on the spiralling ceramic substrate during the firing procedure. Nevertheless, elevation of the samples to high temperatures ($\geq 170^{\circ}C$) for a sufficient period of time ($\simeq$ 4 hrs) will remove these imperfections, regardless of origin, and eventually stablize the resistor species.

REFERENCES

1. Arai, T., J. Phys. Soc. Japan 15, 916 (1960).
2. Kohnke, E.E., J. Phys. Chem. Solids 23, 1557 (1962).
3. Imai, I., J. Phys. Soc. Japan 15, 937 (1960).
4. Fuchs, K., Proc. Cambridge Phil. Soc. 34, 100, (1938).
5. Morgan, D. F. and Wright, D. A., Brit. J. Appl. Phys. 17, 336 (1966).
6. Wolf, H.F., Semiconductors, P. 276, Wiley, New York (1971).
7. Matthews, H. E. and Kohnke, E.E., J. Phys. Chem. Solids 29, 653 (1968).
8. Ishiguro, et al., J. Phys. Soc. Japan 13, 296 (1958).
9. Houston, J. E. and Kohnke, E. E., J. Appl. Phys. 36, 3931 (1965).
10. Fonstad et al., J. El. Chem. Soc. 116, 1271 (1969)
11. Scanlon, W. W., "Stoichiometry in Compound Semiconductors", Proc. Metallurg. Soc. Conf., Boston, 1959.

Table 1

| Component | Amount |
|---|---|
| $H_2O$ | 108 cc. |
| HCl | 250 cc. |
| $H_3BO_3$ | 7.42 gm. |
| $SnCl_4$ | 351 cc. |
| $SbCl_5$ | 2.30 cc. |

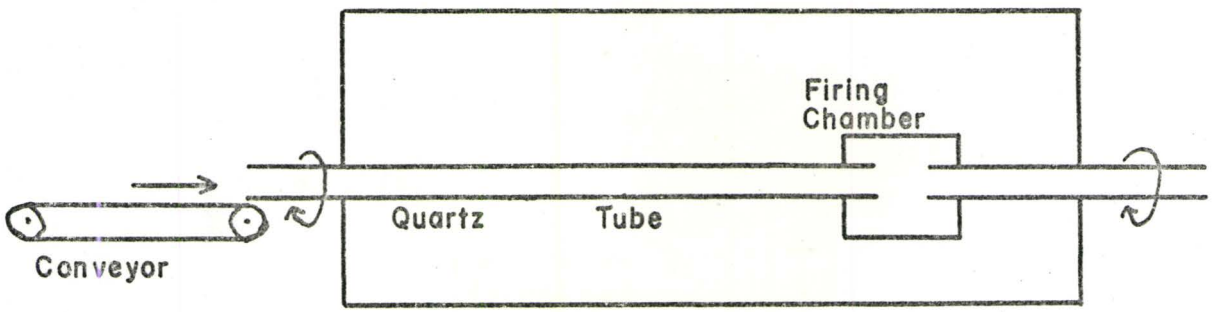
Table 2

| Doping | Activation Energy |
|---|---|
| 6.0%B, .60%Sb-40Ω | .0126 eV. |
| 6.0%B, .90%Sb-20Ω | .0065 eV. |
| 6.0%B, .90%Sb-40Ω | .0132 eV. |
| 9.0%B, .90%Sb-40Ω | .0127 eV. |
| 9.0%B,1.35%Sb-40Ω | .0131 eV. |

Table 3

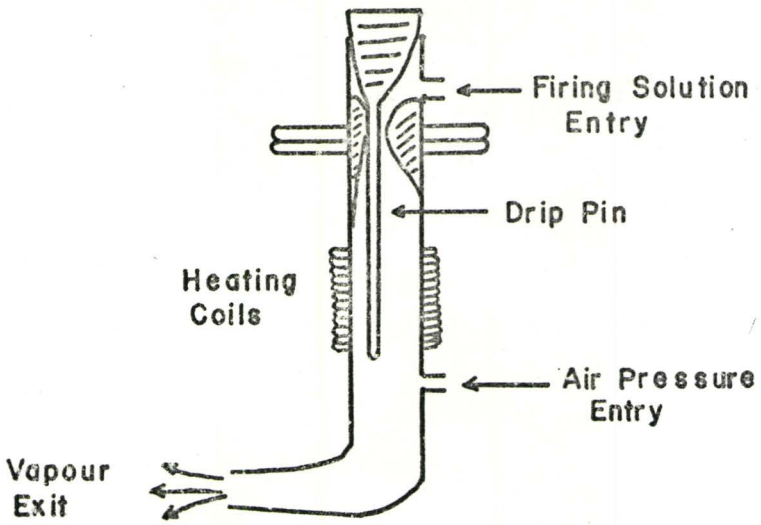
| Doping | Temperature Exponent 14 ohms | 20 ohms |
|---|---|---|
| 2.7%B, .90%Sb | .101 | |
| 4.0%B, .27%Sb | .163 | .118 |
| 4.0%B, .40%Sb | | .097 |
| 4.0%B, .60%Sb | | .063 |
| 6.0%B, .27%Sb | | .086 |
| 6.0%B, .40%Sb | | .063 |
| 6.0%B, .60%Sb | .069 | .035 |
| 6.0%B, .90Sb | .073 | |
| 6.0%B, .40%Sb | | .035 |
| 9.0%B, .90%Sb | .055 | .018 |
| 9.9%B, 1.35%Sb | .038 | .021 |

