RCA Review

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

Contents

3 **Color Signal Inphase and Quadrature Distortion Measurement and Evaluation**
D. H. Pritchard and T. M. Wagner

33 **Performance of Injection Lasers with External Gratings**
H. S. Sommers, Jr.

60 **The Effects of Emitter Impurity Concentration on the High Current Gain of Silicon n-p-n Power Transistors**
R. U. Martinelli and E. Jetter

76 **High-Level Concentration of Sunlight on Silicon Solar Cells**
L. S. Napoli, G. A. Swartz, S. G. Liu, N. Klein, D. Fairbanks, and D. Tamutus

109 **Thermal Analysis of Single-Crystal Silicon Ribbon Growth Processes**
A. E. Bell

139 **Drain Characteristics of Thin-Film MOS FET's**
S. T. Hsu

158 **Technical Papers**

160 **Patents**

163 **Authors**

March 1977 Volume 38 Number 1

RCA Corporation
E. H. Griffiths President and Chief Executive Officer

Editorial Advisory Board
D. M. Cottler Government and Commercial Systems
N. L. Gordon RCA Laboratories
G. C. Hennessy RCA Laboratories
G. B. Herzog RCA Laboratories
J. Hillier RCA Senior Scientist
E. O. Johnson RCA Research Laboratories, Inc.
C. H. Lane Picture Tube Division
D. S. McCoy Consumer Electronics
W. Merz Laboratories RCA, Ltd.
K. H. Powers RCA Laboratories
P. Rappaport RCA Laboratories
J. H. Scott, Jr. RCA Laboratories
L. A. Shotliff International Licensing
T. O. Stanley, RCA Laboratories
F. Sterzer RCA Laboratories
J. J. Tietjen RCA Laboratories
W. M. Webster RCA Laboratories

Secretary, Charles C. Foster RCA Laboratories

Editor **Ralph F. Ciafone**

Associate Editors
D. R. Higgs Missile and Surface Radar Division
W. A. Howard National Broadcasting Company
C. Hoyt Consumer Electronics
H. A. Linke RCA Limited
D. A. Lundgren RCA Americom, Inc.
E. McElwee Solid-State Division
J. C. Phillips RCA Research and Engineering
M. G. Pietz Government and Commercial Systems
C. W. Sall RCA Laboratories
W. S. Sepich Commercial Communicatons Systems Division
J. E. Steoger RCA Service Company

© RCA Corporation 1977 All Rights Reserved Printed in USA

Color Signal Inphase and Quadrature Distortion Measurement and Evaluation

D. H. Pritchard and T. M. Wagner

RCA Laboratories, Princeton, N.J. 08540

Abstract—The total transient response at a color transition in a color TV system is a highly complex function involving a number of system parameters. This report enumerates these factors and provides a brief description of the cause and effects of chroma quadrature distortion. A variety of measurement techniques are discussed, coupled with the results of subjective evaluation of practical limits of color edge distortion. A novel technique for describing color edge reproduction involving colorimetric analysis is included.

1. Introduction

The overall color-television-system amplitude and phase response *symmetry* on both sides of the color subcarrier must be considered as a composite in the determination of the color-signal transient response and in the design to minimize both quadrature and inphase distortion at the edges of colored objects.

The total transient response at a color transition is a highly complex function resulting from a vestigial transmission channel involving linear summations of different specified luminance and chrominance signal bandwidths combined with nonlinear pick-up and display characteristics.

The classic analysis of inphase and quadrature distortion is well documented in the literature, thus the content of this report is limited to a brief description of the causes and effects of chrominance signal quadrature and inphase distortion in a color TV system. A variety of

measurement techniques are available that have particular advantages for specific applications. Discussions of these measurement techniques coupled with the results of subjective evaluations of practical limits for color edge distortion are included. Last, we describe a colorimetric measurement method, particularly useful for overall receiver evaluation, that has been developed here and that takes the total color edge distortions into account.

2. Background and Definitions

2.1 Background

The composite NTSC color television signal consists of a luminance signal and the interleaved color subcarrier signal. The chrominance information is provided by simultaneous modulation of the color subcarrier in both amplitude and phase. The amplitude of the subcarrier relative to the corresponding luminance signal amplitude is a measure of color saturation, while the instantaneous phase of the subcarrier, relative to a reference phase, is a measure of the hue of the scene.

The composite color-signal vector, at any one instance, may be described in terms of two orthogonal components consisting of individually amplitude modulated subcarriers of the same frequency and locked in phase quadrature.

The color demodulation process at the receiver usually consists of at least two synchronous detectors in the video domain operated in phase quadrature from the reference signal locked to the incoming reference burst. The overall reference signal phase is adjusted to obtain two color-difference signals directly (usually R-Y and B-Y), and a simple linear resistive matrix is employed to produce the third color-difference signal (G-Y).

In the case of certain modern corrective matrices (or in the case of I, Q operation) the two demodulators may not produce R-Y and B-Y directly and may not be operated in phase quadrature. In any case, linear resistive matrices are usually employed to eventually derive the proper R-Y, B-Y, and G-Y signals. For the purpose of analyzing the effects of quadrature distortion, the demodulators may be considered to be operated in phase quadrature.

A transition in the color signal produces sidebands that extend outward on both sides of the subcarrier in the frequency domain. Consider a case where the composite chrominance channel response of an overall system is nonsymmetrical in amplitude and/or phase around the subcarrier in the desired frequency range. The resulting instantaneous vector additions of the existing sidebands result in both an effective phase shift

and amplitude distortion of the color subcarrier as a function of the degree of nonsymmetry. This is interpreted by the two synchronous demodulators as a variable phase during the transition and is reproduced as an erroneous edge color. The two demodulators are effectively in phase quadrature; therefore a color transition will produce an incorrect amplitude response in the inphase channel and a spurious output in the quadrature channel when sideband nonsymmetry occurs. A complete mathematical analysis of quadrature distortion is available in numerous textbooks (see, e.g., Ref. [1]).

It must be clearly recognized that two types of color-signal distortion can occur as a consequence of the relative sideband response on either side of the color subcarrier; namely, *inphase* and *quadrature* distortion. Inphase distortion results from nonlinear but symmetrical amplitude and/or phase response, while quadrature distortion results from nonsymmetry (sometimes referred to as "skew-symmetry") of amplitude and/or phase response around the subcarrier, which effectively introduces a carrier phase shift.

The color-signal quadrature distortion introduced by virtue of the vestigial nature of the luminance carrier, the color subcarrier being a sideband of that carrier, is usually considered negligible with respect to other more predominant typical sources of distortion of the color subcarrier modulation sidebands.

The basic "haystack" shape of modern i-f response characteristics dictates a specific chroma channel band-pass and "peaker" circuit compensation of both amplitude and phase characteristics in the circuitry preceding the color demodulators. In addition, the FCC standards incorporate a color correction filter at the transmitter whose envelope delay characteristic has a slope over the double-sideband portion of the color-signal spectrum that is intended to compensate for the average envelope delay of a "typical" i-f amplifier response in that part of the band (see Fig. 1).

The necessity to provide adequate rejection of the sound signal at 4.5 MHz relative to the color subcarrier frequency of 3.58 MHz also imposes severe limits upon the practical ability to maintain symmetrical amplitude and phase response for at least ±500 kHz (the transmitted double-sideband color-signal bandwidth) around the color subcarrier. Thus, the ability to provide adequate circuit designs and to quantitatively measure and subjectively evaluate overall system performance in the area of color-signal transient response is a highly important and complex task.

This paper identifies at least five approaches for obtaining useful information in this area and describes a colorimetric method in detail. Experimental work is discussed that has provided an indication of the

practical limits of a comparative basis by subjective evaluation of chroma channel overall bandwidth versus the degree of sideband nonsymmetry that might be tolerated. Reasonable correlation between the subjective results and the quantitative measurements was obtained.

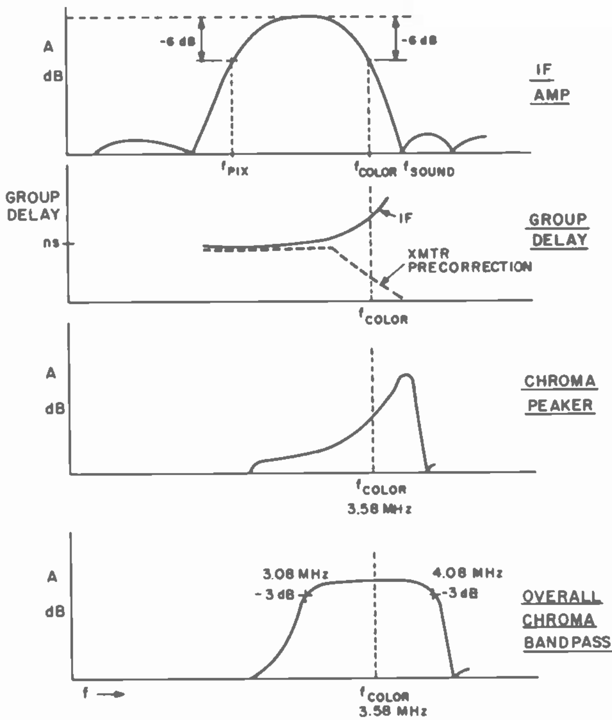


Fig. 1—Typical color TV receiver response characteristics.

2.2 Definitions

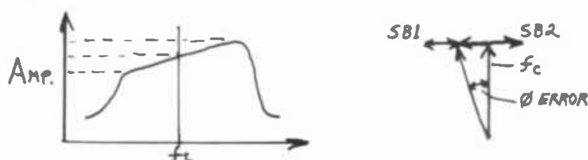
A composite edge transition in color is a complex combination of the following factors:

1. Luminance-channel transient response.
2. Chrominance-channel inphase transient response.
3. Chrominance-channel quadrature response.
4. Linear combination of luminance signal and chrominance signals having different transition rise times (the specific combination being a function of the hue change and the degree of saturation).

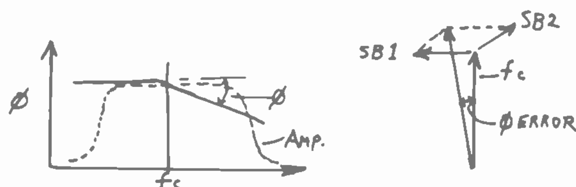
5. Time-delay match between luminance and chrominance signal channels.
6. Registry of R, G, and B signals.

The factors considered as being primary items in this discussion are the chrominance channel inphase and quadrature edge distortions. The sources of these types of distortion are:

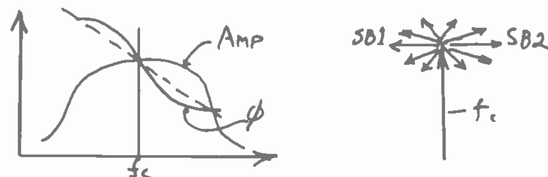
1. Amplitude nonsymmetry: The amplitudes of the upper and lower sideband components differ over the band and generate both inphase and quadrature errors.



2. A phase lead or lag of one set of sideband components with respect to the other that introduces a resultant phase shift of the color subcarrier.



3. Symmetrical, but not linear, phase relationship of one sideband signal component versus the other.



As previously discussed, any or all of these conditions may exist simultaneously in an overall system. However, the major contributing items are the chroma band-pass and peaker filter amplitude and envelope delay characteristics, the effects of the 4.5-MHz sound trap, and the i-f response in the vicinity of the color subcarrier (approximately 3.08 to 4.08 MHz) in relation to the color-correction filter employed at the transmitter in compliance with NTSC specifications.

The linear combination of luminance signal and chrominance signals

having different transition rise times, referred to above, is sometimes overlooked by the TV receiver circuit designer and thus deserves some discussion at this point. The composite transition rise time is a function of color saturation. Since the NTSC system is not linear (gamma of 2.2), a degree of luminance information is contained within the chrominance channel and is maximum in fully saturated colors. Because of the large difference between typical luminance bandwidth (3.6 MHz) and typical chrominance bandwidth (0.5 MHz), this has considerable influence on the overall picture subjective "sharpness," which is a function of color saturation. Thus, the system designer should take this into account when considering the optimum peaking function to be employed.[2,3]

In 100% saturated primary colors in the NTSC system, the relative contributions of the luminance channel and the chrominance channel to the transition response are as follows:

Luminance:	R = 30%	G = 59%	B = 11%
Chrominance:	R-Y = 70%	G-Y = 41%	B-Y = 89%

Fortunately, typical scenes average less than 50% saturation, thereby allowing the luminance channel to be a controlling factor in determining picture "sharpness." However, high-saturation colors do occur often enough to be of considerable importance and cannot tolerate additional edge distortions in amplitude or hue such as may result from excessive color quadrature distortion.

Fig. 2 is a drawing of the idealized transition shapes for a 100% saturated color in the three primary color channels of red, green and blue that points up the importance of this factor. Figs. 3 and 4 are photographs taken of the luminance, red, green and blue channels, respectively, with 100% saturated colors for two cases of a test pattern and a typical scene that provides a subjective indication of this system parameter. This situation is also included in the composite edge measurement technique that is described later in this paper and that utilizes CIE colorimetric error calculations.

3. Subjective Evaluation Tests

A series of subjective tests have been conducted in order to determine the practical limits of color quadrature distortion in terms of amplitude and phase nonsymmetry around the color subcarrier as related to overall chrominance channel bandwidth. Before giving the results of these tests, however, it is appropriate to review some basic and well-documented facts about color quadrature distortion relating color channel resolution to subjective visibility of color edge distortion.

1. The tests involved only equal-band color decoding systems and, therefore, do not necessarily relate directly to I, Q decoding approaches.
2. Equal-band systems of necessity are limited to approximately 500 kHz chroma signal overall bandwidth. The available double-sideband frequency spectrum is limited to the range of from about

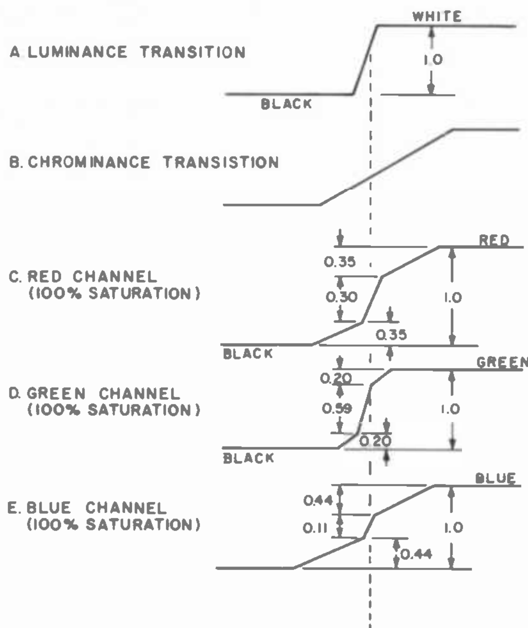
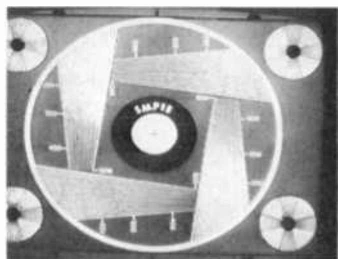
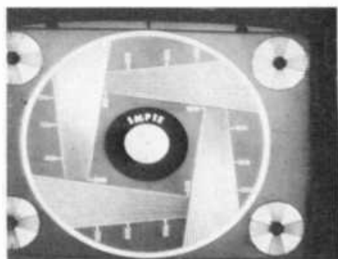


Fig. 2—Liminance plus chrominance transitions.

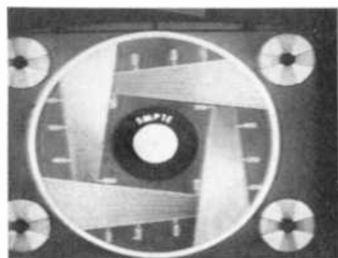
3.08 to 4.08 MHz (3-dB points) due to the necessity for a deep rejection trap at the sound frequency of 4.5 MHz. Chroma channel decoding bandwidths in excess of about 500 to 600 kHz will always result in some degree of quadrature distortion in equal-band systems due to the loss of amplitude of the upper sideband frequencies, even though the phase linearity might be retained.
3. The total energy in the quadrature signal component remains essentially constant as the chrominance channel bandwidth is varied. However, the subjective visibility of the quadrature distortion changes drastically with bandwidth. Wide bandwidth produces high peak amplitude quadrature components, but having short duration. In the extreme, this condition may be vi-



LUMINANCE



GREEN



RED

BLUE

Fig. 3—Test pattern showing Y, R, G, and B channel resolution (sharpness).

LUMINANCE

GREEN

RED

BLUE

Fig. 4—Typical scene showing Y, R, G, and B channel resolution (sharpness).

sually mistaken for misregistry that varies as a function of scene color content. Quadrature distortion that occurs at relatively low frequencies (500 kHz and below) has reduced peak values but is often much more subjectively objectionable due to the relatively long time duration (large spatial dimensions) of the peaks. Thus, the subjective evaluation of quadrature distortion visibility versus resolution in this region has considerable practical importance in developing and evaluating equalband color receiver decoding systems.

For calculation of the specific response waveforms in both amplitude and frequency content, see, e.g., Refs. [1] and [4].

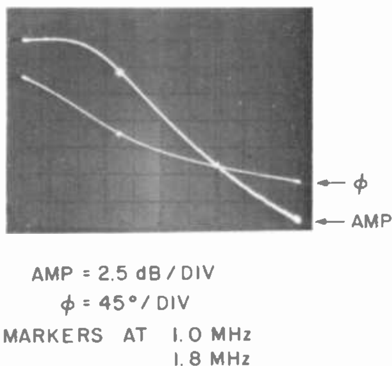
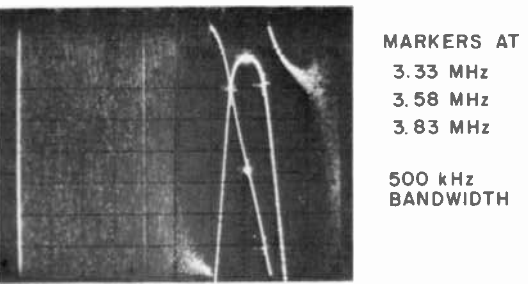


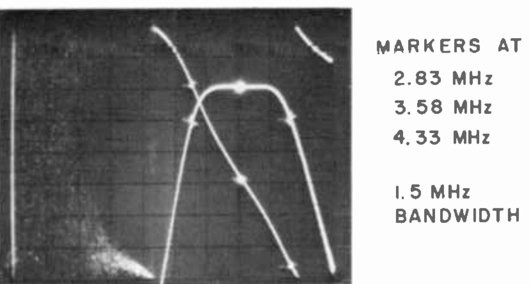
Fig. 5—Decoder low-pass output filter.

3.1 Test Results

The subjective tests recently conducted involved an equal-band decoding system operating at video level with plug-in chroma band-pass and decoder low-pass filters to facilitate the choice of chrominance channel overall bandwidth. In all cases, the decoder low-pass filters were held constant at a bandwidth of 1.0 MHz at 3 dB (see Fig. 5). Thus, the decoder input band-pass filter becomes the resolution (and quadrature) determining factor. In the practical case this is also a desirable situation to minimize spurious intermodulation and 920-kHz beat problems in relation to the color decoding process.

Input band-pass filters of 500 kHz, 1.0 MHz, 1.5 MHz, and 2.0 MHz (3-dB points) were used initially in the tests (see Fig. 6), resulting in chroma pass-band resolution of 250 kHz, 500 kHz, 750 kHz, and 1.0 MHz, respectively.





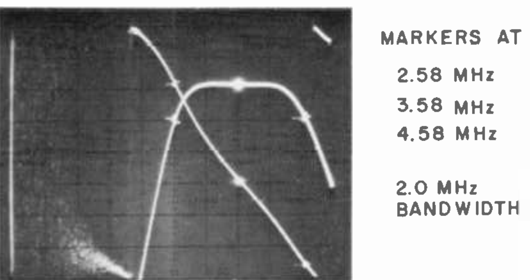


Fig. 6—Decoder input band-pass filters.

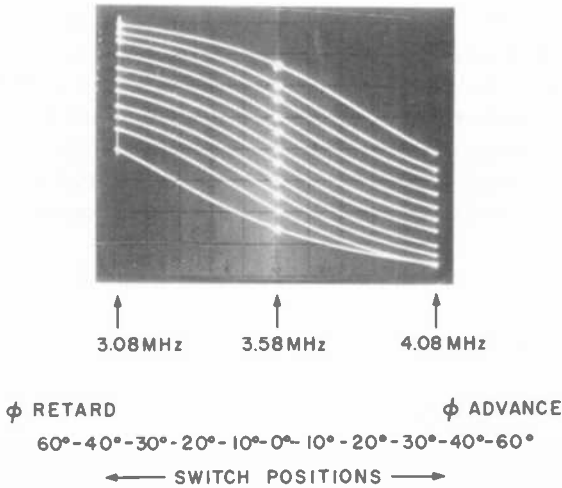
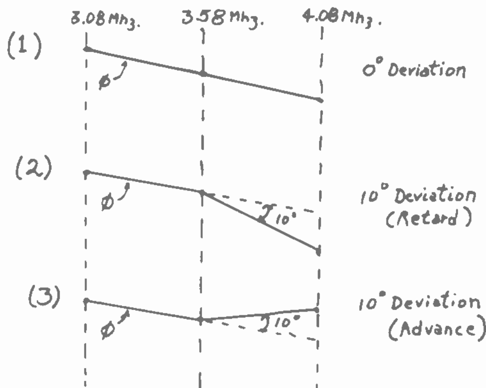


Fig. 7—Phase nonsymmetry calibration curves (centered around 3.58 MHz) for chroma channel.

A circuit was provided in the chrominance channel that was capable of introducing calibrated amounts of phase nonsymmetry, centered around 3.58 MHz, and measured with respect to ±500 kHz (3.08 and 4.08 MHz). Fig. 7 indicates the phase curves for each of the eleven switch positions that allowed the introduction of 0°, 10°, 20°, 30°, 40°, and 60° of phase *retardation* or *advance* of the frequencies above 3.58 MHz with respect to those below 3.58 MHz as measured at the ±500 kHz points.

For the test purposes the degree of phase nonsymmetry was defined as the phase deviation at 4.08 MHz relative to the phase slope between 3.08 and 3.58 MHz, i.e. ±500 kHz relative to 3.58 MHz.



Video signal feed to the decoder simulates the "ideal" condition with a perfect transmitter and receiver rf and i-f situation, such that the degree of quadrature distortion purposely introduced could be controlled accurately. However, the subjective tests also included an actual rf-if link for verification purposes. The overall amplitude and phase response of the Conrac receiver at the second-detector output and the overall response following the color band-pass filter are indicated in Figs. 8a and 8b.

Observers were asked to judge the degree of subjective objectionability on a relative scale for a variety of test signals as well as typical slides. The tests were not intended to be statistical in nature and the subjects in-

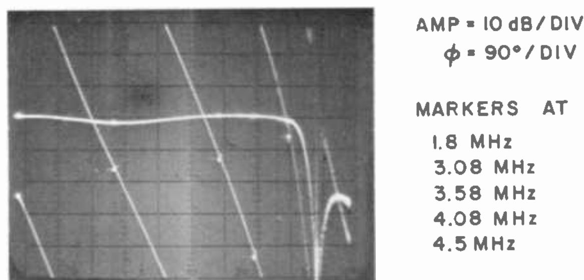


Fig. 8a—RF-IF amplitude and phase response (Conrac receiver).

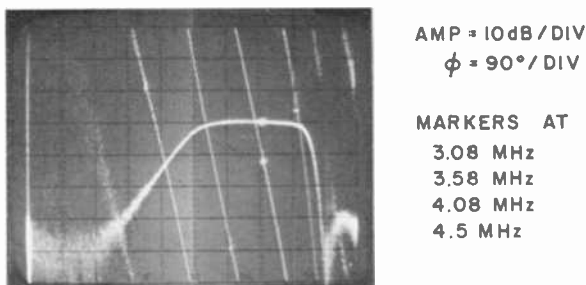


Fig. 8b—Overall chroma channel amplitude and phase response.

volved ranged from average to those with extensive experience in TV engineering. The results constitute practical *design information* and do not necessarily represent typical field cases of quadrature distortion. Unanimous agreement was reached that with the most critical test slide (referred to as the "Indian Blanket" slide), as well as with critical scenes, an overall system phase nonsymmetry measured at ±500 kHz bandwidth of about 15° was just barely perceptible and certainly acceptable. At the other end of the scale, a nonsymmetry of about 40° was objectionable

on almost any typical scene, with degrees of nonsymmetry of 60° or more being totally unacceptable under any conditions. It should be remembered that these figures represent *total* system nonsymmetry and should not be applied as a tolerance to only one particular portion of the system.

Figs. 9 and 10 summarize the results of these tests within the conditions discussed above. The two charts represent the same information, but expressed from two different points of view. Fig. 9 indicates the shape

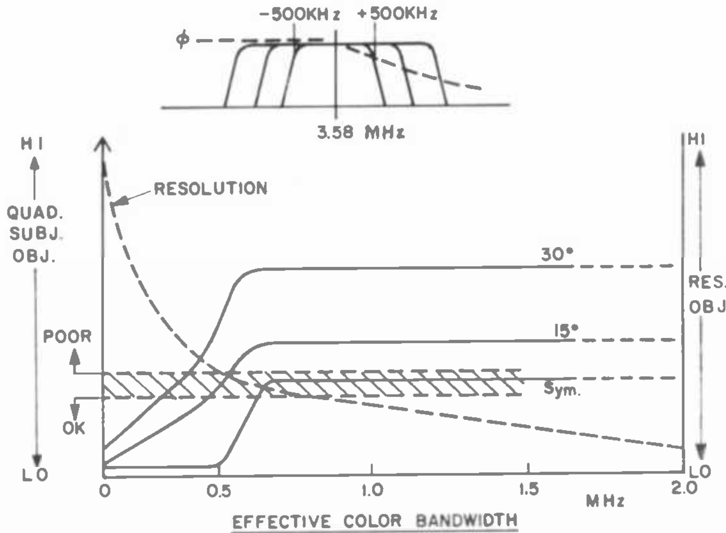


Fig. 9—Subjective test results; chroma channel resolution versus quadrature. Here, the degree of nonsymmetry, expressed as the shift in degrees of the color subcarrier, *is always measured at 500 kHz*. (See Sec. 2.2 for illustrations of amplitude and phase nonsymmetry.)

of the curves that result if the degree of nonsymmetry is always measured at ±500 kHz regardless of actual bandwidth. Fig. 10 indicates the shape of the curves if the degree of nonsymmetry is always measured at the band edges as the bandwidth is varied. In either case it is obvious that the quadrature distortion degree of *objectionability* relative to color resolution is critical in the 500-kHz range and is generally more objectionable as the frequency (bandwidth) is reduced, i.e., the color error becomes smeared, occupying larger spatial dimensions, and is thereby more visible, even though the absolute peak amplitude of the quadrature component decreases.

As the color bandwidth is made very narrow (200–300 kHz), the practical situation results in the likelihood of better symmetry in both

amplitude and phase around the subcarrier with the attendant reduction in quadrature distortion. However, the overall degree of objectionability is still high due to a lack of color resolution and the resulting color edge de-saturation and smear. It appears that in *equal-band decoder systems* the practical compromise in bandwidth is about ±500 kHz. This assumes that reasonable care is taken in the design of critical filters to ensure an order of amplitude symmetry of no worse than about a 1.5 dB differential and that phase symmetry is no worse than about a 15° differential measured over the ±500-kHz band.

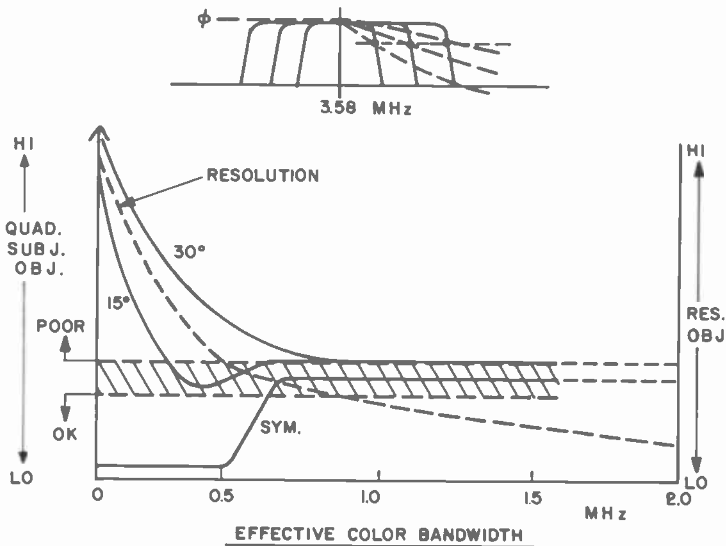


Fig. 10—Subjective test results; chroma channel resolution versus quadrature. Here, the degree of nonsymmetry, expressed as the shift in degrees of the color subcarrier, is always measured at the *band edge*.

To document the signal waveforms resulting from the phase non-symmetry values discussed above, a series of photographs were taken of the inphase and quadrature signal waveforms at the output of the decoder with a 100-kHz repetition rate square-wave burst of 3.58 MHz fed to the color encoder as a test signal. Typical results are shown in Figs. 11, 12, and 13. It should be remembered that the transient response of the inphase waveforms represent changes in the *intensity* of the desired edge color only, while the signal existing in the quadrature output rep-

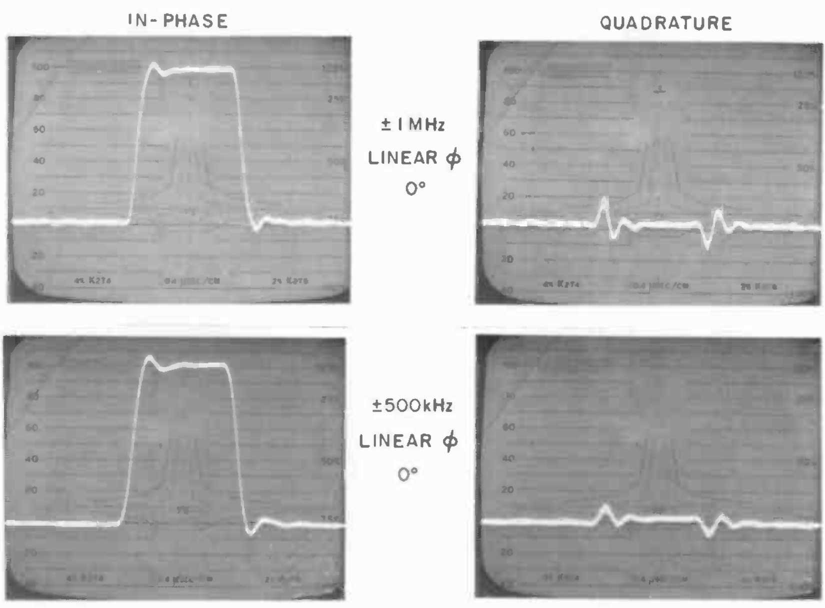


Fig. 11—Chroma channel response, rf-if (bandpass filter ±1 MHz and ±500 kHz).

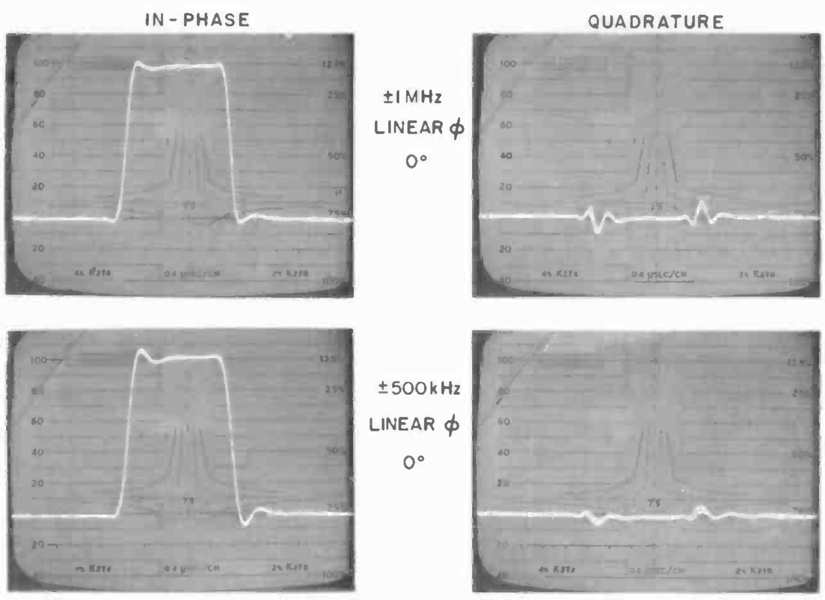


Fig. 12—Chroma channel response, video (no rf-if) (bandpass filter ±1 MHz and ±500 kHz).

resents *erroneous hue* edge content existing at the same time in the composite color-signal output. It may be estimated from this, as well as from other methods of measuring quadrature distortion, that a peak-to-peak value of about 15% quadrature signal amplitude, by a coincidence of numbers, roughly corresponds to a phase nonsymmetry of 15° in a 500-kHz bandwidth system. This number, it should be noted, is a direct function of bandwidth and, therefore, should not be quoted as a general rule except under specific conditions as stated.

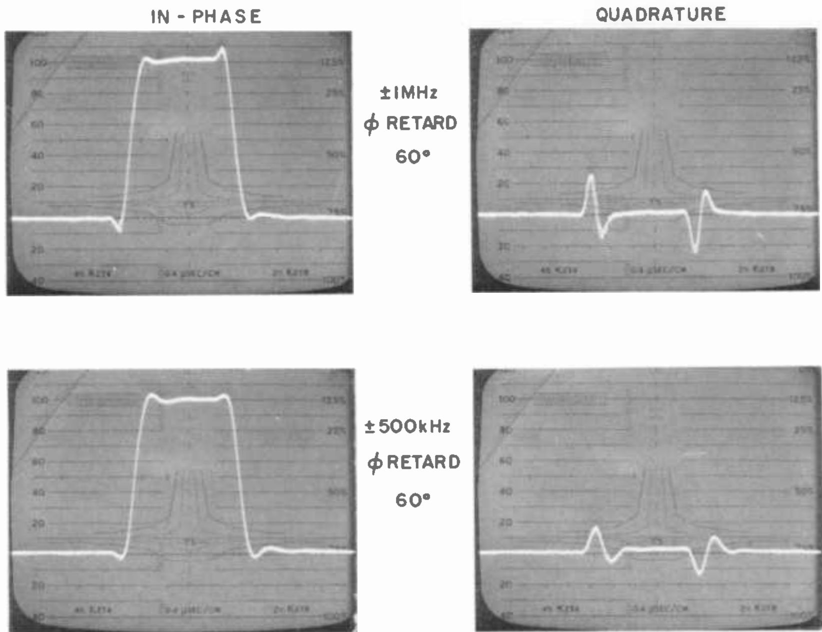


Fig. 13—Chroma channel response, video (no rf-if) (bandpass filter ±1 MHz and ±500 kHz, phase retard 60°).

4. Measurement Techniques

The chrominance signal inphase and quadrature distortion may be measured and evaluated by several methods other than by subjective means. Five of these techniques are briefly described below.

4.1 Amplitude and Phase Sweep Method

A video sweep is used as the input signal to a 3.58-MHz balanced modulator (or color encoder) with the sweep range adjusted to cover the

chroma band. This signal can then be fed at video level directly to the color demodulators, or used to modulate a double-sideband rf (or i-f) modulator to include the receiver rf and i-f pass-band characteristics. The output of the color demodulators is observed with an oscilloscope, while the 3.58-MHz reference phase (hue adjust) is first set for 0° phase (maximum inphase component), then shifted by 90° to indicate the quadrature component. Fig. 14 is a photograph of two examples of in-phase and quadrature response curves.

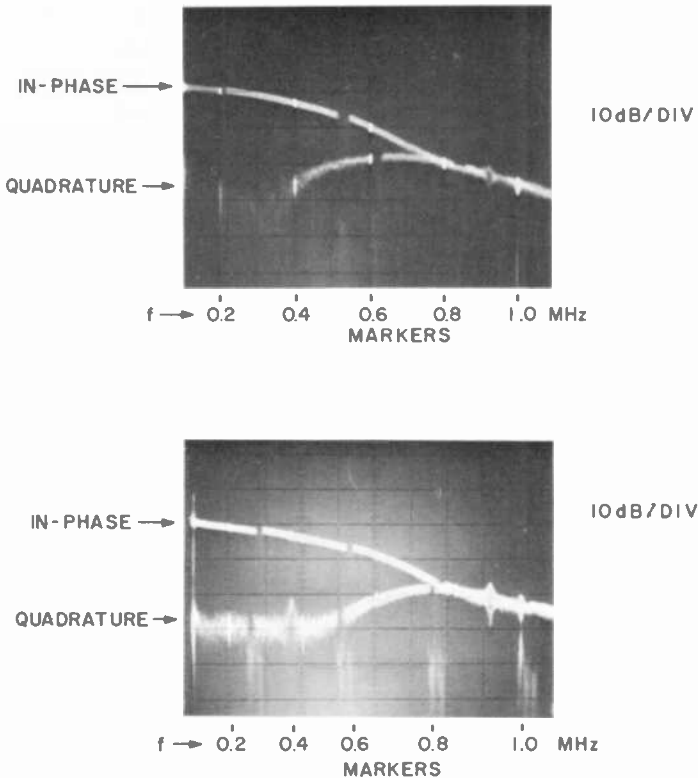


Fig. 14—Amplitude and sweep measurement technique.

This is an excellent method for measuring color-signal response, since a composite picture of both the maximum amplitude of the quadrature distortion, relative to the inphase component, and the extent of the frequency content of the quadrature distortion is immediately visible on the oscilloscope display. One limitation exists, however, in that no

indication is provided as to whether the distortion occurred as a result of amplitude or phase nonsymmetry, or both.

4.2 EIA Recommended Measurement Technique

The Electronics Industries Association has compiled recommended methods of testing monochrome and color television receivers* that contains a section entitled, "Chrominance Signal Quadrature Crosstalk," which includes Definition, Method of Measurement, and Presentation of Data.

This proposed method of measurement is essentially the same as the Sweep Method except that it suggests a point-by-point measurement as a function of frequency within the chroma pass band. If only one of the chrominance components, Q for example, is modulated and subsequently demodulated by a synchronous detector referenced to I phase, the detector output signal should be zero for all modulating frequencies if no quadrature distortion is present in the overall system. The presence of an output signal indicates quadrature distortion at some point within the system. The amount of the output signal, at a given modulating frequency, relative to the normal amount of output signal obtained when the detector is referenced to the desired phase (Q phase), is a measure of the quadrature distortion at that particular frequency. The ratio of the quadrature component to the inphase component of the signal can be determined by the use of a synchronous detector. A graph of this ratio, expressed in dB, versus modulation frequency may be used to depict the overall degree of quadrature distortion. Fig. 15 is an example of such data.

The reader is referred to the EIA proposed measurement techniques document for details of the procedure.

4.3 Vector Display Method

A visual display method for measuring quadrature distortion is to use 75% NTSC color bars as an input signal and feed the demodulated R-Y and B-Y signals to the x and y inputs, respectively, of an oscilloscope adjusted for equal bandwidths and equal gains in both the vertical and horizontal deflection amplifiers. This procedure yields a vector display of the color-bar signal components where the transitions from one color to the next appear as lines connecting the color points. If no quadrature distortion exists, the transition follows a straight line from one color to

---

* Published by IEEE, EIA TV Standards Subcommittee R-4.4.

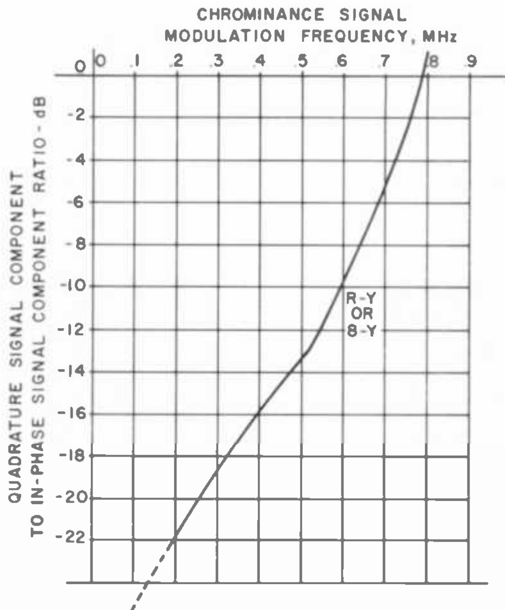


Fig. 15—EIA quadrature crosstalk.

the next. If, however, quadrature distortion does exist in the system, the transition path will be curved, with the degree of deviation from a straight line being indicative of the amount of quadrature distortion present. Fig. 16 is a drawing of an example of such a display. In this example a yellow-to-cyan color bar transition is used. The solid line indicates the ideal transition. A small amount of quadrature distortion (5%) will produce a transitional shift either toward yellow green or through white, while larger amounts of distortion (15% to 50%) will produce color

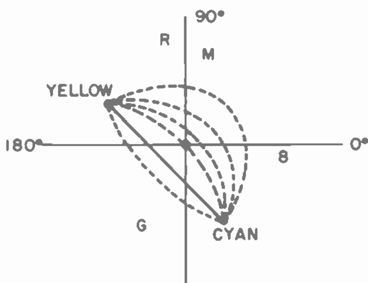
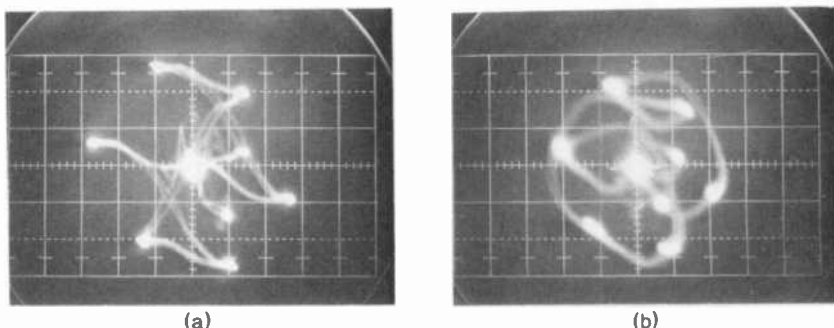


Fig. 16—Vector display method.

shifts toward the purple as indicated in the drawing. Figs. 17(a) and 17(b) indicate actual photographs of such a vector display with a small amount of quadrature distortion and larger amounts, respectively.

Since any point on the vector display represents some combination of R, G and B signal values, the equivalent subcarrier vector angle (hue) may be calculated. Thus, it is possible to compare points on the vector



(a) (b)

Fig. 17—Scope displays as vector diagrams, R-Y/B-Y signal: (left) little and (right) moderate chroma quadrature distortion.

display with calculated CIE colorimetry coordinates. Figs. 18 and 19 are vector plots of color subcarrier phase and amplitude as computed from relative R, G, and B waveform values as measured at the kinescope cathodes. They represent the color transitions from the white to the yellow color bar and from the yellow to the cyan color bar, respectively.

4.4 Square-Wave Subcarrier-Encoded Method

Another method for evaluating color quadrature distortion has already been discussed in the section describing the documentation of the subjective tests. This method makes use of a 3.58-MHz subcarrier-encoded square-wave signal with a basic repetition rate of about 100 kHz. A synchronous detector operating at 0° phase reference will produce the inphase response, while a detector operating at 90° reference phase will indicate the quadrature components. In this method the quadrature output is representative of the amount of distortion, and the time duration is a function of the chroma-signal bandwidth. The overall subjective impression, however, depends upon both amplitude and time duration. Therefore, the amplitude alone is not a sufficient measurement

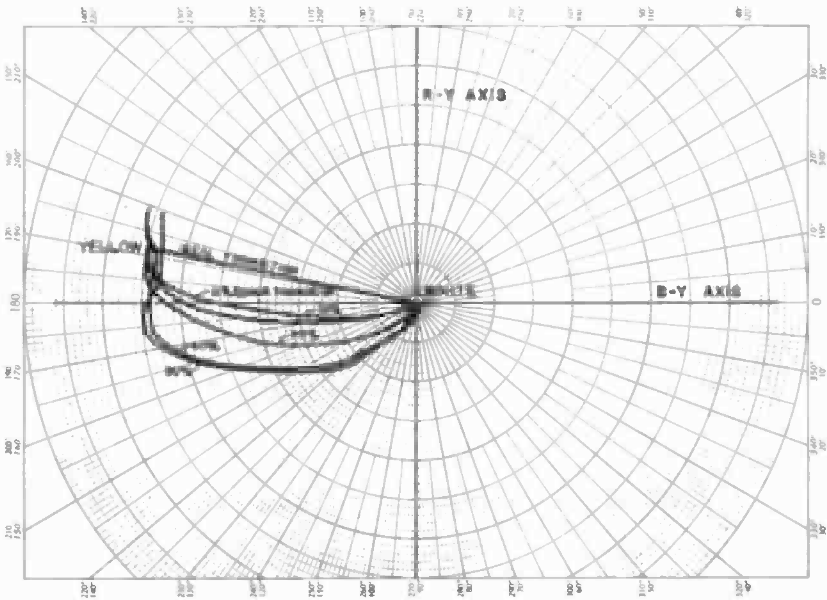


Fig. 18—Subcarrier vector plot for white-to-yellow transition, with quadrature transient as the parameter.