Furnace Schematic

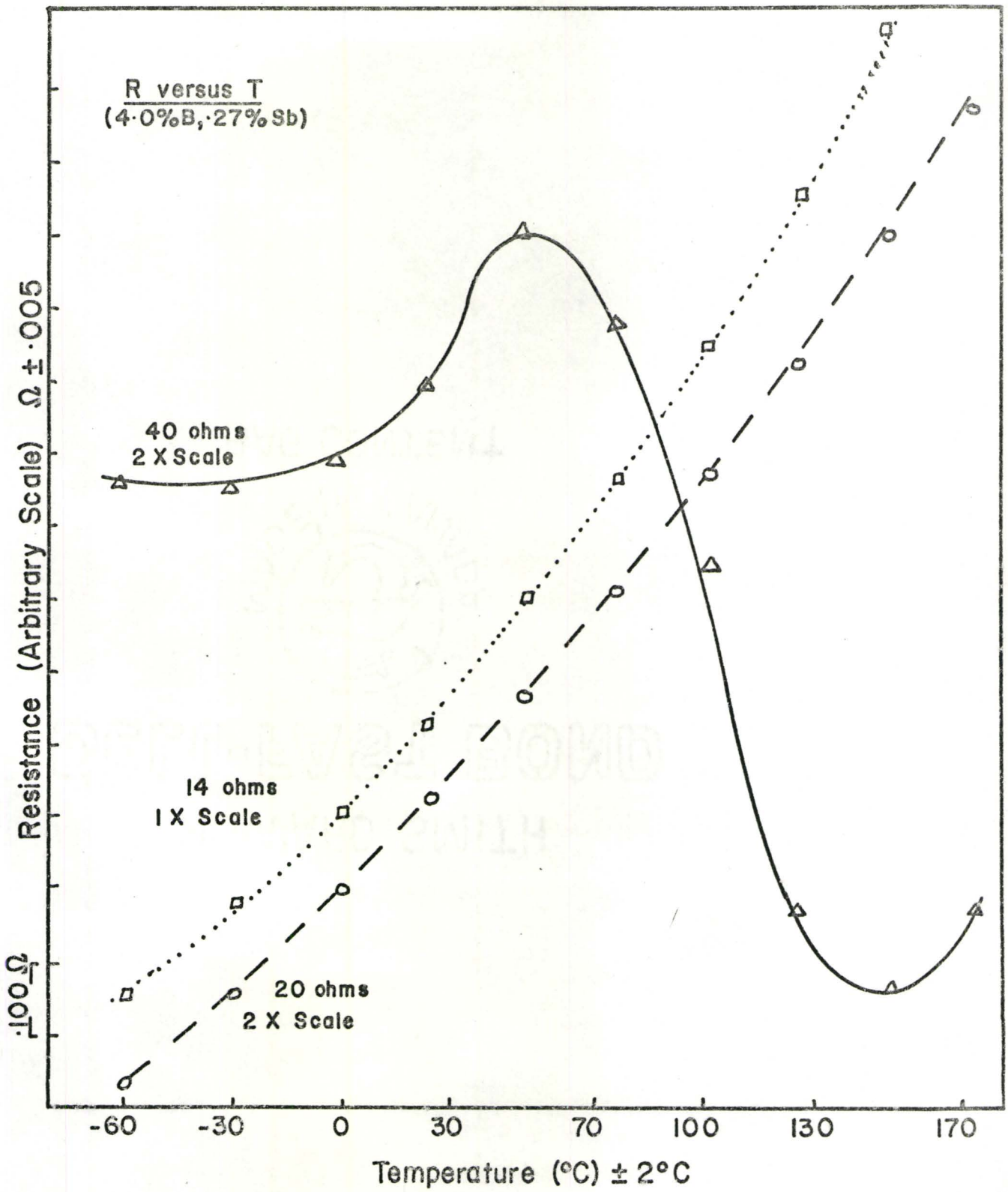



FIG· Ia

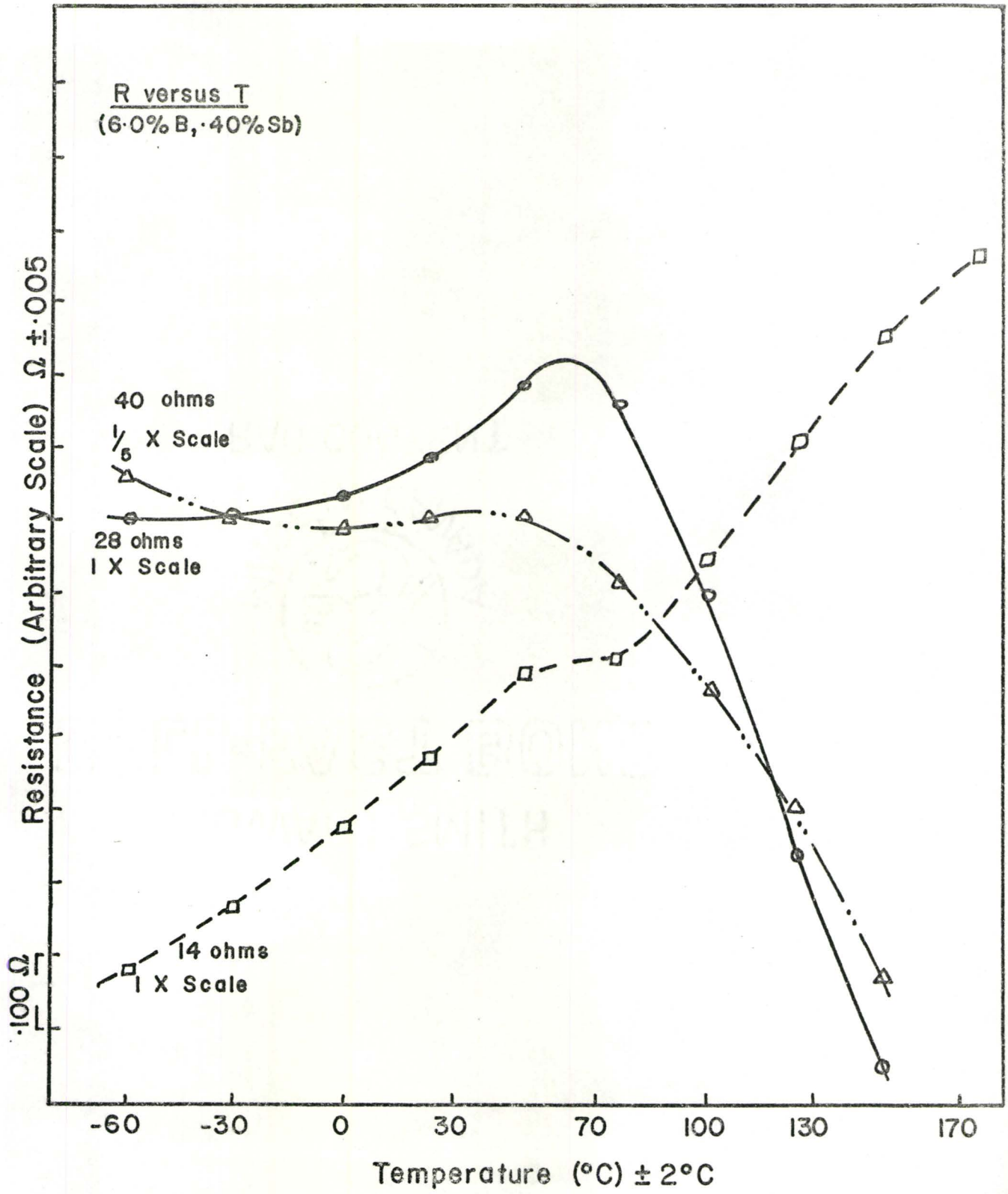
Firing Nozzle



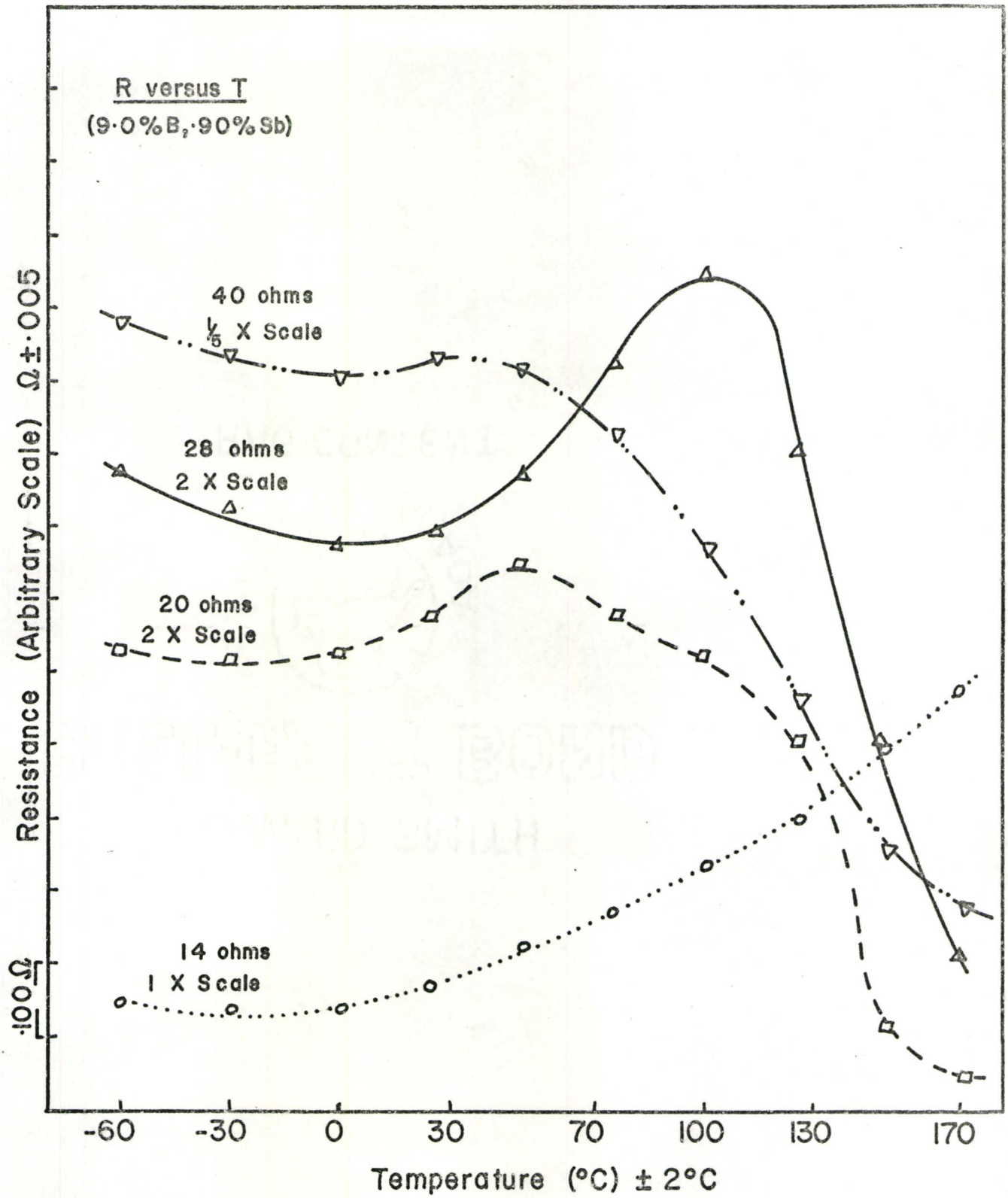

FIG· Ib


R versus T
(4·0%B, ·27% Sb)
40 ohms
2 X Scale
14 ohms
1 X Scale
20 ohms
2 X Scale
Resistance (Arbitrary Scale) Ω ± ·005
±100Ω
-60
-30
0
30
70
100
130
170
Temperature (°C) ± 2°C

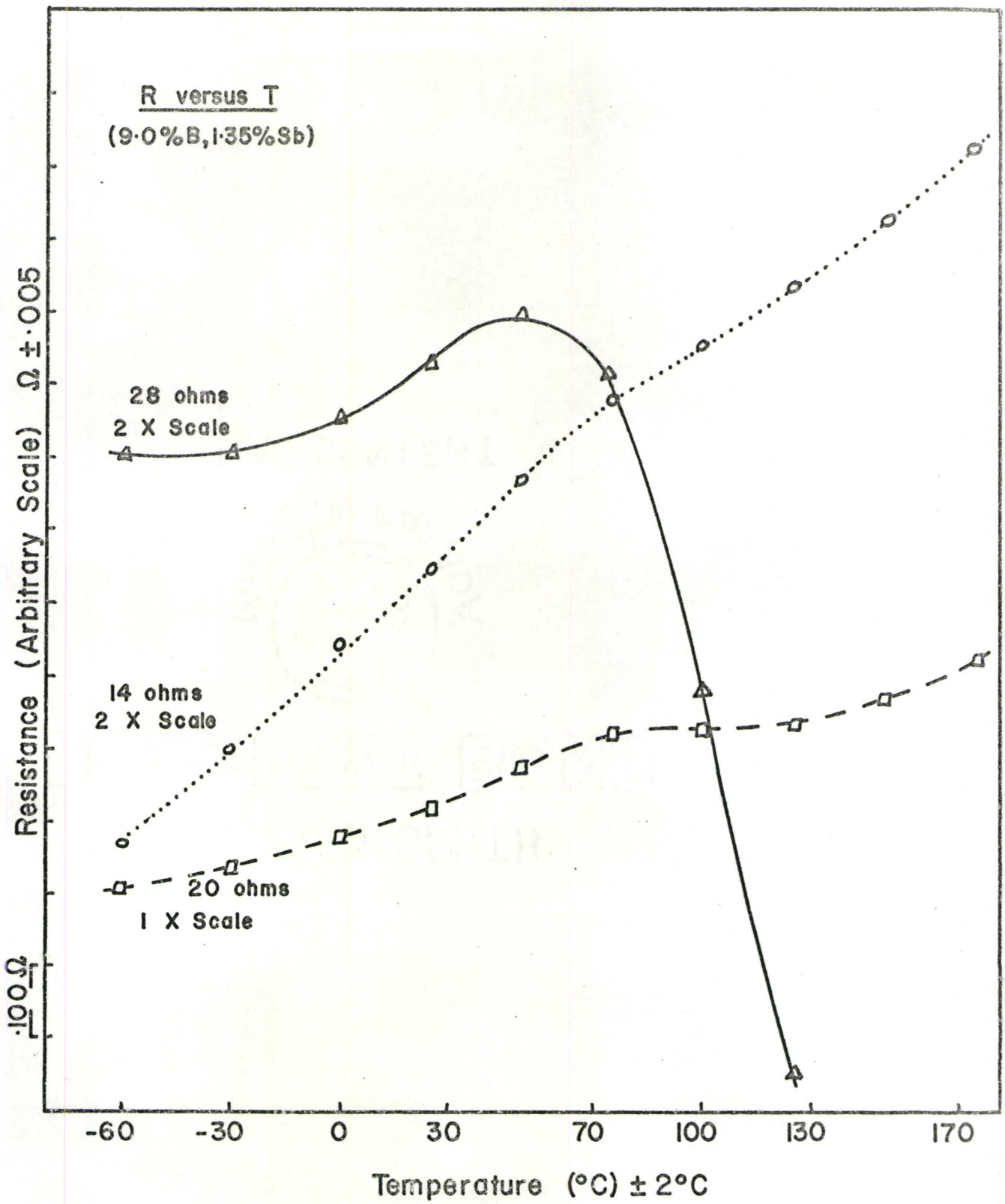
FIG· 2


R versus T
(6·0% B, ·40% Sb)
Resistance (Arbitrary Scale) Ω ±·005
40 ohms
1/5 X Scale
28 ohms
1 X Scale
14 ohms
1 X Scale
·100 Ω
-60
-30
0
30
70
100
130
170
Temperature (°C) ± 2°C

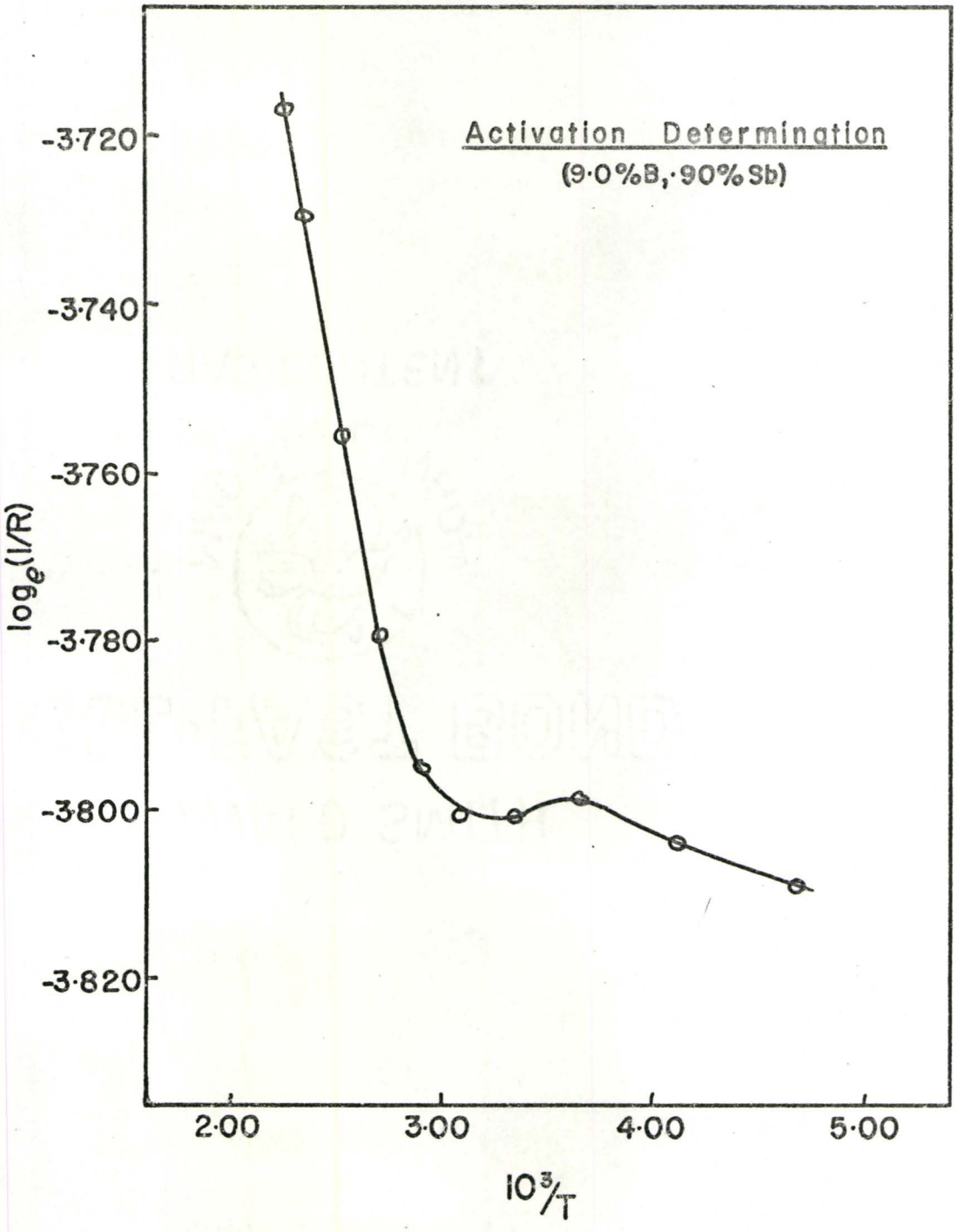
FIG· 3


R versus T
(9·0%B₂·90%Sb)
40 ohms
1/5 X Scale
28 ohms
2 X Scale
20 ohms
2 X Scale
14 ohms
1 X Scale
Resistance (Arbitrary Scale) Ω ± ·005
·100 Ω
-60
-30
0
30
70
100
130
170
Temperature (°C) ± 2°C