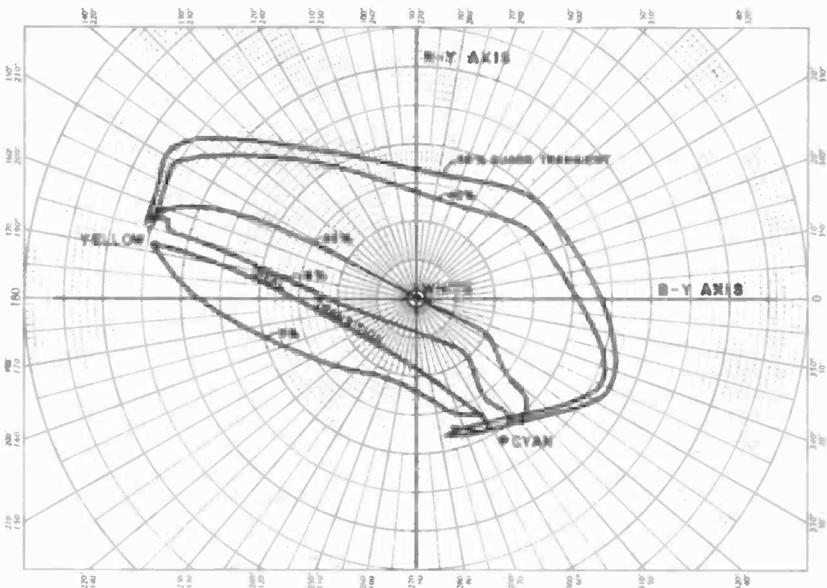


Fig. 19—Subcarrier vector plot for yellow-to-cyan transition, with quadrature transients as the parameter.

criteria. Figs. 11, 12, and 13 are examples of this method of quadrature and inphase response measurement.

4.5 Color Error Evaluation in Terms of the CIE Chromaticity Diagram

This novel approach to the evaluation of color transitions in a complete TV system involves expressing the color distortion in terms of colorimetric color errors (C_E) and/or in terms of CIE coordates (x, y or u, v) to include a plot of these values for a given transition on a CIE diagram. The technique involves measurement of the relative values of the R, G, and B waveforms during the transition at the kinescope cathode and computing the resulting color shifts.

Fig. 20—CIE diagram for yellow-to-white transition (dotted-line triangle is for NTSC gamut and solid-line triangle is for field phosphor gamut).

Examples of the results of this method of measurement are shown in Figs. 20 and 21, and the procedure is described below.

Measurement Technique

One method of demonstrating the effect of chroma quadrature distortion is to display 75% NTSC color bars and observe the color shifts which

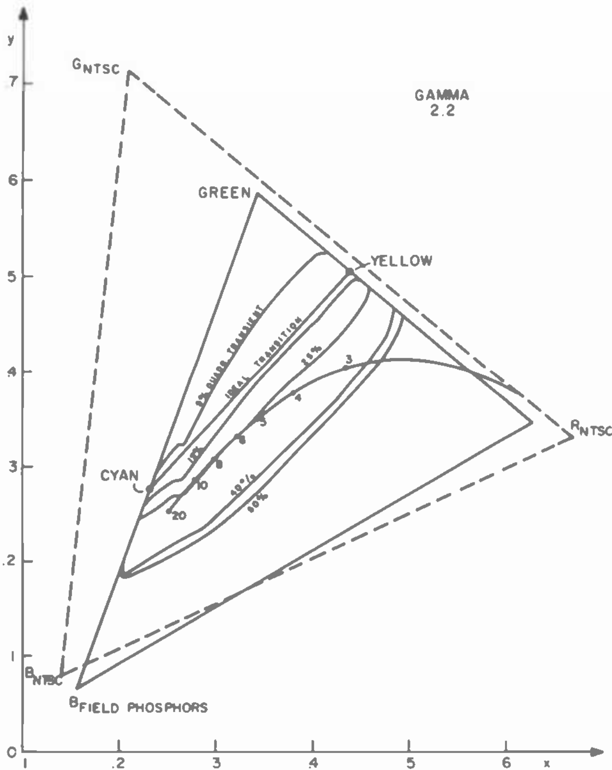


Fig. 21—CIE diagram for yellow-to-cyan transition (dotted-line triangle is for NTSC gamut and solid-line triangle is for field phosphor gamut).

occur around the transition interval. These color shifts can be measured and the result may be displayed on a CIE diagram (see Fig. 20). While it is possible to use a spot-brightness meter to measure the color shifts on the face of the color kinescope, it is more accurate and convenient to measure the R, G, B waveforms at the kinescope cathodes and compute

the displayed colors during the transition from one color bar to the next. A digital processing scope[†] can be used instead of Polaroid photos of the R, G, B transients to obtain a sufficient number of data points. Phosphor primary coordinates must be known and must be measured if absolute comparisons are desired. The transitions from the white to the yellow and from the yellow to the cyan color bars were used as examples.

Quantitative Analysis

The approach taken to evaluate color transition distortion quantitatively and to include a convenient measurement method is as follows:

1. Measure a TV receiver adjusted for normal operation in terms of overall amplitude and phase characteristic. Record the R, G, B waveform values during a transition for 75% NTSC color bars at the kinescope cathode within a 2-μsec. range. A minimum of 20 points within this interval is required for accuracy.
2. Average the transition response over two fields to eliminate subcarrier interlace offset. Compute and plot color shifts for the transition areas of white to yellow and yellow to cyan.
3. Reference the computed color shifts to an ideal transition and compute (UVW)* color errors. The phosphor primaries must be known if absolute comparisons are desired.
4. In the CIE diagram, the area bounded by the ideal transition locus and the locus of the actual color shifts is representative of any existing chroma transient distortion. This area is computed for the loci for all color shifts. The substitution of (UVW)* color errors instead of x,y coordinates in one dimension and time segments in the second dimension must now be introduced.

   The chroma transition distortion errors are analyzed by correlating them with the results of the subjective tests. The computation is done at selected intervals and the CIE coordinates are referenced to an ideal transition from one color to the next to obtain (UVW)* color errors. When reading the signal values at the kinescope cathodes, black is considered to be zero and the R, G, B amplitudes for the white bar are normalized to be 1.00. Negative signal values below the black level are neglected due to kinescope cutoff. The color-error computation process is as follows. After obtaining the Red, Green and Blue signal values and applying a gamma correction of 2.2, the tristimulus values are computed for each point.

[†] For example, the Tektronix 7000 series or the WP1100/WP1200 series.

$$R' = R^{2.2} \qquad G' = G^{2.2} \qquad B' = B^{2.2}$$

$$\begin{vmatrix} \overline{X} \\ \overline{Y} \\ \overline{Z} \end{vmatrix} = \begin{vmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{vmatrix} \cdot \begin{vmatrix} R' \\ G' \\ B' \end{vmatrix}$$

The matrix "C" is the display matrix. For field phosphors and white of illuminant C, it is

$$C = \begin{vmatrix} 0.3440 & 0.4220 & 0.2145 \\ 0.1890 & 0.7197 & 0.0913 \\ 0.0148 & 0.0886 & 1.0778 \end{vmatrix}$$

From the tristimulus values $\overline{X}$, $\overline{Y}$, and $\overline{Z}$, one obtains CIE x,y and u,v coordinates as follows:

$$x = \frac{\overline{X}}{\overline{X} + \overline{Y} + \overline{Z}}, \quad y = \frac{\overline{Y}}{\overline{X} + \overline{Y} + \overline{Z}}.$$

By coordinate transformation,

$$u = \frac{4x}{-2x + 12y + 3} \quad \text{and} \quad v = \frac{6y}{-2x + 12y + 3}.$$

The (UVW)* color error (CE) is the vector length of the distance between color 1 (referenced e.g. to Ill.C) and a second color (referenced also to Ill.C). The brightness information is taken into account by the use of a three-dimensional vector space where

$$(\mathrm{UVW})^*\mathrm{CE} = \sqrt{A^2 + B^2 + C^2}$$

$A = (U_1) - (U_2)$
$B = (V_1) - (V_2)$
$C = (W_1) - (W_2)$

and $U_1 = 13W_1 \cdot (u_1 - u_0)$
$V_1 = 13W_1 \cdot (v_1 - v_0)$
$W_1 = 25\,(\overline{Y}_1)^{1/3} - 17$

u_1, v_1 = Color (1) coordinates
u_2, v_2 = Color (2) coordinates
$\overline{Y}_1$ = Color (1) brightness
$\overline{Y}_2$ = Color (2) brightness
u_0, v_0 = Ill. C coordinates
$u_0 = 0.20086$
$v_0 = 0.30732$

and in the same way U_2, V_2 and W_2 are derived

$$U_2 = 13W_2 \cdot (u_2 - u_0)$$
$$V_2 = 13W_2 \cdot (v_2 - v_0)$$
$$W_2 = 25\,(\overline{Y}_2)^{1/3} - 17$$

Different values of chroma distortion were simulated by changing the fine tuning of the TV receiver.

Measurements were taken with AFT *ON* which results in less than 5% square-wave transients in the quadrature channel. The rf fine tuning (AFT *OFF*) was adjusted to obtain 15% transients in the quadrature channel for purposes of correlating with subjective results. Data was

Table 1—Summary

TA	XC	YC	YBAR	N	COL.ER	T-SORT	C-SORT	C-AREA	NSORT
-1.00	0.310	0.316	1.000	1	0.0	0.0000	0.0	0.021	1
-0.90	0.311	0.321	0.988	2	2.29	0.0183	2.291	0.065	2
-0.80	0.311	0.326	0.976	3	4.59	0.0372	4.586	0.345	3
-0.70	0.316	0.338	0.980	4	7.31	0.0951	7.309	0.516	4
-0.60	0.322	0.351	0.986	5	10.13	0.1543	10.128	0.935	5
-0.50	0.332	0.368	0.986	6	11.27	0.2417	11.266	1.045	6
-0.40	0.342	0.380	0.987	7	12.38	0.3300	12.384	2.171	7
-0.30	0.358	0.417	0.977	8	15.79	0.4841	15.794	1.151	8
-0.25	0.365	0.431	0.974	9	17.46	0.5533	17.457	1.215	9
-0.20	0.372	0.445	0.970	10	18.96	0.6201	18.964	1.551	10
-0.15	0.380	0.461	0.972	11	20.75	0.6981	20.748	1.556	11
-0.10	0.387	0.476	0.975	12	22.36	0.7703	22.364	1.325	12
-0.05	0.395	0.487	0.953	13	21.90	0.8302	21.900	0.960	13
0.0	0.401	0.496	0.933	14	21.36	0.8746	21.361	1.238	14
0.05	0.409	0.506	0.905	15	19.94	0.9345	19.938	0.825	15
0.10	0.416	0.514	0.879	16	18.32	0.9776	18.316	0.205	16
0.15	0.422	0.517	0.846	17	15.87	1.0000	0.000	0.0	24
0.20	0.425	0.516	0.815	18	13.35	1.0000	0.000	0.0	25
0.25	0.427	0.514	0.802	19	11.47	1.0000	0.000	0.0	26
0.30	0.429	0.512	0.789	20	9.63	1.0000	0.000	0.001	27
0.40	0.431	0.511	0.793	21	8.28	1.0008	1.401	0.009	23
0.50	0.435	0.507	0.805	22	4.18	1.0018	15.867	0.006	17
0.60	0.438	0.505	0.813	23	1.40	1.0024	4.179	0.015	22
0.70	0.439	0.504	0.817	24	0.00	1.0047	8.284	0.007	21
0.80	0.439	0.504	0.817	25	0.00	1.0055	9.635	0.012	20
0.90	0.439	0.504	0.817	26	0.00	1.0067	11.470	[illegible]	19
1.00	[illegible]	0.504	0.817	27	0.00	1.0078	13.353	0.0	18

CEAVG= 10.70 CMAX= 22.364 Total C-AREA = 15.135

xc, yc = x, y CIE coordinates

COL.ERR - (UVW)* color error

T-SORT = arranging of all time segments in consecutive order to minimize small measurement errors

C-Area = Color Error Area i.e. CE-segments * time segments

Table 1—Measurements of White-to-Yellow Transition (Automtic Fine Tuning AFT was Employed, Which Produces Less Than 5% Quadrature Transients)

taken for 25%, 40%, and 50% transients to obtain a sufficient range of data points.

Tables 1 and 2 show the data computed for two examples of color transients in tabular form. Table 3 is a summary of the average color

errors and error areas. Figs. 20 and 21 show the results plotted on a CIE diagram. Fig. 22 is a block diagram of the measurement set-up, and Fig. 23 is a flow chart of the computer program developed to evaluate quadrature distortion errors from measured R, G, B waveforms. These

Table 2—Summary

TA	XC	YC	YBAR	N	COL.ER	T-SORT	C-SORT	C-AREA	NSORT
-1.00	0.310	0.316	1.000	1	0.00	0.0000	0.000	0.001	1
-0.90	0.309	0.317	1.004	2	1.56	0.0008	1.561	0.069	2
-0.80	0.311	0.323	1.004	3	3.75	0.0267	3.752	0.195	3
-0.70	0.315	0.332	1.004	4	5.82	0.0675	5.820	0.618	4
-0.60	0.321	0.348	0.991	5	10.75	0.1420	10.749	1.109	5
-0.50	0.328	0.367	0.988	6	16.38	0.2238	16.376	1.896	6
-0.40	0.336	0.389	0.984	7	22.36	0.3217	22.356	0.0	7
-0.30	0.336	0.389	0.984	8	22.36	0.3217	22.356	6.113	8
-0.25	0.360	0.437	0.998	9	29.81	0.5561	29.814	2.104	9
-0.20	0.367	0.451	1.001	10	31.80	0.6243	31.800	2.887	10
-0.15	0.377	0.469	1.002	11	33.37	0.7130	33.371	2.339	11
-0.10	0.385	0.482	1.002	12	34.37	0.7820	34.374	1.711	12
-0.05	0.392	0.492	0.992	13	34.12	0.8320	34.116	1.517	13
0.0	0.398	0.500	0.983	14	33.71	0.8767	33.715	1.873	14
0.05	0.407	0.509	0.947	15	31.78	0.9339	31.785	1.276	15
0.10	0.415	0.515	0.913	16	29.64	0.9754	29.642	0.579	16
0.15	0.420	0.518	0.880	17	27.08	0.9994	27.076	0.008	17
0.20	0.424	0.517	0.850	18	24.39	1.0000	0.000	0.000	27
0.25	0.426	0.515	0.845	19	22.74	1.0002	0.999	0.000	26
0.30	0.427	0.514	0.840	20	21.19	1.0005	2.141	0.003	25
0.40	0.431	0.511	0.851	21	17.95	1.0012	5.239	0.005	24
0.50	0.437	0.506	0.839	22	11.88	1.0020	8.449	0.008	23
0.60	0.441	0.503	0.840	23	8.45	1.0028	11.880	0.022	22
0.70	0.444	0.500	0.844	24	5.24	1.0043	17.951	0.016	21
0.80	0.447	0.498	0.850	25	2.14	1.0051	21.195	0.009	20
0.90	0.448	0.497	0.848	26	1.00	1.0055	22.735	0.008	19
1.00	0.449	0.496	0.847	27	0.00	1.0059	24.393	0.0	18

CEAVG= 17.91   CMAX= 34.374   Total C-AREA = 24.467

Table 2—Measurements of White-to-Yellow Transition (Data Taken for 15% Quadrature Transients)

figures use a yellow–white transition and a yellow–cyan transition as examples.

5. Summary

A tutorial description of the causes and effects of chroma-channel color edge distortion, emphasizing inphase and quadrature components, has been presented.

Five methods for measurement and evaluation of color edge distortions are described that correlate with results obtained from subjective tests that set practical limits on the nature and extent of color quadrature

Table 3—Summary of Average Color Errors and Error Areas

White-to-Yellow Transition

Quadrature transient	Avg. Color Error	Color Error Area
< 5%	10.7	15.2
15%	17.9	24.5
25%	28.5	37.4
40%	32.3	40.9
50%	41.4	51.9

Yellow to Cyan Bar Transition

Quadrature transient	Avg. Color Error	Color Error Area
< 5%	9.2	18.0
15%	12.4	15.0
25%	24.4	39.2
40%	47.4	84.2
50%	53.5	98.7

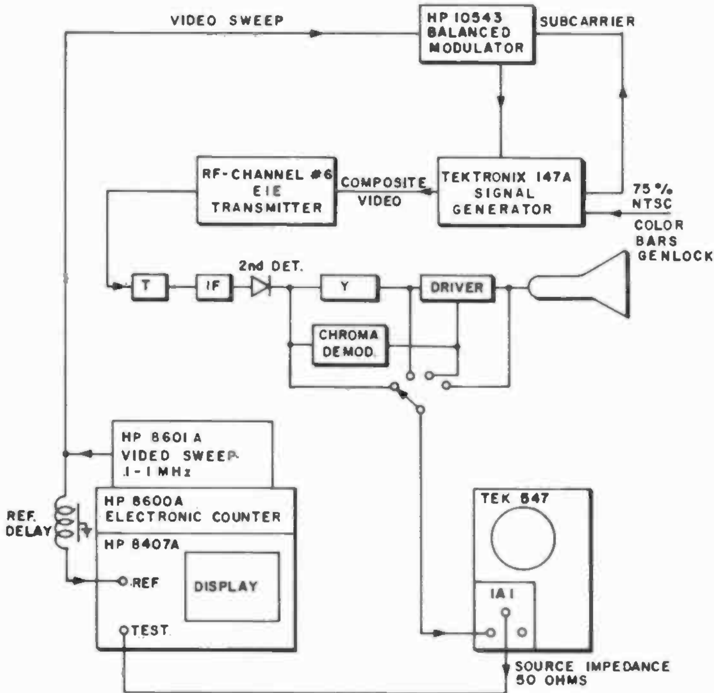


Fig. 22—Color transition measurement set-up.

distortion. No one method of measuring color edge distortion that includes quadrature effects is a panacea for all applications. Each of the methods discussed has features that are attractive for particular purposes.

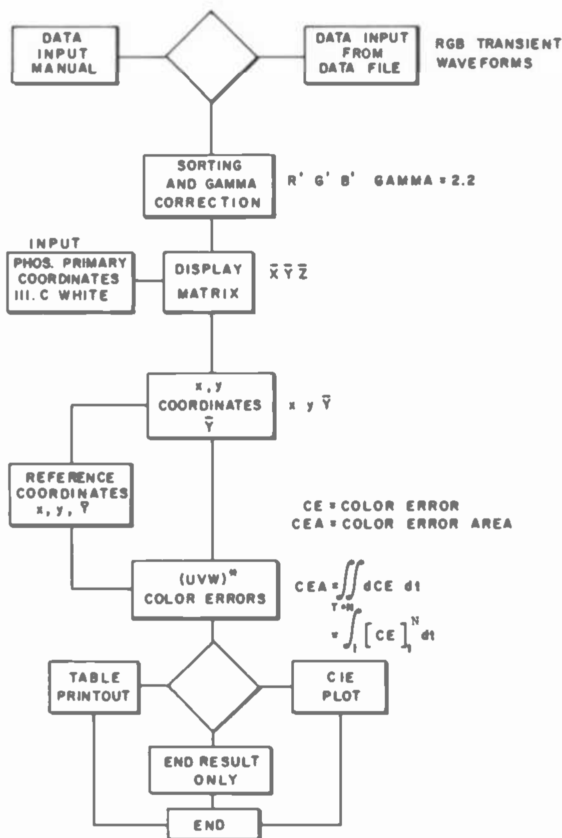


Fig. 23—Chroma quadrature distortion evaluation program (CQD Fortran AI).

The measurement methods included a novel procedure that renders a chroma transient performance figure in terms of colorimetric criteria as it is observed on the face of the display kinescope. This figure includes all sources of color edge-distortion errors such as luminance to chrominance delay, registry, quadrature distortion, etc. A data set corresponding to 15% quadrature transient amplitude was included to cor-

relate with the subjective tests. A Fortran IV computer program was developed for this method of evaluation. Measured data, in general, confirmed the subjective results in both direction and magnitude.

References:

[1] D. Fink, *NTSC Color Standards*, McGraw-Hill Book Co., Inc., New York (1955).

[2] J. B. Chatten, "Transition Effects in Compatible Color Television," *Proc. IRE*, **42**, p. 221, Jan. 1954.

[3] W. Gibson and A. C. Schroeder, "Color TV Luminance Detail Rendition," *Proc. IRE*, **43**, p. 918, Aug. 1955.

[4] *Principles of Color Television*, Hazeltine Corporation, John Wiley and Sons, New York (1956).

Performance of Injection Lasers with External Gratings

H. S. Sommers, Jr.

RCA Laboratories, Princeton, N. J. 08540

Abstract—The operating characteristics of large-area injection lasers in a cavity composed of laser, lens, and grating are studied in detail. Comparison of performances with various combinations of gratings, lenses, and internal absorbers is given. All microscope lenses of N.A. $\geq$ 0.3 perform well, but camera lenses are strikingly inferior. Both reflection and transmission gratings give good results. With reflection gratings, the threshold and external efficiency approximate the bare laser with cleaved uncoated facets. The transmission grating has a similar differential efficiency but higher threshold. Its projected beam is highly collimated, with spread equal to the angle subtended at the lens by the near field on the facet; differential efficiency approaches 15% and the spectral radiant intensity reaches 10^9 W steradian^{-1} μm^{-1}. The power sharing between modes of the bare laser and the compound resonator, as well as the spectral width of the tuned line, follow the same laws as the bare laser; no evidence is found for spectral interference directly associated with a fundamental optical line width. Extensive tables summarize data of help in systems design.

1. Introduction

This paper summarizes an extensive study of the performance of a double-heterojunction injection laser in a tuneable dispersive cavity of the type described by Ludecke and Harris[1] and used extensively by Rossi and co-workers.[2-6] In addition to presenting the first evaluation of the effective reflection coefficients and efficiencies of various combinations of lenses and gratings and of the use of a transmission grating, it amplifies previous reports of spectral tuning range and spectral width. Quanti-

tative data on the power sharing between bare laser and compound resonator modes are discussed in terms of the modal power distribution of conventional lasers in an attempt to increase the understanding of the dynamics of the lasing state and determine the predictability of the modal power distribution from knowledge of the cavity geometry.

The paper is arranged by sections on the instrumentation and individual components, on engineering aspects of the tunable source, and on fundamental studies. The background literature is summarized at the points where it reads on the text.

2. Instrumentation

2.1. Geometry and Terminology of External Cavity

The external cavity has a lens and grating in the autocollimation alignment.[1] Figure 1 is a schematic that shows the optical components and defines the symbols. The laser has length L and facet reflectivities R_1 and R_2, radiating powers P_1 and P_2. A compound lens is placed D_a

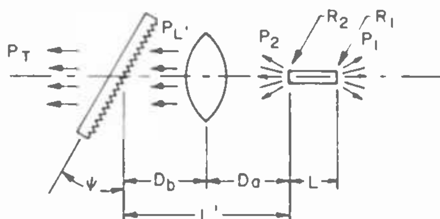


Fig. 1—Schematic of laser in dispersive cavity.

from the second facet, ideally one focal length so that the transmitted power ($P_{L'}$) is focused at infinity. The grating is a distance L' from the laser. Preferably the grating would be at the back focal plane of the lens to give confocal geometry, but for convenience L' was fixed at about 14 cm. The grating was tuned by rotation through the angle ψ.

For maximum selectivity, the grating lines were parallel to the junction plane, since the spread of the beam transverse to the junction is considerably larger than the lateral spread. With a reflection grating, the useful output was the diverging beam P_1, while with the transmitting grating it was the highly collimated beam P_T. We use two other terms,

P_{eff} the net power through facet 2, and R_{eff} an equivalent lumped reflectivity at position 2 of the external cavity plus facet.

In general, two distinct groups of longitudinal modes are excited, the q modes of the bare laser and the q' modes of the compound cavity. The dispersion of the q' modes will be much less than for the q modes because of the much greater length of the compound resonator.[1] With optimum tuning, the q' modes center on a select q mode, producing a fine structure that is not resolved by our spectrometer.

2.2 Alignment

The following routine was used to align the various elements, a process that can be time consuming if the grating or lens is changed. Important tools are an He–Ne sighting beam to define the optical axis of the system and removable holders for the lens and grating with dowel pins for precise replacement.

(1) The laser is aligned on and perpendicular to the optic axis. This requires only standard micromanipulators and angular rotators. The junction plane is aligned with respect to the mechanical axis of scan by manual rotation of the laser holder, monitored with a low-power microscope.

(2) The laser is removed from the holder and the lens aligned approximately on the optic axis. This is monitored by observing the direction of the projected He–Ne beam. If desired, the lens axis can be adjusted parallel to the optic axis by observing the reflected beam. We found no improvement from this alignment.

(3) The lens is removed and the grating inserted and aligned perpendicular to the optic axis. It is then rotated about the optic axis to give the required angle with respect to the junction plane of the laser (perpendicular or parallel). This is monitored by observing the plane of the higher-order reflected beams.

(4) The grating is removed, the laser inserted and excited, and the lens positioned precisely on the optic axis so that the projected laser beam lies on the axis. This is monitored with an infrared image converter. The lens is focused so that the output beam is parallel, maintaining constant cross section independent of distance. Only a crude focusing is possible because the beam cross section is not well defined.

(5) The grating is now replaced and the output power P_1 (OR P_T) monitored with a p-i-n diode. Grating alignment is improved by fine rotations from its nearly correct original alignment. If the current is near the threshold of the bare laser, a complete alignment of grating and lens will produce a dramatic increase in power. Since

the lens focus is not yet precise, the changes with grating angles will probably be small, and care must be taken to distinguish between modulation of P_1 from shadows of the semiconductor and coherent effects. When a reduction in threshold is observed, the lens should be focused, producing a large change in threshold. Trimming is done at currents below the original threshold. The cross section of the beam at the grating should be checked to see that the alignment has produced a parallel beam, not the other resonant condition of focus on the grating.

(6) The grating is now rotated to the tuned position. This is a crude adjustment because the lasing spectrum can be tuned over such a wide range; careful manual rotation without any screw adjustment is adequate to produce lasing with the desired grating order. Finally the focus and grating alignment are adjusted to optimize the most important feature, such as minimum threshold at the desired wavelength.

Sometimes step 5 with the reflection grating is frustrating and no alignment seems to give constructive reflection. In such cases, monitoring a part of the output with a vidicon camera can help to return the reflection from the grating to the laser facet rather than bypassing it. With a little experience, the TV camera is no longer necessary.

The procedure is the same with the transmission grating. Now in step 4, the grating need not be removed. The effects of shadowing are avoided in step 5 by monitoring P_T.

The critical adjustments are grating alignment, lens position, and focus. For grating mount we used a two-axis high precision mirror mount with 360° rotation about the polar axis for selecting the grating order and for coarse and fine adjustments about both axes. The claimed resolution is 0.1 arc sec. The grating lines were oriented by rotation in the grating holder. The lens mount was adequate though less satisfactory, a micromanipulator with three translations driven by homemade differential micrometer screws. The main trouble with this was hysteresis in the focus control, probably from imperfect motion on the ways.

2.3 Components

All the lasers were cut from wafer 638N-LOC, a five-heterojunction structure of the LOC type[7] designed to give a beam with small transverse spread (perpendicular to the junction). The composition and thickness of the various layers is shown in Table 1. The transverse beam profile was around 25° half width, a pattern that can be collected with reasonable efficiency by a lens of numerical aperture 0.3 (f/1.4). The application

Table 1—Wafer 638N-LOC Section through Heterojunctions

HETEROJUNCTION		①		②		③		④		⑤	
DOPANT	Te		Te		Te		Ge		Ge		Ge
CARRIER DENSITY (cm^{-3})	$n = 10^{18}$		$n = 5 \times 10^{17}$		$n = 10^{18}$		$p = 10^{18}$		$p = 5 \times 10^{18}$		$p > 10^{19}$
% AlAs	0		9.5		2		0		28		0
THICKNESS (μm)			1.5		1.1		1.2		1.5		

required an AR (antireflection) coating with very high transmissivity on one facet, a condition favoring the excitation of internally trapped nonradiative modes.[8] These were avoided by using a wide wing structure in which a 100-μm metal contact was evaporated down the center of a sawed pellet about 500 μm wide. The end facets were cleaved and given various coatings.

AR coatings were evaporated SiO deposited to greater than $\lambda/4$ and thinned by successive etchings in a 10% solution of HF. A proper thickness raised the threshold typically from 4 to 7 A and gave a ratio P_2/P_1 exceeding 10/1. The reflectivity was under 0.7% according to the usual expression[9]

$$\frac{P_2}{P_1} = \left[\frac{R_1}{R_2}\right]^{1/2}\left[\frac{1-R_2}{1-R_1}\right] \quad [1]$$

When reflecting coatings were applied to the other end, they consisted of gold over SiO.

Table 2 lists properties of the different lasers. The wafer was very uniform and the principal differences among the units are from length, width, and facet coating. The columns are laser number, length, width,

Table 2—Properties of Bare Lasers (Lasers Are All of Series 4/75-8 Cut From Wafer 638N-LOC)

			I_{th} (A)		η_{ext}[a] (%)			
Laser	L μm	W μm	Cleaved	Coated	Cleaved	Coated	R_1	R_2
1	235	100	4.0		25		0.31	0.31
41	315	100		7.3		28	0.31	0.003
43	350	100		7.5		25	0.31	0.004
61	315	100		5.4		30	0.9[b]	
75	275	100	4.1	6.4	38	28	0.31	0.003
83	480	100[c]	4.9	7.5	34	28	0.31	<0.003

(a) Total from both facets.
(b) Metal reflector, estimated reflectivity.
(c) Sawed sidewall.

threshold before and after coating, external differential quantum efficiency before and after coating, and reflectivity of the two facets.

Table 3 gives the properties of the lenses, gratings, and neutral density filters that were used successfully. The lenses are listed under microscope or camera lenses, and the columns are make, identifying number (∞ means infinity corrected), magnification, numerical aperture or (*f*) number, and focal length. The grating data are type (R or T means reflection or transmission), line spacing, blaze angle, and order used. All are high-quality replica gratings. The neutral density filters are Wratten gelatin filters; the losses are measured at the lasing wavelength.

Table 3—Optical Components

Lens				Gratings			
	Mag.	N.A.	F.L. (mm)	Type	Spacing (mm^{-1})	Blaze	Order
A. Microscope Lenses				R	600	30° 40′	1,2,3
Unitron MP_L	10X	0.3	15	R	625	61° 10′	1,2,3
Unitron MP_L	20X	0.4	8	R	1200	26° 45′	1
Unitron MP_L	40X	0.65	3.5				
Spenser ∞	45X	0.66	3.0	T	600	20° 00′	1,2,3
Leitz ∞	50X	0.85	5.0				
B. Camera Lenses		*f*/		Neutral Density Filters			
Dahlmeyer		0.99	25	No.		One-Way Loss	
Wollensak		1.9	25	NDF-0.6		3 dB	
Wollensak		1.5	45	NDF-1.0		5 dB	
Schneider-Kreuznach		2.0	50	NDF-2.0		10 dB	

In general, the microscope lenses were far easier to align and far more efficient than any of the camera lenses. It seems probable that the latter have serious spherical aberration in this configuration. All microscope lenses gave roughly equivalent results except that the numerical aperture of 0.3 had somewhat lower efficiency. Similarly a change of grating made only a nominal difference, the efficiency and resolution usually increasing with the order of the reflection. The best performances were with the coarser gratings in the highest (third) order.

2.4 Measurements

All measurements were made with current pulses 100 ns nominal length, and peak detection of current and output power. Power was measured with a calibrated large-area p-i-n photodiode. Spectra were measured

with a 0.5-meter spectrometer. The spontaneous emission at 8500 Å was monitored by the transmission through a pair of 100 Å interference-type band-pass filters that rejected the coherent output even when viewed directly in the forward beam.

The entire cavity assembly was mounted on a rotary table with axis through the laser itself to facilitate study of the profiles of P_1. The profile of P_T was studied with fixed cavity and scanning detector placed at various distances from ¼ to 7 meters. Details of the instrumentation have been published.[10]

3. Performance of Various Combinations of Components

The collection efficiency of the lens in the resonant cavity and the grating efficiency and reflectivity are important system data for engineering design that have not been available. Because of the poor knowledge of aberrations of lenses used in geometries for which they were not intended and of grating reflectance away from the blaze angle, it is necessary to empirically determine how well various combinations of lenses and gratings actually work. The direct measurement, using a pellicle beam splitter to measure P_L and the return from the grating, proved to be very unsatisfactory (probably from resonant reflections in the thin pellicle), and we resorted to indirect deductions. The comparisons are sufficient to indicate general trends in performance, but the accuracy is limited by an appreciable lack of repeatability in the mechanical adjustments and an uncertainty in the criteria determining optical alignment. Another factor that hampers detailed comparisons is that the data that best illustrate a particular point were frequently from measurements at different wavelengths.

3.1 Microscope Versus Camera Lens

A camera lens should give higher spectral resolution than a microscope lens because its longer focal length gives a beam covering more grating lines. However, our general experience with camera lenses was very disappointing, and without exception we found any microscope lens to be far easier to align and much more efficient than any camera lens.

Table 4 is representative of our results. It compares an 8-mm microscope lens with N.A. = 0.4 with a 25-mm camera lens of equivalent aperture. The bare laser had a threshold of 7.4 A, and differential efficiencies of 1.8 and 26% through the uncoated and coated facets, giving

---

* Jarrell Ash.

Table 4—Comparison of Microscope and Camera Lenses (Laser #83; Grating: 625/mm, Reflection, 3rd Order)

Lens	I_{th} (A)	η_1 (%)	η_2 (%)	η_{ext} (%)	η_{eff} (%)	R_{eff}
None	7.4	1.8	26	28	100	0.003
MP_L— 20X	4.7	18.4			66	0.53
Dahlmeyer 25 mm	7.4	3.3			12	0.01

a total efficiency η_{ext} = 28%. With the microscope lens, threshold 4.7 A, the efficiency through the cleaved facet was 18.4%, or effective efficiency η_{eff} of lens plus grating of 66%. With the approximation that the total efficiency of the semiconductor is not changed by addition of the external cavity, the ratio of net emissions from the two facets is 18.4/(28 – 18.4) = 1.9. This gives an effective reflectivity R_{eff} for the lens plus grating of 0.53, Eq. [1].

In comparison, the threshold with the camera lens was 7.4 A, the forward efficiency η_1 was 3.3%, and η_{eff} and R_{eff} were 12% and 0.01. We believe the poor performance of the camera lenses is due to bad spherical aberration, since they are not used in the geometry for which the corrections were designed.

3.2 Reflection Gratings

Table 5, the performance of various microscope lenses with two reflection gratings, shows only minor differences in performance with different combinations. In general, a lens aperture of 0.4 was sufficient to collect over 85% of the emitted light, and the performance was then essentially independent of the speed or quality of the lens. Coarse gratings in high order gave the best efficiency and threshold, but were not as good in first order as the fine grating. None of the gratings were designed for our laser light, and the size of the blaze angle did not seem to matter.

Table 5—Characteristics of Lasers With Reflection Gratings (Laser #41)

Lens	I_{th} (A)	η_{ext} (%)	η_{eff} (%)	R_{eff}	I_{th} (A)	η_{ext} (%)	η_{eff} (%)	R_{eff}
Grating:	1200/mm, Reflection, 1st order				625/mm, Reflection, 3rd order			
None	7.3	28.	100	0.003				
MP_L 10×	4.7	12.8	46	0.27				
MP_L 20×					4.8	14	50	0.31
Leitz 50×	4.4	11.7	42	0.22	4.2	12	44	0.24

3.3 Transmission Grating

We made an extensive study with a transmission grating. This geometry has the feature that the lens that projects the beam on the grating also forms the external beam. Table 6 shows the performance of two different lasers and four lenses with a 600/mm transmission grating in third order. Since both the forward and backward beams can be monitored, the effective reflectivity and efficiency of the optical components are given directly, but their interpretation still involves assumptions about the constancy of the efficiency of the semiconductor under different operating conditions. Notice that the threshold is higher for the transmission grating, due to its lower reflectivity (0.05 to 0.1 compared to over 0.3 for the reflection gratings), but that the differential efficiencies are very high.

Table 6—Behavior With Transmission Grating (Laser #43; Grating: 600/mm, Transmission, 3rd Order)

Lens	I_{th} (A)	η_1 (%)	η_T (%)	η_{eff} (%)	R_{eff} (%)
None	7.5	23.5	1.9	100	0.003
Unitron 10×	6.3	3.0	8.7	46	7.1
Unitron 20×	6.4	3.9	11.3	56	7.1
Leitz 50×	6.0	4.9	12.4	68	9.1
AO 45× [a]	6.6	4.0	13.2	68	5.1

[a] 2nd order of grating.

We also achieved good performance with a laser with one facet AR coated and one metallized. The threshold and efficiency of the bare laser were 6.8 A and 26%; with the 50 × lens* and transmission grating in third order, they were 5.6 A and 15%. This efficiency is appreciably higher than for lasers without the reflector on facet 1, Table 6.

The half width of the lateral profile of the transmitted beam was 1.0°, which is essentially the angle subtended by the facet at the lens (100 μm width at 5 mm), while the transverse width was 0.03°, giving a solid angle of about 10^{-5} steradian. At 10 A, 540 mW power, the radiant source intensity was 6×10^4 W/steradian into a single line of 0.8 Å half width.

4. Performance as a Tunable Oscillator

Previous studies have found the laser tuning range to be about 2%,[1-5] with a line width of about 0.5 Å, which was approximately the resolution

* Leitz

of the fully illuminated grating.[1,2,5] We report a more detailed study of the performance at spot frequencies, of the fidelity with which the spectrum tracks the grating setting, and of the change of line width with power and grating order. We also offer a simple theory of the line width and a speculation on the meaning of the smooth tuning between q modes.

4.1 Tuning Range

In their first study, Rossi et al reported a tuning range of 150 Å over which the power was within half its peak value.[2] A study on a single-heterojunction laser showed curious examples of power saturation that depended on wavelength, perhaps associated with loss of modal confinement.[6] We made a more detailed study at spot frequencies and found our LOC cavity had a power that was linear in current at all wavelengths, which is consistent with the above idea; the double heterojunctions preserve the confinement at all currents.

Table 7—Performance at Spot Frequencies (Laser #41, Leitz 50X Lens, 1200/mm Reflection Grating)

Condition	λ (Å)	I_{th} (A)	η_1 (%)	η_2 (%)	η_{ext} (%)	η_{eff} (%)	R_{eff}	P_1 (@ 10 A) (mW)
Bare laser	8860	7.3	1.8	26	28	100	0.003	1060
Focused lens	8860	6.7	3.8	17.2	21		0.019	970
Tuned cavity	8997	5.1	10.7			38	0.17	740
Tuned cavity	8985	4.7	11.7			42	0.22	870
Tuned cavity	8948	4.2	9.1			32	0.12	740
Tuned cavity	8946	4.9	8.3			30	0.10	590
Tuned cavity	8790	6.3	9.4			34	0.13	480

The data of Table 7 give new information about the changes with tuning and the limitations on tuning range. Placing the lens in the focused position lowers the threshold from 7.3 to 6.7 A. The efficiency drops from 28 to 21%, which indicates that the one-way efficiency of the lens is about 75%. Notice that constructive reflection from the lens has increased the effective reflectivity of the coated facet from 0.003 to 0.02 (Sec. 3.1). Addition of the grating tuned at 8997 Å lowers the threshold to 5.1 A and increases the forward efficiency to 10.7%; the power at 10 A has dropped to 740 mW. The corresponding effective reflectivity and efficiency are 0.17 and 38%. The tuning band at 10 A for output of half the maximum is −70 to +140 Å, from the bare cavity lasing, or 2.4%.

The precise values of effective reflectivity and efficiency and output

power are quite sensitive to optical alignment. A different alignment at 8932 Å gave a threshold of 4.4 A, η_1 = 14%, and output of 1120 mW at 10 A. The difference is probably due to better positioning of the lens, whose efficiency was abnormally low in the study tabulated. For the new alignment, the effective reflectivity and efficiency were about 0.33 and 50%, quite similar to the bare uncoated lasers.

In summary:

(a) At the fixed current of 10 A, the forward power varies moderately over the tuning range, from 500 to 850 mW. With agc controlling the drive current, a constant output of 500 mW is reasonable.

(b) The constancy of the forward efficiency indicates that the internal losses have little dependence on λ.

(c) The gain has a stronger dependence on λ as indicated by the change of threshold with λ.

(d) The cutoff at the two ends probably arises from different effects. At large λ, the energy states involved in the lasing transitions must be so nearly filled that the gain is independent of current. This produces an abrupt end to the tuning range that is independent of current because of the ceiling on gain, but should depend on the effective reflectivity. At short λ, the practical limit is set by the tuned threshold exceeding the threshold for bare laser modes and the resulting powering of the q modes.

4.2 Tuning Rate

The cavity was tuned by a micrometer that coupled to the grating through a lever with large mechanical advantage. The wavelength is plotted against drum position in Fig. 2, which shows smooth tuning with little indication of a periodicity of 1.7 Å from frequency pulling by the bare laser cavity. Between successive q modes, the variation is only ±20%. Such a small variation indicates that the lateral modes are changing with a periodicity of 1.7 Å. The gap between the successive peaks of the q spectrum, whose half width is about 0.3 Å, is bridged by the lateral dispersion of the normal modes.[10]

4.3 Selectivity

A single mode radiated by an injection laser with cleaved facets is very sharp. At low power, Picqué et al[11] measured a spectral width of 3×10^{-4} Å, while Mironov et al[12] found a width at 80°K that was an order of magnitude larger and almost independent of power in the range of interest to us, 1–4 mW per mode. While the resolution of the grating may

be slightly larger than that of the bare laser, it should not greatly affect the width of a pure mode.

At the operating levels of our studies, many of these pure modes are excited and the actual shape and width of the tuned line will be determined by the selectivity of the grating. Following North,[13] the current and spectral dependence of the output is

$$P(\lambda) = NP^* \left[\frac{g}{\alpha} - 1\right]; \qquad [2]$$

P^* is a constant of the material, g is the current-dependent gain coefficient of the laser at the tuned wavelength, and α is the wavelength-

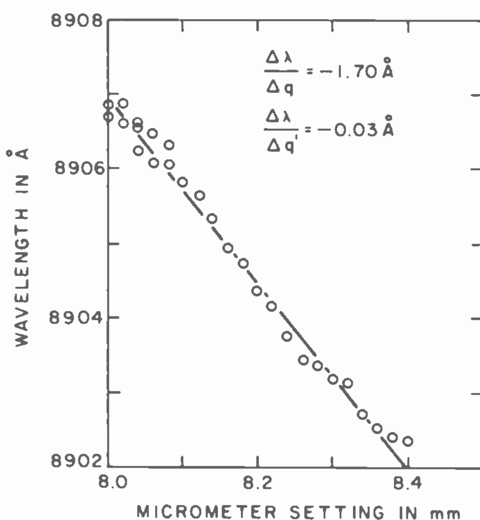


Fig. 2—Tuning curve. (Laser 83; lens MP_L 20X; grating 625/mm, reflection, 3rd order.)

dependent loss coefficient, which is controlled by the grating resolution. Both coefficients are prorated over a round trip of the cavity. N is the number of excited lateral modes, which we assume is independent of λ in the following calculation. The loss coefficient can be written

$$\alpha \equiv \alpha_0 + \alpha(\lambda); \qquad [3]$$

$$\alpha_0 = \alpha_i + \frac{1}{2L} \ln \left[\frac{1}{R_1 R_{eff}}\right],$$

$$\alpha(\lambda) = \frac{1}{2L} \ln \left[\frac{1}{F(\lambda)}\right], \qquad [4]$$

where α_i is the internal loss coefficient of the laser body of length L, and $F(\lambda)$ is the reflectivity of the grating normalized to unity at λ_0. This neglects as unimportant the reflectivity of the coated facet and the angular dependence of the loss.[13]

Introducing the overdrive parameter[13]

$$X \equiv \frac{g_0}{\alpha_0} - 1 \qquad [5]$$

the power at λ_0 is

$$P_0 = NXP^*. \qquad [6]$$

Simple algebra gives the condition for the wavelengths λ' at which the power is reduced to half P_0,

$$\alpha(\lambda') = \frac{\alpha_0 X}{X+2} \qquad [7]$$

For a bare laser of this material with uncoated facets, the principal loss is through the facets and $\alpha_0 = (1/L)\ln(1/R)$. Since the threshold of the tuned laser is about that of the uncoated bare laser, α_0 is about the same. Substituting the typical value for these units of $X = 0.6$ at one watt,

$$F(\lambda') = 0.6. \qquad [8]$$

We see that the theory of modal power distribution predicts the frequent observation that the spectral width is about the resolution of the grating, as well as a slow broadening with X or power.