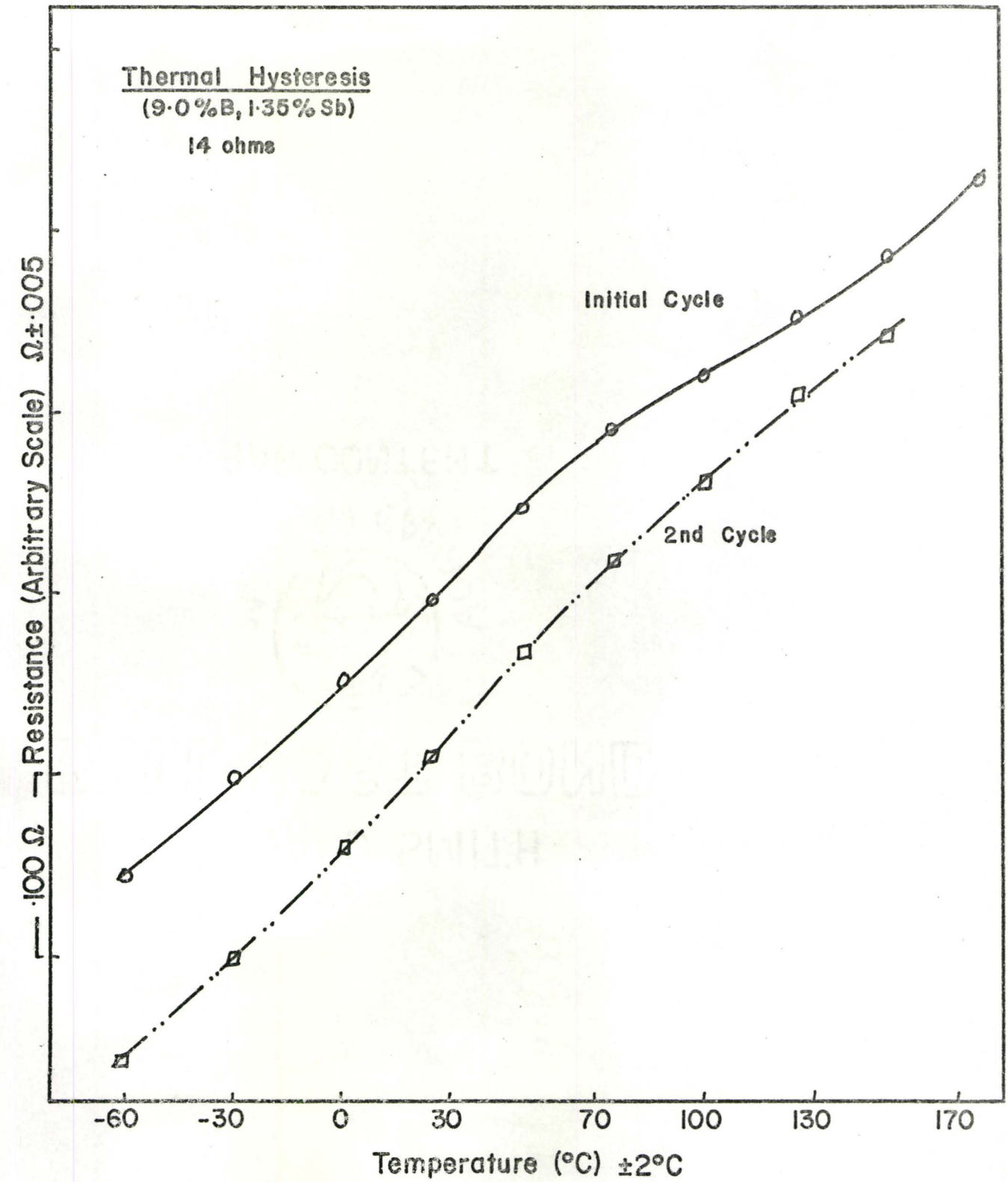
FIG· 4


R versus T
(9·0%B, 1·35%Sb)
28 ohms
2 X Scale
14 ohms
2 X Scale
20 ohms
1 X Scale
Resistance (Arbitrary Scale) Ω ± ·005
·100 Ω
-60
-30
0
30
70
100
130
170
Temperature (°C) ± 2°C

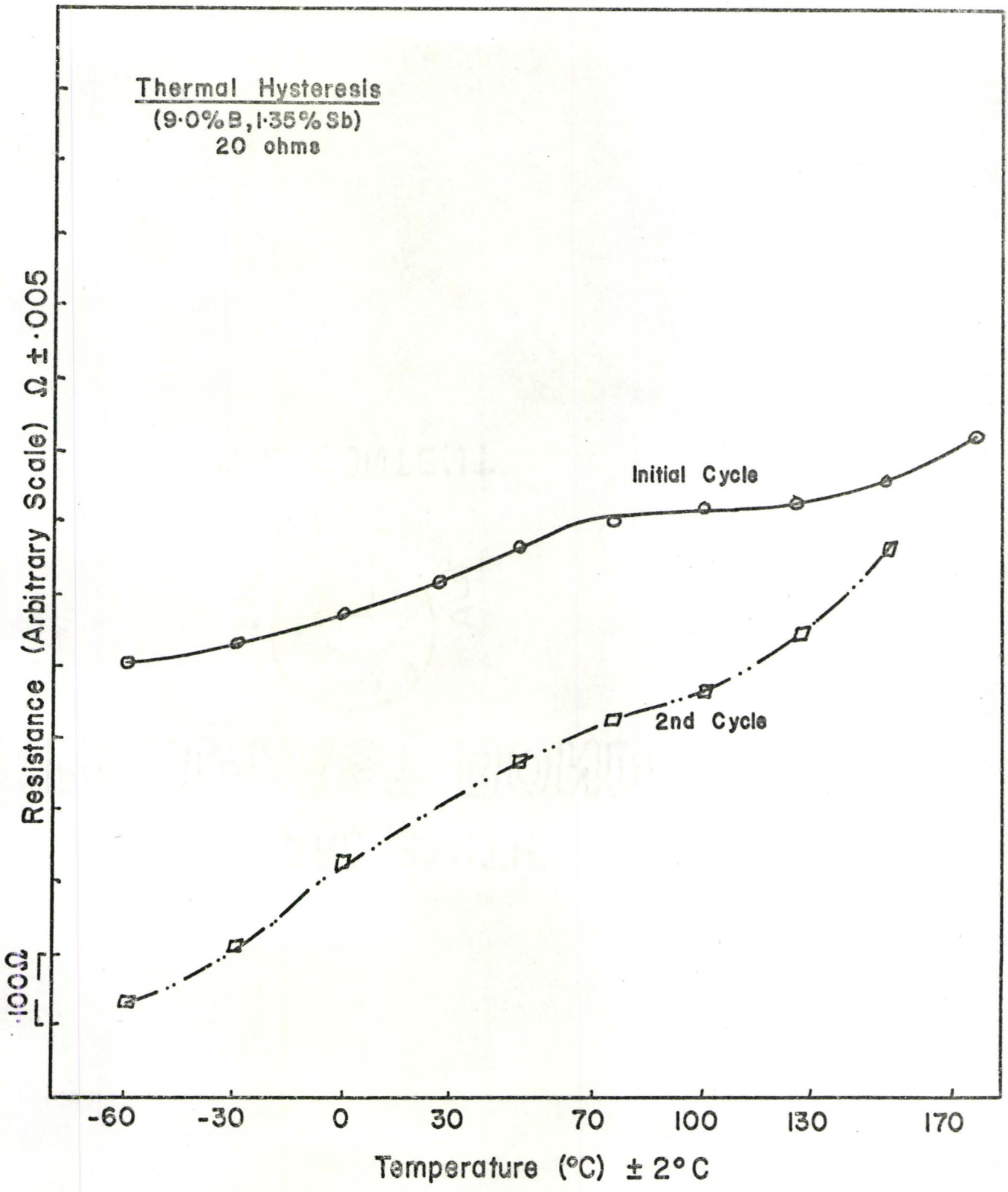
FIG· 5


Activation Determination
(9·0%B,·90% Sb)
-3·720
-3·740
-3·760
-3·780
-3·800
-3·820
$\log_e(I/R)$
2·00
3·00
4·00
5·00
$10^3/T$

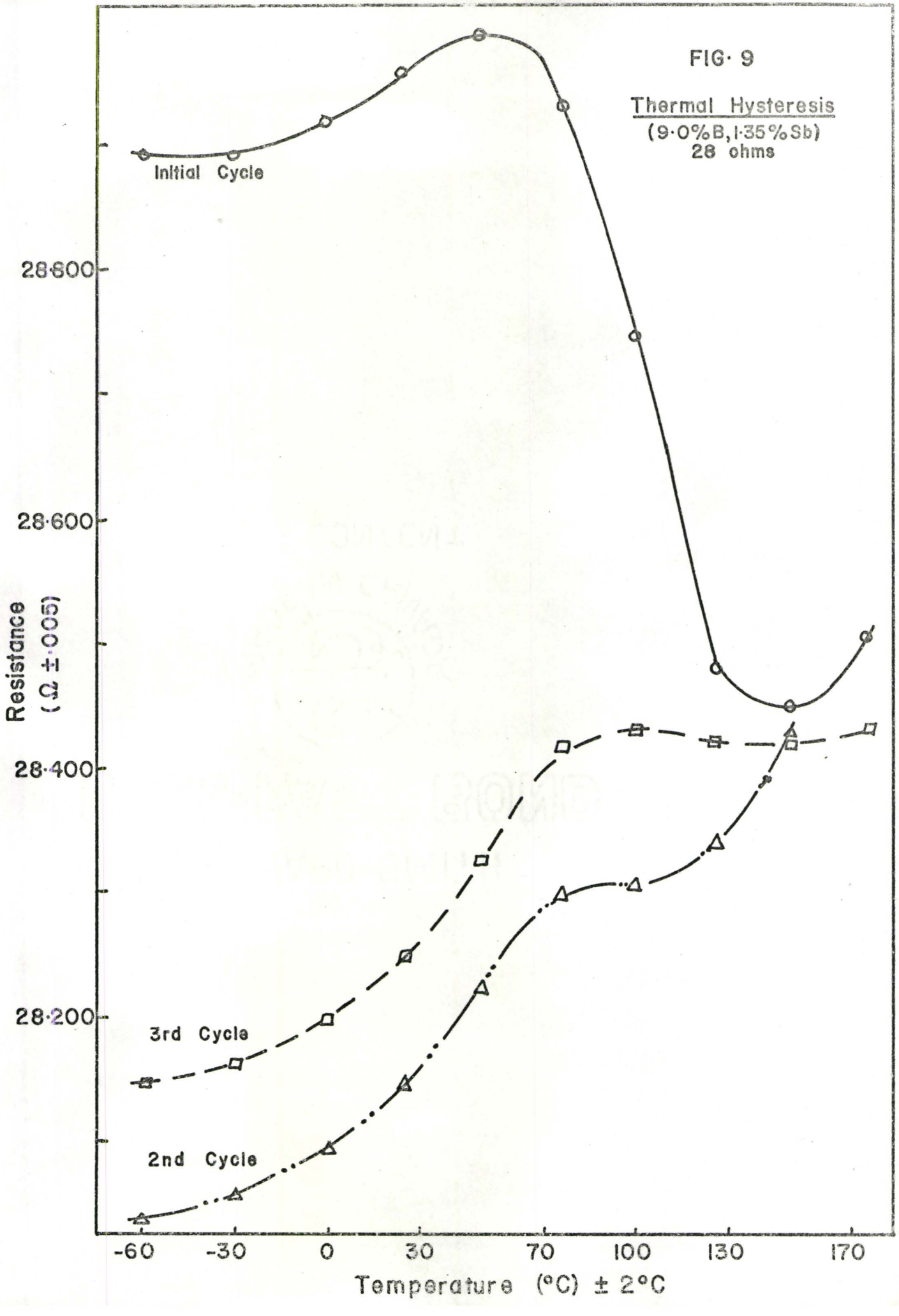
FIG· 6

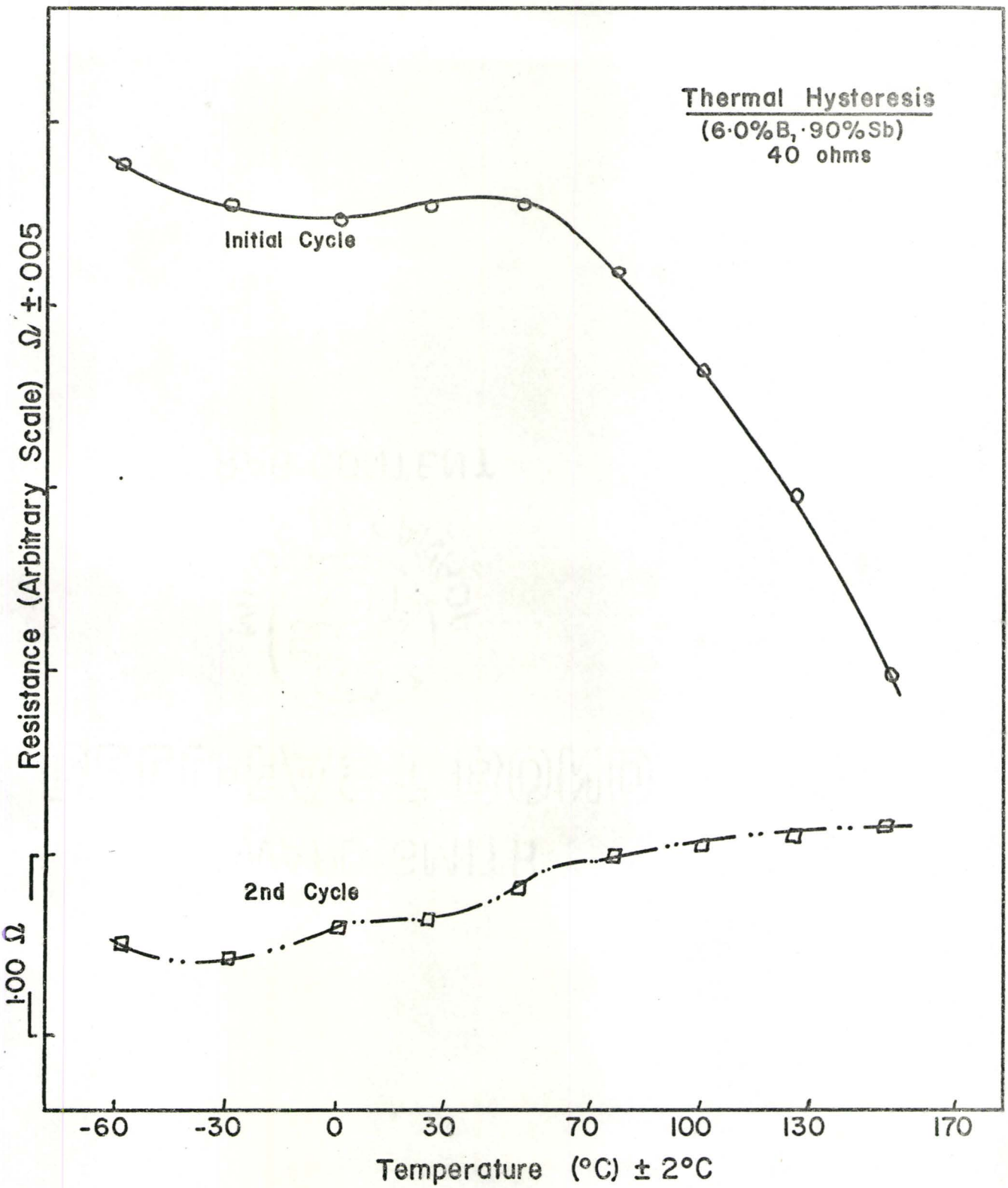

Thermal Hysteresis
(9·0%B, 1·35% Sb)
14 ohms
Initial Cycle
2nd Cycle
100 Ω — Resistance (Arbitrary Scale) Ω±·005
-60
-30
0
30
70
100
130
170
Temperature (°C) ±2°C

FIG· 7


Thermal Hysteresis
(9·0% B, 1·35% Sb)
20 ohms
Initial Cycle
2nd Cycle
Resistance (Arbitrary Scale) Ω ± ·005
·100Ω
-60
-30
0
30
70
100
130
170
Temperature (°C) ± 2° C

FIG· 8


FIG· 9
Thermal Hysteresis
(9·0%B, 1·35%Sb)
28 ohms
Initial Cycle
28·600
28·600
28·400
28·200
Resistance
(Ω ±·005)
3rd Cycle
2nd Cycle
-60
-30
0
30
70
100
130
170
Temperature (°C) ± 2°C