Fig. 3 shows the measured half width as a function of power with a grating of 625 lines/mm in third order and a lens aperture of 8 mm. The solid line is calculated using

$$F(\lambda) = \exp\left\{\frac{-2r(\lambda-\lambda_0)^2}{\lambda_0^2}\right\} \qquad [9]$$

and experimental values[13,18] of X against P in Eqs. [4] and [7]. r is the resolving power of the combination of laser cavity plus grating. It was taken as 2.7×10^4, the resolving power of the grating alone for this lens aperture. Two nearly similar lasers are shown, and one of them on two different alignments. Part of the scatter is due to the complexity of the line shape, probably from poor control of the lateral modes. The agreement with the simple theory is satisfactory, confirming the prediction that the width is about that of the grating function and grows slowly with power.[2]

Fig. 4 is a related curve showing the broadening as the grating resolution is decreased from third-order reflection to first order. The line shape is further distorted for the lower orders by the adjacent q modes of the laser cavity, spacing 1.7 Å. Again the solid curves are calculated from Eqs. [4], [7], and [9]. In first order, the grating has only a small effect on $F(\lambda)$, its effective Q of 4700 giving only a 25% increase over the value

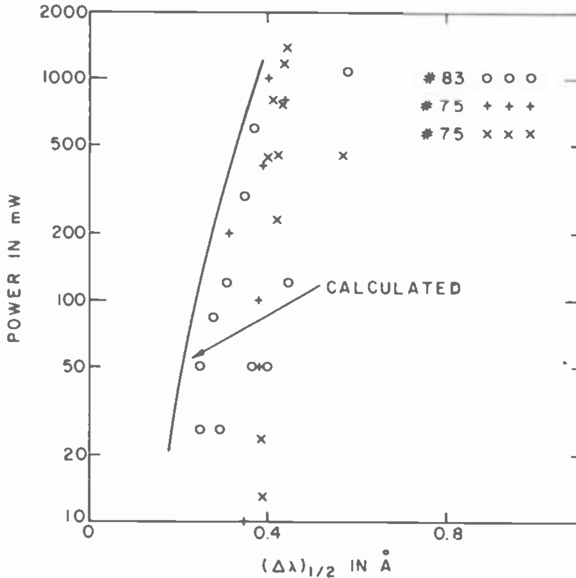


Fig. 3—Spectral resolution at various powers. (Lasers 75 and 83; lens MP_L 20X; grating 625/mm, reflection, 3rd order.) Solid line calculated from Eqs. [4] and [7] and experimental values of X versus P.

without the grating; however, the width at fixed power is appreciably reduced, because the high density of modes in the tuned line lowers the overdrive X.

4.4 Control of Lateral Modes

The ability to insert spatial filters inside the cavity gives the opportunity to control the lateral modes of the external cavity.[19] Although we could drastically modify the lateral profile, we did not demonstrate any real

improvement. We believe that our failure stemmed from the lack of a description of the lateral mode type.

For every lens that we used inside the cavity, the basic geometry of the modes was determined by parasitic reflections from elements of the lens. Their importance is shown by the general observation that the efficiency η_l was always higher with the lens in its focused position than

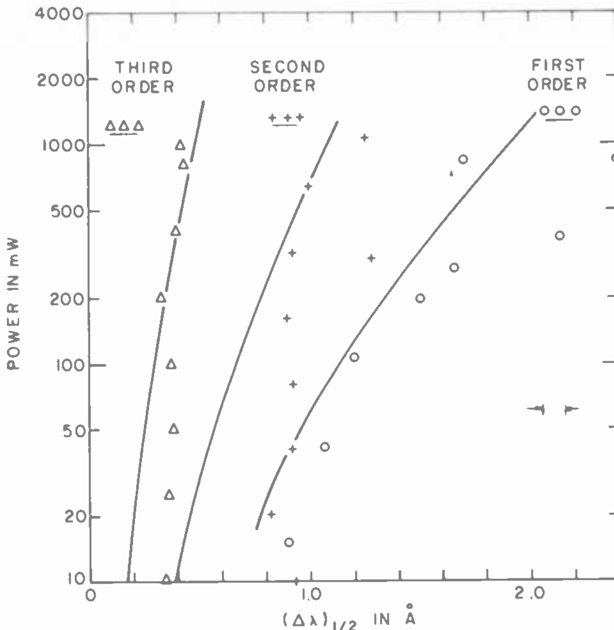


Fig. 4—Spectral resolution for different orders. (Laser 83; lens MP_L 20×; grating 625/mm, reflection.) Solid lines calculated from Eqs. [4] and [7] and experimental values of X versus P.

for the bare laser—see Table 7 for example. Not only was the lateral profile with the lens different from that with the bare cavity, but it changed markedly with very small adjustments of the lens. A narrow profile with high external efficiency requires a spatial filter designed to attenuate all unwanted modes compared to the loss in the desired low-order modes—an unwieldy design problem if the mode type is not known. Although we occasionally achieved an improvement in beam profile, we were never able to design a filter that would give a significant reproducible increase in forward intensity at fixed current.

5. Fundamental Properties

A number of papers have studied changes in powering at one frequency when lasing occurs at another, and proposed various explanations. Paoli and Ripper[14] studied the suppression of the q modes when the grating was added. They interpreted the suppression for a small wavelength separation between the q and q' modes as indicative of homogeneous broadening of the natural emission line. Bogatov et al[15] used a pair of gratings to give two narrow lasing bands and tuned one through the other. The observation that the relative powers depended on the *direction* of tuning they attributed to a stimulated scattering of the lasing light originating in a dynamical inhomogeneity of the carrier population. Wright et al[16] repeated the study of Paoli and Ripper[14] and interpreted their observations of the range of wavelength over which the tuned grating suppressed the q modes in terms of homogeneous and inhomogeneous broadening. Neither of the studies on laser suppression contain the data needed to show whether or not the observed effect is qualitatively different from the modal power sharing observed in simple lasers.

5.1 Power Sharing

We made a detailed study of the power sharing between q and q', including recordings of the complete spectra and total power and observations of changes in population inversion.* The study defines the conditions for suppression of the q resonances, which is a special case of the more general power distribution. To first approximation, we find that the power sharing is determined by the population inversion (or junction voltage), indicating the applicability of the concept of thermal equilibrium throughout each energy band (the "inhomogeneously broadened" atomic levels).[16] In particular, if the intensity of the spontaneous emission at short wavelength indicates that the junction voltage is below the threshold of the bare cavity, no q modes are lasing; if the voltage is well above threshold of both q and q' modes, the relative powering of the two groups can be roughly predicted from the spontaneous emission.

Our general approach is illustrated by studies of the effect of insertion of neutral density gelatin filters between the lens and grating. Fig. 5 shows coherent and spontaneous output against current for several

* The powering of the q modes was observed without altering the cavity by inserting a sheet of black paper between lens and grating. The small reflection from the lens, which controls the lateral modes, is of no consequence in the present context.

combinations. Coherent power without the grating (curve A) has a threshold at about 6.8 A and a forward efficiency of 4%. Reflection from the tuned grating (curve C) reduces the threshold to 4.4 A and raises the efficiency to 14%. With filter NDF-1 (one-way loss of 5 dB) (curve B), the condition is intermediate, 6.0 A and 6%.

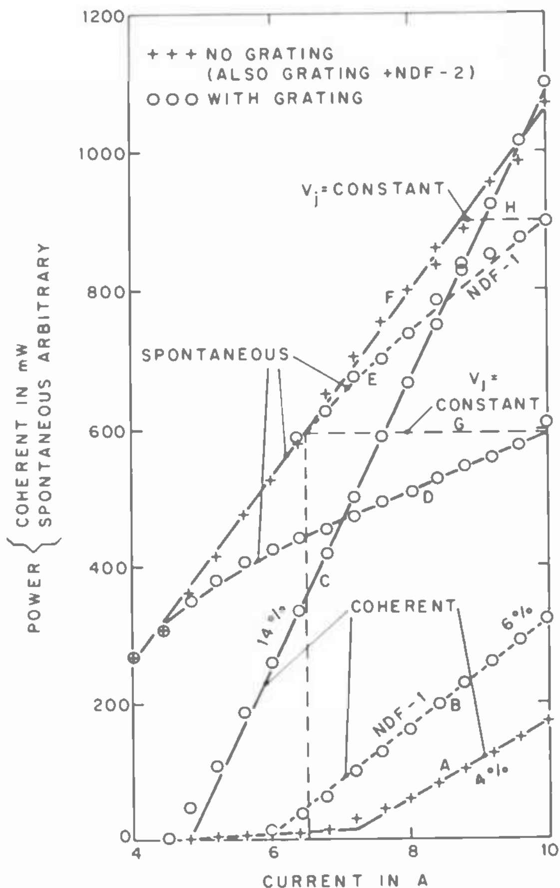


Fig. 5—Output of laser in external cavity. (Laser 41; lens MP_L 20X; grating 625/mm, reflection, 3rd order.)

The accompanying changes in junction voltage are monitored by the spontaneous emission at 8500 Å,[13,17] the upper curves. With no grating, curve F, the spontaneous emission is almost linear in current with only a slight droop above 8 A. Adding the reflection from the tuned grating

produces the marked droop of curve D, which sets in at the new threshold and by 10 A has lowered the intensity from the 1100 units of curve F to 600 units. This drop indicates a difference in junction voltage of about 0.6 KT/q, or 0.6 thermal volts (the natural log of the intensity ratio).[17] With the internal filter NDF-1, curve E, the droop begins near the new threshold and is less pronounced. At 10 A, the junction voltage for E is only about 0.15 thermal volts below curve F. With filter NDF-2 (10 dB loss each way), neither the spontaneous nor the coherent output was measurably different from the no-grating curves F,A.

This graph is a basis for predicting the power distribution between q and q' modes on the "thermal equilibrium" model, since it yields corresponding currents for different cavity conditions (currents producing the same value of junction voltage). Any horizontal line of constant spontaneous intensity is the locus of constant junction voltage. Its intersections with the different spontaneous curves are the equivalent currents. For example, the intersections of the horizontal line G with curves D and F show that 10 A for the tuned cavity is equivalent to 6.8 A with no grating. Reference to the lasing power, curves C and A, predicts that the q modes are still below threshold when the tuned cavity is excited with 10 A and radiates 1100 mW. On the other hand, with NDF-1 at 10 A the equivalent current for no grating is 8.75 A (line H), which is above the threshold of the q modes. From the value of curve A at 8.75 A, we predict that the q modes contribute about 100 of the 320 mW observed with NDF-1 at 10 A (curve B).

Modal analysis of the frequency spectra shows that these predictions are a first-order guide to partitioning between the two mode groups. The q modes form the band at 8860 Å that is found with the bare lasers (Fig. 6, curve A), while the q' modes give the narrow line at 8896 Å in the spectrum of the tuned cavity, curve C. As predicted, the q modes are not excited by 10 A in the tuned cavity.

The intermediate cases, NDF-1 and -2, are exacting tests, since the modal competition is more equal than for the two extremes. With NDF-1, we have already estimated that at 10 A the q modes were slightly above threshold and should contribute ⅓ of the laser power; the high-resolution spectrum (Fig. 7, curve C) shows that actually they are not quite at threshold. Also with NDF-2, the spontaneous emission slightly overestimates the contribution of the q modes. No difference was observed between the current dependences of either the spontaneous or coherent output from the bare laser compared to that from the tuned laser with NDF-2, which suggests negligible excitation of q' modes. The spectra A and B of Fig. 7 show a slight change; curve B has a small decrease in q mode intensity and weak lasing in the q' modes.

In summary, the current dependence of the junction voltage, as

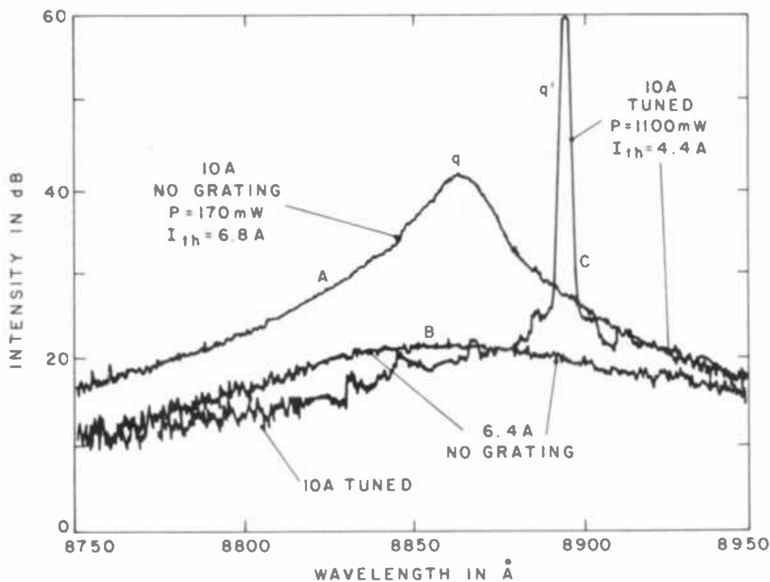


Fig. 6—Spectra of laser in external cavity at full power, with and without grating. (Laser 41; lens MP_L 20×; grating 625/mm, reflection, 3rd order.)

monitored by the spontaneous emission, gives a first-order guide to the power distribution between q modes of the semiconductor body and q' modes of the external resonator, but somewhat overemphasizes the powering of the q modes. A plausible reason for this overestimate may be that a weak reflection from the grating at the wavelength of the q

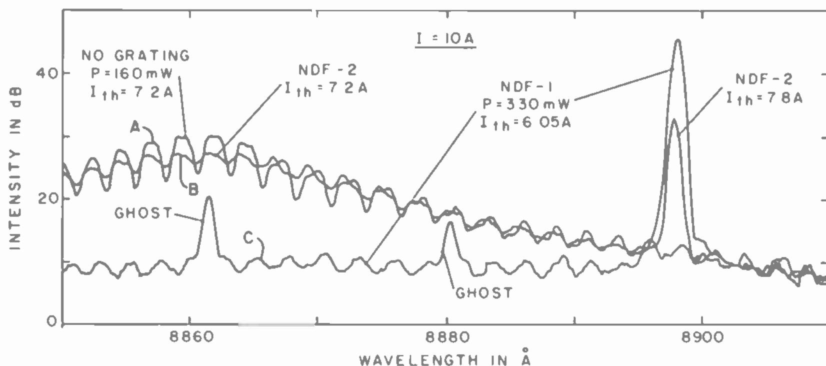


Fig. 7—High resolution spectra of laser in external cavity with internal absorbers. (Laser 41; lens MP_L 20×; grating 625/mm, reflection, 3rd order; absorbers NDF-1,2, one-way transmission -5, -10 dB.)

modes has slightly lowered their Q compared to that of the bare laser cavity, thus increasing the threshold and reducing the output. This explanation has not been documented, but it is consistent with the frequent observation of a lowered peak-to-valley ratio with the tuned cavity (compare curves A and B of Fig. 7).

5.2 Suppression of Lasing

In our analysis of power partitioning in steady state, we assumed that the carrier populations are described by Fermi levels for each energy band that are uniform throughout the bands. The partitioning of the

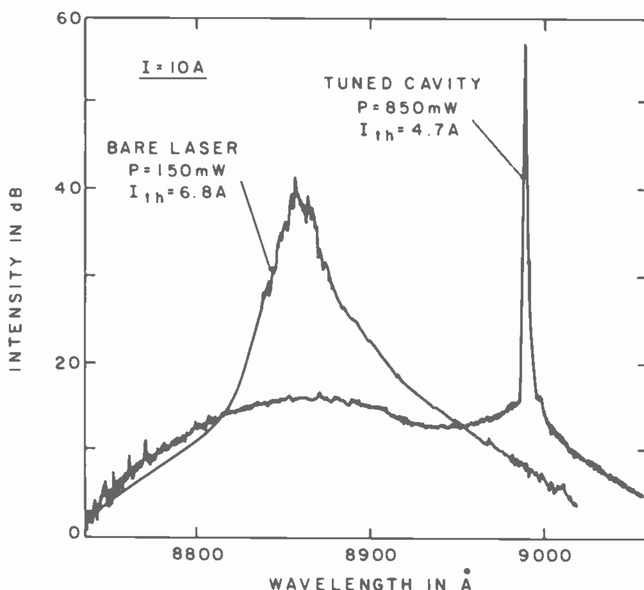


Fig. 8—Extended spectrum at full drive; grating at long wave limit 8990 Å. (Laser 41; lens Leitz 50×; grating 1200/mm, reflection.)

power is described by a gain curve, determined by the uniform Fermi levels, and the losses of each of the cavity modes.[13] The predominant effect of adding the tuned grating is addition of a set of modes with higher Q than the bare laser modes, i.e., modes with lower thresholds, in which the available power is generated at reduced Fermi levels. In this model, the suppression of lasing in the q modes depends on the passive prop-

erties of the modes of the different configurations and the power at which the study is made rather than a fundamental optical linewidth for modal interference.

A contrasting explanation is that there is a fundamental wavelength interval over which lasing in one mode reduces the gain at adjacent wavelengths.[14,16] The maximum separation over which the q' lasing suppresses q lasing has been ascribed to the homogeneous line width of the transitions; the fact that other q' modes were not suppressed was not considered to be a serious contradiction to the interpretation.[14]

The wavelength dependence of the emission is a critical test between the two suggestions. Localized gain suppression within a homogeneous line width of the lasing modes will be evidenced by a dip in emission in this interval. If, in contrast, all the semiconductor excitations are described by the position of a Fermi level that is uniform throughout the energy band, then the lowering of the lasing threshold by addition of the tuned grating should decrease the gain and hence the emission at all wavelengths, with no particular new structure adjacent to the tuned line.

Fig. 7 shows a depression of the gain by addition of the grating that varies smoothly from the tuned line at 8896 Å to 50 Å shorter. Notice the striking similarity between curves A and B to within one q mode of the tuned line. Curve C, which has a far larger increase in Q, tells the same story—no dip in gain localized to 50 Å from the tuned line.

The absence of new structure associated with the tuning wavelength is a characteristic of the spectrum for all tuning wavelengths. Fig. 8 is the spectrum at the longest tuned wavelength we could excite, some 120 Å from the q mode emission. The general shape of the maximum at 8860 Å resembles the normal untuned spectrum below threshold, while the rise on approaching the tuned line is probably from the increase of the very weak reflectivity of the grating away from resonance. In Fig. 9, grating tuned to the other extreme (8790 Å), the same analysis applies. Notice now there is weak lasing in the q modes, indicating that the junction voltage is not lowered as much by tuning to 8790 as to 8990 Å. The respective thresholds, 6.3 and 4.7 A, corroborate this conjecture.

Fig. 10 shows a special case, tuning to the strongest q mode. (This study required a laser with uncoated facet to enhance the q modes.) The spectra are at 10 A, well above the thresholds of 4.5 and 4.0 A for the bare laser and tuned cavity. The tuned spectrum shows lasing in the q modes adjacent to the q' output, with a monotonic fall in intensity on moving away from the center of the grating resonance. Again there is no apparent evidence of a warping of the gain curve due to the strong emission in the tuned line. The power with the grating was 1100 mW, shared about equally between the central peak (q' resonance) and the adjacent q

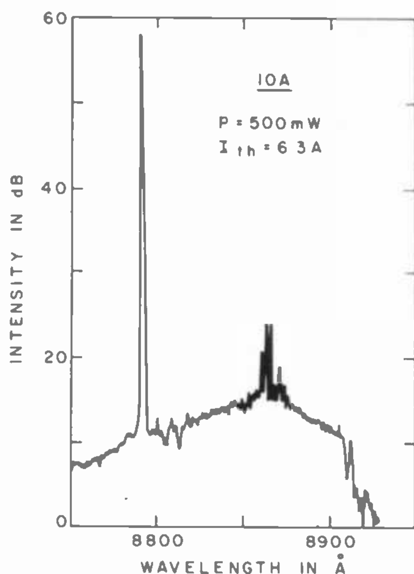


Fig. 9—Extended spectrum at full drive; grating tuned to short wave limit 8790 Å. (Laser 41; lens Leitz 50X; grating 1200/mm, reflection.)

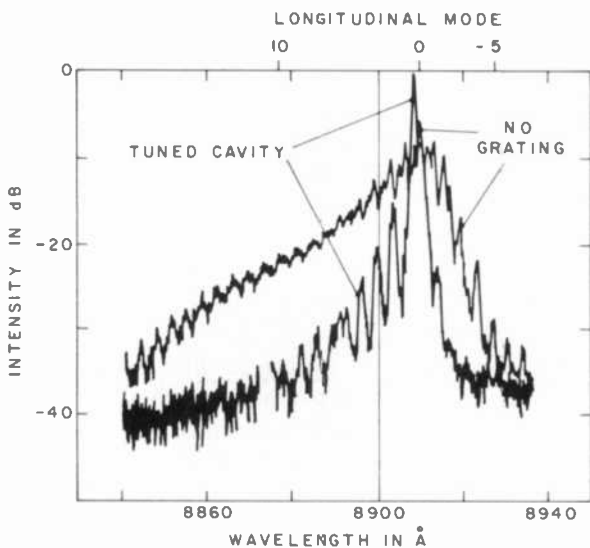


Fig. 10—Extended spectrum at full drive; grating tuned to center of lasing band of bare cavity. (Laser 1; lens MP_L 20X; grating 625/mm, reflection, 3rd order.)

modes. This equal splitting is predicted from the spontaneous outputs at 8500 Å.

Finally, the fact that the powering of both q and q' modes is predicted semiquantiatively from the spontaneous emission 350 Å away from the tuned line shows that the same Fermi level describes the entire emission spectrum, coherent and incoherent, no matter what the separation between tuned and untuned lines.

In summary, all these spectra are consistent with the description of the state of the semiconductor by uniform Fermi levels. Since there is no evidence of a localized change in property close to the lasing lines, the studies of suppression of lasing in the q modes do not seem to bear on the existence or extent of homogeneous broadening.[14,16]

5.3 Dependence of Power on Junction Voltage

One interest in this study is determining the similarity of the performance of the different cavities, and in particular the extent to which the differences arise from changes in the normal modes as opposed to changes in recombination dynamics.

The plot of power against the reduced variable X, the overdrive parameter,[13] gives direct evidence on this. X is approximately the fractional increase from threshold of the spontaneous emission at short wavelength. For values up to unity, it is approximately the change from threshold of the junction voltage in thermal volts,[17] a quantity with direct theoretical link to the emission spectrum.[17,18]

The overdrive and power are connected[13]

$$P \approx N_{1/2} X P^{*}, \qquad [10]$$

with $N_{1/2}$ meaning the total number of cavity modes excited to half the power of the strongest. The critical power P^* describes the strength of the nonlinearity that produces the multimoding. We wish to see how much of the change with cavity configuration is explained by change in $N_{1/2}$, and whether there is evidence of a simultaneous change in P^*.

To deduce the total power from the output P_1 emitted by the front facet we model the composite of back facet, lens, absorber, and tuned grating as an effective reflectivity at the back facet R_{eff} and a net radiation through it P_{eff}.

$$R_{eff} = 0.011 + 0.3\, T^2. \qquad [11]$$

The first term is the reflectivity of the back facet plus lens deduced from the front-to-back power ratio with the lens in place, Eq. [1]. The other term is the grating reflectivity (≈ 0.3) reduced by the round-trip atten-

Table 8—Calculated Parameters of External Cavity

Absorber	I_{th} (A/cm²)	η_1[a] (%)	R_{eff}	P/P_1[a]
∞	6.5	3.1	0.010	9.1
NDF-1.0	6.3	4.6	0.014	6.9
NDF-0.6	5.6	6.9	0.050	3.5
None	4.4	12.0	0.40	1.7

[a] η_1, P_1 – from front facet.

uation T^2 of the absorber. P_{eff} is related to the measured forward emission P_1 by inserting R_{eff} and R_1 in Eq. [1].

Table 8 gives measured and calculated parameters of the compound cavities. The first column is the absorber between lens and grating, with "infinity" and "none" designating black paper or no absorber. Next come the measured thresholds and external efficiencies through the front facet, and last the effective reflectivities of the back facet and the power correction factors $(1 + P_{eff}/P_1)$ deduced as just described.

The plot of P versus X is shown in Fig. 11. The expected dependence of $N_{1/2}$ on cavity parameters should produce the following effects:

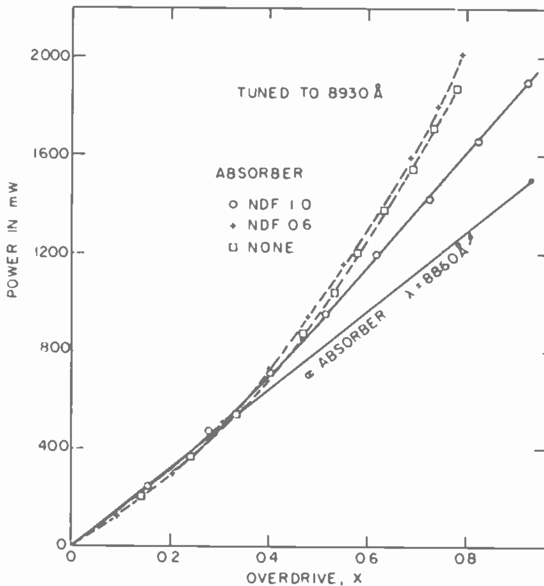


Fig. 11—Power versus overdrive with different internal absorbers. (Laser 41; lens Leitz 50X; grating 625/mm, reflection, 3rd order.) One-way transmissions of NDF-0.6, 1.0, and ∞ are -3 dB, -5 dB, and zero.

(1) Insertion of NDF-0.6 reduces the Q of the external grating, thus broadening the spectrum of q' modes. At fixed X, this increases $N_{1/2}$ and the power, raising the NDF-0.6 curve above the curve for no absorber.
(2) NDF-1 should have higher power than with infinite absorber because the tuned line contains more q' modes than the number of q modes in the bare cavity spectrum.
(3) Comparison between the solid and dashed curves is not appropriate because of the large differences in effective reflectivities. The formalism leading to the explicit form of Eq. [10] has not been examined over extended changes in front-to-back ratios or for extremely low facet reflectivity.

The uncertainties in derivation of R_{eff} and in the correction to deduce the total radiated power prevent definitive conclusions from Fig. 11. Nevertheless, the great similarities of all four curves and the correspondence between observed changes and expectations based only on modal content give evidence of the importance of the types and Q factors of the normal modes. There is no indication that the nonlinearity that controls the modal distribution is seriously affected by changes in the external cavity.

6. Summary

The threshold and efficiency of the compound cavity depend markedly on the internal lens but little on the grating. With a variety of microscope lenses with N.A. of 0.3 or larger, the effective reflectivity of lens plus grating ranged from 10 to nearly 50%, depending on the alignment, while with all camera lenses the effective reflectivities were around 1%.

The efficiencies of the gratings were high. With proper lenses, the differential efficiency of the external cavity was nearly half that of the bare laser.

With the transmission grating, the effective reflectivity of lens plus grating was 5–10% and the threshold appreciably higher than with the reflection grating. The efficiency was comparable.

With the transmission grating, the beam was highly collimated, 1° lateral by 0.03° transverse to the junction. These are approximately the angles subtended at the lens by the facet illumination. At 10 A drive, the spectral radiant intensity was about 10^9 W steradian^{-1} μm^{-1}.

With 10 A drive, the wavelength tuned smoothly over 210 Å (2.4%) between half-powers with little periodicity at the spacing of the q modes and with power exceeding 500 mW.

The observed spectral width is predicted by the grating resolution and the increase in junction voltage above threshold.

The modal power distribution between bare laser and compound cavity modes follows the behavior previously found in simple lasers.

There is no evidence of a spectral interference associated with a fundamental line width of the optical transitions.

Acknowledgment

Wafer 638N-LOC was supplied by H. Lockwood, the facet coatings by M. Ettenberg, the etchings by A. Kan, and the design of the optical bench and rotating platform by R. E. Schell. We acknowledge helpful discussions with H. Kressel and D. O. North and the excellence of the assistance of L. Elbaum, M. G. Harvey, H. V. Kowger, and D. P. Marinelli in the various aspects of device fabrication.

References:

[1] R. Ludeke and E. P. Harris, "Tunable GaAs Laser in an External Dispersive Cavity," *Appl. Phys. Lett.*, **20**, p. 499, 15 June (1972).

[2] J. A. Rossi, S. R. Chinn, and H. Heckscher, "High Power Narrow Linewidth Operation of GaAs Diode Laser," *Appl. Phys. Lett.*, **23**, p. 25, 1 July (1973).

[3] J. A. Rossi, H. Heckscher, G. E. Stillman, and S. R. Chin, "Time Delays in External Cavity Controlled GaAs (GaAl) As Single Heterostructure Diode Lasers," *Appl. Phys. Lett.*, **23**, 254, 1 Sept. (1973).

[4] J. A. Rossi, H. Heckscher, and S. R. Chin, "Threshold, Spectral, and Output Power Characteristics of GaAs/(GaAl)As Single Heterostructure Diode Lasers," *J. Appl. Phys.*, **23**, p. 257, 1 Sept. (1973).

[5] H. Heckscher and J. A. Rossi, "Flashlight Size External Cavity Semiconductor Laser with Narrow Linewidth Tuneable Output," *Appl. Opt.*, **14**, p. 94, Jan. (1975).

[6] J. A. Rossi, J. J. Hsieh, and H. Heckscher, "Gain Profile and Time Delay Effects in External Cavity Controlled GaAs Lasers," *IEEE J. Quantum Electron*, **QE-11**, Part II, p. 538, July (1975).

[7] H. F. Lockwood, H. Kressel, and H. S. Sommers, Jr., "An Efficient Large Optical Cavity Injection Laser," *Appl. Phys. Lett.*, **17**, p. 499, 1 Dec. (1970).

[8] M. Ettenberg, H. F. Lockwood, and H. S. Sommers, Jr., "Radiation Trapping in Laser Diodes," *J. Appl. Phys.*, **43**, p. 5047, Dec. (1972).

[9] M. Ettenberg, H. S. Sommers, Jr., H. Kressel, and H. F. Lockwood, "Control of Facet Damage in GaAs Laser Diodes," *Appl. Phys. Lett.*, **18**, p. 571, 15 June (1971).

[10] H. S. Sommers, Jr., "Experimental Properties of Injection Lasers. IV. Modes of Large Cavities with Sawed Sides," *J. Appl. Phys.*, **44**, p. 3601, Aug. (1973).

[11] J. C. Picqué, S. Roizin, H. H. Stroke, and O. Testard, "CW, Single-Mode, Tuneable GaAs Laser System with Good Frequency Stability," *Appl. Phys.* **6**, p. 373 (1975).

[12] Yu. M. Mirinov, V. I. Molochev, V. V. Nikitin, and A. S. Semenov, "Investigation of Emission Linewidth of an Injection Laser in Strong and Weak Fields," *Sov. J. Quantum Electron*, **6**, p. 123, Jan. (1976).

[13] H. S. Sommers, Jr. and D. O. North, "The Power Spectrum of Injection Lasers: The Theory and Experiment on a Nonlinear Model of Lasing," *Solid-State Electron.*, **19**, p. 675 (1976).

[14] T. L. Paoli and Jose E. Ripper, "Single Longitudinal Mode Operation of CW Junction Lasers by Frequency Selective Optical Feedback," *Appl. Phys. Lett.*, **25**, p. 744, 15 Dec. (1974); T. L. Paoli, J. E. Ripper, A. C. Morosini, and N. B. Patel, "Suppression of Intensity and Self Pulsations in CW Junction Lasers by Frequency Selective Optical Feedback," *IEEE J. Quantum Electron.*, **QE-11**, Part II, p. 525, July (1975).

[15] A. P. Bogatov, P. G. Eliseev, and B. N. Sverdlov, "Anomalous Interaction of Spectral Modes in Semiconductor Laser," *IEEE J. Quantum Electron.*, **QE-11**, Part II, p. 510, July (1975).

[16] P. D. Wright, J. J. Coleman, N. Holonyak, Jr., M. J. Ludowise, and G. E. Stillman, "Homogeneous or

Inhomogeneous Line Broadening in a Semiconductor Laser: Observations on (InGa) (PAs) Double Heterojunctions in an External Grating," *Appl. Phys. Lett.*, **29**, p. 18, 1 July (1976).

[17] D. O. North and H. S. Sommers, Jr., "Saturation of the Junction Voltage in Stripe-Geometry (AlGa)As Double Hetero-Structure Junction Lasers: A Comment," *Appl. Phys. Lett.*, **30**, p. 116, 15 Jan. (1977).

[18] H. S. Sommers, Jr. and D. O. North, "Spontaneous Power and the Coherent State of Injection Lasers," *J. Appl. Phys.*, **45**, p. 1787, April (1974).

[19] E. M. Phillip-Rutz, *IEEE J. Quant. Electron.*, **QE-8**, p. 632, July (1972).

The Effects of Emitter Impurity Concentration on the High-Current Gain of Silicon n-p-n Power Transistors

Ramon U. Martinelli

RCA Laboratories, Princeton, N. J. 08540

Evelyn Jetter

RCA Solid State Div., Somerville, N. J. 08876

Abstract—Using n-p-n power transistors with virtually identical base and collector regions, we demonstrate that devices with a low concentration phosphorus emitter (LCE) have an improved high-current gain compared with devices having a high concentration phosphorus emitter (HCE). The LCE has a surface concentration in the range 5 to 7 $\times$ 10^{19} cm^{-3}; that of the HCE is about 5 $\times$ 10^{20} cm^{-3} or greater. From the dependence of the cutoff frequency f_T on collector current at a given collector emitter voltage, we show that the improved high-current gain of the LCE results from the improved injection efficiency of the periphery and sidewalls of the LCE.

1. Introduction

A reduction of the surface impurity concentration in the emitter from levels beyond 5 $\times$ 10^{20} cm^{-3} to levels around 5 to 7 $\times$ 10^{19} cm^{-3} is known to increase the emitter efficiency and, hence, the dc current gain of bipolar transistors at low collector currents.[1-3] A lower surface concentration in the emitter implies a smaller areal charge density in the emitter for a given emitter–base junction depth. Compared with a high-concentration emitter (HCE), the low-concentration emitter (LCE) has an effectively higher minority carrier lifetime, which may result from less

damage in the emitter,[4] a longer Auger recombination lifetime,[5] or the absence of bandgap narrowing at high doping levels.[6] Irrespective of the specific mechanism involved, the LCE is more effective than the HCE in reducing hole currents from the base into the emitter.

Calculations by Mock[7] suggest that the LCE is also more efficient at high collector currents, where base-current crowding and base-widening effects predominate. At high collector currents, the periphery and sidewalls of the emitter comprise the active injection area, and it is expected that these portions of the LCE will also more efficiently block hole flow from the base into the emitter. The purpose of this paper is to demonstrate the improved injection efficiency of the LCE at high current levels.

We begin with a detailed discussion of the transistors used in the comparative experiments. A model of transistor behavior at high current densities is then outlined, followed by a presentation of our experimental results and a discussion of the data.

2. Device Description

The transistors used in these experiments were patterned after a 3 A, 300 V power transistor. Fig. 1 shows the horizontal geometry. The chip is a 0.46 cm square, the emitter area A_e is 0.097 cm^2, and the emitter

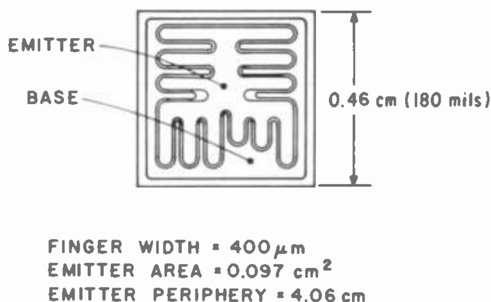


Fig. 1—Geometrical pattern of the base and emitter of the test devices.

periphery, l, is 4.1 cm. The emitter finger widths are 400 μm, which implies that curvature in the periphery can be neglected at high collector currents where the device is severely current crowded. Indeed, we will see in Sec. 4 that typical crowding distances at the onset of base widening are 30 to 60 μm, so the periphery will be assumed to be linear. Contact

by the metallization is made over the entire base and emitter areas, which are shown in Fig. 1.

Fig. 2 schematically depicts the vertical structure of the device. The collector is a 60 μm thick n^- epitaxial layer, doped to about 3×10^{14} cm^{-3}, that is grown on an n^+ substrate doped to 10^{19} cm^{-3}. Part of the base is a 30 μm thick p^- epitaxial layer, doped to 3×10^{14} cm^{-3}, that is grown on the n^- collector. Into this p^- layer a p^+ base is diffused 20 μm deep.

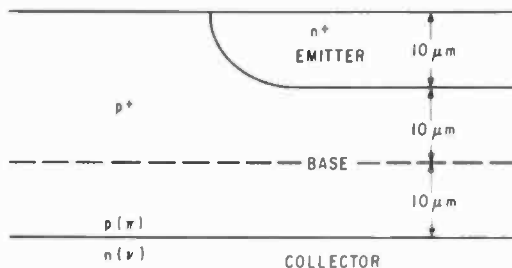


Fig. 2—Schematic diagram of the vertical structure.

At this point in the fabrication process, the given wafer is cut in half. Into one half an LCE is diffused, and into the other half an HCE is diffused. The lower concentration of the LCE is obtained by depositing phosphorus from a $POCl_3$ vapor source at a lower temperature than that used for the HCE deposition. Each emitter is diffused at a given temperature to the same depth, about 10 μm, such that the areal charge density in the base under each emitter, N_G, is the same. This equality of N_G requires, of course, that the LCE, with its lower surface concentration, be diffused some four to five times longer than is the HCE to achieve the same depth. In this way, HCE and LCE devices are fabricated with virtually identical base and collector regions. Except for the longer emitter diffusion time of the LCE, all the units are processed together in identical fashion.

The emitter sidewall area, which is roughly the emitter depth times the emitter periphery, is about 4% of the geometrical area A_e.

Fig. 3 shows the impurity concentration profile of an LCE and an HCE. These data were derived from a two-point-probe spreading-resistance scan of two angle-lapped devices.[8] The HCE has a surface concentration of about 5×10^{20} cm^{-3}. The LCE has a surface concentration of 7×10^{19} cm^{-3}, nearly an order of magnitude less than that of the HCE. The emitter areal charge density, Q_E, of the LCE is also one tenth that of the HCE. These concentrations and densities were derived

from the spreading-resistance data. Notice that the base doping profiles for each device are virtually identical. Each device has a base areal charge density or Gummel number, N_G, of 3.2×10^{13} cm^{-2}, which was determined from an analysis of the $\ln(I_C)$ versus V_{EB} characteristic of each unit, after Gummel.[9]

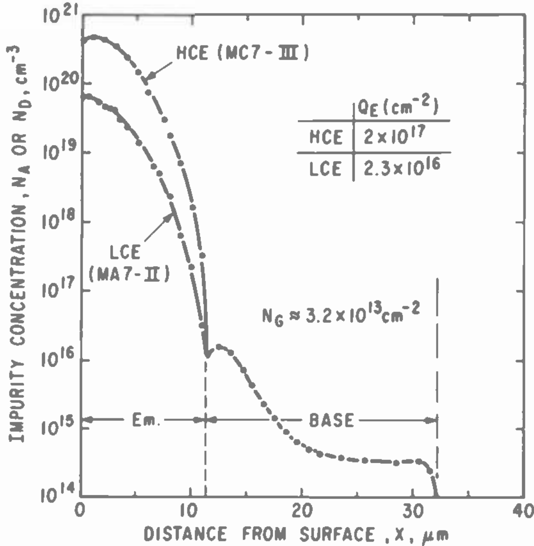


Fig. 3—Comparison of the LCE and HCE emitter and base impurity concentration profile.

Fig. 4 shows the emitter, base, and collector impurity concentration profile for the HCE of Fig. 3. The collector region is about 60 μm wide and has an average collector doping density N_D of about 5×10^{14} cm^{-3}. The areal charge density in the lightly doped portion of the collector is about 0.1 N_G, which fact we shall refer to later on.

All the devices fabricated for these experiments had the general structure presented above. To give a variety of low collector-current gains, N_G was varied over the range 10^{13} to 10^{14} cm^{-2}. Comparisons of current-gain characteristics were made between LCE and HCE devices from split wafers.

3. Models of Transistor Behavior

At low collector currents, where conductivity modulation of the collector and ohmic drops in the base are negligible, the one-dimensional models

of transistor action apply. As the collector current increases, two major effects enter. (1) The collector current crowds toward the emitter periphery, owing to the finite sheet resistance of the base beneath the emitter ρ_{BS}. Fletcher has analyzed this effect.[10] (2) The collector current density J_c increases rapidly as the result of crowding and increased base

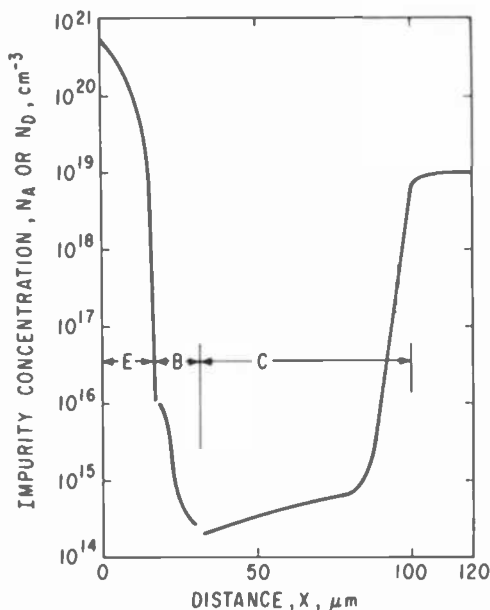


Fig. 4—Vertical impurity concentration profile of an HCE device showing lightly doped collector region.

drive, leading to the onset of base widening in the lightly doped collector at some critical collector current I_{crit}. Since the transit time of injected electrons increases rapidly as the base widens, the onset of base widening is signaled by a decrease in the cutoff frequency f_T, which is inversely proportional to the transit time.[11]

The collector current is described by Gummel's expression[12]

$$I_C = qA_eD_{NB}n_i^2\left[\frac{\exp(qV_{EB}/kT) - \exp(qV_{BC}/kT)}{Q_B}\right]. \qquad [1]$$

D_{NB} is the electron diffusion constant in the base, n_i is the intrinsic carrier density, and Q_B is the areal charge density of holes in the base, which, at low collector currents, equals N_G. We have neglected any base–collector leakage currents.

A similar expression can be obtained for the base current as follows:[13]

$$I_B = qA_eD_{PE}n_i^2\,[\exp(qV_{EB}/kT) - 1]/Q_E + C_B[\exp(qV_{BE}/nkT) - 1]. \qquad [1a]$$

Here D_{PE} is the hole diffusion constant in the emitter, and Q_E is the areal charge density of electrons in the emitter. The first term in Eq. [1a] is the ideal portion of the base current. The second term is the nonideal component of base current, and the coefficient C_B lumps together all the sources of recombination between base and emitter currents, such as bulk and surface recombination currents in the emitter–base depletion layer. The nonideal term has an exponential dependence on V_{EB} that is slightly different from that of the ideal term in that the factor kT is multiplied by n. In general, n varies from 1 to 3. The nonideal term dominates the base current at low values of V_{EB} and, hence, small values[14] of I_C. At low collector currents, the temperature dependence of I_B is, likewise, dominated by the nonideal terms.[13]

For a transistor that has ideal base and collector currents and that is injection-efficiency limited* (that is, minority recombination in the base is negligible), the dc current gain h_{FE} is given by[15]

$$h_{FE} = \frac{D_{NB}Q_E}{D_{PE}N_G}. \qquad [2]$$

It is assumed, of course, that qV_{EB} and $-qV_{BC} \gg kT$. From Fig. 3 typical values of Q_E/N_G are 10^3 to 10^4, so that low-current gains of about 10^3 should be seen for values of $D_{NB}/D_{PE} \approx 0.1$. The much lower values observed have been explained by postulating an *effectively* low lifetime in the emitter for injected holes.[16]

For nonideal base currents, Eq. [2] becomes more complicated, and the simple interpretation of gain being proportional to Q_E cannot be made. When dividing I_C by I_B to obtain h_{FE}, the exponential terms in Eqs. [1] and [1a] do not cancel, and the temperature dependence of h_{FE} is governed by the quotient of the exponential terms in Eqs. [1] and [1a], with the result that h_{FE} increases with temperature at low collector currents where the base currents are nonideal. Alternatively, one can account for the presence of nonideal base currents by assuming that only a small fraction of Q_E is effective in blocking hole current. As we shall see, the LCE is more effective than the HCE in blocking hole current from the base, even though its Q_E is one tenth that of the HCE.

* Our devices give no evidence for recombination losses in the base using George and Clark's criterion,[26] and we assume that our devices are emitter efficiency limited.