Thermal Hysteresis
(6·0%B, ·90%Sb)
40 ohms
Initial Cycle
2nd Cycle
Resistance (Arbitrary Scale) Ω ±·005
1·00 Ω
-60
-30
0
30
70
100
130
170
Temperature (°C) ± 2°C

FIG· 10

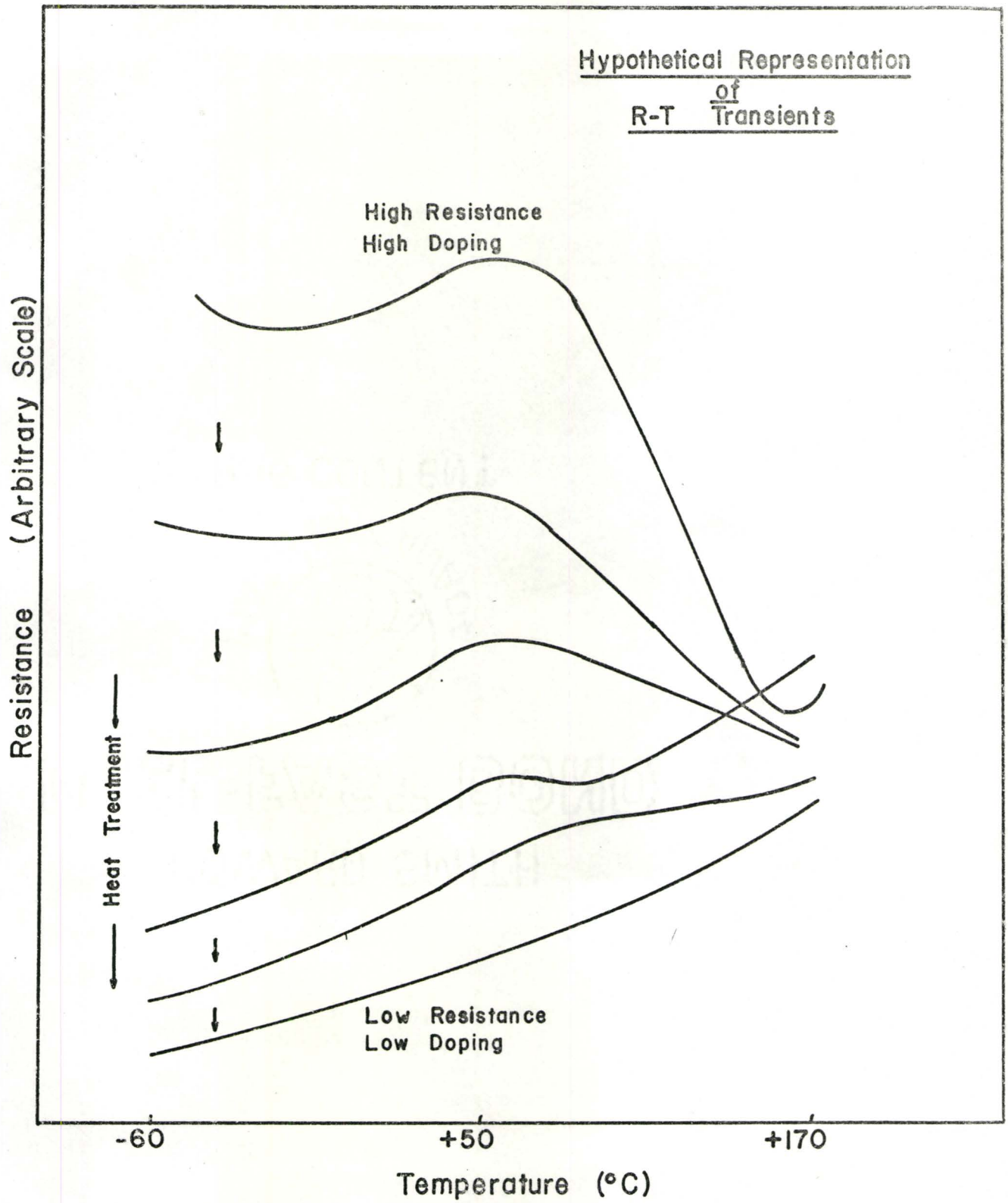
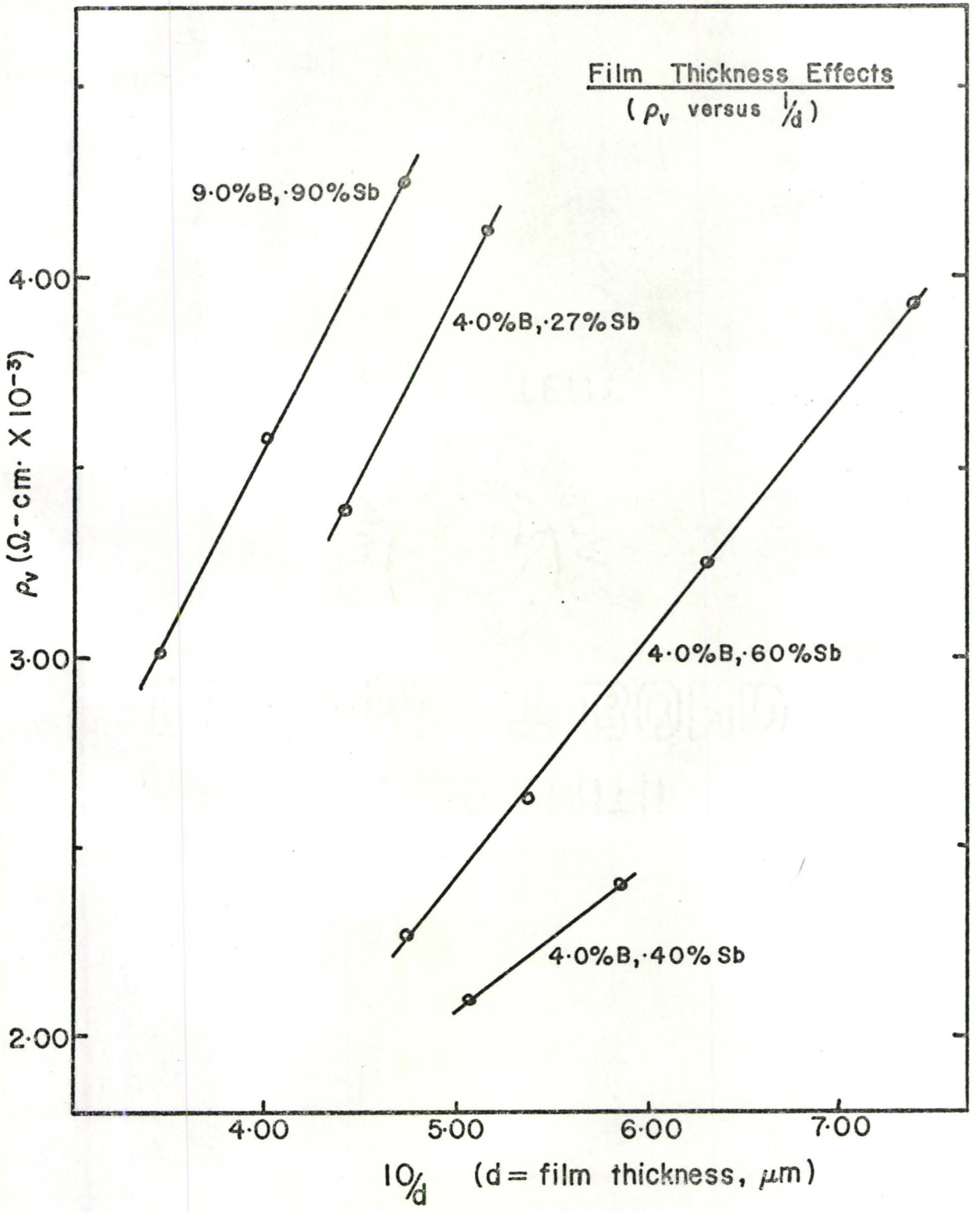
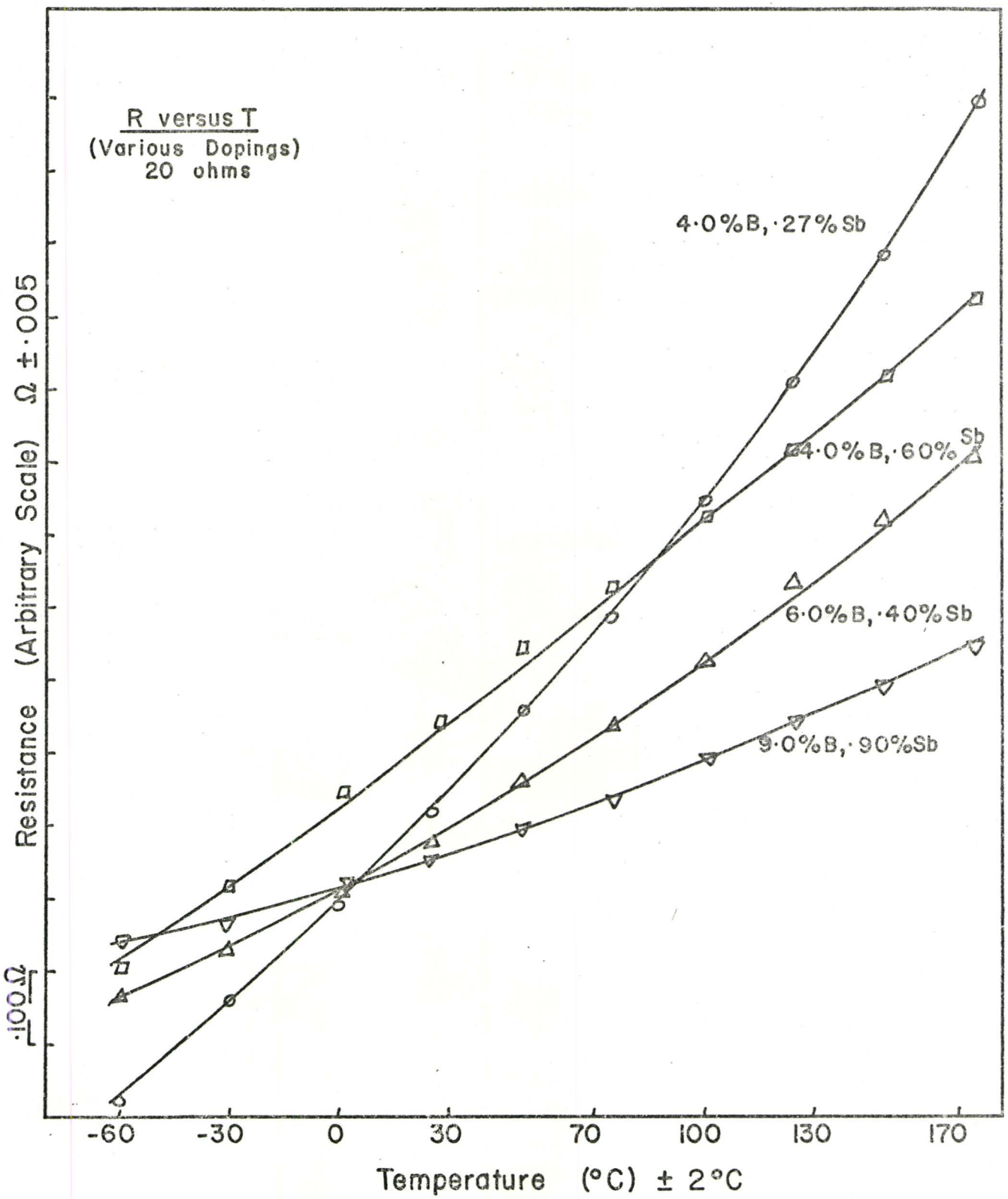