We now turn to a description of current gain at high collector currents, where base-current crowding and base widening occur. At high base and collector currents, the ohmic drops in the base beneath the emitter debias the emitter–base junction causing current to crowd near the emitter periphery. Using Fletcher's analysis of current crowding,[10] the collector current density as a function of distance from the emitter edge is given by

$$J_C(X) = \frac{J_C(0)}{[1 + (X/X_0)]^2}. \quad [3]$$

$J_C(0)$ is the current density at the emitter edge, $X = 0$, which is taken at the point on the periphery where the curved sidewall area meets the plane of the emitter–base junction. X is the distance from the periphery into the interior of the base beneath the emitter. X_0 is the characteristic crowding distance given by

$$X_0 \approx \left[\frac{2kT}{q}\frac{h_{FE}}{(\rho_{BS}J_C(0))}\right]^{1/2}. \quad [4]$$

If X_0 is much smaller than an emitter finger width, Eq. [3] can be integrated to infinity to obtain an expression for the total collector current

$$I_C = lJ_C(0)X_0. \quad [5]$$

Eliminating $J_C(0)$ between Eqs. [4] and [5] and using $h_{FE} = I_C/I_B$, we have

$$X_0 = \frac{2kTl}{q\rho_{BS}I_B}. \quad [6]$$

For our present devices, $l \cong 4$ cm and $\rho_{BS} \cong 2 \times 10^3$ ohms/□, which gives

$$X_0 = I_B^{-1}\ \mu\text{m}. \quad [7]$$

The critical collector current density for the onset of base widening is[17–19]

$$J_{crit} = qN_D\mu_{NC}V_{BC}/W_C, \quad [8]$$

where μ_{NC} is the majority carrier mobility for electrons in the lightly doped collector region and V_{BC} is the applied collector–base voltage. J_{crit} is the current density at which the mobile electron space-charge density carrying the current equals the collector doping density N_D. J_{crit} is also the current density at which the collector–base junction is at zero volts applied bias with all of the applied base–collector potential dropped across the lightly doped collector region. It is assumed that the electric

field in the collector, V_{BC}/W_C, is small in the sense that the electron drift velocity is not saturated. In our case, with $V_{BC}/W_C \approx 500$ V/cm, the devices operate well below the region of saturated drift velocity.[20] At current densities above J_{crit}, the electrical base moves rapidly toward the n^+ substrate.[21]

If we denote the critical collector current for the onset of base widening, I_{crit}, as the current at which $J_C(0) = J_{crit}$ in Eqs. [4] and [5], then from Eqs. [5], [6], and [8]

$$I_{crit} = \left[\frac{2qN_D D_{NC} h_{FE} l^2 V_{BC}}{\rho_{BS} W_C}\right]^{1/2}, \qquad [9]$$

where $D_{NC} = \mu_{NC} kT/q$ is the diffusion constant for electrons in the n^- collector. For typical devices under consideration, $N_D = 5 \times 10^{14}$ cm^{-3}, $\mu_{NC} = 10^3$ cm^2/(V sec), $V_{BC} \approx 2$ V, $W_C = 60$ μm, $\rho_{BS} = 2 \times 10^3$ ohms/□, and $h_{FE} = 30$, $I_{crit} = 0.56$ A, and $X_0 = 54$ μm. Hence, at collector currents in the vicinity of 0.6 A, base widening is expected to begin.

When the base impurity profile concentration can be simply described, the cutoff frequency f_T decreases as W_B^{-2}, where W_B is the base width.[22] As base widening begins, f_T decreases, and the collector current at which this decrease starts is a measure of I_{crit}. For the devices under discussion, the unwidened base width is about 20 μm. When fully widened to the underlying substrate, the base width is 80 μm, or four times wider than it is at low currents. Therefore, f_T is expected to drop to 1/16 of its low-current value. However, if we assume that the mobile space charge equals N_D everywhere in the widened base, the amount of areal charge density added to Q_B when the base is fully widened is small, since $N_D W_C$ is only 10% of N_G. Qualitatively, h_{FE}, which goes as Q_B^{-1}, will remain constant during the initial stages of base widening, but f_T, which is more sensitive to base widening, will decrease as soon as base widening commences at I_{crit}.

The difference between the sensitivities of h_{FE} and f_T to base widening can be seen from their respective dependencies on the areal density of holes, Q_B, in the base. If $N_B(y)$ is the total density of holes in the base as a function of distance, y, normal to the emitter–base junction,

$$Q_B = \int_0^{W_B} dy N_B(y). \qquad [10]$$

h_{FE} goes as Q_B^{-1}, as mentioned above, and $Q_B = N_G$ at current levels well below I_{crit}.

Since f_T is inversely proportional to the transit time for injected electrons in the base,[11]

$$f_T \propto \left[\int_0^{W_B} \frac{dy}{D_{NB} N_B(y)} \int_y^{W_B} dz N_B(z) \right]^{-1}. \quad [11]$$

f_T depends on the spatial distribution of $N_B(y)$, *not* on any scale factor multiplying $N_B(y)$ that directly changes the magnitude of Q_B as in Eq. [10]. In Eq. [11], D_{NB} may be a function of distance in the widened base.

As current crowding causes an increased fraction of electron and hole injection to occur at the periphery and sidewalls of the device, our one-dimensional picture loses significance. Electron injection along the sidewalls of the emitter occurs at regions where the emitter–base junction is farther from the collector. These effects imply that sidewall-injected electrons see a larger areal density of holes in the base, and hence for a given V_{EB}, I_C is lower. Effectively, N_G increases for sidewall injection. The transit time for sidewall-injected electrons also increases, and hence f_T decreases. Under severe current-crowding conditions (i.e., $I_C \gg I_{crit}$), the simple one-dimensional model outlined above breaks down, and crowding distances of the order of the emitter depth (10 μm) become meaningless. Nevertheless, it is qualitatively clear that both f_T and h_{FE} will continue to decrease in this high-current regime beyond I_{crit}, as the effects of sidewall injection become increasingly important.

The conclusions to be drawn from our qualitative model are: (1) as I_C increases, f_T is the first to decrease, signaling the beginning of classical base widening in the presence of current crowding; (2) h_{FE} will roll off as base widening advances and as sidewall injection becomes significant; and (3) under comparable conditions, the emitter with the highest efficiency will show the highest h_{FE} at all current levels.

At elevated temperatures, the LCE is expected to continue having higher gains than the HCE at large values of I_C. For each type of emitter, the gain will decrease with increasing temperature at large values of I_C owing to the increase in the resistivity of the lightly doped collector region and to the decrease in minority carrier mobility in the widened base. The low-current gain, however, generally increases with the increasing temperature as a result of the temperature dependence of the nonideal base currents. These nonideal components of the base current dominate the temperature dependence of h_{FE} at low collector currents.[13,14] As a function of increasing temperature, then, the LCE is expected to maintain its improved performance over the HCE, even though at high collector currents, both emitters will display a decreased h_{FE}.

4. Experimental Results

A comparison of the LCE and HCE at low collector currents is shown in Fig. 5. The peak dc gain, h_{FE}, taken at $V_{CE} = 50$ V is plotted versus the Gummel number. This large value of V_{CE} insures that the devices are operating deep in the active mode, far from the current-crowded regime. Gains were measured using low-duty-cycle (10 to 30 Hz) 100 to 300 μs pulses to eliminate heating effects. Peak gains occurred at I_C =

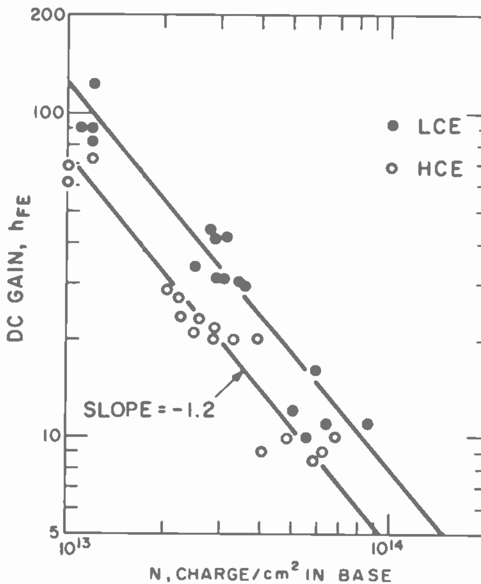


Fig. 5—Comparison of low-current gain for LCE and HCE devices.

50 to 100 mA. To obtain N_G, ln(I_C) was measured as a function of V_{EB} and was extrapolated to $V_{EB} = 0$. N_G was then calculated from Eq. [1], assuming $D_{NB} = 10$ cm²/sec. In all cases, Eq. (1) described the data for I_C to better than one percent from $I_C = 10^{-10}$ to 10^{-3} A.

Fig. 5 shows that for a given value of N_G, the LCE produces a 60% higher gain than does the HCE. This higher value of h_{FE} results, in turn, from the more efficient blocking of holes from the base into the emitter by LCE. The particular source of the improvement, whether it be longer hole lifetime, less physical damage, or less induced strain in the LCE, has not been ascertained. However, Fig. 5 summarizes our experience. The LCE produces a higher gain for a given N_G at low values of I_C.

The straight-line fits to the data of Fig. 5 have a logarithmic slope of -1.2, which means $h_{FE} \propto N_G^{-1.2}$. According to Eq. [1], h_{FE} is proportional to N_G^{-1}, not $N_G^{-1.2}$. In calculating N_G, we assumed D_{NB} = 10 cm^2/sec, and at large values of N_G, D_{NB} may decrease sufficiently rapidly with increased base doping density to add to the exponent of N_G an additional term[23] of -0.2. In any case, the difference in the exponent does not change our conclusion that at low collector currents, the LCE is more efficient than the HCE.

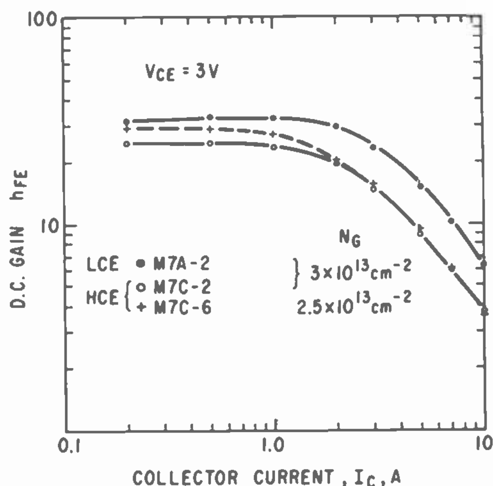


Fig. 6—Comparison of LCE and HCE devices over a broad range of collector currents.

The high current gain is compared in Fig. 6, where the h_{FE} versus I_C characteristics for three devices that have virtually identical base and collector regions are shown. (V_{CE} = 3 V, as it is for all subsequent data of this type.) Devices M7A-2 and M7C-2 have measured Gummel numbers of 3×10^{13} cm^{-2}, and device M7C-6 has $N_G = 2.5 \times 10^{13}$ cm^{-2}, so that the three devices are very nearly identical. The two C units are HCE devices, the A unit is an LCE. The data of Fig. 6 show that the LCE device M7A-2 is more efficient than its comparable HCE companion M7C-2. At both low and high currents the h_{FE} versus I_C characteristic for the LCE is greater, and the gain at I_C = 10 A is about a factor of 1.5 greater for the LCE than for the HCE. Device M7C-6 was chosen for comparison because its N_G is lower than the N_G of M7C-2 and, hence, its low-current gain is nearly equal to that of the LCE. Even so, at high

collector currents, where h_{FE} decreases with increasing I_C, the gain of unit M7C-6 is lower than that of the LCE, and it is identical to the gain of HCE unit M7C-2. Irrespective of the value of N_G, the higher current gain of an LCE will be higher than that of an HCE for a given base and collector profile. Of course, this statement holds for modest variations in N_G. If the value of N_G is greatly reduced for the HCE device, the base width decreases, and the doping profile changes. The devices are then no longer comparable. For minor variations in N_G of about 30%, the LCE has the higher h_{FE} at larger values of I_C.

The data of Fig. 6 illustrate a second major point. Notice that the higher current h_{FE} of the HCE is independent of N_G, whereas the low current gain shown in Fig. 5 varies inversely as a power of N_G. These different behaviors of h_{FE} with respect to N_G are simply explained. At low collector currents, the entire emitter area is injecting into the base and the simple one-dimensional model holds; the sidewalls of the emitter are, recall, only 4% of A_e. In this regime we expect N_G to affect the gain. At higher collector currents, the injected emitter current is crowded toward the emitter periphery. Sidewall injection plays an important role, and injection into the base is taking place in regions where the charge density in the base is greater than that given by N_G, which is measured in the interior of the emitter area. There are actually two parallel devices operating over the range of collector currents shown in Fig. 6. At low currents, the one-dimensional device whose emitter area is A_e is dominant. At high currents, it is the crowded, base-widened transistor at the emitter periphery that carries the greater fraction of I_C. Fig. 6 shows that the LCE is more efficient than the HCE in blocking hole currents from the base, even when injection takes place at the emitter sidewalls and periphery.

Fig. 7 shows dc gain characteristics for an LCE and an HCE device, both of which have higher low-current gain than the devices of Fig. 6. Again we see that at high I_C the LCE has the higher gain, even though the HCE has the higher gain at lower values of I_C, which results from its lower value of N_G. This once more exemplifies the point that for a given base and collector structure, the LCE has the higher gain at large values of I_C, where h_{FE} does not depend on N_G.

We have argued that at large values of I_C, injection occurs in the peripheral region of the emitter and that the base is widened. We also calculated in Sec. 3 that the critical collector current for the onset of base widening was, for our devices, typically about 0.6 A. Data that corroborate these notions are shown in Fig. 8, where measured values of h_{FE} and f_T are plotted against I_C. For both the HCE and the LCE, f_T begins to fall with increasing I_C and $I_{crit} \cong 0.5$ A, which is about the value of I_{crit} given by Eq. [9]. Note that h_{FE} for the LCE of Fig. 8 is higher than that

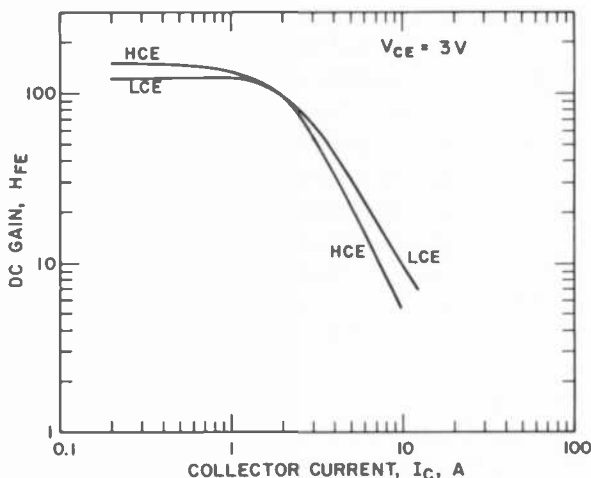


Fig. 7—Comparison of LCE and HCE devices, which have relatively high low-current gain.

of the HCE, so that I_{crit} for the LCE should be about 14% higher. From our data, we cannot distinguish a difference this small.

Beyond I_{crit} = 0.5 A, base widening and current crowding develop rapidly, as can be inferred in Fig. 8 from the rapid drop of f_T with I_C at collector currents above I_{crit}. In the vicinity of I_{crit}, h_{FE} is, however, constant with I_C, and not until I_C is 1.5 to 2.0 A does h_{FE} begin to fall. At these collector currents, current crowding and base widening are advanced to the point where, as discussed in Sec. 3, it is meaningless to

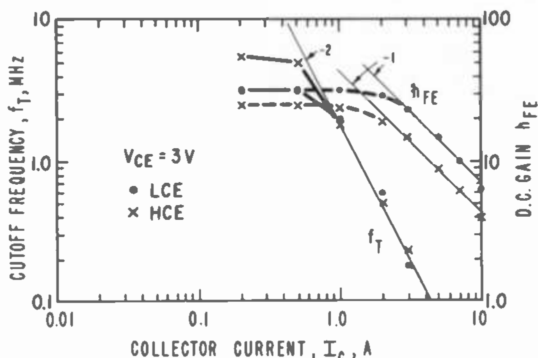


Fig. 8—f_T and h_{FE} characteristics for comparable LCE and HCE transistors.

ascribe a crowding distance to the situation, and virtually all injection takes place in the peripheral regions of the emitter.

The reason h_{FE} remains constant as the base widens is that the electrical base can move well into the lightly doped collector without adding significantly to the total charge Q_B. That is, if, qualitatively speaking, the entire collector were exactly compensated by the mobile charge

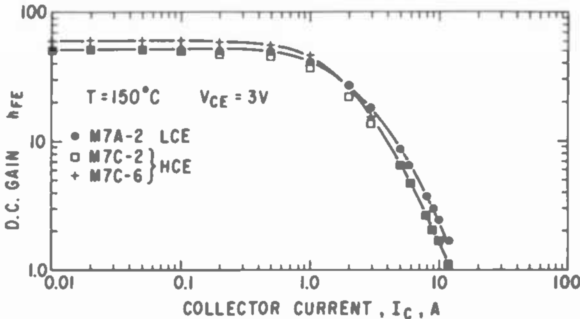


Fig. 9—Improved high-current performance of the LCE at elevated temperatures.

carrying the current, the additional holes in the base would be $N_D W_C \cong 0.1\ N_G$, which would increase the base charge by 10%.

It is clear from Fig. 8 that at the collector current where h_{FE} begins to decrease with increasing I_C, the periphery of the device is injecting all the current, and therefore the peripheral regions of the LCE are more efficient than those of the HCE.

In Fig. 8, we have indicated the dependence of h_{FE} and f_T on I_C. f_T rolls off as I_C^{-2} beyond I_{crit}. If the base width W_B widens as I_C^{-1} and if f_T decreases as W_B^{-2}, then f_T is expected to decrease as I_C^{-2}. h_{FE} decreases as I_C^{-1} for both the LCE and the HCE, as shown in Fig. 8. This behavior occurs in a regime where base widening is well advanced and where the peripheral portions of the emitter are a significant fraction of the emitter area.

Similar results have been found by other workers. Clark[24] concludes that high-current h_{FE} is limited by emitter efficiency. Whittier and Tremere[25] find f_T dropping off rapidly beyond I_{crit}, whereas h_{FE} does not begin to decrease until I_C is well above I_{crit}. Kumar and Hunter[19] also find f_T decreasing rapidly once I_{crit} is exceeded.

We have taken data of h_{FE} versus I_C at elevated temperatures for the devices shown in Fig. 6. The h_{FE} versus I_C characteristics for these transistors, taken at 150°C, are shown in Fig. 9. Although the difference

in gain is smaller, the h_{FE} for the LCE at $I_C = 10$ A is still higher than the gains for the HCE devices, which are identical, as was the case at room temperature (see Fig. 6). Measurements at temperatures between 150°C and room temperature give similar results.

The low-current gains behave anomalously with temperature. In Fig. 9 the HCE device M7C-6 has a low-current gain that is greater than the gain of M7A-2, the LCE transistor, while the gain of M7C-2 is equal to that of M7A-2. (Note that, in general, the low-current gain increases with temperature.) Since the exact values of the Gummel numbers could not be accurately measured at 150°C, it is difficult to pinpoint the source of the anomaly. Nonetheless, based on the high-current results, we infer that the LCE is again more efficient than the HCE at elevated temperatures up to 150°C.

5. Summary and Conclusions

In this paper we have studied the effects of a reduced emitter impurity concentration on the dc current gain of n-p-n power transistors. To do this, devices were made with high and with low concentration emitters, each type having virtually identical base and collector structures. HCE devices had $Q_E = 3 \times 10^{17}$ cm^{-2}, LCE devices had $Q_E = 3 \times 10^{16}$ cm^{-2}. Examining comparable devices over the temperature range 25 to 150°C, we have shown that the LCE is the more efficient emitter at both low and high collector currents. These results imply that the entire emitter area of the LCE, sidewalls and planar area, is more effective in reducing hole current from the base to the emitter than is the HCE.

Acknowledgments

The authors gratefully acknowledge the technical assistance of R. Stolzenberger and W. Ordille.

References:

[1] R. B. Fair, "Optimum Low-Level Injection Efficiency of Silicon Trransistors with Shallow Arsenic Emitters," *IEEE Trans. Elec. Dev.*, **ED-20**, p. 642 (1973).

[2] R. P. Mertens, H. J. deMan, and P. J. Van Overstraeten, "Calculation of the Emitter Efficiency of Bipolar Transistors," *IEEE Trans. Elec. Dev.*, **ED-20**, p. 772 (1973).

[3] P. J. Kannam, "Effect of Emitter Doping on Device Characteristics," *IEEE Trans. Elec. Dev.*, **ED-20**, p. 845 (1973).

[4] M. C. Duffy, F. Barson, J. M. Fairfield, and G. H. Schwuttke, "Effects of High Phosphorous Concentration on Diffusion into Silicon," *J. Electrochem. Soc.*, **115**, p. 84 (1968).

[5] W. W. Sheng, "The Effect of Auger Recombination on the Emitter Injection Efficiency of Bipolar Transistors," *IEEE Trans. Elec. Dev.*, **ED-22**, p. 25 (1975).

[6] H. J. J. deMan, "The Influence of Heavy Doping on the Emitter Efficiency of a Bipolar Transistor," *IEEE Trans. Elec. Dev.*, **ED-18**, p. 833 (1971).

[7] M. S. Mock, "On Heavy Doping Effects and the Injection Efficiency of Silicon Transistors," *Sol. State Elec.*, **17**, p. 819 (1974).

[8] N. Goldsmith, R. V. D'Aiello, and R. A. Sunshine, "The Experimental Investigation of Two-Point Spreading Resistance Correction Factors for Diffused Layers," NBS Special Publication 400-10, p. 223 (1974).

[9] H. K. Gummel, "Measurement of the Number of Impurities in the Base Layer of a Transistor," *Proc. IRE*, **49**, p. 834 (1961).

[10] N. H. Fletcher, "Some Aspects of the Design of Power Transistors," *Proc. IRE*, **43**, p. 551 (1955).

[11] J. Lindmayer and C. Y. Wrigley, *Fundamentals of Semiconductor Devices*, D. Van Nostrand Company, New York, p. 171 (1965).

[12] H. K. Gummel, "A Charge Control Relation for Bipolar Transistors," *Bell Syst. Tech. J.*, **49**, p. 115 (1955).

[13] R. U. Martinelli, "The Temperature Dependence of the DC Base and Collector Currents in Silicon Bipolar Transistors," *IEEE Trans. Elec. Dev.*, **ED-23**, p. 1218 (1976).

[14] J. E. Iwersen, A. R. Bray, and J. J. Kleimack, "Low-Current Alpha in Silicon Transistors," *IRE Trans. Elec. Dev.*, **ED-9**, p. 474 (1962); also Ref. [13] above.

[15] A. S. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley and Sons, New York, Chapt. 7 (1967).

[16] R. J. Whittier and J. P. Downing, "Simple Physical Model for the Injection Efficiency of Diffused pn-Junction," *IEDM*, paper 12.4, 1968.

[17] D. L. Bowler and F. A. Lindholm, "High Current Regimes in Transistor Collector Regions," *IEEE Trans. Elec. Dev.*, **ED-20**, p. 257 (1973).

[18] C. T. Kirk, Jr., "A Theory of Transistor Cutoff Frequency (f_T) Falloff at High Current Densities," *IRE Trans. Elec. Dev.*, **ED-9**, p. 164 (1962).

[19] R. Kumar and L. P. Hunter, "Prediction of f_T and h_{FE} at High Collector Currents," *IEEE Trans. Elec. Dev.*, **ED-22**, p. 1031 (1975).

[20] E. J. Ryder, "Mobility of Holes and Electrons in High Electric Fields," *Phys. Rev.*, **90**, p. 766 (1953).

[21] H. C. Poon, H. K. Gummel, and D. L. Scharfetter, "High Injection in Epitaxial Transistors," *IEEE Trans. Elec. Dev.*, **ED-16**, p. 455 (1969).

[22] J. L. Moll and I. M. Ross, "The Dependence of Transistor Parameters on the Distribution of Base Layer Resistivity," *Proc. IRE*, **44**, p. 72 (1956).

[23] M. B. Prince, "Drift Mobilities in Semiconductors, II. Silicon," *Phys. Rev.*, **93**, p. 1204 (1954).

[24] L. E. Clark, "High Current-Density Beta Diminution," *IEEE Trans. Elec. Dev.*, **ED-17**, p. 661 (1970).

[25] R. J. Whittier and D. A. Tremere, "Current Gain and Cutoff Frequency Falloff at High Currents," *IEEE Trans. Elec. Dev.*, **ED-16**, p. 39 (1969).

[26] W. George and L. Clark, "Experimental Determination of Gain Degradation Mechanisms," *IEEE Trans. Nuc. Sci.*, **NS-18**, p. 387 (1971).

High Level Concentration of Sunlight on Silicon Solar Cells*

L. S. Napoli, G. A. Swartz, S. G. Liu, N. Klein, D. Fairbanks, and D. Tamutus

RCA Laboratories, Princeton, N.J. 08540

Abstract—Silicon solar cells have achieved 15.5% conversion efficiency at 330 suns concentration level. Design rules regarding the fabrication and use of these solar cells at high concentration levels are derived. Finally, a system consisting of lens arrays and solar cells designed to generate 100 watts is described in detail. Preliminary data shows that the system efficiency is 8.3% during conditions of 74 mW/cm^2 insolation level.

1. Introduction

The cost of power from present-day silicon solar cells is in the range of $10 to $30 per peak watt, precluding their use in all but a very limited number of terrestrial applications. The manufacture of solar-cell-grade polycrystalline silicon, the production of single-crystal silicon, and the processing and assembly of these into solar cell arrays are being studied at RCA and other laboratories in an effort to reduce this cost.[1] In parallel with these efforts, there are programs[2] to reduce the cost of photovoltaic based electric power by the use of concentration methods. These programs aim to reduce the quantity of silicon by using lenses to collect the sunlight and focus the energy on smaller solar cells.

The disadvantages of such an approach are that sophisticated structures are required, since tracking of the sun must be maintained, and

* This work was supported in part by Sandia Laboratories, Albuquerque, New Mexico, under Contract #02-7996, as a part of the Photovoltaic Systems Definition Project at the Energy Research and Development Administration.

a fraction of the energy (that which is diffuse) is not utilized. The advantages are, of course, that solar-cell cost per unit watt is reduced, additional energy is collected over the day (since normal incidence is maintained), and higher solar cell efficiency can be achieved.[3]

This paper concerns itself with one particular approach, namely high levels of concentration (500–1000 suns) on small solar cells (0.180 × 0.180 inch). These small cells generate as much electricity as would two 3 inch wafers, namely one watt. The first part of the paper discusses solar-cell design with regard to metalization patterns, thickness, diffusion profile, and anti-reflection coating. The performance of these cells under high levels of concentration is shown. Second, a thermal analysis is performed, and finally, a system is described that is designed to generate 100 watts.

2. Solar Cell Design

There are several important parameters in solar-cell design; active layer thickness and resistivity, junction layer thickness and profile, and metalization pattern. The active layer thickness and resistivity is a compromise between current collection efficiency and resistive losses; the junction layer thickness and profile is a compromise between violet response and fill factor; and the metalization pattern is a compromise between shadowing and resistive losses. A full-fledged analysis requires a sophisticated computer program; however, in this paper, we simplify the analysis so that the essence of the physics remains without a significant compromise to the validity of the results.

2.1 Active Layer Thickness

Most of the photon excited carriers are generated within the first few microns of the solar cell and for that reason the simplifying assumption is made that all the carriers are generated in a thin sheet in the vicinity of the junction. Under optimum load conditions, most of the minority carriers diffuse across the junction and generate power, while the rest diffuse into the active layer, recombine, and provide conductivity modulation of the active layer. The density of carriers under these conditions is

$$n(x) = N_0 e^{-x/L_D}. \qquad [1]$$

N_0 is evaluated by allowing all these carriers to recombine at a rate of $I_\phi \Gamma$, where I_ϕ is the available photon generated current and Γ is that fraction that recombines

$$I_\phi \Gamma = \frac{q}{\tau} \int_0^\infty n(x) dx \quad [2]$$

Note the simplifying assumption in the limits of integration.

The internal voltage drop due to carriers flowing through the load is

$$V_D = I_\phi (1 - \Gamma) \int_0^d \frac{dx}{q(\mu_e + \mu_h) n(x) + \sigma_0}, \quad [3]$$

where

L_D	is the ambipolar diffusion length,
τ	lifetime
μ_e, μ_h	mobility of electrons and holes
d	the active layer thickness
σ_0	base conductivity.

The voltage drop is

$$V_D = I_\phi \frac{(1 - \Gamma)}{\sigma_0} \left(d + L_D \ln \left[\frac{k + e^{-d/L_D}}{k + 1} \right] \right) \quad [4]$$

where

$$k = \frac{\sigma_0 L_D}{(\mu_e + \mu_h) I_\phi \Gamma \tau}. \quad [5]$$

Note that the voltage drop increases with active layer thickness, but so does the collection efficiency. Clearly, for a given lifetime, and concentration level, an optimum thickness exists. As the lifetime increases and concentration level decreases, the optimum thickness increases. This is demonstrated in Fig. 1 where collection efficiency and voltage drop are plotted. Note that at very high levels of concentration for the assumed 1 ohm-cm, 10 μsec lifetime material, the voltage drop curves tend to converge. For higher resistivity, the curves converge at lower intensities. Hovel's[4] collection-efficiency curves are for a specific set of conditions on lifetime, junction depth, and air mass. The significance, however, is that there is not much to be gained in increasing the base-layer thickness over 25 μm when a back surface field exists. For 25 μm thickness and 300 to 3000 suns, about 1% degradation in voltage is experienced due to base layer resistance.

2.2 Solar-Cell-Grid Design

Designing an optimum ohmic-contact top grid pattern for solar cells entails several tradeoffs: (1) grid shadowing (the metal contact grid is

opaque to light), (2) sheet resistance losses (current collected above the solar cell p-n junction must travel in the thin sheet of semiconductor material before it reaches the contact grids), and (3) metalization resistance losses (current collected by the grid must travel along the grid before it reaches the output terminal of the solar cell).

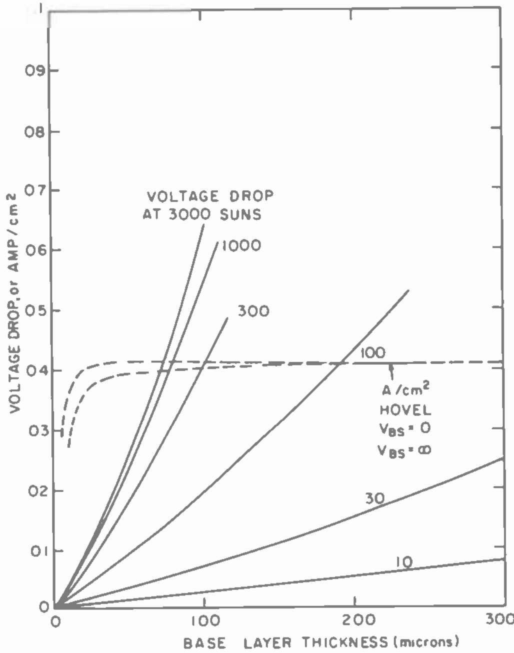


Fig. 1—Voltage drop and collection efficiency in base layer.

(a) Optimum Grid Topology

To design an optimum grid pattern, one must first choose an optimum grid topology. Two obvious choices exist: a cross-hatched pattern where grid lines cross each other at right angles or a simple line pattern where grid lines do not cross each other.

The two patterns are shown in Fig. 2. Note that the cross-hatched pattern has twice the spacing of the line pattern, while each has the same line width. This is to keep the shadowing of each similar, thereby having the solar cell capable of collecting an identical number of photons. The line width is kept similar, because the sheet resistance of the metalization pattern is a function of shadowing and metalization thickness. The

metalization thickness is technologically limited to the line width. Thus, in order to keep the metalization sheet resistance the same in each pattern, the line width must be kept the same.

What will be compared in the following analysis is the voltage drop in the semiconductor sheet above the junction, for both the line and

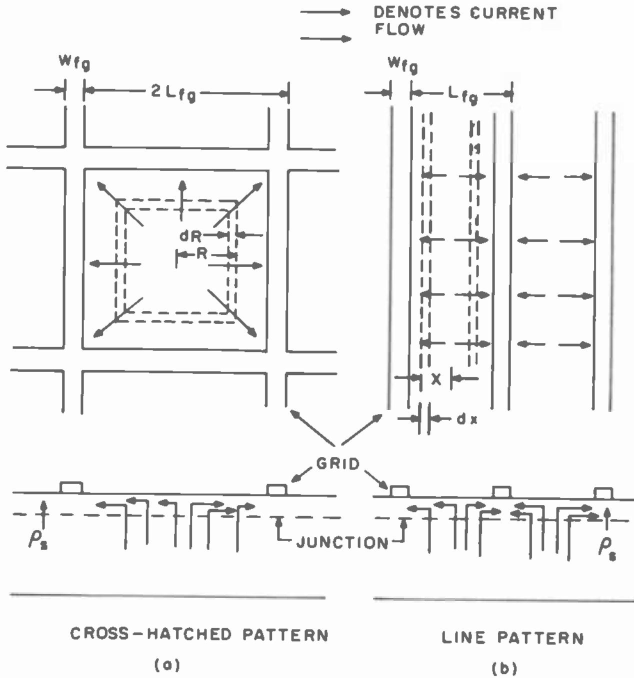


Fig. 2—Metalization patterns for solar cells.

cross-hatched patterns. Assume an incident flux such that the current density arriving above the junction is J_0. In the cross-hatched pattern, the sheet current I_s in the element dR is

$$I_s = J_0 4R^2. \quad [6]$$

The incremental voltage drop is

$$dV_{CH} = \frac{I_s}{8R} \rho_s dR. \quad [7]$$

Thus the voltage drop from the junction to the fine grid is

$$V_{CH} = \frac{1}{2}\int_0^{L_{fg}} J_0\rho_s R dR = \frac{J_0\rho_s}{4} L_{fg}{}^2. \quad [8]$$

In the line pattern, the sheet current per unit length is

$$I_s = 2J_0 x, \quad [9]$$

whereas the incremental voltage drop is

$$dV = \frac{I_s}{2}\rho_s dx. \quad [10]$$

Thus the voltage drop from the junction to the fine grid is

$$V_L = \int_0^{L_{fg/2}} J_0\rho_s x dx = \frac{J_0\rho_s}{8} L_{fg}{}^2. \quad [11]$$

Clearly, the voltage drop for the line grid is half that for the cross-hatched grid while they both have similar shadowing and metal sheet resistance.

(b) Optimum Line-Grid Design

In the analysis, we consider power loss due to shadowing S (fraction of cell covered by metal), contact resistance (ρ_c (ohm-cm^2)), sheet resistance (ρ_s (ohm/□)) and metalization resistance ρ_M (ohm-cm). We then maximize the output power to arrive at an optimum cell design.

The voltage drop as a result of sheet resistance, according to the previous analysis for the line pattern, is

$$V_{DS} = \frac{J_0}{8}\rho_s L_{fg}{}^2. \quad [12]$$

The voltage drop as a result of contact resistance is

$$V_{DC} = \frac{J_0\rho_c}{S}, \quad [13]$$

since the current density in the contact area is magnified by $1/S$.

The voltage drop in the metalization resistance can be arrived at by using Eq. [12]. We substitute $\rho_M/t_M S$ for ρ_S in Eq. [12], since the sheet resistance of the metal is dependent on the metal thickness t_M and the shadowing S. We replace t_M by ¼ of the grid width ($W_{fg}/4$), since this is a practical limit, and further, replace the grid width by its equivalent in terms of shadowing and grid spacing, i.e., $W_{fg} = SL_{fg}$. The current is collected at the limits of the lines, or at both boundaries of the cell, so we replace L_{fg} by L_c, the dimensions of the solar cell:

$$V_{DM} = \frac{J_0}{8}\frac{\rho_M L_c{}^2}{t_M S} = \frac{J_0}{8}\frac{\rho_M L_c{}^2}{1/4 W_{fg} S} = \frac{J \rho_M L_c{}^2}{2 L_{fg} S^2}. \quad [14]$$

The output voltage, then, is

$$V_0 = V_M - V_{DM} - V_{DC} - V_{DS},$$

where V_M is the maximum voltage with no resistive voltage drop. The output current is approximately

$$I_0 \cong J_0 L_c{}^2 (1 - S)$$

Thus the output power is

$$P_0 = J_0 L_c{}^2 (1 - S)\left(V_M - \frac{J_0 L_c{}^2 \rho_M}{2 S^2 L_{fg}} - \frac{\rho_c}{S} - \frac{J_0}{8}\rho_s L_{fg}{}^2\right). \quad [15]$$

Maximizing the output power by varying the grid spacing and shadowing gives

$$\left.\frac{\partial P}{\partial L_{fg}}\right|_{S=Const} = 0, \quad \text{and}$$

$$\left.\frac{\partial P}{\partial S}\right|_{L_{fg}=Const} = 0.$$

Thus

$$L_{fg}{}^3 = \frac{2\rho_M}{\rho_s}\frac{L_c{}^2}{S^2} \quad [16]$$

$$\frac{V_M}{J} = \frac{\rho_c}{S^2} + \frac{\rho_s}{8} L_{fg}{}^2 + \frac{L_c{}^2 \rho_M}{S^3 L_{fg}}\left(1 - \frac{S}{2}\right). \quad [17]$$

We can rearrange Eqs. [16] and [17] for simple calculations by substituting [16] into [17] and neglecting $S/2$ compared to 1:

$$L_{fg}{}^3 \left[A + \sqrt{L_{fg}}\, B\right] = \frac{V_M}{J} \quad [18]$$

$$\frac{1}{S^2} = \frac{L_{fg}{}^3 \rho_s}{2 \rho_M L_c{}^2}, \quad [19]$$

where

$$A = \frac{\rho_s \rho_c}{2 \rho_M L_c{}^2}$$

and

$$B = \frac{\rho_s{}^{3/2}}{2\sqrt{2}\rho_M{}^{1/2} L_c}.$$

The method of solution is as follows. Assign values to V_M, J, ρ_S, ρ_M, ρ_C, L_c. Try a value of L_{fg} to get the term in the brackets of Eq. [18], solve for L_{fg}, and re-do. The system converges within 3 or 4 calculations. Finally, solve for S. An example is shown in Table 1.

In the case of the standard nonconcentrator cell, the grid spacing is a strong function of sheet resistance, the shadowing is a weak function of both sheet resistance and contact resistance, and the shadowing and grid spacing are both a very weak function of contact resistance. This can be summarized by a further simplification of Eq. [18]. Letting $A \ll B\sqrt{L_{fg}}$,

Table 1—Optimum Line Grid Design

	Nonconcentrated Cell	Concentrated Cell
ρ_M	2.5×10^{-6} Ω-cm (gold)	2.5×10^{-6} Ω-cm
ρ_S	100 Ω/□	100 Ω/□
ρ_c	10^{-4} Ω-cm²	10^{-4} Ω-cm²
J	0.040 A/cm²	20 A/cm²
V_M	0.5 volts	0.7 volts
L_c	2 cm	0.4 cm
A	10^3	1.25×10^4
B	3.162×10^5	5.6×10^5
L_{fg}	0.074 cm	0.82×10^{-2}
S	0.022	0.12
W_{fg}	1.65×10^{-3}	9.8×10^{-4}

$$L_{fg} \cong \left(\frac{2\sqrt{2}\,V_M L_c}{J}\right)^{2/7} \frac{\rho_M{}^{1/7}}{\rho_s{}^{3/7}} \qquad [20]$$

$$S \cong \frac{\sqrt{2}\,L_c{}^{4/7}\rho_M{}^{2/7}\rho_s{}^{1/7}}{\left(2\dfrac{\sqrt{2}\,V_M}{J}\right)^{3/7}}. \qquad [21]$$

These approximations are fairly good for the standard nonconcentrator cell only,

(c) Power Output Variation with Grid Pattern Changes

It is of interest to calculate equal power contours on a shadowing-grid spacing plot for two reasons. First, the technology may not allow the optimum grid design and, second, processing deviations can alter grid spacing and shadowing. A power matrix was calculated from Eq. [15] varying both grid spacing and shadowing. Given the matrix, equal power contours are plotted in Fig. 3. Note that this is the condition for the concentrator cell with conditions as stated in Fig. 3. The equal-power

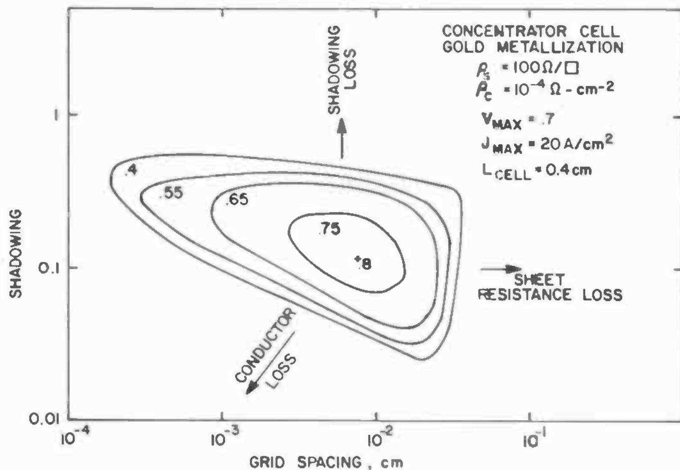


Fig. 3—Equal power contours for various metalization patterns at ~1000 suns.

contours are normalized to the maximum power one would obtain from an ideal cell (no shadowing, grid loss, or sheet resistance loss). The maximum normalized power is 80% with a grid spacing of 85 μm and 12% shadowing. Interestingly enough, one can deviate from the optimum design by factors of 2 and decrease the normalized power to only 75%.

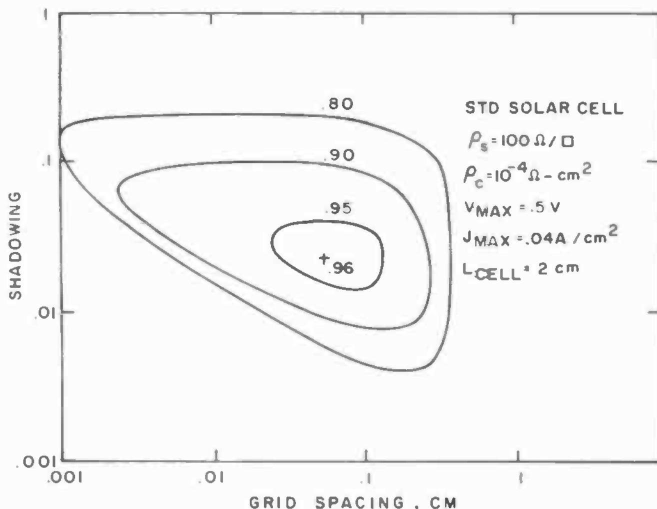


Fig. 4—Equal power contours for various metalization patterns at ~1 sun.

A similar power matrix was calculated for the standard nonconcentrator cell and equal power contours are shown for this in Fig. 4.

3. Solar Cell Fabrication and Experimental Results

A solar cell designed for operation at high light intensity must incorporate all of the following features:

(1) A metal grid pattern with a sufficiently thick metalization to reduce the potential drop along the grid to a negligible quantity.
(2) Low contact resistance between the metal grid and the silicon.
(3) A long minority carrier lifetime to insure a high plasma density and, thus, low series resistance throughout the thickness of the cell.
(4) Base layer relatively thin so as to minimize bulk resistance.

Solar cells from epitaxial and bulk starting material were fabricated according to guidelines defined earlier, and tested at high intensity. The majority of the epitaxial cells were n-type, and all of the bulk cells were p-type. The advantage of the epitaxy over bulk cells is that the thin base layers required for good performance with high concentration are easily achieved. The disadvantages are that minority carrier lifetime in epitaxial wafers is not uniformly high and, also, that epitaxial wafers are more costly.

3.1 Fabrication

In fabricating the solar cells, starting with both epitaxial and bulk material, we adhered to the guidelines regarding emitter sheet resistance, metalization pattern, and base layer thickness. There were, however, some processing differences, as outlined in the Table 2.

Slow cooling after high temperature processing is used to preserve minority carrier lifetime, and annealing of the electron beam evaporated antireflection coating was required to reduce transmission loss. Photographs of the finished solar cells are shown in Fig. 5. All of the bulk wafers with the exception of Sample K3 were processed using the metalization patterns with the circular aperture, while the epitaxial cells were made using the metalization pattern with the rectangular aperture or with a cross-hatched pattern (not shown).

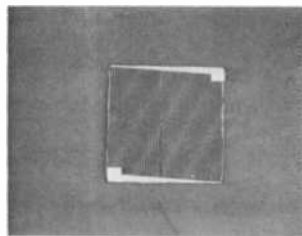
3.2 Spectral Response Measurements

The spectral response was measured by using a GE ELH lamp and narrow-band optical filters with a calibrated EG&G PV 44A photodetector. The response of the cell under test was compared to that of the

Table 2—Processing Difference Between Bulk and Epitaxial Solar Cells

	Bulk	Epitaxy
Type	P	P or N
Thinning	125 μM	—
Back surface diffusion	Borosilicate glass	—
Deep diffusion under metal grid	Phosphorosilicate glass	—
Emitter diffusion	$POCl_3$	Boron nitride or boron ion implants for N-type Phosphorus ion implant for P-type
Metalization	Chromium-gold	Chromium-gold
Anti-reflection coating	SiO_2 on TiO_2	SiO_2 on TiO_2 or SiO_2 or Al_2O_3

EG&G standard. The measured performance of several cells is shown in Fig. 6. All the cells had the SiO_2 on TiO_2 anti-reflection coating. Correction has been made for grid shadowing. The device from wafer K14 had a relatively low emitter sheet resistance, which resulted in a poor violet response, while the device from wafer K3 had a relatively high sheet resistance and a good violet response.



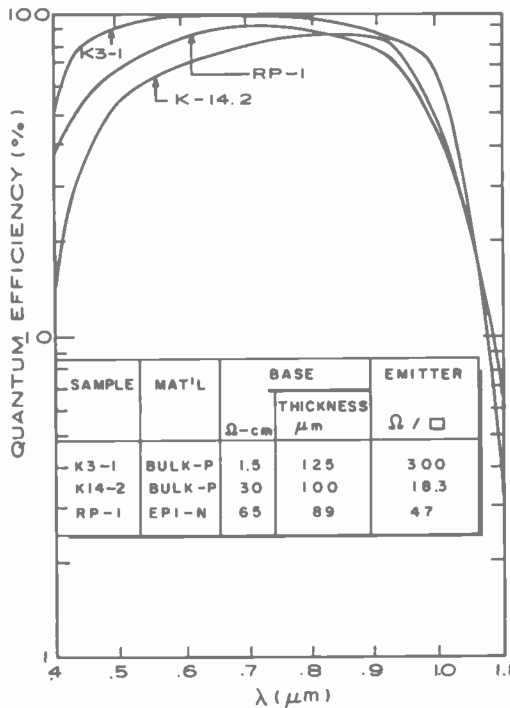
0.17"x0.17" EPI N-TYPE CELL

0.22" APERTURE BULK P-TYPE CELL

Fig. 5—Finished solar cells.

3.3 High Insolation Testing of Solar Cells

Solar cells were tested outdoors with concentration from a 4½ inch or 6 inch diameter lens or, in the laboratory, using a 300 watt Xenon short arc lamp. The experiments are shown schematically in Fig. 7. Both experiments gave essentially the same results provided the short-circuit current and ambient temperature was the same. The solar cells were



SAMPLE	MAT'L	BASE		EMITTER
		Ω-cm	THICKNESS μm	Ω/□
K3-1	BULK-P	1.5	125	300
K14-2	BULK-P	30	100	18.3
RP-1	EPI-N	65	89	47

Fig. 6—Spectral response of solar cells with TiO_2 and SiO antireflection coating.

soldered to T0-3 transistor headers and the header was tightly mounted to a transistor heat sink. Insolation with no concentration was determined by monitoring the short-circuit current of an EG&G cell calibrated for AM1 insolation by NASA-LEWIS. Under high levels of concentration, it was assumed that the short circuit current was proportional to insolation.

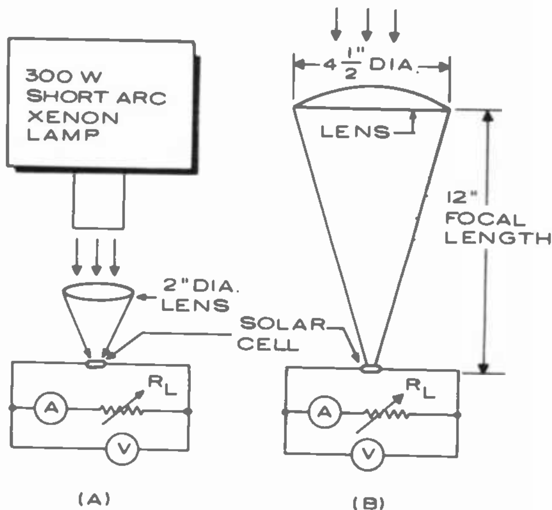


Fig. 7—Schematic of solar cell experiments: (a) laboratory short-arc high-intensity simulator and (b) outdoor lens-solar cell apparatus.

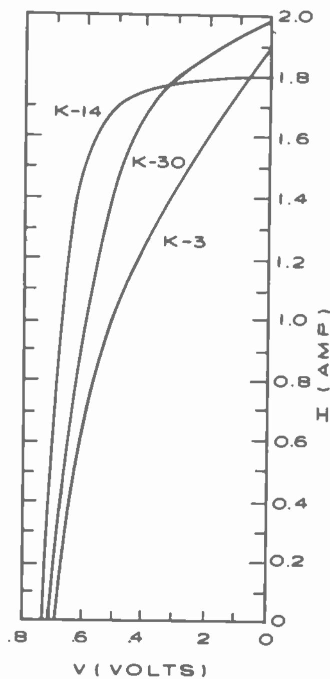


Fig. 8—500-sun current-voltage characteristics of cells.

3.4 Curve Factor at High Insolation Levels

Current–voltage characteristics of bulk material solar cells were measured under nearly the same short-circuit current conditions. The curve factor of these cells was grossly different, as is shown in Fig. 8. Sample K3 has the highest sheet resistance; thus one would think that this accounts for the poor fill-factor. However, using Eq. [12] to calculate the voltage drop in the emitter sheet, only 0.053 volts can be accounted for, not nearly enough to account for the poor curve factor. The base resistivity of K3 is much lower than that of Eqs. [4] and [5], and to evaluate the base voltage drop, we must assume less than 1 μsec lifetime in order to fit the current voltage curve of K3, while measurements of open circuit voltage decay after a short Xenon flash illumination indicates a 5.5 μsec lifetime. The tendency at this point, however, is to believe that the lifetime does play a major role in determining curve factor, since K14 did have a lifetime nearly twice that of K3, and the lifetime of K30 was greater than that of K3 and less than that of K14. Therefore, it is believed that the analysis that resulted in Eqs. [4] and [5] for the base voltage drop is overly optimistic.

3.5 Cell Performance at High Insolation Levels

One of the most frequently asked questions regarding the performance of solar cells under such high levels of insolation is whether the temperature of the cell gets excessively high. The experiment described below answers this in a relatively straightforward way. Two lenses were used to focus the sun on an identical solar cell, namely the epitaxial cell RP1. Both lenses had a 12 inch focal length, so the theoretical sun spot size should be 0.12 inch in both cases. (The apparent spot size with the 4½ inch lens was 0.17 inch and the spot appeared to spill over the cell with the 6 inch lens.) In any case, one would expect that the open-circuit voltage would severely degrade using the 6-inch lens, and as can be seen in Fig. 9, no degradation in open circuit voltage occurred. Of added significance is the apparent increase in conductivity modulation with the higher current curve. This is noted by comparing the slope of the I-V curves at the zero-current intercept.

Tables 3 and 4 provide a complete summary of both epitaxial and bulk solar-cell parameters as well as their performance at high levels of insolation. One must be cautioned, however, that the incident power was calculated assuming that insolation and short-circuit current are proportional. There is some indication that this may not be the case, especially for the cells processed from bulk material. Even though there is a well-defined knee in the illuminated I-V curve, there is a non-zero slope

Table 3—Summary of Solar Cell Characteristics

Cell no.	Type	ρ (Ω-cm)	t (μm)	Source of Impurity	AR Coating	τ (μs)	Cell Dimension	Grid Pattern	Sheet Resistance ($\Omega/\square$)
RP-1	p^+nn^+	65	89	BN	SiO_2 + TiO_2	12.0	0.17" x 0.17"	line	47
NSO33	p^+nn^+	41-46	60-68	BN	SiO_2	4.0	0.17" x 0.17"	line	~50
NSA21	p^+nn^+	30-40	70~77	$^{49}BF_2^+$	Al_2O_3	9.2	0.18" x 0.18"	crosshatched	~80
NSA18	p^+nn^+	30-40	70~77	$^{49}BF_2^+$	SiO_2	13.3	0.18" x 0.18"	crosshatched	~30
NSO10	p^+nn^+	64-90	48	B^{11}	SiO_2	4.5	0.15" x 0.15"	crosshatched	—
PSO8	n^+pp^+	4.6	55	P^{31}	SiO_2	1.3	0.18" x 0.18"	crosshatched	—
K14-1	n^+pp^+	30-40	75	PHO_3CL	SiO_2 + TiO_2	9.5	0.17 dia.	Line	18
K39-1	n^+pp^+	30-40	100	PHO_3CL	SiO_2 + TiO_2	16	0.22 dia.	Line	27
K13A	n^+pp^+	1.5	100	PHO_3CL	SiO_2 + TiO_2	~5	0.17" dia.	Line	23

Table 4—Solar Cell Performance at High Levels of Insolation

Cell No.	Type	Short Circuit Density (A/cm²)	Incident Power Density (W/cm²)	Power Output (W)	Efficiency (%)	V_{OC} (V)	Curve Factor	Optimum R_L (Ω)
RP-1	p^+nn^+	14.5	53.7	1.00	12.8	0.75	0.63	0.27
RP-1	p^+nn^+	18.3	67.6*	1.43	11.4	0.75	0.56	0.18
NSO-33	p^+nn^+	12.7	48.2	0.842	12.0	0.67	0.68	0.31
NSA-21	p^+nn^+	12.4	49.6	0.928	11.9	0.73	0.65	0.28
NSA-21	p^+nn^+	7.0	27.5†	0.560	12.8	0.72	0.71	0.63
NSA-18	p^+nn^+	11.0	51.6	0.839	11.1	0.72	0.73	0.39
NSO-10	p^+nn^+	10.2	56.9	0.794	9.6	0.73	0.73	0.49
PSO-8	n^+pp^+	9.17	49.6	0.729	9.4	0.72	0.70	0.44
K14-1	n^+pp^+	10.3	31.3	0.78	17.1	0.74	0.70	0.46
K39-1	n^+pp^+	6.28	19.5	0.864	17.9	0.77‡	0.72	0.48
K13A	n^+pp^+	11.5	33.6	0.764	15.5	0.75	0.61	0.46

* A 6" diameter lens was used to concentrate the sun in this experiment.
† A neutral density filter was inserted in the focused beam to reduce the incident power.
‡ Measured on an extremely cold day, 20°F.

of the curve at the zero-voltage intercept. This may be evidence that there is an increase in the fractional loss of photon generated carriers as the intensity increases. The phenomenon (and its consequences) is not presently understood and is undergoing investigation. Several things however, are very encouraging and worthwhile pointing out. The power

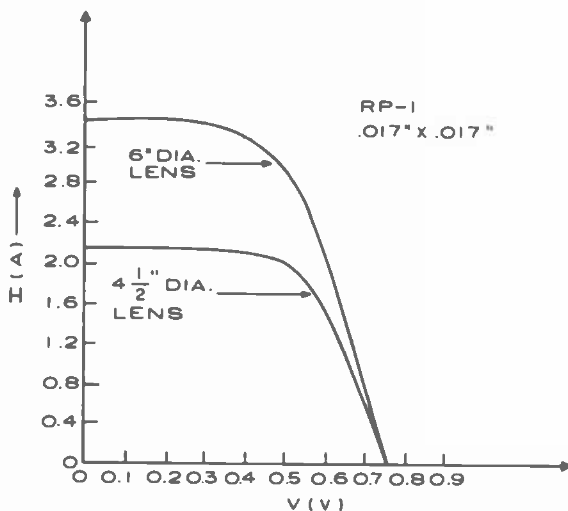


Fig. 9—Illuminated I-V curves of RP-1.

generated ranged from 0.73 and 1.0 watts, while the cell was illuminated from a 102 cm^2 lens. In all cases the direct insolation on the lens was less than 100 mW/cm^2. The open circuit voltage was as high as 0.773 volts and the curve factor was as high as 0.73. The high apparent efficiencies of 17.1% and 17.9% show that we can achieve quite respectable fill-factors and open-circuit voltages at current densities corresponding to 195 and 313 suns, respectively.

4. Anti-reflection Coatings

Anti-reflection (AR) coatings on solar cells are typically single-layer coatings one quarter wavelength thick centered at the wavelength of maximum useable solar flux. The refractive index of the AR coating is chosen to be approximately 2, a good match to the average refractive index of silicon (approximately 4). The limitation of this AR coating is that it matches perfectly for only one wavelength, yet the optical bandwidth of interest is more than one octave.

For a broadband match, a double-step transformation (double AR coating) is a better choice than the single-step transformation (single AR coating). The design rules[5] for such a coating can be expressed as

$$\left(\frac{\eta_2}{\eta_1}\right)^2 = \eta_s \qquad [22]$$

and

$$\eta_1 \eta_2 = \eta_s, \qquad [23]$$

where

η_1 = refractive index of the coating exposed to air
η_2 = refractive index of the coating in contact with silicon
η_s = refractive index of silicon

The thickness of each coating is one quarter wavelength.

Since silicon has a refractive index of approximately 4 over a good portion of the visible spectrum (it actually varies from 7 to 3.5 over the spectrum 3700 Å to 12,000 Å),[2] then $\eta_1 = 1.414$ and $\eta_2 = 2.828$. The refractive index of SiO_2 makes it a good choice for η_1 and TiO_2 for η_2.

SiO_2 and TiO_2 have been sputter deposited* and electron beam deposited† on silicon both singly and in multiple layers. Transmission data on thin layers are shown in Figs. 10, 11, and 12. Included in these curves is the calculated transmission coefficient of silicon from Runyon.[6] In the actual measurements, the reflection coefficient was measured relative to bare silicon. Loss or absence of loss in the coating can be determined when the thickness of the coating is one half wavelength. At this wavelength, the transmission of the coated wafer should be identical to the bare wafer if the coating is lossless. For the TiO_2 coatings this is evidenced at λ = 3000 Å for a thickness t = 750 Å, λ = 4050 Å for t = 800 Å, and at λ = 4350 Å for t = 850 Å. The refractive index of the material can be determined by the transmission coefficient and refractive index of silicon at odd quarter wavelengths (points of maximum transmission).

The equation relating transmission coefficient to refractive index at $\lambda/4$ is

$$V = \left[\frac{2}{T} - 1\right] + \sqrt{\frac{4}{T}\left(\frac{1}{T} - 1\right)} \qquad [24]$$

where

---

* J. J. O'Neill of RCA Laboratories sputter deposited the SiO_2-TiO_2 layers.

† F. Tams of RCA Laboratories electron beam evaporated SiO_2-TiO_2 layers.

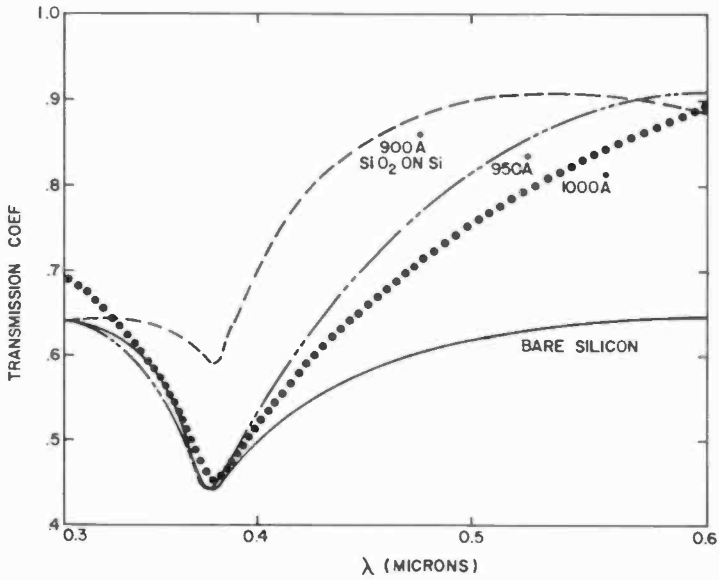


Fig. 10—Transmission coefficient for SiO_2 on Si.

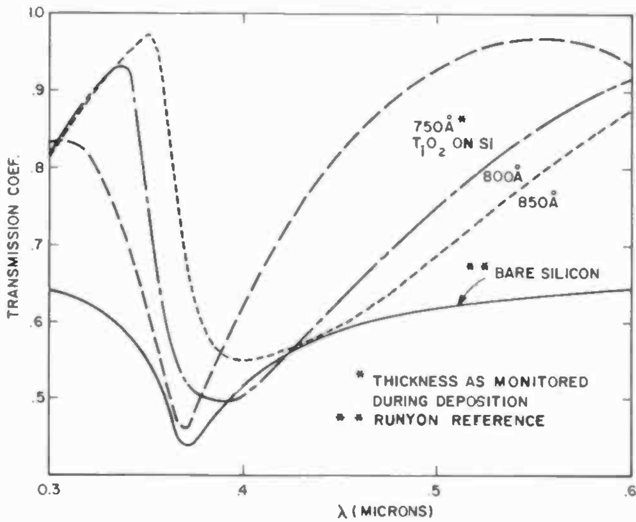


Fig. 11—Transmission coefficient for TiO_2 on Si.

$$V = \frac{\eta_c^{\,2}}{\eta_s} \quad \text{or} \quad \frac{\eta_s}{\eta_c^{\,2}} > 1 \qquad [25]$$

η_c = refractive index of coating

η_s = refractive index of silicon

T = transmission coefficient

Using Figs. 10 and 11 and Eqs. [3] and [4], the refractive index and thickness have been tabulated and are shown in Table 5.

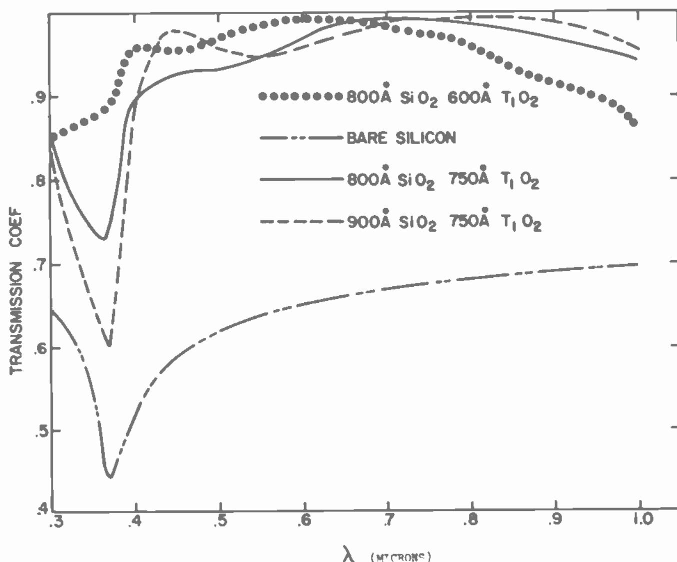


Fig. 12—Transmission coefficient for SiO_2 on TiO_2 on Si.

Several combinations of SiO_2 on TiO_2 on silicon have been deposited and measured. The best combinations are shown in Fig. 12. The 800-Å SiO_2–600-Å TiO_2 seems to be best for short wavelengths, whereas the thicker coating is best for the longer wavelengths. Since most of the energy exists at the shorter wavelengths, the better choice is the thinner layers. In appearance, the thin layers have a deep green appearance and also appear less reflective.

5. Thermal Effects of Highly Concentrated Sunlight

There are three fundamental parts to the thermal analysis. First, the concentrated solar energy must spread from the silicon chip through the

heat sink. Second, the thermal energy must convect from the heat sink, or heat exchanger. Third, the thermal energy must radiate from the heat exchanger. The problem becomes difficult to solve precisely, since the total heat flux from the solar cell decreases as we move away from the cell because of convection as well as radiation. Both convection and radiation are nonlinear functions of temperature, and practical heat geometries are not amenable to analysis.

Table 5—Refractive Index and Thickness of Anti-Reflection Coatings

	Materials	
	SiO_2	TiO_2
Thickness as Deposited	900 Å	750 Å
Transmission Max.		
$\lambda/4$	5200 Å	5500 Å
$3\lambda/4$	—	3050 Å
Thickness (calculated)		
$\lambda/4$	880 Å	565 Å
$3\lambda/4$	—	740 Å
Refractive Index (calculated)		
$\lambda/4$	1.47	2.43
$3\lambda/4$	—	3.1

In order to make the problem tractable, we have made the following assumptions:

(1) The *absolute* temperature does not vary drastically from 300°K. (In this way, radiation can be approximated as a linear function of temperature.)

(2) The convective film coefficient, which varies as the fourth root of the rates of characteristic length and temperature difference will be assumed a constant. The constant will be evaluated at an average temperature difference of 30°C, and a typical heat sink length of 10 cm. Convection can also be approximated as a linear function of temperature.

(3) Heat will be convected and radiated from both sides (upward and downward) of the heat sink.

The heat flow equation that results is in the form of a Bessel function of the second kind with an imaginary argument. Based on the solution, the following conclusions are reached:

(1) The temperature rise of the silicon solar cell will be a function of the thermal conductivity and thickness of the heat sink, the coefficients

of radiation and convection, and the solar concentration ratio and solar cell size.

(2) A practical heat-sink size will be calculated so as to convect or radiate all of the heat while not significantly increasing the solar cell temperature.

5.1 Approximate Relationships for Radiation and Convection

Heat radiates from a body as:

$$q_r = \epsilon\sigma\,(T^4 - T_0{}^4) \qquad [26]$$

where

ϵ = emissivity = .94 for flat black enamel
$\sigma = 5.67 \times 10^{-12}\ \mathrm{W cm^{-2} {}^\circ K^{-4}}$
T = local heat sink temperature
T_0 = ambient temperature
q_r = power density in watts/cm^2

Eq. [26] can be linearized as follows:

$$q_r = \epsilon\sigma\,(T^2 + T_0{}^2)(T + T_0)(T - T_0) \qquad [27]$$

$$\simeq 4\epsilon\sigma \overline{T}^3 \Delta T$$

$$\simeq k_r \Delta T \qquad [28]$$

if $\overline{T}$ is assumed to be 310°K, $k_r = 6.35 \times 10^{-4}$.

Heat convects from a body as[7]

$$q_c = 4.3 \times 10^{-4} \left(\frac{T - T_0}{L}\right)^{1/4} (T - T_0)\ \text{watts/cm}^2 \qquad [29]$$

where the coefficient $4.3 \times 10^{-4}\,[(T - T_0)/L]^{1/4}$ is an empirically derived relation, and is a weak function of temperature difference $(T - T_0)$ and L, the characteristic length. Eq. [29] has been linearized as follows:

$$q_c \cong 4.3 \times 10^{-4} \left(\frac{\Delta\overline{T}}{L}\right)^{1/4} \Delta T \qquad [30]$$

$$\cong k_c \Delta T. \qquad [31]$$

If $\Delta\overline{T}$ is assumed to be 20°C and L is 10 cm, k_c is 5.11×10^{-4}. It should be noted that factors of two in $\Delta\overline{T}$ only change k_c by 18%.

5.2 Heat Flow Equation

Referring to Fig. 13, the differential temperature drop at r in a distance dr is

$$dT = \frac{Q}{2\pi\kappa Wr} dr, \tag{32}$$

where T is the local temperature difference between the heat sink and ambient, Q is the total solar energy in watts concentrated by the lens onto the cell minus the energy converted into electricity and minus the energy

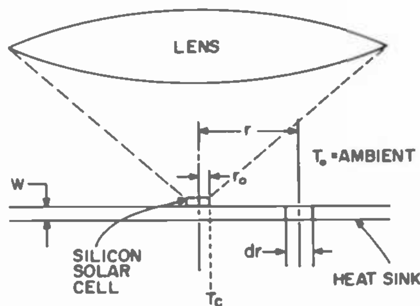


Fig. 13—Thermal schematic of lens–heat sink structure.

radiated and convected away in the distance between r_0 and r, k is the thermal conductivity in watts/°C-cm, and W is the heat sink thickness in cm (k = 2.1 watts/°C-cm for aluminum).

The total heat convected and radiated from both sides of the heat sink at r in a region dr, is

$$dQ = -2\pi r dr(2k_c + 2k_r)T \tag{33}$$
$$= -2\pi\kappa T r dr,$$

where $\kappa = 2k_c + 2k_r = 1.88 \times 10^{-3}$ Wcm^{-2}°C^{-1}. Combining Eqs. [32] and [33] we obtain

$$\frac{d}{dr}\left(\frac{rdT}{dr}\right) - \frac{\kappa}{kW} T. \tag{34}$$

Eq. [34] is a form of Bessels equation,[8] and its solution is a zero order Bessel function having an imaginary argument and is as follows:

$$T = AK_0\left(\sqrt{\frac{\kappa}{kW}}r\right) + BI_0\left(\sqrt{\frac{\kappa}{kW}}r\right). \tag{35}$$

The two boundary conditions are: (1) at the solar cell, $r = r_0$, the total power flow is the total input of the solar energy and therefore $Q = Q_0$; (2) at the edge of the heat exchanger, all the power has been radiated or convected away and therefore $Q = 0$ at $r = r_f$. Thus the temperature rise above ambient is

$$T_0 - T_{AMB} = \frac{Q_0}{2\pi k W} \frac{\left\{ \dfrac{K_1 \sqrt{\frac{\kappa}{kW}} r_f}{I_1 \sqrt{\frac{\kappa}{kW}} r_f} - \ln \sqrt{\frac{\kappa}{kW}} r_0 \right\}}{1 - \dfrac{K_1 \sqrt{\frac{\kappa}{kW}} r_f}{I_1 \sqrt{\frac{\kappa}{kW}} r_f} \left(\frac{\kappa}{kW} \right) \dfrac{r_0^2}{2}} \quad [36]$$

where kW = sheet thermal conductance of heat exchanger
Q_0 = total thermal energy
κ = radiative and convective coefficient
r_f = radius of heat exchanger
r_0 = radius of focused sunlight

For infinite size heat exchanger and normalizing to the solar concentration ratio, Eq. [36] simplifies to

$$T_0 - T_{AMB} = -q \frac{M r_1^2}{kW} \ln \sqrt{\frac{\kappa}{kW}} r_0 \quad [37]$$

where r_1 is the lens radius and $M = (r_1/r_0)^2$ is the concentration ratio.

Eq. (37) is plotted in Fig. 14 as a function of magnification ratio for the case of an infinite aluminum heat sink 0.1-inch thick and a solar-cell diameter (spot diameter of sun) of 0.12 inch. Note that there is no serious temperature rise even at 1000 suns.

Finally, Eq. [37] is plotted as a function of heat exchanger diameter in Fig. 15, keeping $M = 1000\times$, $r_0 = 0.12$ inch, and the heat sink thickness = 0.0625 inch. Note that above 7 to 8 inch diameter, there is no significant thermal advantage. This is essentially equivalent to a heat sink area three to four times the area of the lens.

6. System Design

6.1 Tracking

The obvious first step in designing a system that requires high levels of concentration is to decide on the required system of tracking. There are

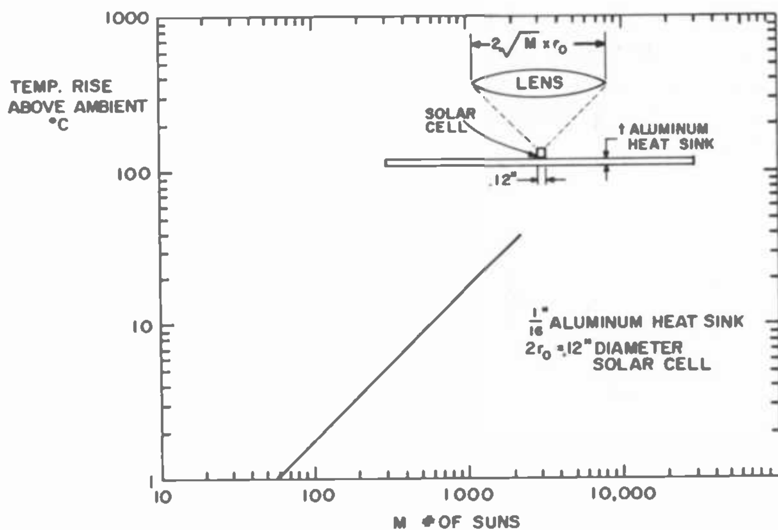


Fig. 14—Temperature rise versus concentration.

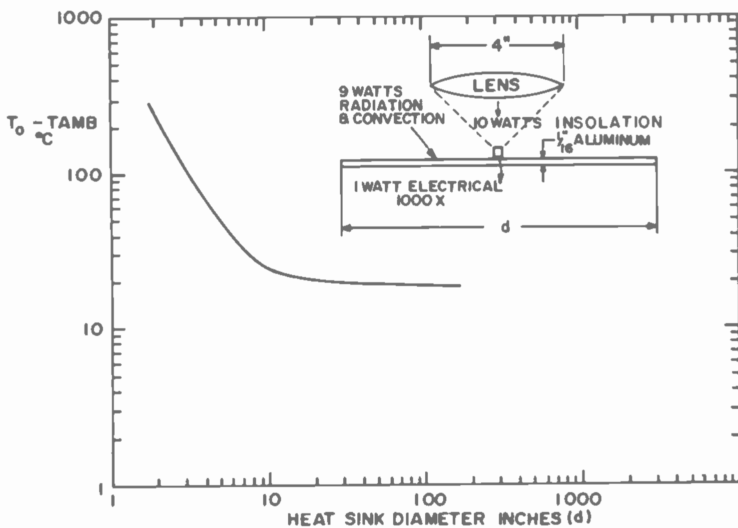


Fig. 15—Temperature rise versus heat-sink diameter.

two extremes that one could implement with regard to tracking. The first would be to have the system totally programmed both daily and seasonally, and the second would be to have the system totally controlled by a sensing system that zeros in on the sun. The first has problems associated with set-up accuracy and long-term stability, and the second has problems associated with decision making in conditions of short-term or long-term cloud cover, as well as at sunrise and sunset.

The tracking system chosen is a compromise. A 4-MHz crystal-controlled clock drives a stepping motor at ½ Hz to compensate for the daily rotation of the earth about its axis. In addition an east-west sun-sensing system fine tunes this motion. Seasonal motion, i.e., adjustments to compensate for the change in the earth's axis in relation to the sun, is analog driven and controlled by a north-south sun-sensing system actuated at mid-day.

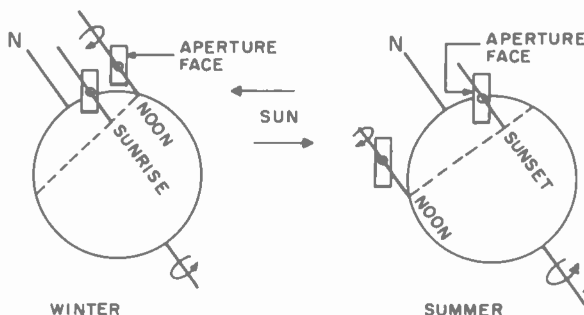


Fig. 16—Concept of polar axis tracking.

The line drawing in Fig. 16 indicates the desired degrees of motion, and a schematic of an implementation of the desired degrees of motion is shown in Fig. 17. The panels are allowed to tilt within the axis of daily rotation (a gimbaled motion), but since the axis of daily rotation is parallel to the earth's axis, very little seasonal motion is required during the course of one day.

6.2 Sun Sensing

From experience we have learned it is important to sense only the direct rays of the sun, since conditions of haze, bright objects, or clouds can generate intolerable inaccuracies in a global sensor. For direct ray sensing, we have used the system shown schematically in Fig. 18. A closely packed quadrature arrangement of solar cells is placed at the focal

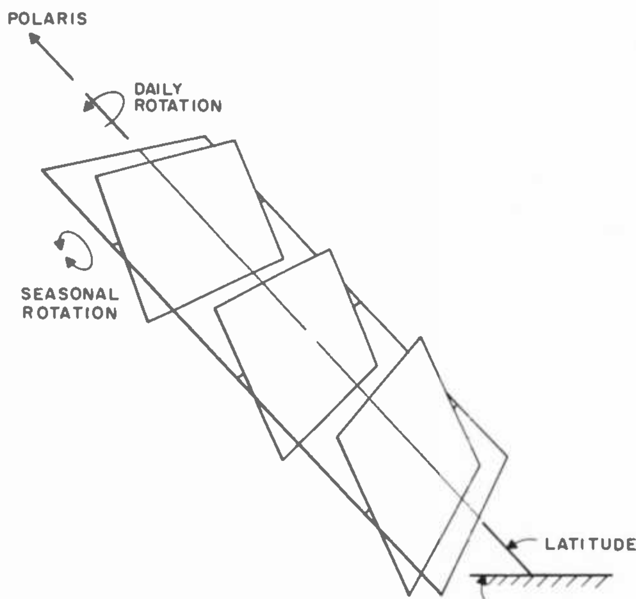


Fig. 17—Schematic of polar tracking implementation.

point of a lens. As the sun moves relative to the system, the image of the sun moves from one solar cell to the other. Current levels from each of the cells provide decision making information for the logic circuits. The mechanics of the system are such that high winds can overcome the available torque and stop the motion but cannot drive it ahead faster

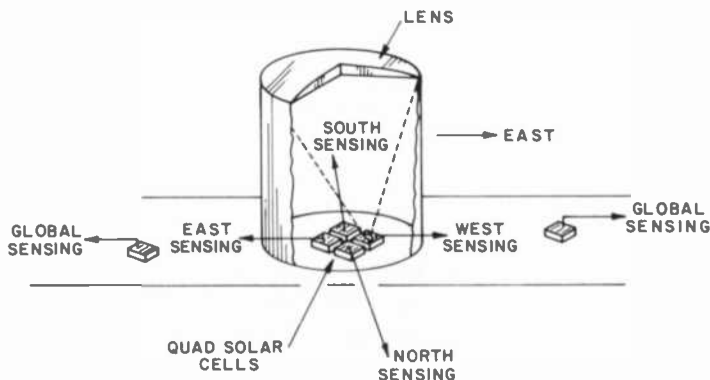


Fig. 18—Direct ray solar sensor.

than its pre-determined speed. Thus, if the image of the sun moves completely off the cells, it is assumed the tracker is lagging the sun. A pair of solar cells with global sensing provides decision making information for that condition. That is, if the global cells are generating current while none of the cells within the lens are generating current, the tracker is lagging the sun. The total power required for logic and drive is 0.25 to 0.4 watts and is provided by a 2.5 ampere-hour 12.5 volt battery. The battery is charged by the photovoltaic power.

6.3 Lens Array

Arrays of 4 × 4 inch square plano-convex lenses were fabricated by a unique method of vacuum forming* a thin sheet of acrylic over a master mold then later filling these shells with an acrylic epoxy. Epoxy shrinkage created nonuniformities on the "flat" side of the lens, array so this had to be rough machined, and another thin flat sheet of acrylic was epoxy cemented to complete this side of the lens array. Photographs of the shell and the completed lens array are shown in Fig. 19(a) and (b). Lens aberrations were such that the sun converged to within a 0.125 inch diameter of the focal point as is shown in Fig. 20. The repeatability from focal point to focal point was within 0.030 inch.

7. System Fabrication

A photograph of the completed system is shown in Fig. 21. The frame is constructed of lightweight tubular steel conduit and is supported by tubular steel tripods. The height of the tripods is adjustable to account for latitude position at the installation site. The boxes within the frame are fabricated of sheet steel and are made to support the lens array, solar cells, and heat sinks. The boxes are ganged and they tilt for seasonal adjust within the rotating frame. Forty foot-pounds of torque applied to the frame is such that the system can withstand 70 mph winds without stalling.

7.1 System Assembly, Adjustments and Performance

Solar cells from 3 wafers, (13A, 14A, and 14, which were described in Sec. 3) were mounted on T0-3 transistor headers, bolted to transistor heat sinks through mica insulators, and interconnected with #18 wire in a series parallel arrangement so as to obtain 4 parallel sets of 27 series-

* Vacuum forming was done by Analytic Plastics, Bristol, Pa.

connected solar cells. In this manner, we should achieve nominally 13.5 volts at 8 amps when the solar intensity is 100 mW/cm^2. Since each of the three panels consists of 36 solar cells, the top and bottom panels were broken into a 27-cell series connection and a 9-cell series connection, while the center panel was broken into two 18-cell series connected units.

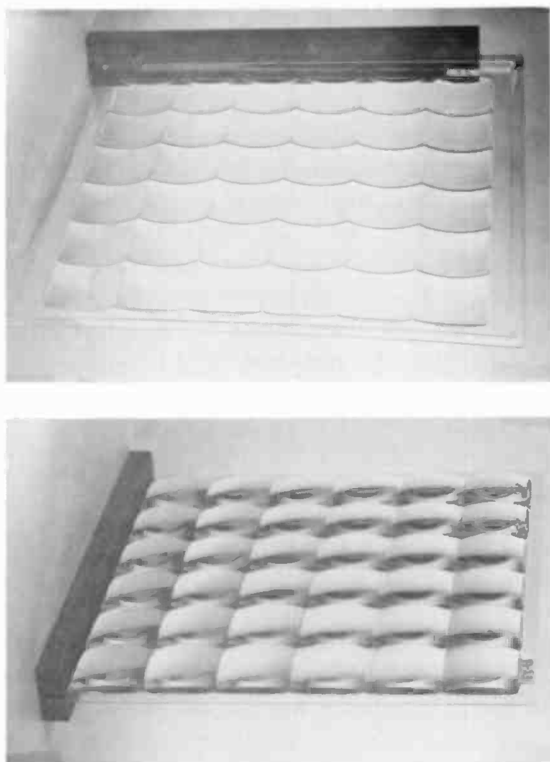


Fig. 19—(a) Lens array shell and (b) lens array.

Interconnections between the panels give the desired four parallel sets of 27 series-connected cells.

Initial tests were made on a cold mid-December day in Princeton, N.J., when the solar intensity was approximately 74 mW/cm^2. The results in Table 6 were obtained from each of the series connected units in each of the three panels.

The solar cells were placed in position in a random manner. As a result, the cell that generates the lowest current will dominate its particular

series string. A more efficient use of these cells would be to characterize them as to their current capability and populate the photovoltaic generator accordingly. Another group of cells from several different wafers were tested using a high intensity, focused xenon arc. The voltage generated across a 0.25-ohm resistor was measured and the distribution of 111 cells is as follows:

Voltage	.48	.47	.46	.45	.44	.43	.42	.41	.4
Number of Cells	1	4	8	26	19	22	13	10	8

Two cells measured only 0.28 volt and one measured 0.35 volt. The cells of 0.4 volt and above will be grouped according to their current capability to populate the second 100-watt unit now being assembled. Presumably, the efficiency of the second unit will be greater than that of the first unit.

Table 6—Power and Efficiency of Sun-Concentrated Solar Cells (Princeton, N.J., Mid-December 1975)

		Readings at Max. Pwr.			
		Current (A)	Voltage (V)	Pwr. (W)	Efficiency (%)
Top Panel	27-cell set	1.05	16.0	16.8	8.3
	9-cell set	1.13	5.0	5.7	8.4
Center Panel	18-cell set	1.00	11.5	11.5	8.5
	18-cell set	1.00	11.0	11.0	8.1
	9-cell set	1.13	5.0	5.7	8.4
Bottom Panel	27-cell set	1.20	17.4	17.4	8.6

8. Summary and Conclusion

Silicon solar cells have been fabricated and tested at 336 times air-mass one insolation at a solar-cell efficiency of 15.5%. The design incorporated optimum grid pattern, careful consideration of contact resistance, junction depth and impurity density, and base-layer thickness. The cell temperature rise was limited to 20°C by adequate heat sinking and natural convection. These silicon solar cells, together with plastic lens arrays, were assembled into a 12-square-foot photovoltaic generator with polar axis and seasonal tracking. The solar cells are 0.18 inch square and lenses within each of the three 36-lens arrays are 4 inches square.

Although there are no firm cost estimates on the production of systems such as the one described, certain conclusions can be made. First, the

solar cells have been reduced in size such that the silicon required to generate one watt at on-sun illumination can now generate several hundred watts at high concentration levels, while still maintaining greater than 10% aperature efficiency. We can therefore ignore the solar cell costs in such a system and concentrate instead on the cost of lenses, support structure, and tracking drive. A first step in this direction is an accounting of material cost. This is certainly not sufficient, but at least it provides a lower cost limit.

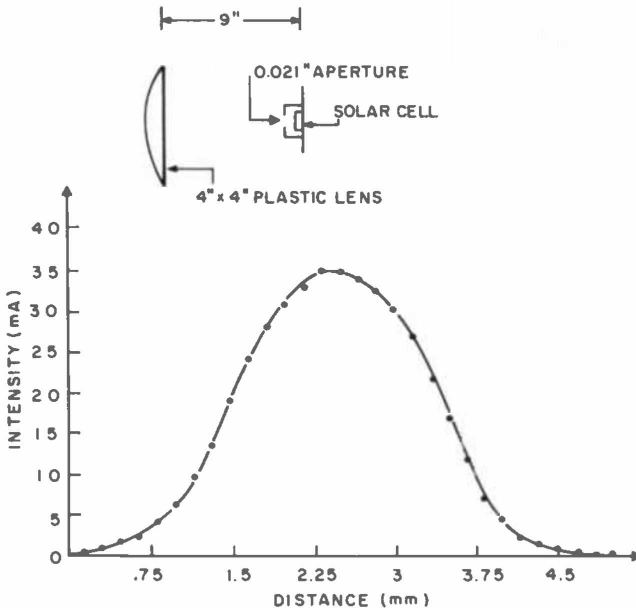


Fig. 20—Solar intensity at focal plane of plastic lens.

We used approximately 140 pounds of steel in the tubular tripods, tubular frame, and lens solar-cell support enclosures. Sixty pounds of plastic were used in the lenses. At 18¢/lb for steel and 57¢/lb for plastic, we are in the neighborhood of $60 for material cost for the 100-watt system described. Considerable cost savings can be achieved by using essentially the same structure for a 300-watt unit and by use of molded fresnel lenses rather than the plano-convex plastic lenses. With these modifications, a 300-watt system can be made with $63 worth of materials. The most costly part of the system as it presently is designed is in

the set-up and alignment time. Lens–solar-cell support and interconnects *must* be made with a sufficient degree of reproducibility and precision such that the labor required for set-up and alignment is negligibly small. In our judgement this can be done.

Fig. 21—Experimental 100-W photovoltaic generator.

Acknowledgments

A program such as this encompasses a varied array of disciplines and the list of people contributing is quite long. Fabrication and assembly utilized the talents of S. Buss, H. Gervasoni, W. Haldane, J. Furch, A. Gombar, and H. Myers. Circuit design assistance was obtained from J. Henderson and J. Avins. The wind analysis was the work of R. Demers. Considerable

processing and diagnostic assistance was provided by R. D'Aiello, C. P. Wu, D. Peterson, J. J. Hughes, and F. Duigon. We especially acknowledge the continuing encouragement of F. Sterzer who conceived the small solar cell approach to high levels of concentration for its thermal and economic advantages in the generation of photovoltaic electric power. Finally, a large fraction of this work was performed under the sponsorship of the Energy Research and Development Administration through Sandia Laboratories under the management of D. Shueler.

References:

[1] B. F. Williams, *Automated Array Assembly*, Quarterly Report No. 1, ERDA/JPL-95352-76/1, March 1976.

[2] J. G. Fossum, E. L. Burgess, "Silicon Solar Cell Development for Concentrated Sunlight, High Temperature Applications," 12th Photovoltaics Specialist Conf., Nov. 15–18, Baton Rouge, La. Proceedings to be published by IEEE.

[3] F. Sterzer, "The Conversion Efficiency of Ideal P-N Junction Photovoltaic Converters in Concentrated Sunlight," *RCA Review*, **36**, p. 316, June 1975.

[4] H. J. Hovel, *Semiconductors and Semimetals Vol. II, Solar Cells*, p. 97, Academic Press, N.Y. (1975).

[5] T. Moreno, *Microwave Transmission Data*, p. 53, Dover Publications, Inc., N.Y. (1948).

[6] W. R. Runyon, *Silicon Semiconductor Technology*, p. 200, McGraw-Hill Book Co., N.Y. (1965).

[7] Baumeister and Marks, *Standard Handbook for Mechanical Engineers*, p. 4–103, McGraw-Hill Book Co., N.Y. (1967).

[8] Watson, *Theory of Bessel Functions*, Cambridge University Press, N.Y. (1958).

Thermal Analysis of Single-Crystal Silicon Ribbon Growth Processes*

A. E. Bell

RCA Laboratories, Princeton, N.J. 08540

Abstract—The thermal aspects of processes for the continuous growth of single-crystal silicon ribbon are investigated theoretically by the numerical solution of the one-dimensional heat-diffusion equation. Two separate thermal models are developed. The first considers processes, such as edge-defined growth or the inverted Stepanov technique, where the ribbon geometry is maintained through the use of a shaping die. The second model represents the ribbon-to-ribbon growth process in which no die is used, and the molten zone is maintained by heat absorbed from a highly focused carbon dioxide laser beam incident on the ribbon surface.