Hypothetical Representation
of
R-T Transients
High Resistance
High Doping
Low Resistance
Low Doping
Heat Treatment
Resistance (Arbitrary Scale)
-60
+50
+170
Temperature (°C)

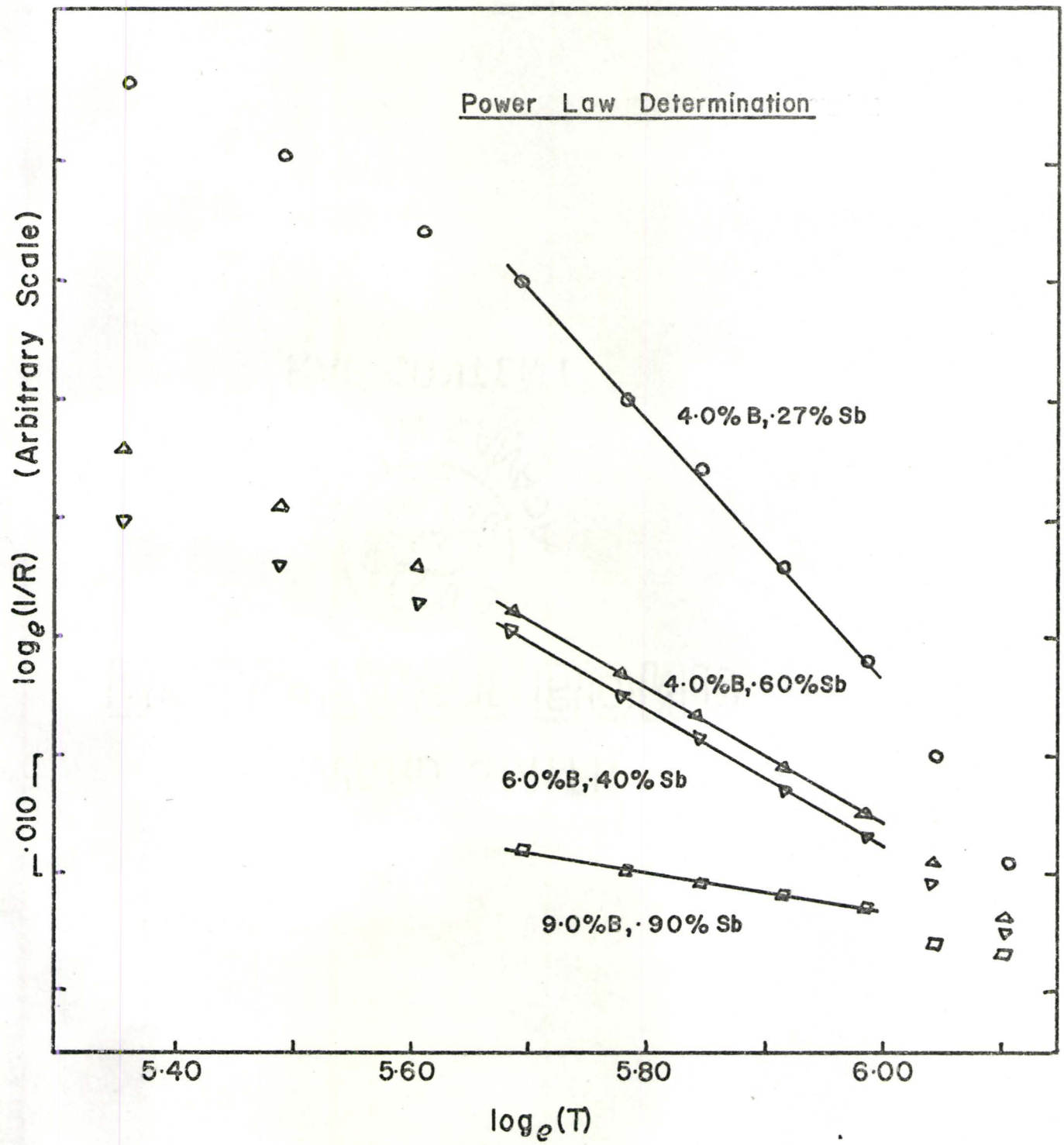
FIG. II


Film Thickness Effects
($\rho_v$ versus $1/d$)
9·0%B,·90%Sb
4·0%B,·27%Sb
4·0%B,·60%Sb
4·0%B,·40% Sb
4·00
3·00
2·00
$\rho_v$ ($\Omega$-cm· X $10^{-3}$)
4·00
5·00
6·00
7·00
$10/d$ (d = film thickness, $\mu$m)