1. Introduction

There are several techniques currently under development in which single-crystal silicon can be continuously grown from the melt with a predetermined ribbon or sheet geometry. Such processes are especially attractive for the production of single-crystal silicon for solar-cell applications, since, unlike the Czochralski technique,[1] there is no requirement for subsequent machining or slicing of the grown crystal before incorporation into the solar-cell array. Among the techniques for growing single-crystal silicon with ribbon geometry are the edge-defined film-fed growth (EFG) process[2,3] [Fig. 1(a)], the Stepanov[4] and inverted

* The research reported herein was sponsored jointly by Jet Propulsion Laboratories, Pasadena, CA, under Contrast No. 954465, and RCA Laboratories, Princeton, N.J. 08540.

Stepanov (IS)[5] process [Fig. 1(b)], and the ribbon-to-ribbon growth (RTR) process[6] [Fig. 1(c)].

The EFG and IS processes are similar to the extent that in both cases the geometry of the single crystal is maintained by surface tension in conjunction with a shaping die through which the liquid silicon is drawn from the melt reservoir. In the EFG process the die material is chosen so that it is wetted by the liquid silicon, and capillary rise draws the

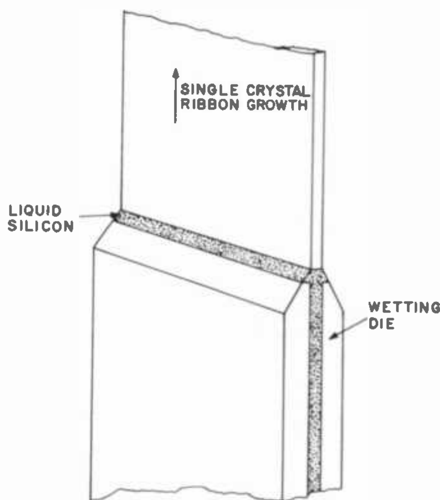


Fig. 1(a)—Schematic of silicon ribbon growth by the edge-defined growth (EFG) process.

molten silicon through the die aperture. In the IS process the die material is not wetted by the silicon and a supplementary force, e.g., a gravitational head or differential gas pressure, is required to push the liquid silicon through the die aperture. The RTR process, on the other hand, requires no shaping die; a high-power carbon dioxide laser supplies the thermal energy needed to maintain an unsupported molten zone between the polycrystalline feed ribbon and the growing single crystal. The forces of surface tension alone support the molten zone and preserve the ribbon dimensions as the single crystal is extracted. The RTR process has the advantage that the possibility of chemical contamination of the silicon by the die material is avoided; however, it does require that the polycrystalline feedstock have the same ribbon geometry as required in the final single crystal.

In all three processes an understanding of the thermal characteristics

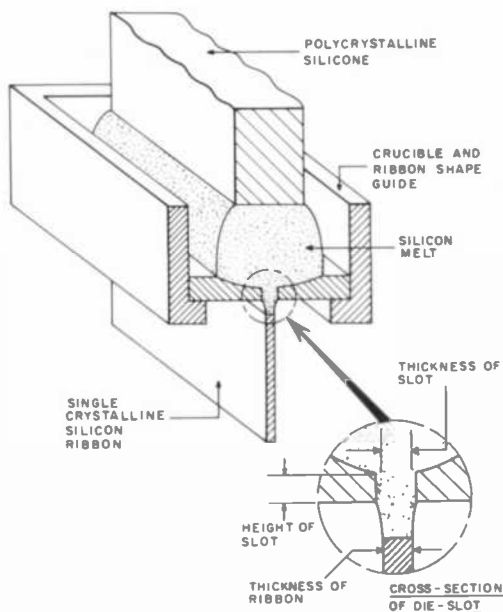


Fig. 1(b)—Schematic of silicon ribbon growth by the inverted Stepanov (IS) technique.

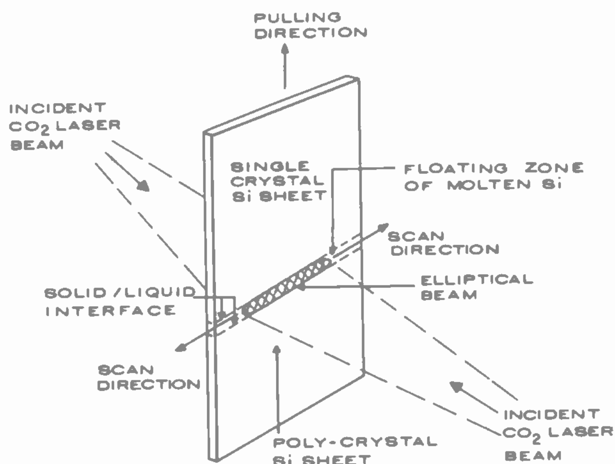


Fig. 1(c)—Schematic of ribbon-to-ribbon (RTR) growth utilizing a scanning carbon dioxide laser beam.

of the growth process is an important factor in guiding experiment and in assessing the impact of ambient conditions and die temperature or laser power on the maximum growth velocities that may be obtained. In the following, we develop a simple thermal model for the growing ribbon and discuss the results we have obtained. The essential similarity from a thermal point of view of the EFG and IS processes allows a single model to account for both. The thermal model for the RTR process is discussed separately in the Sec. 3.

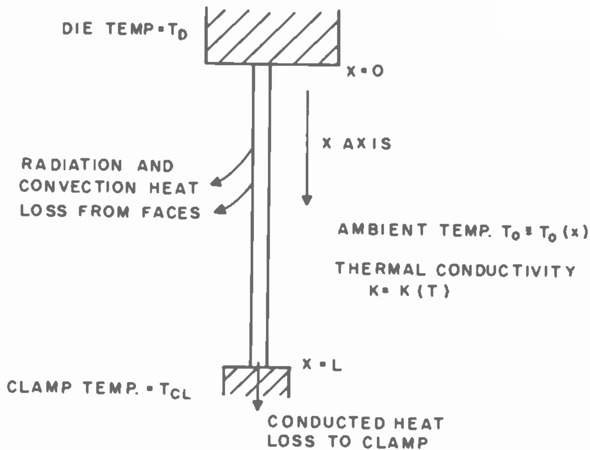


Fig. 2—One dimensional thermal model for the IS and EFG processes.

2. EFG and IS Processes

2.1 Mathematical and Computational Aspects

In Fig. 2 we illustrate the dominant factors that determine the heat transfer and thus the temperature profile within the growing silicon ribbon in the IS process. In order to form the molten zone the temperature of the die, T_D, is maintained a few degrees in excess of the melting point of silicon. Heat is drawn from the die by conduction and also by convective transport* since the ribbon is growing with a constant velocity V cm/sec and, consequently, molten silicon must flow from the die. This heat flux, together with the latent heat of fusion generated at the freezing

* We use the term convective transport to refer to that heat energy transport due solely to the motion of the hot silicon ribbon relative to the die and the growth chamber.

interface of the molten zone, passes down the ribbon where dissipation to the ambient occurs via radiation and gas convection. In addition to these heat losses there are conductive and convective losses to the clamp that supports and withdraws the growing crystal. We will assume that the clamp temperature is held at a constant value T_{CL}.

In general, the ribbon will be grown with a width in excess of 2 cm and a thickness of 0.05 cm or less. In this case, the surface-area contribution from the ribbon edges, and therefore the associated surface heat loss, will represent less than a few percent of the total surface area of the ribbon. This implies that we can safely neglect the ribbon edge losses and solve the heat-diffusion equation in two dimensions only—along the growth direction (x-axis, Fig. 2) and through the ribbon thickness (y axis). The dissipation of heat through radiation and convection from the surface of the ribbon to the ambient results in a temperature gradient normal to these faces given by

$$K\left(\frac{dT}{dy}\right)_{surf} = \dot{Q}_R + \dot{Q}_C, \qquad [1]$$

where K is the thermal conductivity of silicon and $\dot{Q}_R$ and $\dot{Q}_C$ are the heat loss rates by radiation and gas convection, respectively. As we shall see, the radiative term dominates, so that the maximum value of the temperature gradient normal to the surface can be estimated to be

$$\left(\frac{dT}{dy}\right)_{surf} \lesssim \frac{\sigma\epsilon}{K}(T_M{}^4 - T_0{}^4), \qquad [2]$$

where T_M is the melting point of silicon, ϵ is the emissivity, σ is the Stefan-Boltzmann constant, and the ambient temperature T_0 is given the lowest practical value of room temperature, 300°K. Under these conditions

$$\left(\frac{dT}{dy}\right)_{surf} \lesssim 115\text{°K/cm}. \qquad [3]$$

Since the thickness of the ribbon is 0.05 cm or less, we can now estimate the temperature differential between the ribbon center and its surface as

$$\Delta T \lesssim \frac{t}{2}\left(\frac{dT}{dy}\right)_{surf} = 3\text{°K}.$$

In view of this small maximum value for the temperature differential we can reasonably approximate the ribbon to be isothermal through its thickness and thus simplify the model further to a solution of the heat-diffusion equation[7] in one dimension only, along the growth axis of the ribbon, i.e.,

$$\frac{d}{dx}\left[K\frac{dT}{dx}\right] = \frac{2(w+t)}{wt}\,[\sigma\epsilon(T^4 - T_0^4)$$

$$+ G(T - T_0)] + VS\frac{dT}{dx} \qquad [4]$$

where w is the width of the ribbon, t is the thickness of the ribbon, G is the gas convection heat transfer coefficient, and S is the volume specific heat of silicon. The final term $VSdt/dx$ in Eq. [4] arises from the convective transport of heat resulting from the motion of the growing single-crystal ribbon relative to the die and the stationary mesh-point grid (x-axis) upon which the temperature profile is computed. (This terminology should not be confused with the convective heat loss from the ribbon surfaces to the ambient of the growth chamber.)

The numerical values of the physical constants in Eq. [4] above are well known[8,9] except for the gas convection heat transfer coefficient, G. We have estimated G using an expression appropriate to natural convection heat loss from a vertical plate in air[10] and find $G \approx 1.5 \times 10^{-3}$ W cm^{-2} K^{-1} for $T_0 = 300°$K and $G \approx 0.5 \times 10^{-3}$ W cm^{-2} K^{-1} for $T_0 = 1500°$K. Both of these values are uncertain to within a factor of two, and we have used a typical figure $G = 1.10^{-3}$ W cm^{-2} K^{-1} for the calculations that follow. Fortunately, the gas convection heat loss is a small fraction of the total dissipation so that this uncertainty does not have a large impact on the results of the calculations.

In order to correctly account for the discontinuous change in thermal conductivity that occurs on melting ($K = 0.22$ W cm^{-1} K^{-1} for solid silicon and $K = 0.64$ W cm^{-1} K^{-1} for liquid silicon at the melting point), we have used a heat flux conserving discretization scheme in the numerical analysis. Rewriting Eq. [4] with the heat flux $\phi = KdT/dx$,

$$\frac{d\phi}{dx} = \frac{2(w+t)}{wt}\,[\sigma\epsilon(T^4 - T_0^4) + G(T - T_0)] + VS\frac{dT}{dx}. \qquad [5]$$

On the ith mesh point of a grid running from $x = 0$ to $x = L$ (Fig. 2) with point spacing ΔX Eq. [5] becomes

$$\frac{\phi_{i+1/2} - \phi_{i-1/2}}{\Delta X} = \frac{2(w+t)}{wt}\,[\sigma\epsilon(T_i^4 - T_{0i}^4) + G(T_i - T_{0i})]$$

$$+ VS\frac{(T_{i+1} - T_{i-1})}{2\Delta X} \qquad [6]$$

where

$$\phi_{i+1/2} = K_{i+1/2}\frac{(T_{i+1} - T_i)}{\Delta X}, \qquad [7]$$

$$\phi_{i-1/2} = K_{i-1/2} \frac{(T_i - T_{i-1})}{\Delta X}, \qquad [8]$$

and $K_{i\pm 1/2}$ is the thermal conductivity of the silicon at a temperature midway between that on the ith and the $(i \pm 1)$th mesh points. Substituting Eq. [7] and Eq. [8] into Eq. [6], the temperature on the ith mesh point is given by

$$T_i^4 + \alpha T_i - \beta = 0, \qquad [9]$$

where

$$\alpha = \frac{C\Delta X^2 + K_{i+1/2} + K_{i-1/2}}{R\Delta X^2}, \qquad [9a]$$

$$\beta = T_{0i}^4 + \frac{C}{R} T_{0i} + \frac{K_{i+1/2}T_{i+1} + K_{i-1/2}T_{i-1}}{R\Delta X^2}$$

$$- VS \frac{(T_{i+1} - T_{i-1})}{2R\Delta X}, \qquad [9b]$$

$$R = \frac{2(w+t)}{wt} \sigma\epsilon, \qquad [9c]$$

$$C = \frac{2(w+t)G}{wt}. \qquad [9d]$$

We have solved Eq. [9] iteratively on each mesh point using the Gauss–Seidel method with an over-relaxation factor of 1.8 to 1.95 to speed the convergence.[11] The liberation of latent heat of fusion at the freezing boundary of the molten zone was taken into account in the following manner. On each successive iteration of the mesh the temperature profile was examined to determine the location of the freezing boundary and the index of the mesh point closest to this boundary position. A quantity of heat just equal to $\dot{Q}_F$, the rate of liberation of latent heat, was then included as a heat gain term at this mesh point during the subsequent iteration of the computation. $\dot{Q}_F$ depends only on the growth velocity and the cross-sectional area of the ribbon, and is easily calculated from

$$\dot{Q}_F = VwtL, \qquad [10]$$

where L is the latent heat of fusion in joules/cm^3 for silicon.

The convergence of the computer temperature profile to its steady-state value was monitored by evaluating the energy imbalance,

$$\Delta = \frac{Q^+ - Q^-}{Q^+}, \qquad [11]$$

where Q^+ is the heat input to the ribbon via conduction and convective transport at the die plus the latent heat of fusion generated at the molten zone and Q^- is the total heat loss from the ribbon surface and also through conduction and convective transport to the ribbon clamp. Since Δ is computed directly from the calculated temperature profile, it provides a useful means for physical validation of the profile by ensuring that conservation of energy is maintained. In general we were able to obtain a convergence to $\Delta \lesssim 0.02$.

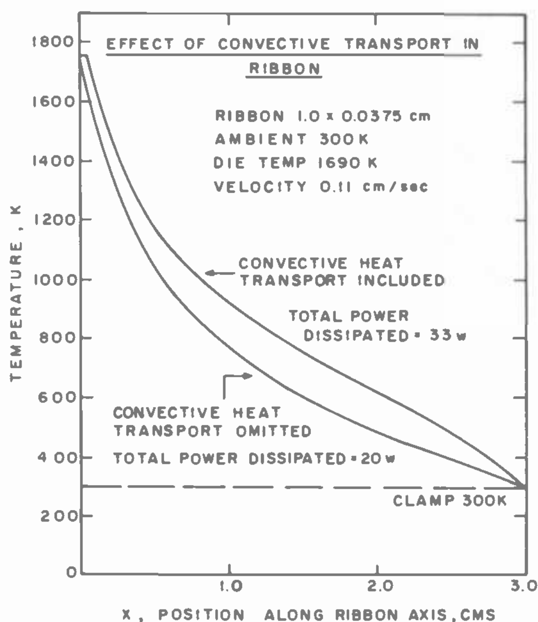


Fig. 3—Effect of convective heat transport on the temperature profile of a growing ribbon.

2.2 Results of the Model

To illustrate the effect of the convective heat transport term, $VSdT/dx$, on the temperature profile, we computed the profile for a 3.0-cm-long, 1.0×0.0375 cm ribbon growing with a velocity of 0.11 cm/sec in a 300°K uniform ambient, both with and without the inclusion of the convective term. The results are shown in Fig. 3, and we see that when the convective term is included there is a general elevation of the temperature profile along the ribbon in order to dissipate the additional heat input arising from convective transport from the die in the form of the liquid silicon

Table 1—Effect of the Convective Transport Term on the Total Power Dissipation and its Distribution Between the Various Modes of Dissipation

Parameter	No convective transport (W)	Convective transport included (W)
Total power dissipated	19.7	33.3
Radiation loss	15.88	22.64
Convection loss	2.44	3.50
Clamp conduction loss	1.38	4.29
Clamp convective loss	0.00	2.86

which is drawn from the die. The net convective heat input to the ribbon is given by

$$P_C = SV(T_D - T_{CL})wt, \qquad [12]$$

where T_D is the die temperature and T_{CL} is the clamp temperature. P_C is 11.6 W for the case shown in Fig. 3, and the distribution of the heat loss by mode of dissipation is shown in Table 1.

In Fig. 4 we show the temperature profile in the neighborhood of the molten zone as a function of the ribbon growth velocity. In all cases the

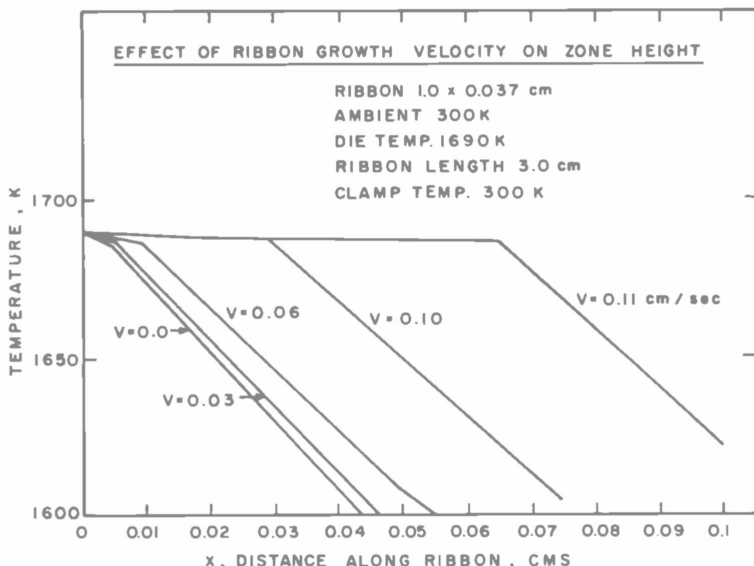


Fig. 4—Temperature profile in the vicinity of the molten zone as a function of growth velocity in uniform 300°K ambient.

ribbon geometry (0.0375 × 1.0 cm, 3.0 cm long), die temperature (1690°K) and ambient temperature (300°K, uniform along ribbon) remain unchanged and only the growth velocity is varied. The position of the freezing boundary of the molten zone is given by the point at which the gradient of the temperature profile dT/dx undergoes an abrupt change in value. This feature is a direct consequence of the liberation of the latent heat of fusion Q_F (Eq. [10]) and the discontinuous change in the thermal conductivity of silicon at the melting point, since conservation of energy across the freezing boundary demands that

$$-K_S \left(\frac{dT}{dx}\right)_{x_B+\delta x} = -K_L \left(\frac{dT}{dx}\right)_{x_B-\delta x} + \dot{Q}_F, \qquad [13]$$

where x_B is the position of the freezing interface. Since the heat conducted away from the freezing interface $-K_S(dT/dx)_{x_B+\delta x}wt$ is limited by the rate at which heat can be dissipated from the remainder of the ribbon ($x > x_B$) to the ambient and to the clamp, we expect that as the velocity is increased, the magnitude of the latent heat term in Eq. [13] will increase, and the temperature gradient in the liquid silicon $(dT/dx)_{x<x_B}$ must decrease. This effect is evident in Fig. 4 and leads to a reduction in the heat drain on the die due to conduction into the molten zone. On the other hand, increasing the growth velocity will lead to increased heat drain on the die due to the convective transport of heat retained in the molten silicon supplied from the die aperture. The dependence of these quantities on the ribbon growth velocity for the case shown in Fig. 4 is indicated in Table 2.

The most notable feature of Fig. 3 is the increase of the zone height with the ribbon growth velocity. The molten zone height, normalized to the ribbon thickness, is shown in Fig. 5 directly as a function of growth velocity for the data of Fig. 4, and we see an initially slow rate of increase of the zone height, which becomes ever more rapid as the growth velocity approaches and exceeds about 0.1 cm/sec (140 inches/hr). Clearly, the surface tension of the liquid silicon cannot support an indefinitely large molten zone height, and even before a catastrophic collapse of the zone occurs a zone height that is large relative to the ribbon thickness is expected to affect the geometry of the growing single crystal. Thus the divergence of the zone height above the critical velocity in effect limits the maximum growth rate for the ribbon for a given set of ambient conditions.

In Fig. 6 we show the temperature profile in the neighborhood of the molten zone as a function of the growth velocity for a 0.0375 × 1.0 cm ribbon, 3.0 cm long in an elevated uniform ambient of 1300°K for a die temperature of 1690°K (T_M = 1683°K). The elevated ambient tem-

Table 2—Influence of Growth Velocity on the Thermal Characteristics of Growing Ribbon for 300° K Ambient and 1690° K Die Temperature

Die Temp T_D (K)	Ambient Temp. (K)	Growth Velocity V (cm/s)	Growth Velocity V (in./min)	Zone Height H (cm)	Zone Height $\frac{H}{t}$	Latent Heat Input Q_c (W)	Net Convective Input P_c (W)	Power Conducted From Die P_D (W)	Total Power Dissipated (W)	Mode of Power Dissipation % Radiation	Convection	Conduction	Convective Transport
1690	300	0.0	0.0	0.006	0.17	0.0	0.0	19.5	19.0	80	13	7	0
1690	300	0.03	0.71	0.007	0.18	4.66	3.17	13.8	21.7	76	12	9	3
1690	300	0.06	1.42	0.012	0.31	9.31	6.32	8.46	24.9	73	11	10	6
1690	300	0.10	2.36	0.037	0.99	15.5	10.52	2.76	30.4	71	10	12	7
1690	300	0.11	2.60	0.067	1.79	17.1	11.6	2.88	33.3	71	10	12	7

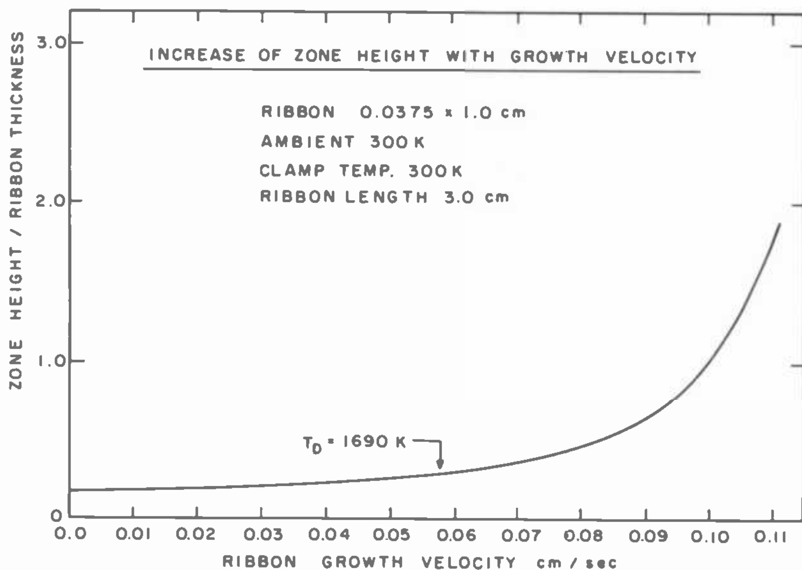


Fig. 5—Zone height, normalized to the ribbon thickness, as a function of growth velocity taken from the data of Fig. 4.

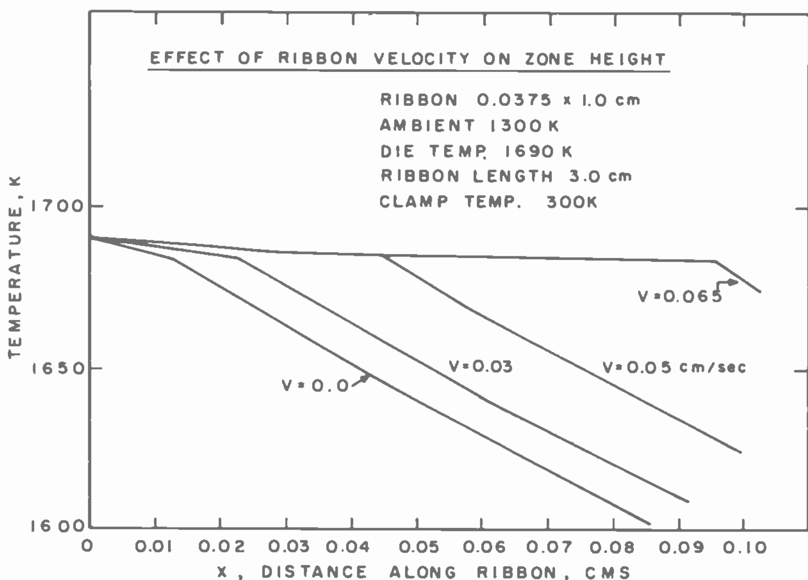


Fig. 6—Temperature profile in the vicinity of the molten zone as a function of growth velocity in a uniform 1300°K ambient (die temperature 1690°K).

perature reduces the rate at which heat can be dissipated from the ribbon to the ambient and the ribbon clamp to about 11 W ($V = 0$) compared to about 19 W ($V = 0$) for the 300°K ambient. Moreover, we note that the mode of dissipation is shifted even more toward the radiative mode (compare Tables 2 and 3). The reduction in power dissipation in the 1300°K ambient reduces the temperature gradient in the solid adjacent to the molten zone, $(dT/dx)_{x_B+\delta x}$, and in view of Eq. [13] we expect the reduction of the temperature gradient in the liquid, and the consequent increase in the zone height, to occur at lower growth velocities. This is apparent in Fig. 6, where the zone height has already grown quite large for a velocity of only 0.065 cm/sec.

The zone height is related to the temperature gradient in the liquid zone and the temperature of the die by

$$H \approx \frac{T_D - T_M}{(dT/dx)_{x<x_B}}; \qquad [14]$$

therefore some reduction in the zone height under conditions of constant growth velocity and ambient temperature should be possible by reducing the elevation of the die temperature above the melting point. In Fig. 7 we show the temperature profile for the same ribbon and ambient as in Fig. 6, but with the die temperature reduced to 1685°K. The change in the zone height at constant growth velocity is shown most clearly in Fig. 8, where we note that despite the apparent reduction in the zone height at the lower velocities, the maximum practical growth rate is increased by perhaps only 20% due to the rapid divergence of the zone height in both instances as the critical velocity is approached. As the critical velocity is approached the temperature gradient in the liquid zone tends to zero so that Eq. (13) becomes

$$-K_S \left(\frac{dT}{dx}\right)_{x_B+\delta x} = Q_F = \hat{V}wtL, \qquad [15]$$

where $\hat{V}$ is the critical velocity. $-K_S(dT/dx)_{x_B+\delta x}$ is equal to the rate at which heat is dissipated from the ribbon at $x > x_B$, so that the maximum growth rate is seen to be that velocity that generates latent heat of fusion at a rate equal to the rate at which heat can be dissipated from a stationary ribbon, P_S. (The increase in dissipated power seen in Tables 2–4 for a moving ribbon serves only to balance the extra convective heat input, and does not aid in the dissipation of latent heat.) Thus, for given ambient conditions

$$V \lesssim \frac{P_S}{Lwt}. \qquad [16]$$

Table 3—Influence of Growth Velocity on Thermal Characteristics of a Growing Ribbon for 1300° K Ambient with Die Temperatures of 1690° K and 1685° K

Die Temp T_D (K)	Ambient Temp (K)	Growth Velocity V		Zone Height		Latent Heat Input Q_c (W)	Net Convective Input P_c (W)	Power Conducted From Die P_c (W)	Total Power Dissipated (W)	Mode of Power Dissipation (%)			
		(cm/s)	(in./min)	H (cm)	$\frac{H}{t}$					Radiation	Convection	Conduction	Convective Transport
1690	1300	0.0	0.0	0.014	0.36	0.0	0.0	11.0	10.7	98	2	0	0
1690	1300	0.03	0.71	0.024	0.63	4.66	0.94	6.3	14.5	78	2	0	20
1690	1300	0.05	1.18	0.046	1.23	7.76	1.48	3.95	17.6	70	2	0	28
1690	1300	0.065	1.54	0.096	2.57	10.1	1.94	3.37	21.3	68	2	0	30
1685	1300	0.0	0.0	0.0007	0.02	0.0	0.0	10.4	10.2	98	2	0	0
1685	1300	0.03	0.71	0.0036	0.10	4.66	0.87	5.57	13.7	76	2	0	22
1685	1300	0.06	1.42	0.019	0.50	9.31	1.75	1.12	17.7	65	2	0	33
1685	1300	0.07	1.65	0.064	1.70	10.9	2.03	1.52	20.6	65	2	0	33

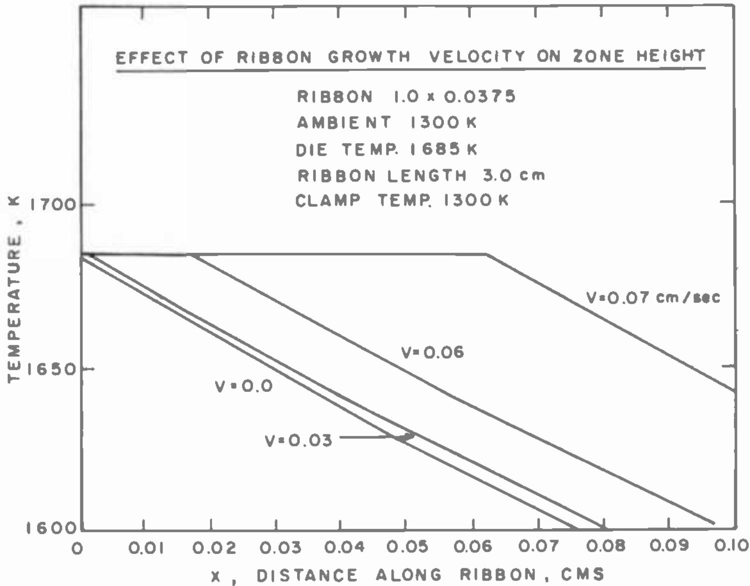


Fig. 7—Temperature profile in the vicinity of the molten zone as a function of growth velocity in a uniform 1300°K ambient (die temperature 1685°K).

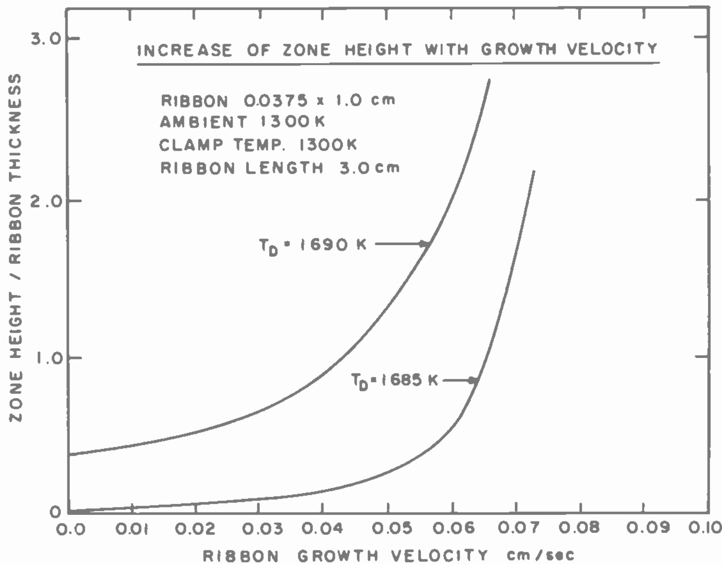


Fig. 8—Zone height, normalized to the ribbon thickness as a function of growth velocity taken from the data of Figs. 6 and 7.

Reducing the die temperature will help to maintain a lower zone height at subcritical growth rates, but since reducing the die temperature under otherwise constant ambient does not greatly affect P_S (see Table 3), the critical growth rate remains essentially unchanged.

Eq. [16] suggests that some increase in the critical growth rate may be achieved by reducing the ribbon thickness, t, thereby reducing the rate of liberation of latent heat at constant growth velocity. In Fig. [9] we investigate this possibility by showing the temperature profile near the molten zone for a ribbon whose thickness is reduced to one-half (0.0187 cm) of the previous ribbon thickness. The ambient temperature is still maintained at 300°K, and we see that, for a given growth velocity, the zone height for the thinner ribbon is indeed smaller than that for the 0.037-cm-thick ribbon shown in Fig. [4]. When the reduced zone height is normalized to the reduced ribbon thickness as is shown in Fig. [10] and the maximum growth velocities of the 0.0375-cm-thick and 0.0187-cm-thick ribbons are determined for the same H/t criterion (for example $H/t \approx 2$), then we see that the 50% reduction in ribbon thickness gains ~35% increase in growth velocity, to yield a figure 0.15 cm/sec for the thinner ribbon. In Table 4 we observe that the reduction in ribbon thickness has lowered the thermal conductance of the ribbon so that P_S, the power dissipated to the ambient by a stationary ribbon, has been reduced from 19 W where t = 0.0375 cm (Table 2) to 13.3 W where t = 0.0187 cm in the same ambient conditions. Thus, while it is evident that a thinner ribbon will have a higher maximum growth velocity, some of the advantage of the redced generation of latent heat will be offset by the reduced thermal conductance of the ribbon and the consequently smaller value of heat dissipation from ribbon to ambient.

In summary, our model for the EFG and IS processes leads to the following points:

(1) For a given ribbon geometry and ambient conditions, the zone height increases with the growth velocity and does so more and more rapidly as the critical (maximum) growth velocity is approached.

(2) The critical growth velocity is given approximately as that velocity that leads to the generation of latent heat of fusion at a rate just equal to the rate at which heat can be dissipated from the ribbon when stationary in the same ambient.

(3) The highest growth rates will be achieved with the coolest ambient temperatures, by improving the convective heat transfer coefficient, G, or reducing the separation of the die from the ribbon clamp so as to increase the conducted heat loss from the ribbon.

(4) At a given growth rate and under otherwise constant conditions, reduction of the die temperature will reduce the zone height.

(5) The maximum growth rate for thinner ribbon is increased, but by

Table 4—Influence of Growth Velocity on Thermal Characteristics of 0.0187 × 1.0 cm Growing Ribbon for 300° K Ambient and 1690° K Die Temperature

Die Temp T_D (K)	Ambient Temp (K)	Growth Velocity V (cm/s)	Growth Velocity V (in./min)	Zone Height H (cm)	Zone Height $\frac{H}{t}$	Latent Heat Input Q_L (W)	Net Convective Input P_c (W)	Power Conducted from Die P_D (W)	Total Power Dissipated (W)	Mode of Power Dissipation % Radiation	Mode of Power Dissipation % Convection	Mode of Power Dissipation % Conduction	Mode of Power Dissipation % Convective Transport
1690	300	0.00	0.00	0.005	0.27	0.00	0.00	13.7	13.3	83	14	3	0
1690	300	0.06	1.42	0.005	0.27	4.64	3.15	8.06	10.1	77	14	5	4
1690	300	0.10	2.36	0.010	0.53	7.74	5.26	4.67	18.3	75	13	6	6
1690	300	0.15	3.54	0.043	2.30	11.60	7.90	2.35	22.9	74	12	7	7
1690	300	0.16	3.78	0.063	3.37	12.4	8.39	2.72	24.2	74	11	7	8

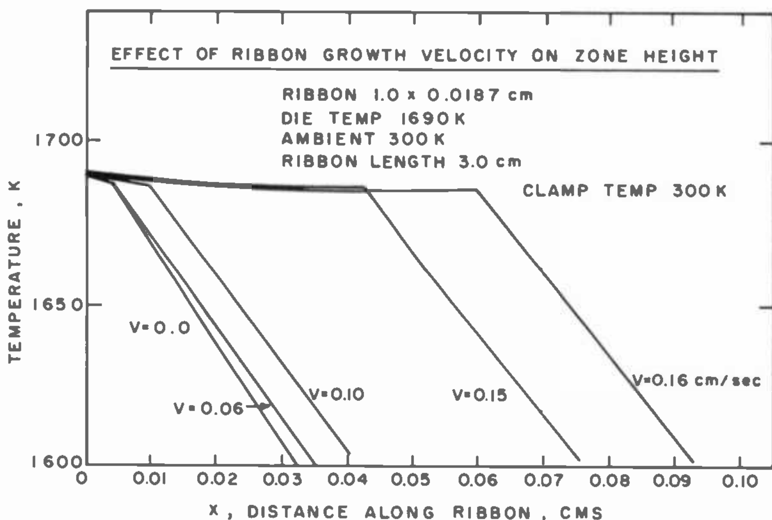


Fig. 9—Temperature profile in the vicinity of the molten zone as a function of growth velocity for a ribbon with 50% reduction in thickness (uniform ambient 300°K).

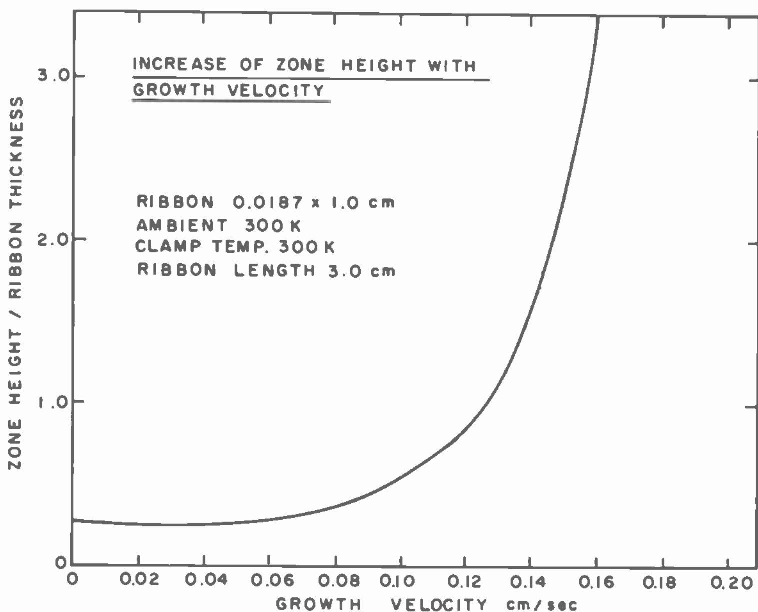


Fig. 10—Zone height, normalized to ribbon thickness, as a function of growth velocity, taken from the data of Fig. 9.

an amount that is less than proportional to the reduction in the thickness.

3. Ribbon-to-Ribbon (RTR) Process

Fig. 1(b) is a schematic of the RTR growth process. A laser beam, focused into the shape of an ellipse, is scanned across the width of the silicon ribbon with just sufficient beam power to maintain a zone of molten

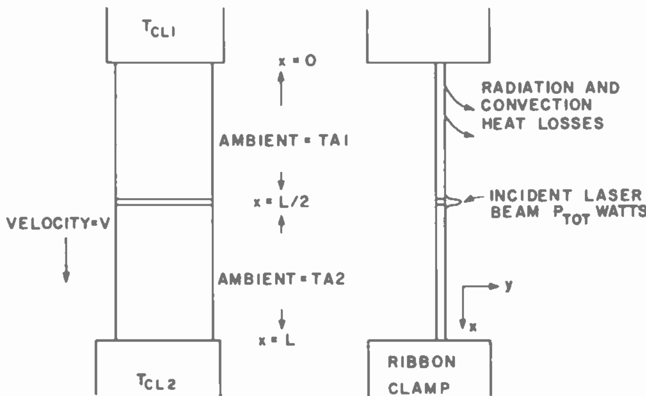


Fig. 11—One-dimensional thermal model for the RTR growth process.

silicon between the polycrystalline ribbon and the single-crystal ribbon. During the growth process the preshaped polycrystalline ribbon is fed continuously into the molten zone in order that the single-crystal ribbon, of essentially similar dimension, may be withdrawn at the same constant velocity. One theoretical model[12] estimated that in order to maintain a stable zone and ribbon geometry for the growing ribbon, the height of the molten zone should be limited to a dimension approximately equal to the ribbon thickness. Experimentally stable growth configurations have been reported[13] where the zone height was two or three times the ribbon thickness.

The simplified thermal model that we use to describe the RTR process is shown in Fig. 11. The ribbon is considered to be supported between two clamps of specified temperatures T_{CL1} and T_{CL2}, separated by a distance L. Provided the laser beam scan is sufficiently rapid, the excursions of the ribbon temperature profile about its mean value will be small so that we may replace the scanning beam with a stationary beam focused midway between the clamps to have uniform intensity across

the ribbon width and a Gaussian intensity along the ribbon axis (x axis Fig. 11) given by

$$P(x) = \frac{P_{TOT}}{\sigma_s w} \sqrt{\frac{2}{\pi}} \exp\left[-2\left(\frac{x - (L/2)}{\sigma_s}\right)^2\right] \text{ W cm}^{-2}, \qquad [17]$$

where P_{TOT} = total incident beam power in watts and $2\sigma_s$ = width of the beam between $1/e^2$ intensity points.

Using the same arguments as those applied in the previous section we may also approximate the dominant thermal aspects of the RTR process by a numerical solution of the one-dimensional heat-diffusion equation;

$$\frac{d}{dx}\left[K\frac{dT}{dx}\right] = \frac{2(w + t)}{wt}\left[\sigma\epsilon(T^4 - T_0{}^4) + G(T - T_0)\right]$$

$$- (1 - R)\frac{P(x)}{t} + VS\frac{dT}{dx}, \qquad [18]$$

where R is the reflectivity of liquid silicon at the wavelength of the carbon dioxide laser beam, 10.6 μm ($R = 0.30$).[8] Eq. [18] for the RTR process differs from that applied to the IS and EFG methods only by the term that represents the heat absorbed from the laser beam. The temperature profiles along the axis of the growing ribbon and feeder ribbon were obtained using essentially the same numerical procedure as that already detailed in Sec. 2. The computer program allows specification of a variety of ambient conditions and clamp temperatures, including those corresponding to preheating the polycrystalline feed ribbon before it enters the region of the molten zone.

3.1 Results of the Model

As an example, in Fig. 12, we show the computed temperature profile along a ribbon of length 5.52 cm and cross-section 0.0375 × 1.0 cm in a uniform ambient temperature ($T_{A1} = T_{A2}$ in Fig. 11) of 300°K. The incident laser power is 62 W and the beam is focused so that $2\sigma_s = 0.04$ cm, approximately the ribbon thickness. As would be expected, the profile for the stationary ribbon ($V = 0$) is symmetrical about the laser beam, and we see that when the ribbon moves with $V = 0.11$ cm/sec the effect of the convective term is to shift the temperature profile in the direction of motion, reducing the temperatures along the feeder ribbon and increasing those along the growing single crystal.

In Fig. 13 a more detailed view of the temperature profile of the same ribbon is shown for a range of incident laser powers in order to demonstrate the increase in the height of the molten zone as the laser power

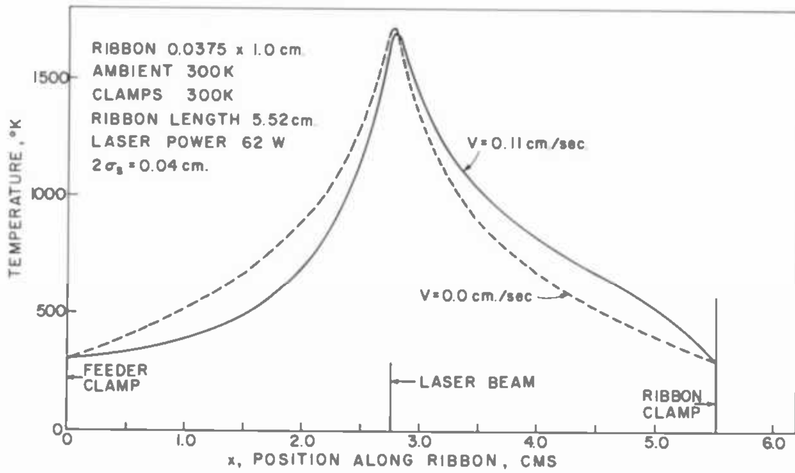


Fig. 12—Temperature profile along both feeder ribbon and single-crystal ribbon in uniform 300°K ambient.

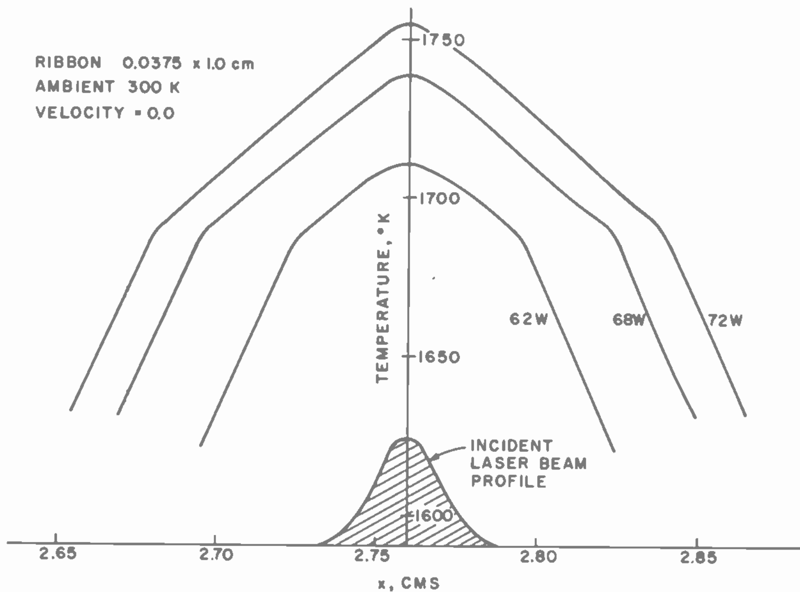


Fig. 13—Temperature profile in the neighborhood of the molten zone for various incident laser powers (uniform 300°K ambient).

is increased. In all cases shown in Fig. 13 the growth velocity is zero. When the incident laser power is equal to 62 W, the height of the molten zone is equal to about twice the ribbon thickness, and we expect the zone to be stable under these conditions. Since the reflectivity of the silicon was set to 0.30, this incident power corresponds to a heat input of about 43 W, so that under these conditions both the feeder ribbon and the

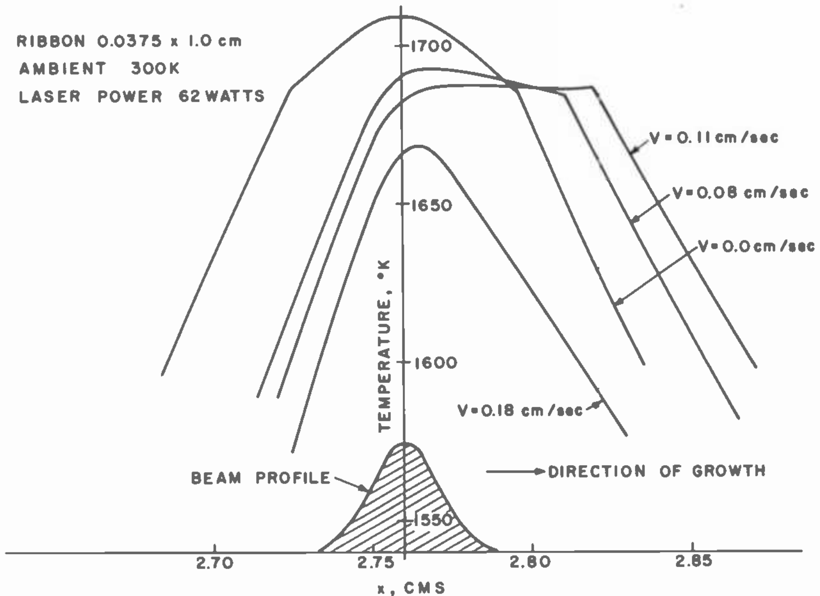


Fig. 14—Effect of increasing the ribbon growth velocity on the temperature profile in the vicinity of the molten zone (uniform 300°K ambient).

single-crystal ribbon dissipate 21.5 W to the ambient and by conduction to their respective clamps. In Table 2 for the EFG or IS process we see that for a stationary ribbon of similar geometry also in a 300°K uniform ambient, the power dissipated from the ribbon is 19.5 W. In this regard it is apparent now that for a ribbon of given dimensions, in a prescribed ambient configuration, the laser serves only as an intense, localized source of heat to provide the energy that would otherwise be supplied by the die in the EFG and IS schemes.

In Fig. 14 we show the effect on the temperature profile in the neighborhood of the molten zone when the ribbon growth velocity is increased under otherwise constant conditions with the uniform 300°K ambient. In all cases the laser power is 62 W and we have chosen to focus the beam for $2\sigma_s = 0.04$ cm, approximately the ribbon thickness. As the velocity

is increased, the position of the molten zone is progressively shifted in the direction of the ribbon growth, and at the same time the peak temperature reached within the zone is steadily reduced. Both these effects are a direct result of the convective term, $VS(dT/dx)$, included in Eq. [18]. At a growth velocity of 0.11 cm/sec, the position of the melting boundary of the liquid zone has advanced to a position directly beneath the laser beam, which is still able to supply sufficient energy to the boundary to satisfy the demands of the latent heat of melting. If the growth velocity is increased still further, then it is apparent that the molten zone can no longer be maintained, since the melting boundary will have passed completely through the laser-illuminated region, indicating that insufficient power is available to both elevate the feeder ribbon to the melting point and supply the latent heat of fusion. Eventually, for high enough velocity ($V = 0.18$ cm/sec in Fig. 14) the effect of the convective term is so pronounced that the peak temperature achieved in the ribbon remains below the melting point. This thermal model for the RTR growth of the 0.0375×1.0 cm ribbon in 300°K uniform ambient places the maximum growth velocity at $V \approx 0.11$ cm/sec, which is essentially the same as that already determined in Sec. 2 for similar ribbon geometry and ambient condition in the EFG or IS process (See Fig. 4). There is, however, one distinct difference in the behavior of the temperature profile as the critical velocity is approached, and that is that no divergence of the zone height is observed in the RTR process. This difference in the limiting behavior of the two thermal models and their respective growth processes is due to the differing nature of the thermal boundary conditions imposed by each of the experimental configurations. In the case of the EFG and IS model the molten zone is subject to the constant-temperature boundary condition of the die aperture, whereas in the RTR model the molten zone is subject only to the constant power input of the laser beam as the boundary condition. The reduction of the peak temperature in the molten zone as velocity is increased in the RTR model compensates for the reduction in the temperature gradient within the zone at the freezing boundary (Eq. [13]), so that the zone height remains essentially constant.

For growth velocities in excess of 0.11 cm/sec, but less than 0.17 cm/sec, the one-dimensional thermal model for the RTR process fails to converge to a solution for the thermal profile, except when the latent-heat parameter is reduced to some fraction of the actual value appropriate for silicon. This is a shortcoming of the one-dimensional treatment in that a two-dimensional treatment, which would include the temperature distribution through the ribbon thickness (y axis Fig. 11), would be able to represent qualitatively the partial melt-through of the molten zone, which is indicated by the one dimensional model for $0.11 < V < 0.18$

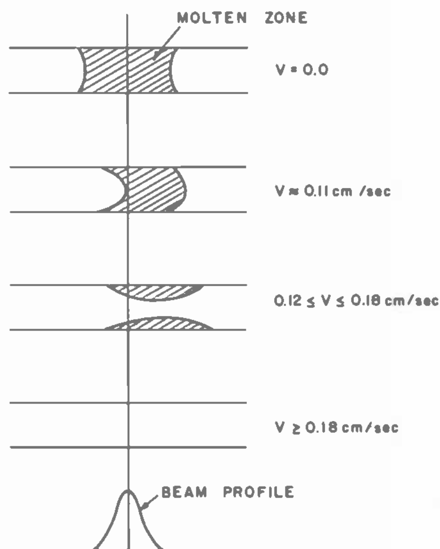


Fig. 15—Schematic representation of molten zone geometry in two dimensions—along the ribbon axis and through the ribbon thickness. The laser power is held constant.

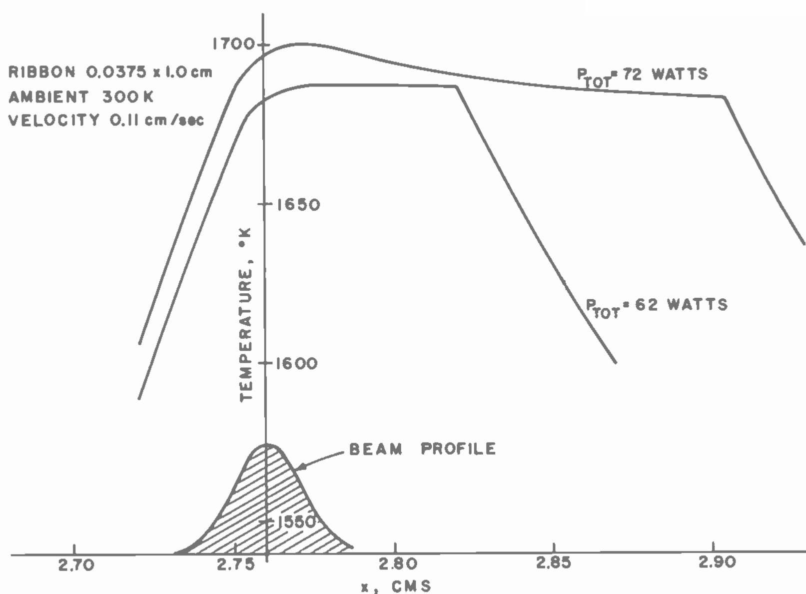


Fig. 16—Effect of increasing the laser beam power at constant velocity 0.11 cm/sec in a uniform 300°K ambient.

cm/sec. Obviously, this range of velocities depends on the actual ribbon thickness and ambient configuration, but for the present case the zone geometry as a function of growth velocity is shown schematically in Fig. 15.

In an attempt to increase the maximum growth velocity under given ambient conditions, it might at first seem reasonable to simply increase the power in the laser beam in order to provide the increased requirement for both latent and convective heat demanded by the increased growth velocity. Fig. 16 shows that the effect of increasing the laser power from 62 to 74 W is to allow only a minimal increase in the maximum growth velocity, but to cause a considerable increase in the zone height to a value several times the ribbon thickness, where the stability of the zone and the single-crystal geometry will be questionable. The observation that the maximum velocity is hardly affected by increasing the laser power is consistent with the conclusion reached in Sec. 2 that the rate of crystal growth is limited to that velocity that generates latent heat of fusion at the freezing boundary at a rate just equal to the rate of heat dissipation from the surfaces of the growing crystal and by conduction to the clamp. This rate of dissipation is determined solely by the ambient conditions and the point at which the single-crystal ribbon is clamped, and is not affected by the heat input due to the laser. Furthermore, since the ribbon geometry and ambient conditions of the present RTR example are very similar to those of the first example considered in the IS model [See Figs. 4 and 5], it is quite consistent that the predicted maximum growth velocity should be the same, i.e., ~0.11 cm/sec.

In Fig. 17 we show the temperature profile along a stationary ribbon of cross section 0.0375 × 1.0 cm in an elevated uniform ambient temperature of 1300°K. The increased ambient temperature reduces the power dissipated from the surfaces of the ribbon so that the incident laser power required to establish the molten zone is correspondingly reduced to about 33 W. Fig. 18 shows the shift in the location at the molten zone with increasing growth velocity, and the effect of the reduced dissipation from the surfaces of the single-crystal ribbon is to limit the maximum growth velocity to about 0.06 cm/sec. Once again, this value is the same as that predicted by the model for the IS process for a ribbon of the same geometry growing under the same ambient conditions.

When the ribbon growth velocity is equal to the maximum value for the specified ambient conditions, the heat of fusion liberated at the freezing boundary of the molten zone is conducted into the single-crystal ribbon and then dissipated to the ambient and the clamp. Thus, at maximum velocity, the laser power serves essentially to provide the energy necessary to heat the silicon feed ribbon to the melting point, supplying the required amount of latent heat of melting and also making

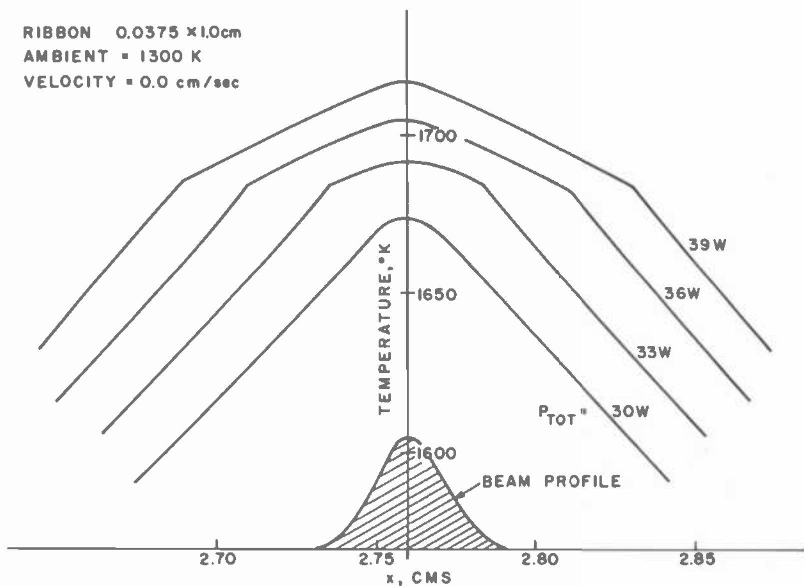


Fig. 17—Temperature profile in the vicinity of the molten zone for various laser powers (uniform 1300°K ambient).

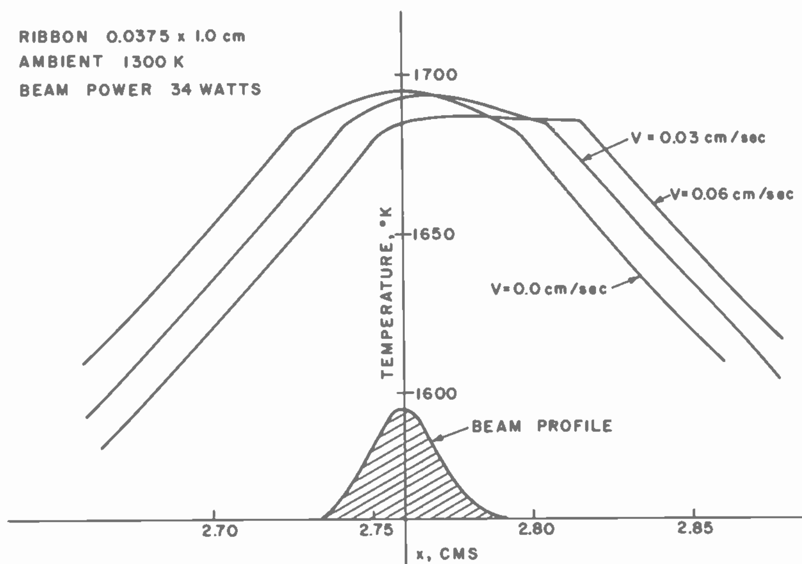


Fig. 18—Effect of increasing the ribbon growth velocity on the temperature profile in the vicinity of the molten zone (uniform 1300°K ambient).

up the radiation convection and conduction losses to the ambient and the clamp from the polycrystalline feeder ribbon. In other words, while the thermal ambient conditions on the single-crystal side of the molten zone dictate the maximum allowable growth velocity, those on the feeder side of the zone will determine the laser power required to maintain the molten zone. This observation implies that in terms of achieving the highest growth rates with the minimum laser power, one should arrange for high ambient temperatures at the feeder ribbon and low ambient temperatures around the growing crystal. In Fig. 19 we show the asymmetric temperature profile computed along the 0.0375 × 1.0 cm ribbon when the feeder clamp and ambient temperatures were maintained at 1300°K, while the single-crystal clamp and ambient temperatures are held at 300°K. The profile in the region of the molten zone is shown in more detail in Fig. 20 for an incident laser power of 48 W and a range of crystal growth velocities up to a maximum of 0.11 cm/sec.

Table 5—Effect of Ambient Conditions on Laser Power and Growth Rates in the RTR Process.

T_{CL1} (°K)	T_{CL2} (°K)	P_{TOT} (W)	V_{max} (cm/sec)	P_F (W)	P_A (W)
300	300	62	~0.11	28.6	43.4
1300	1300	34	~0.06	11.1	22.4
1300	300	48	~0.11	20.3	33.6

T_{CL1} = Feeder ribbon clamp temperature.
T_{CL2} = Single crystal ribbon clamp temperature.
P_{TOT} = Incident laser power.
V_{max} = maximum growth velocity for 0.0375 cm × 1.0 cm ribbon 5.52 cm between clamps
P_F = sum of latent and specific heats supplied to feeder ribbon (Eq. [19]).
P_A = Power absorbed from the laser beam if R = 0.30.

The data obtained from our model of the RTR process under the three ambient temperature distributions that we have computed are compared in Table 5. We see that the predicted maximum growth velocity, for the asymmetric 1300°K/300°K ambient is equal to that for the 300°K/300°K uniform ambient; however, the incident laser power has been reduced by 25% due to the preheating effect of the elevated ambient and clamp temperatures for the feeder ribbon. Compared to the predicted growth velocity of the uniform 1300°K/1300°K ambient, that for the 1300°K/300°K ambient has been increased by a factor of almost two with only a 50% increase in the laser power required. This additional heat input from the laser is required to provide for the increased rate at which latent and specific heat must be supplied to the moving feeder ribbon

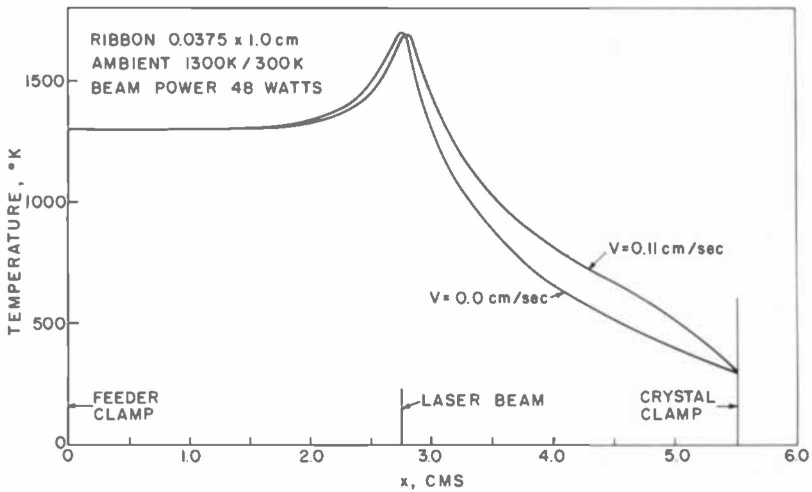


Fig. 19—Temperature profile along both feeder ribbon and single-crystal ribbon for asymmetric ambient temperature. Ambient is 1300°K for $0 < x < L/2$ and 300°K for $L/2 < x < L$.

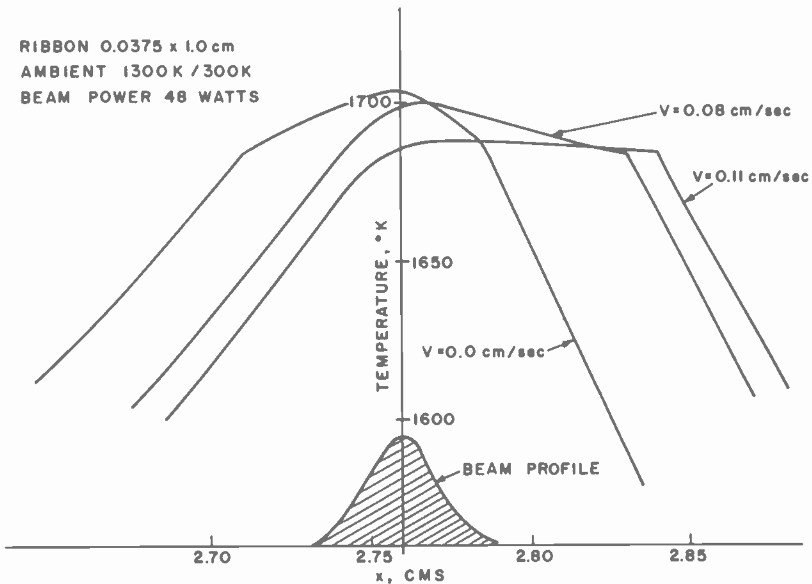


Fig. 20—Effect of increasing the ribbon growth velocity on the temperature profile in the vicinity of the molten zone for the asymmetric ambient of Fig. 19.

as it is transported more rapidly to the melting boundary of the molten zone. The sum of the rates at which latent and specific heats must be supplied is

$$P_F = Vwt[L + S(T_M - T_{CL1})], \qquad [19]$$

which is tabulated in Table 5, along with the heat absorbed from the laser beam in each instance. The data of Table 5 shows that when the maximum velocity is increased from 0.06 cm/sec in the 1300°K/1300°K ambient to 0.11 cm/sec in the 1300°K/300°K case, P_F is increased by about 9 W, which corresponds well with the extra 11 W required in the absorbed power input from the laser.

In summary, our model for RTR process leads to the following points:

(1) For equal ribbon geometry and thermal conditions in the single-crystal ribbon side of the growth chamber, the predicted maximum ribbon growth velocities are the same as those for the EFG or IS process.

(2) The laser power required to maintain the molten zone at the maximum growth velocity is determined by the ambient conditions on the polycrystalline feed ribbon side of the growth chamber.

(3) An asymmetric ambient with a high temperature in the vicinity of the feed ribbon, but a low temperature around the growing single crystal, will improve the growth rate per unit of laser-beam power delivered to the ribbon. The absolute lower limit for the laser-beam power corresponds to the case where the feeder ambient heats the polycrystalline ribbon almost to the melting point, so that the laser is required to provide only the latent heat of melting, i.e.,

$$P_{LASER} > \frac{VwtL}{R}. \qquad [20]$$

Conclusions

Despite the contrasting experimental configurations represented in the EFG, IS, and RTR single-crystal silicon ribbon growth processes, the prediction of the two thermal models presented above indicate that as far as the purely thermal aspects of the ribbon growth are concerned these processes are actually very similar. The crystal growth velocity is found to be limited to a maximum value that is independent of the type of ribbon growth process employed, being determined only by the desired ribbon geometry and the ambient thermal conditions in the single-crystal ribbon growth chamber.

Acknowledgment

The author would like to acknowledge the assistance of R. W. Klopfenstein in the mathematical and computational aspects of this work, and also G. W. Cullen, K. M. Kim, and S. Berkman for other technical discussions.

References:

[1] H. E. Buckley, *Crystal Growth*, John Wiley and Sons, Inc., New York (1951).

[2] B. Chalmers, et al, *Continuous Silicon Solar Cells*, NSF Grant GI-37067X, Final Report, 1975.

[3] J. C. Schwartz, T. Surek, and B. Chalmers, *J. Electronic Mat.*, 4, 255 (1975).

[4] A. V. Stepanov et al., *Bull. Acad. Sci. USSR, Phys. Series*, 33, 1826 (1969).

[5] K. M. Kim, G. W. Cullen, S. Berkman, and A. E. Bell, *Silicon Sheet Growth by the Inverted Stepanov Technique*, Quarterly Progress Report No. 1, June 1976, ERDA/JPL/954465-76/1.

[6] R. W. Gurtler and A. Baghdadi, *Laser Zone Growth in a Ribbon-to-Ribbon (RTR) Process*, Quarterly Report No. 1, March 1976, ERDA/JPL/954376-76/1.

[7] H. S. Carlsaw and J. C. Jaeger, *Conduction of Heat in Solids*, Clarendon Press, Oxford, England (1959).

[8] M. Neuberger, et al., *Silicon*, Hughes Aircraft Co., Culver City, Colorado, AD698342.

[9] Y. S. Touloukian et al., Eds., *Thermophysical Properties of Matter*, IFI/Plenum Press, New York (1970).

[10] A. I. Brown and S. M. Marco, *Introduction to Heat Transfer*, McGraw-Hill Book Co., New York (1942).

[11] B. Carnahan, H. A. Luther, J. O. Wilks, *Applied Numerical Methods*, John Wiley and Sons, New York (1969).

[12] A. Mustachi, private communication.

[13] R. W. Gurtler and A. Baghdadi, Third Low Cost Silicon Solar Array Project Integration Meeting, Pasadena, Calif., July 1976.

Drain Characteristics of Thin-Film MOS FET's

S. T. Hsu

RCA Laboratories, Princeton, N. J. 08540

Abstract—Theoretical expressions for the static *I-V* characteristics of thin film MOS FET's have been derived in terms of doping concentration, current carrier mobility, device geometry, and gate voltage. The drain saturation current of a deep-depletion type thin-film MOST operated in depletion mode depends strongly on the doping and carrier mobility profiles in the semiconductor film. At large gate biases, the drain saturation current increases with the square of $V_G - V_{FB}$. The drain saturation current of an enhancement-type thin-film MOST is proportional to the square of $V_G - V_T$. The threshold voltage, V_T, is a function of the thickness of the gate insulator, the doping concentration, and the thickness of the silicon film. The theoretical results are shown to be in good agreement with experimental data. It is shown that the anomalous drain characteristics of thin-film MOST's are due to the modification of charge density at the space-charge region by the drain biases, while those of thick-film MOST's are due to the floating-substrate effect.

1. Introduction

SOS technology has been greatly advanced in recent years. However, most of the investigations are of SOS integrated circuits. The device physics investigators focus their interest on the current kinks in SOS MOST's.[1-4]

Thin-film deep-depletion type SOS MOST's were first described by Heiman[5] and were theoretically analyzed by Hofstein[6] in terms of generalized voltages. In this paper we present an alternative analysis of the drain characteristics of thin-film MOS field-effect transistors operated at drain saturation voltages. The results show that if current carriers are injected into the space-charge region due to multiplications of elec-

tron–hole pairs at the drain depletion region, or if the trapped electron density in the space-charge region is modified by the voltage applied to the drain, the drain current exhibits anomalous characteristics. The drain saturation current, the transconductance, and the channel conductance of the MOST's are expressed in terms of the doping concentration, the carrier mobility, the gate voltage, and the thickness of the semiconductor film and the gate insulators.

2. Theory

For simplicity, the depletion approximation will be made in this calculation. We will calculate the drain characteristics of deep-depletion type thin-film MOST's first. Deep-depletion type MOST's can be operated in either the depletion or enhancement modes, while enhancement-type thin-film MOST's can be operated only in the enhancement mode. We shall assume n-channel MOST's. The results can easily be applied for p-channel MOST's.

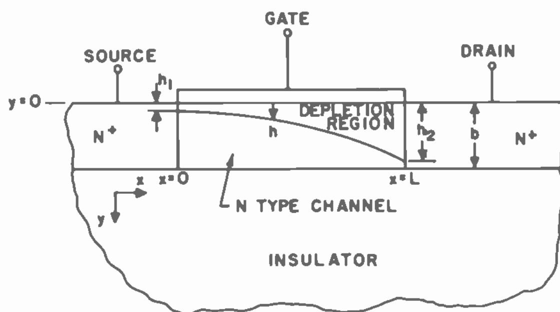


Fig. 1—N-channel deep-depletion type thin film MOS FET operated in depletion mode.

2.1 Deep-Depletion Type Thin-Film MOS FET's

We first derive the drain characteristics of deep-depletion type thin-film MOS FET's. The doping concentration of the semiconductor film is assumed to be arbitrarily distributed in the direction of the film thickness, i.e., in the y direction as shown in Fig. 1.

(a) *Depletion-Mode Operation of Deep-Depletion Type Thin-Film MOST's*

The gate voltage of the MOST's when part of the semiconductor film is depleted, is given by

$$V_G = V(x) + \varphi_{MS} - \frac{Q_{SS}}{C_{ox}} - \frac{1}{\epsilon_S}\int_0^h y\rho_{SC}\,dy - \frac{1}{C_{ox}}\int_0^h \rho_{SC}\,dy, \qquad [1]$$

where:

φ_{MS} = the work function difference between gate metal (or polysilicon for silicon gate structure) and semiconductor,
Q_{SS} = fixed charge density at semiconductor-gate insulator interface,
C_{ox} = unit area gate insulator capacitance,
ϵ_S = permitivity of the semiconductor,
ρ_{SC} = charge density at the space-charge region, and
$V(x)$ = channel voltage at x.

The other notations are shown in Fig. 1. The first integration term on the right-hand side of Eq. [1] is the voltage drop across the space-charge region and the second integration term is the voltage drop across the gate insulator due to the charge in the space-charge region.

Differentiating Eq. [1] with respect to h yields

$$\frac{dV_G}{dh} = \frac{dV(x)}{dh} - \frac{1}{C_{ox}}\frac{dQ_{SS}}{dh} - \left(\frac{h}{\epsilon_S} + \frac{1}{C_{ox}}\right)\rho_{SC} - \int_0^h \left(\frac{y}{\epsilon_S} + \frac{1}{C_{ox}}\right)\frac{d\rho_{SC}}{dx}\frac{dx}{dh}\,dy. \qquad [2]$$

It can easily be shown that the drain current of the MOST is given by

$$I_D = W\,\mu\rho dy\,\frac{dV(x)}{dx}, \qquad [3]$$

where μ is the current carrier mobility, ρ is the current carrier density, and W is the width of the device. From Eqs. [2] and [3], the drain current is found to be

$$I_D = \frac{W}{L}\int_{h_1}^{h_2}\left(\int_h^b \mu\rho dy\right)\left[\left(\frac{h}{\epsilon_S} + \frac{1}{C_{ox}}\right)\rho_{SC} + \frac{1}{C_{ox}}\frac{dQ_{SS}}{dh} + \int_0^h\left(\frac{y}{\epsilon_S} + \frac{1}{C_{ox}}\right)\frac{d\rho_{SC}}{dx}\frac{dx}{dh}\,dy\right]dh. \qquad [4]$$

The drain current is saturated when $h_2 = b$. We assume that both Q_{SS} and ρ_{SC} are independent of x, as is the case when the drain bias voltage is small. The drain saturation current $I_{D,sat}$ is

$$I_{D,sat}\,\big|_{small V_D} = \frac{W}{L}\int_{h_1}^{b}\left(\int_h^b \mu\rho dy\right)\left(\frac{h}{\epsilon_S} + \frac{1}{C_{ox}}\right)\rho_{SC}dh. \qquad [5]$$

It is shown later that at larger drain voltages the trapped electron density becomes field dependent. Also, electron–hole pair multiplication may take place at the drain depletion region. Therefore, the last term on the right-hand side of Eq. [4] cannot be neglected. The drain saturation voltage, $V_{D,sat}$, is obtained from Eq. [1]:

$$V_{D,sat} = V_G - \varphi_{MS} + \frac{Q_{SS}}{C_{ox}} + \frac{1}{\epsilon_S}\int_0^b y\rho_{SC}dy + \frac{1}{C_{ox}}\int_0^b \rho_{SC}dy. \quad [6]$$

The saturation voltage, therefore, is also a function of the thickness of the semiconductor film.

The transconductance of the device is given by

$$g_{m,sat} = \frac{\delta I_{D,sat}}{\delta V_G} = \frac{\delta I_{D,sat}}{\delta h}\frac{\delta h}{\delta V_G} = \frac{W}{L}\int_{h_1}^b \mu\rho dy. \quad [7]$$

It can easily be shown that the transconductance is equal to the channel conductance.

The pinch-off voltage, V_P, is defined as the gate voltage required to completely deplete the semiconductor film when $V_D = 0$, or

$$V_P = \varphi_{MS} - \frac{Q_{SS}}{C_{ox}} - \frac{1}{\epsilon_S}\int_0^b y\rho_{SC}dy - \frac{1}{C_{ox}}\int_0^b \rho_{SC}dy, \quad [8]$$

when

$$\frac{1}{\epsilon_S}\int_0^b y\rho_{SC}dy \le 2\,\varphi_F.$$

Here φ_F is the Fermi energy of the semiconductor film measured from the center of the forbidden gap. The pinch-off voltage increases with the thickness of the gate insulator, the doping concentration, and the thickness of the semiconductor film.

The gate voltage dependence of the drain current is very complex in the depletion mode of operation because the current carrier mobility is a function of y. It has been shown by Hsu and Scott[7] that in thin SOS films the mobility of current carriers in the silicon film decreases exponentially with distance from the Si-SiO_2 interface. Therefore, the drain current of MOST's operated in this mode increases almost exponentially with gate bias voltage.

Let us consider a special case by assuming that the donor density in the silicon film is uniformly distributed and that electron mobility decreases exponentially with y, or

$$\rho(y) = q N_D$$

and [9]

$$\mu(y) = \mu_0 e^{-\alpha y},$$

where α is a constant that depends on the quality of the silicon film and μ_0 is the electron mobility at the $Si\text{-}SiO_2$ interface when the surface is at the flat-band condition. At that condition the effect of the surface scattering mechanism on the current carrier mobility is small.[7] Substituting Eq. [9] into Eq. [5], yields

$$I_{D,sat} = \frac{Wq^2N_D{}^2}{\alpha L}\left[\mu(h_1)\left(\frac{1+\alpha h_1}{\alpha^2\epsilon_S} + \frac{1}{\alpha C_{ox}}\right) - \mu(b)\left(\frac{1+\alpha b}{\alpha^2\epsilon_S} + \frac{1}{\alpha C_{ox}} + \frac{b^2 - h_1{}^2}{2\epsilon_S} + \frac{b-h_1}{C_{ox}}\right)\right]. \quad [10]$$

The transconductance, $g_{m,sat}$, and hence the channel conductance G_0 is given by

$$g_{m,sat} = G_0 = \frac{W_q N_D}{L\alpha}\left[\mu(h_1) - \mu(b)\right]. \quad [11]$$

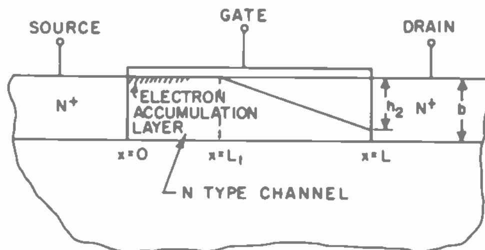


Fig. 2—N-channel deep-depletion type thin film MOS FET operated in enhancement mode.

(b) *Enhancement-Mode Operation of Deep-Depletion Type Thin-Film MOST's*

When the gate bias voltage is larger than $\varphi_{MS} - (Q_{SS}/C_{ox})$, electrons are accumulated at the semiconductor–gate-insulator interface when $V_D = 0$. At larger drain biases, part of the semiconductor–gate-insulator interface is depleted as indicated in Fig. 2. The MOST is operated in the enhancement mode. We divided the channel into two regions. In region I, electrons are accumulated at the semiconductor–gate-insulator interface. In region II, the interface is depleted. Therefore, in region I,

$$V_G - \varphi_{MS} + \frac{Q_{SS}}{C_{ox}} - V(x) = -\frac{Q_n}{C_{ox}},$$

where Q_n is the charge density of electrons accumulated at the semiconductor–gate-insulator interface. At the boundary of region I and region II,

$$V(x) = V_G - \varphi_{MS} + Q_{SS}/C_{ox}.$$

The current equation in region I is

$$I_D L_1 = W\left[\frac{1}{2}C_{ox}\mu(0)\left(V_G - \varphi_{MS} + \frac{Q_{SS}}{C_{ox}}\right)^2 + \left(V_G - \varphi_{MS} + \frac{Q_{SS}}{C_{ox}}\right)\int_0^b \mu(y)\rho(y)dy\right]. \quad [12a]$$

The current equation in the region II is the same as that given by Eq. [4] provided that L is replaced by $L - L_1$ and $h_1 = 0$. The drain current therefore, is, given by

$$I_D = \frac{W}{L}\left[\frac{1}{2}C_{ox}\mu(0)\left\{V_G - \varphi_{MS} + \frac{Q_{SS}}{C_{ox}}\right\}^2 + \left\{V_G - \varphi_{MS} + \frac{Q_{SS}}{C_{ox}}\right\}\int_0^b \mu(y)\rho(y)dy + \int_0^{h_2}\left(\int_h^b \mu\rho(y)dy\right)\left\{\left(\frac{h}{\epsilon_S} + \frac{1}{C_{ox}}\right)\rho_{SC} + \frac{1}{C_{ox}}\frac{dQ_{SS}}{dh} + \int_0^h \left(\frac{h}{\epsilon_S} + \frac{1}{C_{ox}}\right)\frac{d\rho_{SC}}{dx}\frac{dx}{dh}dy\right\}dh\right], \quad [12]$$

where $\mu(0)$ is the mobility of the accumulated electrons. Due to the surface scattering effect, $\mu(0)$ is, in general, smaller than μ_0. When h_2 in Eq. [12] is equal to b, the drain current is saturated. The drain saturation voltage is the same as that given by Eq. [6]. At larger gate bias voltage, the first term on the right-hand side of Eq. [12] is the dominant current term. Therefore, the drain current increases with the square of $V_G - V_{FB}$, where V_{FB} is the flat-band gate voltage and is equal to $\varphi_{MS} - (Q_{SS}/C_{ox})$. The channel conductance is also equal to the transconductance at the drain saturation region, or

$$g_{m,sat} = G_0 = \frac{W}{L}\left[C_{ox}\mu(0)(V_G - V_{FB}) + \int_0^b \mu(y)\rho(y)dy\right]. \quad [13]$$

2.2 Enhancement-Type Thin-Film MOST's

We assume that the thickness of the semiconductor film is thin enough that when an inversion channel is formed at the semiconductor–gate–insulator interface, the semiconductor film is completely depleted. When the film is very thick, there is a good conductive neutral layer between the depletion region and the silicon–sapphire interface when the silicon surface is strongly inverted. The floating substrate effect must be taken into account. This effect has been studied by many investigators[2–4] and is not discussed here. We assume that the thickness of the insulator substrate is orders of magnitude larger than that of the semiconductor film and the gate insulator. The typical thickness of sapphire substrates in SOS devices is larger than 400 μm, while the thickness of the silicon film is less than 1 μm. (The typical silicon film thickness of devices is about 0.5 μm.) Therefore, this approximation is typical for practical SOS MOS devices. Since the substrate insulator is so thick, we may neglect the substrate effect.

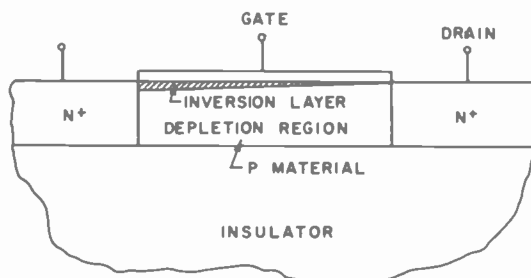


Fig. 3—An n-channel enhancement-type thin-film MOS FET.

An enhancement-type n-channel thin-film MOST operated in the saturation region is sketched in Fig. 3. The voltage and the field intensity at the semiconductor– substrate-insulator interface, $V_b(x)$ and $E_b(x)$, respectively, are not necessarily zero. It can be easily shown that the channel voltage at x, $V(x)$, is given by

$$V(x) = V_b(x) - \frac{1}{\epsilon_S} \int_b^0 \int_b^y \rho_{SC}\, dy' dy + b\, E_b(x). \qquad [14]$$

The gate voltage is then given by

$$V_G = \varphi_{MS} + 2\,\varphi_{FP} - \frac{Q_{SS} + Q_S + Q_N}{C_{ox}} + V(x)$$

$$= V_T + V(x) - \frac{Q_N}{C_{ox}}. \qquad [15]$$

Here Q_N is the charge density of the free current carrier at the inversion layer, Q_S is the unit area space-charge density in the semiconductor film, φ_{FP} is the Fermi level of the p-type semiconductor film with respect to the center of the forbidden gap, and V_T is the threshold voltage. The drain current equation is standard:

$$I_D = W\mu C_{ox}[V_G - V_F - V(x)] \frac{dV(x)}{dx} \qquad [16]$$

or

$$I_D = \frac{W}{L} \mu C_{ox} \left[(V_G - V_T)V_D - \frac{1}{2} V_D{}^2 \right].$$

From Eq. [15] the drain saturation voltage, $V_{D,sat}$, is given by

$$V_{D,sat} = V_G - V_T. \qquad [17]$$

Therefore, the drain saturation current is

$$I_{D,sat} = \frac{W}{2L} \mu C_{ox} (V_G - V_T)^2. \qquad [18]$$

This expression is exactly the same as that for a bulk MOST, the only difference being the threshold voltage, V_T, which is given by

$$V_T = \varphi_{MS} + 2\, \varphi_{FP} - \frac{Q_{SS} + Q_S}{C_{ox}}. \qquad [19]$$

The threshold voltage is proportional to the unit area impurity in the semiconductor films.

From Eqs. [16] and [18] the channel conductance is found to be equal to the transconductance of the device,

$$G_0 = g_{m,sat} = \frac{W \mu C_{ox}}{L} (V_G - V_T). \qquad [20]$$

2.3 Anomalous Drain Characteristics

We shall now qualitively examine the drain-current equations obtained previously. It is well understood that an edge dislocation in an Si crystal creates a dangling bond per unit cell. Each dangling bond provides an attractive site for trapping an extra electron to complete the bond and

for creating a localized negative charge.[8] This effect has been experimentally identified.[9] When there is a large electric field intensity at the edge dislocation, the trapped electron may escape. As a result the space-charge density becomes less negative toward the drain junction. When electron–hole pair multiplications occur at the drain space-charge region, electrons flow to the drain and holes flow toward the source along the space-charge region, which again creates a positively charged carrier gradient toward the drain electrode. The release of trapped electrons will cause the last term of Eq. [4] to have the same sign as the first term on the right-hand side of Eq. [4] for n-channel MOST's. For p-channel MOST's, these two terms have different signs. Consequently, the drain current of n-channel thin-film deep-depletion type MOST's increases and that of p-channel devices decreases with the drain bias voltage, as has been experimentally observed.[1]

When the space-charge density gradient is caused by avalanche multiplication, the last term of Eq. [4] has the same sign as the first term of the same equation for both n- and p-channel thin-film MOST's. Hence, the drain current increases with the multiplication current. The multiplication current increases very drastically with the drain bias voltage. Therefore, the drain current also increases very drastically with the drain bias voltage. Since the first two terms of Eq. [12] are independent of charge density in the space-charge region, these two terms will not be affected by charges injected into the space-charge region from the drain junction. Consequently, the current kink can best be seen at lower drain-current density.

At large drain-current density no current kink can be observed. Eq. [8] shows that when ρ_{SC} is increased the pinch-off voltage is larger. Therefore, if the current injected from the drain can reach the source, the turn-on voltage will be smaller. In enhancement-type thin-film MOST's, the release of trapped electrons increases V_T for p-channel devices and decreases V_T for n-channel MOST's. This again increases the drain current of n-channel MOST's but decreases the drain current of p-channel devices as is seen from Eq. [18]. When there are electron–hole pair multiplications the charge density in the space-charge region decreases. Therefore, the turn-on voltage, V_T, of both n- and p-channel MOST's decreases with the increase of the multiplication rate. We note that the current carriers may be unable to reach the source region before they recombine. The threshold voltage at the source region is not affected. The multiplication process, therefore, increases the drain sauration voltage, as shown by Eq. [17], which, in turn, increases the drain current. From Eq. [18] it can be easily shown that current kink also can best be seen at small drain-current density. At large drain-current density no current kink is apparent.

3. Experiment

The devices used in this experiment are SOS MOST's. The silicon films are epitaxially grown on sapphire substrates. The silicon films are no thicker than 2 μm where the thickness of the sapphire substrates is approximately 400 μm. The thickness of the gate SiO_2 is typically 1000 Å. We present experimental data on the drain characteristics of a typical n-channel depletion-type SOS MOST and a typical p-channel enhancement-type SOS MOST. The length and the width of the gate area of these MOST's are 7.6 μm and 0.5 mm, respectively.

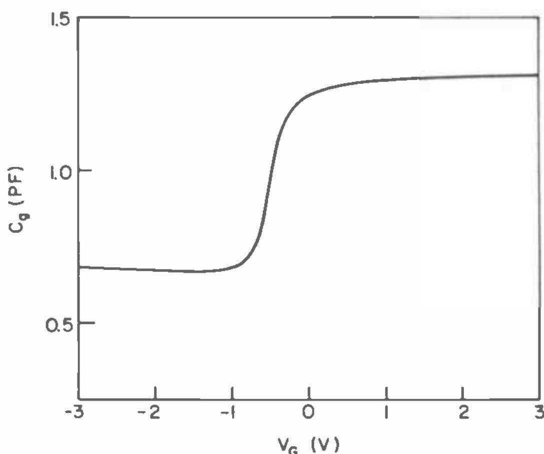


Fig. 4—Gate capacitance versus gate voltage plot of a typical n-channel deep-depletion type SOS MOS FET.

The gate capacitance of the n-channel deep-depletion SOS MOST is plotted as a function of gate supply voltage in Fig. 4. The silicon film of this device is 1 μm thick and is doped with 2×10^{15} cm^{-3} phosphorus. The flatband voltage is approximately[9] -0.5 V. Therefore, when the gate voltage is more negative than -0.5 V the device is operated in the depletion mode. The silicon surface becomes strongly accumulated at $V_G > 0.5$ V, as seen in Fig. 4.