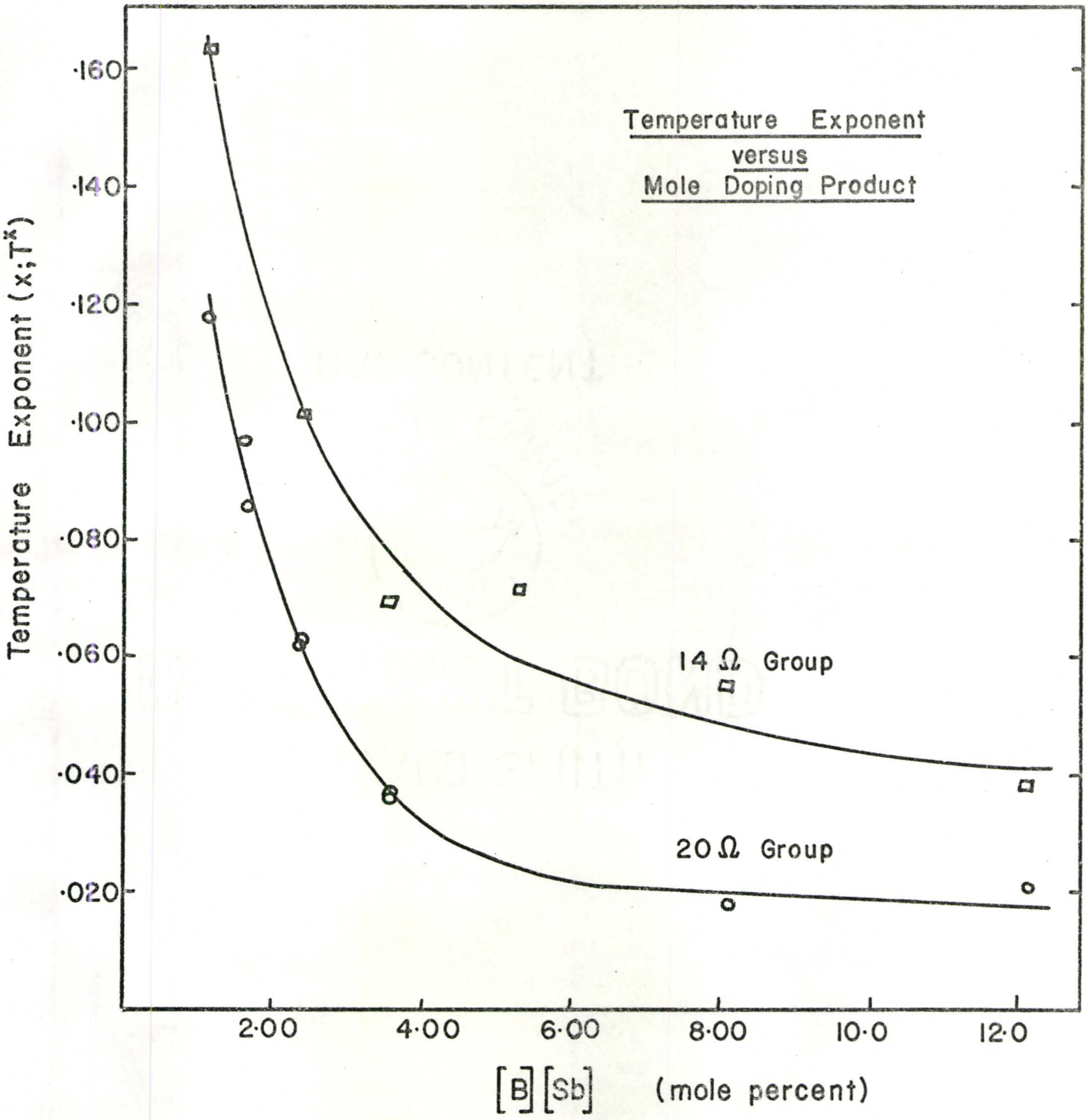
FIG· 12


R versus T
(Various Dopings)
20 ohms
4·0%B, ·27% Sb
4·0%B, ·60% Sb
6·0%B, ·40% Sb
9·0%B, ·90%Sb
Resistance (Arbitrary Scale) Ω ± ·005
·100Ω
-60
-30
0
30
70
100
130
170
Temperature (°C) ± 2°C

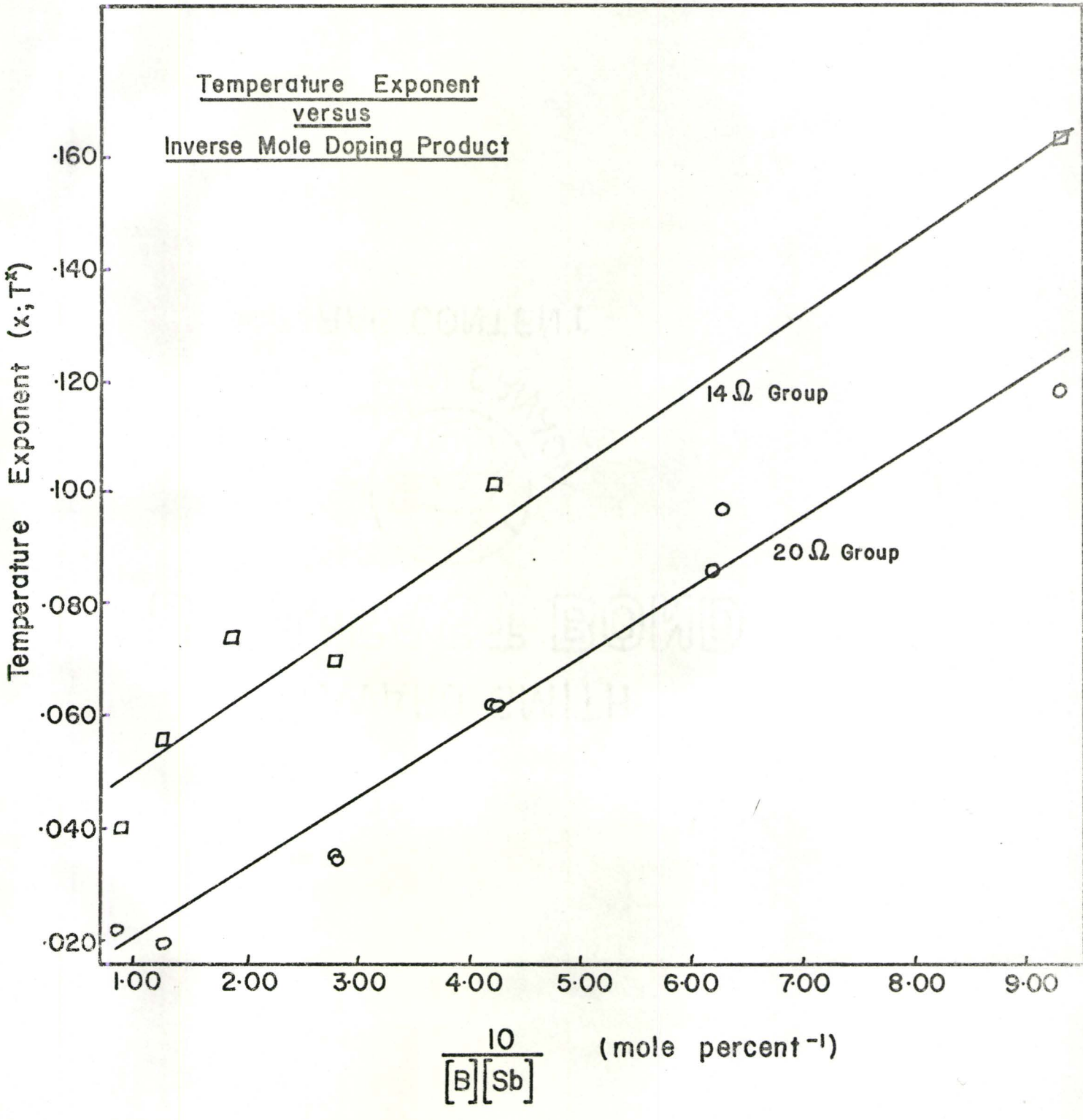
FIG· 13


Power Law Determination
4·0% B, ·27% Sb
4·0%B, ·60%Sb
6·0%B, ·40% Sb
9·0%B, · 90% Sb
log$_e$(I/R) (Arbitrary Scale)
·010
5·40
5·60
5·80
6·00
log$_e$(T)

FIG· 14


Temperature Exponent
versus
Mole Doping Product
Temperature Exponent (x; T^x)
·160
·140
·120
·100
·080
·060
·040
·020
14 Ω Group
20 Ω Group
2·00
4·00
6·00
8·00
10·0
12·0
[B] [Sb] (mole percent)

FIG· 15


Temperature Exponent
versus
Inverse Mole Doping Product
Temperature Exponent (x; T^x)
·160
·140
·120
·100
·080
·060
·040
·020
14 Ω Group
20 Ω Group
1·00
2·00
3·00
4·00
5·00
6·00
7·00
8·00
9·00
10 / [B][Sb]
(mole percent^-1)

FIG· 16