The drain current of this n-channel depletion-type SOS MOST operated in the saturation region is plotted as a function of V_G in Fig. 5, which also shows the measured drain I-V characteristics. The drain voltage used in this plot is smaller than the onset voltage of the current kink. The transconductance of the SOS MOST operated in the saturation region and the electron mobility are also plotted as a function of V_G

in Fig. 5. The mobility in this plot is defined as the electron mobility at the edge of the space-charge region of the given gate bias voltage. This mobility is measured using the method described by Ref. [7]. It is clearly shown that the drain saturation current and the transconductance of the device are proportional to the electron mobility in the depletion mode operation as is expected from Eqs. [10] and [11]. At small V_G, h_1 is lin-

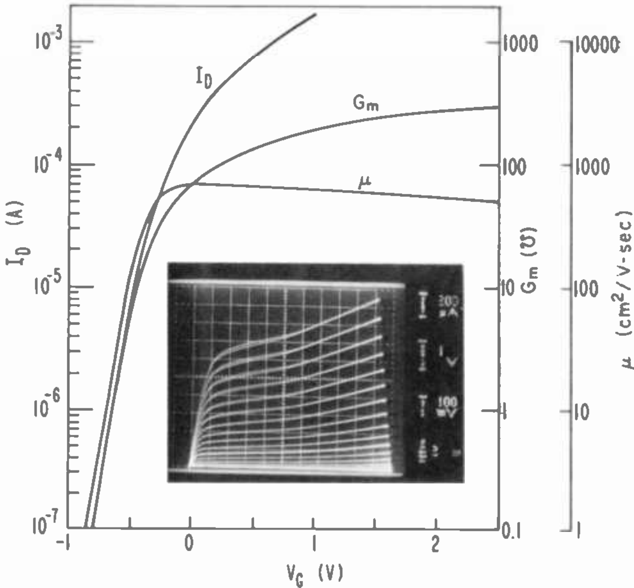


Fig. 5—Drain saturation current versus gate voltage of the device used in Fig. 4.

early proportional to V_G and $\alpha = \beta Q_D/C_{ox}$, where β is a constant that can be obtained from the slope of log μ versus V_G curve at $V_G = 0$. The theoretical value of the drain saturation current at $V_G = 0.5$ V is equal to 3.5 μA, which is about a factor of 2 smaller than the measured value. A possible explanation of this difference could be that the actual free-electron density can be much smaller than 2×10^{15} cm^{-3}, as was doped during the epitaxial process. The out-diffusion of Al from the sapphire substrate, the escape of phosphorus during the high-temperature device fabrication process, and the electron trap states can significantly reduce the free-electron density.

The drain saturation current of this device is replotted as a function of $V_G - V_T$ in Fig. 6. The curve is the measured value and the points are calculated from Eq. [12]. The agreement between the experimental curve and the calculated values is excellent when the device is operated in the

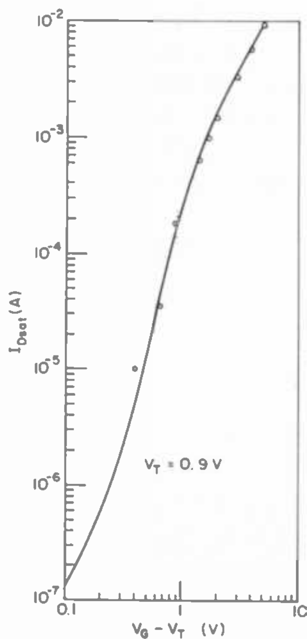


Fig. 6—Drain saturation current versus gate voltage of the device used in Fig. 4 at the large gate biases.

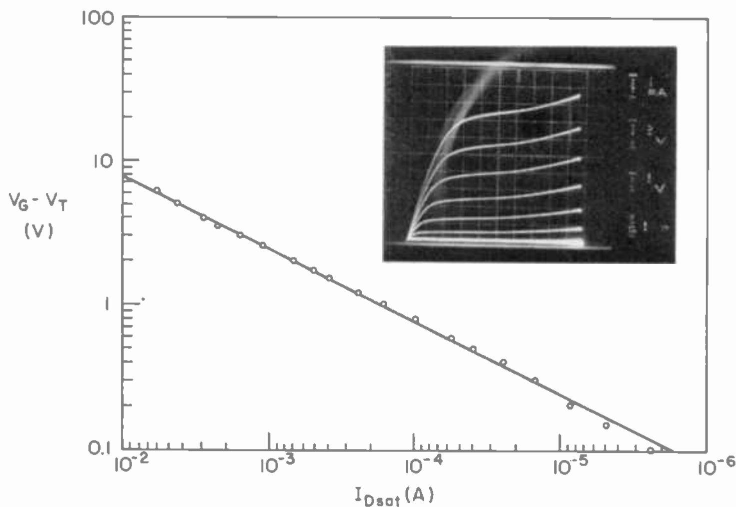


Fig. 7—Drain saturation current versus gate voltage of a typical enhancement-type SOS MOS FET.

enhancement mode. At strong enhancement, the drain current is mainly due to the accumulated electrons. The density of accumulated electrons can be accurately measured.

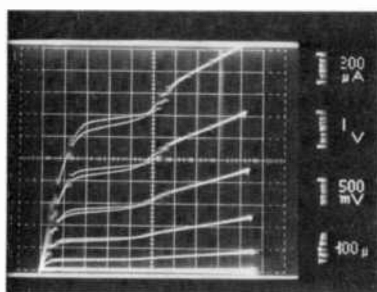
The drain saturation current of a typical p-channel enhancement type SOS MOST is plotted in Fig. 7 as a function of $V_G - V_T$. The drain voltage in this measurement is -10 V. The thickness of the silicon film of this device is 0.6 μm. The doping concentration is 2×10^{15} cm^{-3}. Therefore, during the normal operation condition the film is completely depleted. The points in Fig. 7 are measured data and the line has a slope of 2. Thus, when $I_D > 10^{-5}$ A the drain saturation current increases with the square of $V_G - V_T$, as expected from Eq. [18]. At small drain current the measured drain current is larger than predicted. This is because at small gate biases the hole mobility at the surface inversion layer is larger than that when the surface is strongly inverted.

If we substitute the measured current into Eq. [18], the hole mobility is found to be 145 cm^2/V-sec. This is comparable to 154 cm^2/V-sec as measured from the drain current versus gate voltage curve when the device is operated in the linear region.

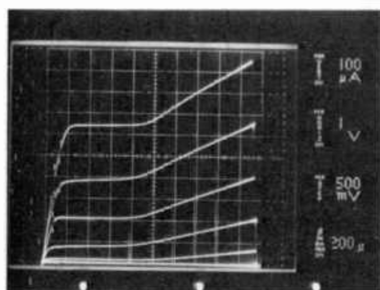
The drain I-V characteristics, as shown in Fig. 5, also exhibit a current kink. Since the current density is large, the current kink is small. Fig. 8a shows the drain I-V characteristics of a deep-depletion type SOS MOST with a very large current kink. Fig. 8b shows the drain I-V characteristics of a typical enhancement-type n-channel SOS MOST. The thickness of the silicon films of these two devices is equal to 0.6 μm.

In general, the onset voltage of the current kink in n-channel SOS MOST's is from 3 to 6 V depending on the gate bias voltage. N-channel SOS MOST's made on 0.4 μm thick silicon films also exhibit a current kink. The drain characteristics of a typical p-channel SOS MOST made on a 0.6-μm-thick n-type silicon film are shown in Fig. 7. Although the slope of the output I-V curve is larger when the drain bias voltage is larger than 10 V, there is no clear current kink. However, if the silicon film is very thick, such that at the normal operation of the MOST's there is a good conductive layer between the channel depletion region and the silicon–sapphire interface, the drain I-V characteristics of the p-channel enhancement-type MOST's also exhibit a current kink as shown in Fig. 9. The thickness of the n-type silicon film of this device is 2 μm. The onset voltage of the current kink is approximately 10 V. This onset voltage is about two times larger than that of n-channel SOS MOST's. If an electrode is deposited on the silicon film of this thick-film SOS MOST, and if the current flow through this electrode is called "substrate current", the substrate current is very small at small drain biases.

At larger drain biases the substrate current drastically increases with



(A)



(B)

Fig. 8—Drain I-V characteristics: (a) deep-depletion type n-channel SOS MOST and (b) typical enhancement type n-channel SOS MOST.

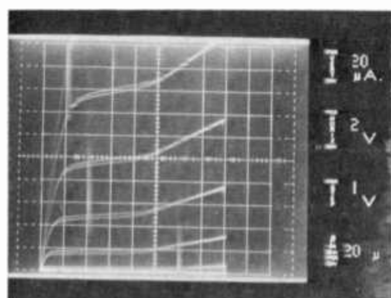


Fig. 9—Drain I-V characteristics of a typical thick-film p-channel SOS MOST.

the drain bias voltage. This is shown in Fig. 10. The gate bias voltage is kept constant for these three curves. Curve (a) of Fig. 10 is the substrate current. The scale of the substrate current is 10 nA per division. Curve (b) is the drain I-V curve when the substrate terminal is not connected, and curve (c) is the drain I-V curve when the substrate terminal is shorted to the source. The drain current exhibits a current kink when the substrate terminal is floated. The onset voltage of the drain current

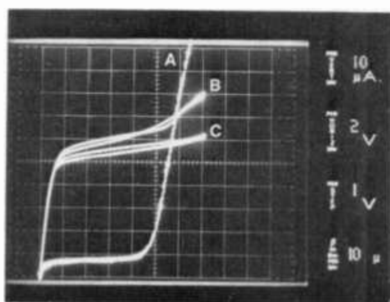
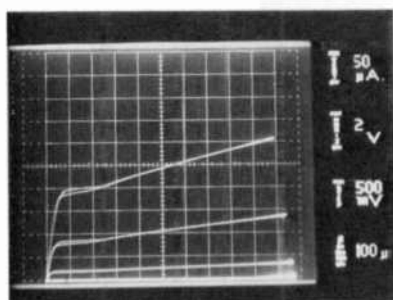


Fig. 10—Drain and substrate currents of a typical thick-film four-terminal p-channel SOS MOST.

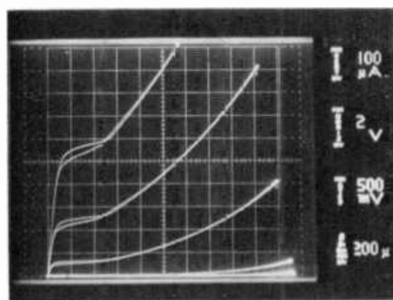
kink is slightly larger than that of the substrate current. The current kink in this case is obviously caused by the floating-substrate effect.

The substrate current is due to avalanche multiplication at the drain space-charge region. The same phenomenon also has been observed in thick film n-channel enhancement type SOS MOST's. The onset voltage of the current kink in thick film n-channel SOS MOST's, however, is about 3–5 V.

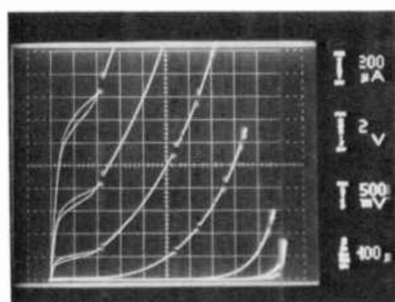
Eq. [8] shows that if ρ_{SC} at $x = 0$ is increased, the pinch-off voltage of a depletion-type thin-film MOST is increased. Fig. 11 demonstrates this effect. The three photographs of Fig. 11 are the drain characteristics of three adjacent n-channel deep-depletion type P+ silicon gate SOS MOST's. The width of their gates is 0.32 mm. The length of the gates for the MOST's used for (a), (b), and (c) of Fig. 11 are 25.4 μm, 10 μm, and 5.1 μm, respectively. It is evident that at drain bias voltages larger than the onset voltage of the current kink, the turn-on voltage of short-channel MOST's decreases with the increasing drain bias voltage. The turn-on voltage of long-channel MOST's is not affected by the drain bias voltage. Because of the built-in potential of the P+ silicon gate the decrease in turn-on voltage is equivalent to the increase in channel pinch-off voltage.



(A)



(B)



(C)

Fig. 11—Drain I-V characteristics of three adjacent deep-depletion n-channel SOS MOST's. The channel widths are 0.32 mm. The channel lengths are 25.4 μm, 10 μm, and 5.1 μm for curves (a), (b) and (c), respectively.

The lifetime of holes in the space-charge region is small. Holes injected from the drain into the space-charge region cannot be collected by the source electrode unless the length of the channel is short. This result also indicates that for thin-film SOS MOST's the current kink is caused by the increase of the drain current saturation voltage with drain bias voltage. The electrical properties at the source junction can be affected by the drain bias voltage only if the channel length is very short or if the silicon film is very thick, so that there is a good conductive layer between the channel depletion region and the silicon–sapphire interface.

The slope of the drain I-V curves, shown in Fig. 11, cannot be explained by the channel shorting effect. It is obvious that the large drain conductance of short channel SOS MOST's is due to the increase of pinch-off voltage caused by holes that flow to the electrode through the space-charge region.

4. Discussions and Conclusions

Theoretical expressions for the drain static characteristics of thin-film MOS field-effect transistors with arbitrary doping profile have been derived. When a deep-depletion type thin-film MOST is operated in the depletion mode the drain saturation current is strongly dependent on the doping and current carrier mobility profiles. Since the mobility of electrons and holes in a thin silicon film epitaxially grown on a sapphire substrate decreases exponentially with the distance from the top surface, and assuming the doping concentration is relatively uniform, the drain saturation current of a deep depletion type SOS MOST operated in the depletion mode increases almost exponentially with the gate bias voltage. When a deep-depletion type MOST operates in the enhancement mode, the drain saturation current increases with the square of $V_G - V_{FB}$. The drain saturation current of an enhancement-type thin-film MOST increases with the square of $V_G - V_T$. Physically, the flatband voltage, V_{FB}, is the threshold voltage for a depletion-type thin-film MOST to be operated in the enhancement mode. Therefore, the principle of the enhancement-mode operation of a deep-depletion type MOST is the same as that of an enhancement-type MOST.

The drain saturation currents of deep-depletion type and enhancement-type thin-film MOST's are given by Eqs. [4], [12], and [18]. These equations also predict that if holes (electrons) are injected into the space-charge region of n-channel (p-channel) MOST's the drain saturation current versus drain bias voltage curves may exhibit a kink. The decrease of the trapped electron density at large V_D increases the drain current of n-channel MOST's but decreases the drain current of p-

channel MOST's. Holes injected into the space-charge region of an n-channel SOS MOST may recombine with electrons at trap states. This empty trap state is a positively charged dangling bond. Electrons from adjacent trap states may hop into these empty trap states. As a result the positively-charged empty trap states, i.e., the dangling bonds, appear to flow toward the source. Since the electron density at the space-charge region is nearly zero, the empty electron trap states have relatively long lifetime. As a result a large density of positive charge, i.e., empty electron trap states and holes, can be accumulated in the space-charge region by holes injected from the drain junction. Consequently, both deep-depletion and enhancement-type n-channel thin-film SOS MOST's exhibit drain current kink.

The current carriers injected into the channel depletion region from the drain junction of p-channel MOST's are electrons. The dangling bonds are all filled. Therefore, these electrons have very short lifetimes. Negative charges can hardly be accumulated at the space-charge region. The drain I-V curves of p-channel SOS MOST's may have larger slopes at larger drain bias voltages, but no clear current kinks can be observed.

If the silicon film is very thick, the lifetime of the minority carrier is longer than that for thin-film devices. The electrons injected into the space-charge region from the drain junction flow to the good conductive neutral region between the depletion region and the silicon–sapphire interface. These electrons become majority carriers. The effect is the same as that of the floating substrate in bulk MOST's. Therefore, the drain I-V curves exhibit current kinks. The same phenomenon is also observed from thick-film n-channel SOS MOST's.

The floating-substrate effect can also be explained with the physical model discussed in this paper. When positive charges are injected from the drain into the substrate neutral region, the substrate potential increases. Therefore, the total charge at the space-charge region is decreased. This, in turn, increases the drain saturation voltage and decreases the turn-on voltage, as is indicated by Eqs. [19] and [21], respectively. Consequently, the drain current increases with avalanche multiplication at the drain junction.

Acknowledgments

The author wishes to express his deep appreciation to J. H. Scott, Jr., for many helpful discussions and comments. Thanks are also due to M. C. Jones and J. M. Cartwright for their help in wafer fabrications and electrical measurements.

References:

[1] S. T. Hsu, J. H. Scott, Jr., and W. E. Ham, "Anomalous Current-Voltage Characteristics in SOS MOST's," Int. Elec. Dev. Mtg., Wash., D.C., Dec. 1973, Late Paper #88.

[2] J. Tihanyi and H. Schlotterer, "Influence of the Floating Substrate Potential on the Characteristics of ESFI MOS Transistors," *Solid-State Electron.*, **18**, p. 309, 1975.

[3] J. Tihanyi and H. Schlotterer, "Properties of ESFI MOS Transistors Due to the Floating Substrate and Finite Volume," *IEEE Trans. Electron Devices*, **ED-22**, p. 1017, Nov. 1975.

[4] D. Farrington, "Anomalous Current-Voltage Characteristics of Floating Bulk MOS FET's," p. 345, Extended Abstracts Vol. 74-2, Electrochem. Soc., Fall Meeting, New York, N. Y., Oct. 13–17, 1974.

[5] F. P. Heinman, "Thin-Film Silicon-on-Sapphire Deep Depletion MOS Transistors," *IEEE Trans. Electron Devices*, **ED-13**, p. 855, Dec. 1966.

[6] S. R. Hofstein, "An Analysis of Deep Depletion Thin-Film MOS Transistors," *IEEE Trans. Electron Devices*, **ED-13**, p. 846, Dec. 1966.

[7] S. T. Hsu and J. H. Scott, Jr., "Mobility of Current Carriers in Silicon-on-Sapphire (SOS) Films," *RCA Review*, **36**, No. 2 p. 240, (1975).

[8] A. C. Ipri, "Variations in Electrical Properties of Silicon Films on Sapphire Using the MOS Hall Technique," *Appl. Phy. Lett.*, **20**, No. 1, p. 1 (1972).

[9] A. Goetzberger, "Ideal MOS Curves for Silicon," *Bell Syst. Tech. J.*, **45** 1097 (1966).

Recent Papers by RCA Authors

Listing is alphabetical by name of primary author. For copies of reprints, the reader should contact the publication directly.

B. Abeles and J. I. Gittleman, "Composite Material Films: Optical Properties and Applications," **Appl. Optics**, Vol. 15, No. 10, p. 2328, Oct. 1976.

M. S. Abrahams, "The Microstructure of Silicon on Sapphire by Transmission Electron Microscopy," **Proc. Sixth European Congress on Electron Microscopy**, p. 212, Sept. 1976.

M. S. Abrahams, C. J. Buiocchi, R. T. Smith, J. F. Corboy, Jr., J. Blanc, and G. W. Cullen, Jr., "Early Growth of Silicon on Sapphire. I. Transmission Electron Microscopy," **J. Appl. Phys.**, Vol. 47, No. 12, p. 5139, Dec. 1976.

D. Baeriswyl and R. D. Wehner, "Static Susceptibility of Lattices with Local Double-Well Potentials," **Ferroelectrics**, Vol. 14, p. 653, 1976.

R. A. Bartolini, A. Bloom, and D. J. Channin, "Organic Medium for Integrated Optics," **Appl. Optics**, Vol. 15, p. 2628, Nov. 1976.

J. Blanc, C. J. Buiocchi, M. S. Abrahams, and W. E. Ham, "The Si/SiO_2 Interface Examined by Cross-Sectional Transmission Electron Microscopy," **Appl. Phys. Lett.**, Vol. 30, No. 2, p. 120, Jan. 15, 1977.

M. T. Gale, "Sinusoidal Relief Grating for Zero-Order Reconstruction of Black-and-White Images," **Optics Communications**, Vol. 18, No. 3, p. 292, Aug. 1976.

L. A. Goodman, D. Meyerhofer, and S. DiGiovanni, "The Effect of Surface Orientation on the Operation of Multiplexed Twisted-Nematic Devices," **IEEE Trans. on Electron Devices**, Vol. 23, No. 10, p. 1176, Oct. 1976.

J. S. Helman and B. Abeles, "Tunneling of Spin-Polarized Electrons and Magnetoresistance in Granular Ni Films," **Phys. Rev. Lett.**, Vol. 37, No. 21, p. 1429, Nov. 22, 1976.

J. Klatskin and A. Rosen, "Simple Method for Separating Completed Diodes from Semiconductor Wafers," **J. Electrochem. Soc.**, Vol. 123, No. 12, Dec. 1976.

H. P. Kleinknecht and A. E. Widmer, "Zinc Diffusion in $Al_xGa_{1-x}P$," **Solid-State Electronics**, Vol. 19, p. 1005, 1976.

K. Knop, "Color Pictures Using the Zero Diffraction Order of Phase Grating Structures," **Optics Communications**, Vol. 18, No. 3, p. 298, Aug. 1976.

I. Ladany, M. Ettenberg, H. F. Lockwood, and H. Kressel, "Al_2O_3 Half-Wave Films for Long-Life CW Lasers," **Appl. Phys. Lett.**, Vol. 30, No. 2, p. 115, Jan. 15, 1977.

R. U. Martinelli, "The Temperature Dependence of the DC Base and Collector Currents in Silicon Bipolar Transistors," **Trans. on Electron Devices**, Vol. 23, No. 11, p. 1218, Nov. 1976.

J. H. McCusker, S. S. Perlman, and H. S. Veloric, "Microsonic Pulse Filters–Replacements for Traditional Butterworth Designs," **RCA Review**, Vol. 37, No. 3, p. 389, Sept. 1976.

D. Meyerhofer, "A New Technique of Aligning Liquid Crystals on Surfaces," **Appl. Phys. Lett.**, Vol. 29, No. 11, p. 691, Dec. 1, 1976.

M. D. Miller, "Difference Between Platinum and Gold Doped Silicon Power Devices," **IEEE Trans. on Electron Devices**, Vol. 23, No. 10, p. 1279, Dec. 1976.

D. O. North and H. S. Sommers, Jr., "Saturation of the Junction Voltage in Stripe-Geometry (AlGa)As Double-Heterostructure Junction Lasers: A Comment," **Appl. Phys. Lett.**, Vol. 30, No. 2, p. 87, Jan. 15, 1977.

D. O. North, "Theory of Modal Character, Field, Structure, and Losses for Semiconductor Lasers," **IEEE J. Quantum Electronics**, Vol. 12, No. 10, p. 616, Oct. 1976.

C. J. Nuese, G. H. Olsen, M. Ettenberg, J. J. Gannon, and T. J. Zamerowski, "CW Room-Temperature $In_xGa_{1-x}As/In_yGa_{1-y}P$ 106-μm Lasers," **Appl. Phys. Lett.**, Vol. 29, No. 12, p. 807, Dec. 15, 1976.

A. Okada, "Optical Waveguiding Properties of Sputtered $LiNbO_3$ Single Crystal Thin Films on $LiTaO_3$ Substrates," **Ferroelectrics**, Vol. 14, p. 739, 1976.

J. I. Pankove and D. E. Carlson, "Electroluminescence in Amorphous Silicon," **Appl. Phys. Lett.**, Vol. 29, No. 9, p. 620, Nov. 1, 1976.

J. I. Pankove, "Photoluminescence of Ion-Implanted GaN," **J. Appl. Phys.** Vol. 47, No. 12, p. 5387, Dec. 1976.

K. M. Schleiser, J. M. Shaw, and C. W. Benyon, Jr., "Al_2O_3 as a Radiation-Tolerant CMOS Dielectric," **RCA Review**, Vol. 37, No. 3, p. 359, Sept. 1976.

P. Sheng and P. J. Wojtowicz, "Constant Coupling Theory of Nematic Liquid Crystals," **Phys. Rev. A**, Vol. 14, No. 5, p. 1883, Nov. 1976.

P. Sheng, "Phase Transistions in Surface-Aligned Nematic Films," **Phys. Rev. Lett.**, Vol. 37, No. 16, p. 1059, Oct. 18, 1976.

E. F. Steigmeier, D. Baeriswyl, G. Harbeke, and H. Auderset, "Coupling of the CDW and the Water Mode in $K_2Pt(CN)_4Br_{0.3}\cdot 3.2H_2O$," **Solid State Communications**, Vol. 20, p. 661, 1976.
E. F. Steigmeier, R. Loudon, G. Harbeke, and H. Auderset, "Raman Scattering in $K_2Pt(C)_4Br_{0.3}\cdot 3H_2O$," **Ferroelectrics**, Vol. 13, p. 549, 1976.
E. F. Steigmeier, G. Harbeke, and H. Auderset, "Softening of Charge Density Wave Excitations at the Superstructure Transition in $2HTaSe_2$," **Solid State Communications**, Vol. 20, p. 667, 1976.
C. W. Struck and W. H. Fonger, "Temperature Quenching of Luminescence for Linear and Derivative Nuclear Operators," **J. Luminescence**, Vol. 14, No. 4, p. 253, Sept. 1976.
C. L. Upadhyayula, R. E. Smith, J. E. Wilhelm, S. T. Jolly, and J. P. Paczowski, "Transferred Electron Logic Devices for Gigabit-Rate Signal Processing," **IEEE Trans. on Microwave Theory and Techniques**, Vol. 24, No. 12, p. 920, Dec. 1976.
C. R. Wronski, D. E. Carlson, and R. E. Daniel, "Schottky Barrier Characteristics of Metal-Amorphous Silicon Diodes," **Appl. Phys. Lett.**, Vol. 29, No. 9, p. 602, Nov. 1, 1976.

Patents Issued to RCA Inventors Fourth Quarter, 1976

October

A. A. Ahmed Equalization of Base Current Flow in Two Interconnected Transistor Amplifiers (3,987,368)
A. A. Ahmed Absolute-Value Circuit (3,989,997)
A. A. Ahmed Current Amplifier (3,990,017)
J. Avins Transient Suppression in Television Video Systems (3,984,865)
B. W. Beyers, Jr. Character Generator for Television Channel Number Display with Edging Provisions (3,984,828)
C. J. Busanovich and R. M. Moore Selenium Rectifier Having Hexagonal Polycrystalline Selenium Layer (3,990,095)
S. P. Clurman Reduction of Hunting in Synchronous Motor (3,988,653)
R. Denning and W. G. Einthoven Thermally Balanced PN Junction (3,988,759)
H. J. Digneffe Alternating Current Control System (3,990,000)
F. C. Duigon and S. Liu Planar Trapatt Diode (3,990,099)
M. Ettenberg and H. Kressel Solar Cell Device Having Two Heterojunctions (3,990,101)
J. R. Ford Apparatus for Erecting a Trude Vertical Axis (3,985,033)
J. B. George High Power Remote Control Ultrasonic Transmitter (3,984,705)
E. C. Giaimo, Jr. Method of Increasing the Image Exposure and Developing Sensitivity of Magneto-Electric Printing System (3,986,872)
I. Gorog Flying Spot Scanner Unaffected by Ambient Light (3,984,629)
J. Gross and W. H. Barkow Display System Utilizing Beam Shape Correction (3,984,723)
F. Z. Hawrylo Ohmic Contact (3,984,261)
K. C. Hudson and R. F. Kenville Optical Scanner with Large Depth of Focus (3,989,348)
G. S. Kaplan and A. D. Ritzie Homodyne Communication System (3,984,835)
S. W. Kessler, Jr. Transcallent Semiconductor Device (3,984,861)
W. F. Kosonocky Introducing Signal at Low Noise Level to Charge-Coupled Circuit (3,986,198)
H. Kressel and V. L. Dalal Solar Cell Device Having Improved Efficiency (3,988,167)
S. Larach Apparatus and Method for Analyzing Biological Cells for Malignancy (3,984,683)
S. A. Lipp Chemical Vapor Deposition of Luminescent Films (3,984,587)
M. L. McNeely and H. Rees Apparatus for Producing Injection Molded and Centrally Apertured Disc Records (3,989,436)
P. J. Mikulich, Jr. and J. D. Callaghan Broadband Antenna System with the Feed Line Conductors Spaced on One Side of a Support Boom (3,984,841)
J. E. Miller Tracking Gate Servoed by Relative Range (3,987,441)
L. S. Napoli Method of Forming Conductive Coatings of Predetermined Thickness by Vacuum Depositing Conductive Coating on a Measuring Body (3,987,214)
E. J. Nossen and E. R. Starner Accurate Digital Phase/Frequency Extractor (3,984,771)
J. I. Pankove Insulating Nitride Compounds as Electron Emitters (3,986,065)
R. P. Perry Data Processor Reorder Shift Register Memory (3,988,601)
R. P. Perry Digital Matched Filtering Using a Step Transform Process (3,987,285)
A. J. Pikor Multiple Mesa Semiconductor Structure (3,988,765)
H. L. Pinch and S. T. Opresko Vapor Deposition of Cermet Layers (3,985,919)
J. Rosen and E. J. Denlinger Two-Inductor Varactor Tunable Solid-State Microwave Oscillator (3,984,787)
B. W. Siryj Object Orientation Apparatus (3,986,604)
B. W. Siryj Labeling Apparatus (3,984,279)
B. K. Smith Yoke Mount Assembly (3,986,156)
S. A. Steckler and A. R. Balahan Deflection Waveform Correction Signal Generator (3,984,729)
A. V. Tuma Digital Remote Control for Electronic Signal Receivers (3,987,414)
J. L. Vossen, Jr., F. R. Nyman, and G. F. Nichols Adherence of Metal Films to Polymeric Materials (3,984,907)
M. H. Wardell, Jr. Cathode Ray Tube Assembly Fixture (3,989,233)
P. K. Weimer Charge Transfer Memories (3,986,176)
D. H. Willis Gating Circuit for a Video Driver Including a Clamping Circuit (3,984,864)
H. A. Wittlinger Series Energized Transistor Amplifier (3,986,132)
C. R. Wronski, A. D. Cope, and B. Abeles Low Dark Current Photoconductive Device (3,987,327)
C. T. Wu Display Device Utilizing Magnetic Storage (3,988,738)

November

V. Auerbach Error Cancelling Scanning Optical Angle Measurement System (3,992,106)
L. R. Avery Drive Pulse Generator for a Television Deflection Circuit (3,992,648)
J. F. Balcewicz Minimum Shift Keying Communication System (3,993,868)
D. E. Carlson and L. A. Goodman Method for Forming Electrode Patterns in Transparent Conductive Coatings on Glass Substrates (3,991,227)
D. E. Carlson and C. E. Tracy Deposition of Tin Oxide Films on Glass (3,991,228)
G. A. Cutsogeorge Fast Acquisition Circuit for a Phase Locked Loop (3,993,958)
W. Den Hollander Gating Circuit for Thyristor Deflection System (3,993,931)
D. W. Fairbanks Overhead Signal Pickup Device (3,993,316)
D. A. Gore Binary-Coded Fraunhofer Hologram Recording Technique (3,990,773)
E. H. Griffin Digital-to-Snychro Converter (3,993,993)
R. E. Hanson Acceleration Burst Test Apparatus and Method for Internal Combustion Engines (3,994,160)
G. S. Kaplan and A. S. Clorfeine Respiration Monitor (3,993,995)
L. A. Kaplan Phase Control Circuit Including an Operational Transconductance Amplifier suitable for Use in Audio Frequency Signal Processing Apparatus (3,995,235)
S. L. Knanishu Alternating Current Meter Circuit (3,993,951)
M. B. Knight Current Amplifiers (3,992,676)
M. A. Leedom and J. C. Bleazey Disc Record Groove Skipper (3,993,863)
H. F. Lockwood, M. Ettenberg, H. Kressel, and J. I. Pankove Light Emitting Diode with Reflector (3,991,339)
P. L. Myers Automatic Assembly of Semiconductor Devices (3,992,770)
R. L. Pryor Complementary Field Effect Transistor Differential Amplifiers (3,991,380)
W. F. Reichert Method of Forming Raised Electrical Contacts on a Semiconductor Device (3,993,515)
F. N. Sechi Linear High Power Transistor Amplifier (3,991,381)
E. M. Sutphin, Jr. Fast Automatic Gain Contral Circuit with Adjustable Range (3,995,224)
R. E. Tetrev Detection System for Spatially-Distributed Set of Radiation Beams Manifesting Multibit Binary Code (3,995,146)
C. L. Upadhyayula Planar Transferred Electron Logic Device (3,991,328)
P. W. Wolverton Article Positioning Apparatus (3,990,692)
D. J. Woywood Charge Coupled Parallel-to-Serial Converter for Scene Scanning and Display (3,995,107)

December

D. R. Andrews Tape Cartridge Player with Cartridge Pull-In Mechanism (3,999,219)
L. R. Avery Amplifier Suitable for Use as a Color Kinescope Driver (3,996,609)
W. H. Barkow Deflection Yoke Having Nonradial Winding Distribution (3,996,542)
D. S. Bond Method of Storing Space Satellites in Orbit (3,995,801)
M. E. Breese Monopulse Radar System (3,996,589)
R. L. Camisa, B. F. Hitch, and S. Yuan Metal-Insulator-Semiconductor Device Phase Shifter (3,996,536)
L. A. Cochran and L. A. Harwood Controllable Gain Signal Amplifier (3,999,141)
W. R. Curtice Varactor Tuning Apparatus for a Strip Transmission Line Device (3,996,529)
W. R. Curtice Transferred Electron Device Pulse Train Generator (4,000,415)
R. H. Dean, L. S. Napoli, and S. G. Lui Method of Fabricating a Photovoltaic Device (3,999,283)
W. G. Einthoven and W. Hulstrunk Semiconductor Device with Solder Conductive Paths (3,997,910)
W. G. Einthoven Semiconductor Device Having Parallel Path for Current Flow (3,999,217)
L. A. Harwood Hue Correction Apparatus Having a Restricted Range (3,996,608)
K. G. Hernqvist Gas Laser Optical System (3,996,527)
K. G. Hernqvist and A. H. Firester Method of Aligning a Laser Device (3,999,858)
H. Kawamoto Comb Filter Apparatus for Video PlaybacK Systems (3,996,610)
D. Meyerhofer Controlled Angle Viewing Screens by Interference Techniques (3,996,051)
R. S. Mezrich and D. H. Vilkomerson Ultrasonic Wave Radiation Pattern Display System Incorporating Phase Contrast Means (3,997,717)
L. Muhlfelder and N. U. Huffmaster Backup Wheel for a Three Axis Reaction Wheel Spacecraft (3,999,729)

R. A. Newell, J. H. Bacher, H. W. Bilsky, and P. J. Callen Satellite Battery Reconditioning System and Method (3,997,830)
S. J. Niconienko Multiposition Rotary Switch with Detent Means (3,996,440)
C. S. Oh and E. F. Pasierb Alkoxybenzylidene-Aminobenzonitriles (3,996,260)
R. S. Packman Panel Edge Fastener Clip (3,999,356)
K. J. Phillips Minimization of Residual Spacecraft Nutation Due to Disturbing Torques (3,997,137)
W. Phillips Method of Preparing Optical Waveguides (3,997,687)
J. S. Pistiner Minimization of Spacecraft Attitude Error Due to Wheel Speed Reversal (3,998,409)
D. H. Pritchard Comb Filter for Video Processing (3,996,606)
J. S. Radovsky Bias Current Circuit (3,999,140)
G. H. Riddle and R. R. Demers Mechanically Adjustable Electron Gun Apparatus (3,997,807)
J. Rosen Semipassive Responder Utilizing a Low Voltage, Low Power Drain Reflective Varactor Phase Modulator (3,996,587)
B. D. Rosenthal and A. G. Dingwall Voltage Amplitude Multiplying Circuits (4,000,412)
W. Rosnowski Diffusion of Conductivity Modifiers into a Semiconductor Body (3,997,379)
O. H. Schade, Jr. Current Limiting Circuit and Method (3,996,498)
H. D. Scheffer Ultrasonic Wire Bonding Chuck (3,995,845)
F. N. Sechi Linear Amplifier Utilizing Adaptive Biasing (3,996,524)
R. G. Stewart Rectifier Structure for a Semiconductor Integrated Circuit Device (3,999,205)
B. K. Taylor Disc Record Locked Groove Escape Apparatus (3,997,717)
Z. Turski Microwave Bulk Acoustic Delay Device Having Two Transducers on the Same Surface and Method of Making Same (3,996,535)
H. A. Wittlinger Oscillator Circuit Whose Frequency is Voltage Controllable which Contains a Comparator (3,996,535)

AUTHORS

Alan E. Bell received his B.Sc. (Hons. 1st class) degree in Physics from Imperial College, London University in 1969, becoming at the same time an Associate of the Royal College of Science (A.R.C.S.). His thesis work in the Metal Physics group at Imperial College was an experimental study of the low temperature magnetic and electrical properties of very dilute alloys (the Kondo effect), for which he was awarded his Ph.D. in 1973, becoming at the same time a Diplomate of Imperial College (D.I.C.).

In 1973 he was awarded the Sarnoff Fellowship to visit RCA Laboratories at Princeton to engage in post-doctoral research for a period of nine months. This time was spent working in the field of liquid crystalline materials, using optical Raman scattering to study molecular order in the nematic phase of the material. Since becoming a member of the staff at RCA Laboratories in 1974, he has worked mainly in the area of materials development for the optical video disc recorder project. This work has included a theoretical analysis of the thermal and optical properties of thin film multilayers which has led to new designs for structures that increase both thermal and optical efficiency of the recording process. Dr. Bell has recently extended this experience in analysis of thermal diffusion problems to model the thermal conditions involved in the shaped growth of silicon single crystals using a laser zoning technique, and has also performed a combined experimental and theoretical study of the thermal characteristics of the RCA SelectaVision stylus during playback.

Dr. Bell received an RCA Laboratories Achievement Award for his work in thermal analysis and is a member of the Optical Society of America.

David W. Fairbanks joined RCA Laboratories in 1963, working in the Mechanical Engineering Section. His work includes the design of in-house equipment and production equipment used in other RCA plants, including two years designing SelectaVision Player Systems. He was responsible for the mechanical design of a laser scan mask inspection system used in the RCA Somerville plant. He has also been responsible for the mechanical design of the Photovoltaic Tracker under a contract with Sandia Laboratories, Albuquerque, N.M. From 1960 to 1963, he was associated with Johnson and Johnson as a planner estimator for industrial trades. From 1953 to 1960, he operated his own machine shop, designing and producing high-speed wire forming machinery. He also privately maintained a large experimental machine shop where he was engaged in developing an experimental automobile.

Mr. Fairbanks has received two RCA awards.

Sheng T. Hsu received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1958; the M.S.E.E. degree from National Chiao-Tung University, Hsienchu, Taiwan, in 1960, and the Ph.D. degree in electrical engineering from the University of Minnesota, Minneapolis, in 1966. From 1966 to 1970, he was with Fairchild Semiconductor Research and Development Laboratory, Palo Alto, California, as a Member of the Technical Staff. From 1970 to 1972, he was an Assistant Professor of the Department of Electrical Engineering, University of Manitoba, Winnipeg, Man., Canada. He is now a staff member of RCA Laboratories, Princeton, N.J., working on semiconductor devices and integrated circuit technologies.

Evelyn Jetter received the B.S.E.E. from Cooper Union, N.Y., in 1950 and the M.S. in physics from Rutgers University, New Brunswick, N.J., in 1963. From 1948 to 1953, she worked for the U.S. Atomic Energy Commission in their New York Operations Office developing methods for measuring and interpreting radiation exposures. From 1956 to 1965, she worked at Newark Controls in Bloomfield, N.J., on the design of pressure and density switches for operation in extreme environments, and from 1965 to 1967 at Lionel electronics Laboratories, Hillside, N.J., on radiological instrument calibration.

In 1967, she joined the RCA Solid State Division, Somerville, N.J., working on semiconductor devices that control high current, high voltage for power application, and particularly automotive ignition devices. She is presently working on the design of a computer-controlled system for monitoring and controlling semiconductor processes for the manufacture of integrated circuits.

Norman Klein graduated from the Bulova School of Watchmaking in 1949. From 1952–56, he worked for Bell Telephone Laboratories in Murray Hill, N.J., as a electron device mechanic. In 1956, he joined RCA in Harrison, N.J., where he specialized in the fabrication of traveling wave and magnetron tubes, and special R&D devices used in microwave technology. He was then assigned to the Microwave Satellite Program, supervising the assembly of traveling wave tubes and quality control. In 1963, he transferred to RCA Laboratories, Princeton, N.J., working as a Research Technician in the Microwave Research Laboratories. His work responsibilities included the fabrication of solid-state and vacuum tube devices such as plasma tubes, lasers, integrated circuits, and microminiature transistors, diodes and miniaturized components. Mr. Klein has completed electronics courses at RCA Institutes and RCA sponsored courses in vacuum technology.

In 1969, he became a technical staff member and in 1973 joined the Microwave Technology Center, working on semiconductor processing for Trapatt diodes, PIN diodes, and high intensity solar cells.

Shing-gong Liu received the B.S. degree in electrical engineering from Taiwan University, Taipei, Taiwan, in 1954, the M.S. degree in electrical engineering from North Carolina State College, Raleigh, in 1958, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, Calif., in 1963. From 1958 to 1959, he worked with the IBM Laboratories, Poughkeepsie, N.Y., from 1960 to 1963, he did research work on microwave ferrites in the Hansen Microwave Laboratories, Stanford University. He joined RCA Laboratories, Princeton, N.J., in August 1963, where he has been engaged in research on the use of high-field nonlinear effects in GaAs for microwave devices and optical modulation, beam deflection devices, and avalanche diode microwave oscillators, using both Si and GaAs materials. Since 1974, he has been working on high efficiency solar cells for concentrator application and GaAs Schottky diodes for low noise mixers for satellite ground stations. He received three RCA Outstanding Achievement Awards for his work on high-power high-efficiency avalanche diodes.

Dr. Liu is a member of Phi Kappa Phi, IEEE, Sigma Xi, and the American Physical Society.

Ramon U. Martinelli received the A.B. and M.S. degrees from Dartmouth College, Hanover, N.H., in 1960 and 1962, respectively, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, N.J., in 1966.

He joined RCA Laboratories as a Member of the Technical Staff in 1966 and until 1974 worked on electron emission and transport problems. His present research is in the area of silicon bipolar device physics.

Dr. Martinelli is a member of the American Physical Society.

Louis S. Napoli received his B.S. in 1959, and the M.S. in 1961, in Electrical Engineering, both from Rutgers University. He has pursued further studies in plasma physics at Princeton University and in electro-physics at the Polytechnic Institute of Brooklyn. Since joining the technical staff of RCA Laboratories in June 1959, he has specialized in research relating to microwave phenomena in gaseous plasmas, solid-state microwave devices, and microwave integrated circuits. He is presently Head of the Microwave Components Technology Group in the Microwave Technology Center at RCA Laboratories. His work in collaboration with Dr. George Swartz on amplification at 24 Gc by the interaction of an electron beam with a cesium plasma was cited by Industrial Research Magazine as one of the 100 most important achievements in 1963. He is the recipient of RCA Laboratories Achievement Awards in 1963, 1965, and 1968.

Mr. Napoli is a member of Sigma Xi, IEEE, Tau Beta Pi, and Eta Kappa Nu.

Dalton H. Pritchard received the B.S.E.E. degree in Electronics in 1943, from Mississippi State University. Upon graduation he entered the U.S. Army Signal Corps. In 1946, Mr. Pritchard joined RCA Laboratories at Riverhead, N.Y., as a Member of the Technical Staff where he was engaged primarily in research activities in the area of single sideband, radio-teletype, multiplex communications systems. In 1950, he transferred to the RCA Laboratories in Princeton, N.J., and has been engaged, for a number of years, in research in many aspects of color television systems development, receivers, color kinescopes, transmitting encoders, cameras, magnetic recording of TV, and many associated systems and devices. This work included the planning and testing of systems and circuits proposed for adoption by the National Television Systems Committee (NTSC).

More recently, Mr. Pritchard has been engaged in research in the field of Information Display Systems and Devices which has included activities in the related fields of vacuum preparation of materials, specialized electron-optic devices, lasers, the development of electro-optic materials and techniques for purposes of light control and display and color camera special systems development and evaluation. At present, he is engaged in the development and evaluation of video processing circuitry for color TV receivers, particularly in the areas of colorimetry and decoder matrix techniques.

Mr. Pritchard was appointed Fellow of the Technical Staff at RCA Laboratories in 1975. He is a Fellow of the IEEE, a Fellow of the SID, and a member of Sigma Xi, Tau Beta Pi, and Kappa Mu Epsilon and is listed in American Men of Science and Who's Who in the East.

Henry S. Sommers, Jr., attended Stanford University in 1933 and 1934, and received his B.A. in physics from the University of Minnesota in 1936. He was awarded a Ph.D. in physics (nuclear physics) in 1941 from Harvard University, where he served as Physics Instructor during 1942. From 1942 to 1945, he was a Staff Member at the MIT Radiation Laboratory. In 1946, Dr. Sommers was appointed Assistant Professor of Physics at Rutgers University, where he also conducted research and development work on a precision current control for a large electromagnet. In 1949, he joined the research staff at the Los Alamos Scientific Laboratories in New Mexico, where he carried out fundamental studies on the thermodynamic properties of liquid helium and on the scattering of neutrons by liquid helium.

Dr. Sommers joined RCA Laboratories at Princeton, N.J., in 1954 and was appointed a Fellow of the technical staff in 1969. His work at RCA has been devoted to the study of electrical and optical properties of insulators and semiconductors, including the development of tunnel diodes for which he received an RCA Laboratories Achievement Award and a David Sarnoff Outstanding Team Award in Science. More recently he has been conducting fundamental studies on injection lasers, concentrating on laser dynamics and model excitations. He spent the year 1960–1961 at Laboratories RCA Ltd. in Zurich and as a Guggenheim Fellow and Fulbright Lecturer at the Hebrew University in Jerusalem.

Dr. Sommers is a Fellow of the American Physical Society and a member of the Federation of American Scientists and of the American Association for the Advancement of Science.

George A. Swartz received his B.S. degree from Massachusetts Institute of Technology in 1952, and the M.S. and Ph.D. degrees in physics from the University of Pennsylvania in 1954 and 1958, respectively. He joined the technical staff of RCA Laboratories in 1957. His work has been in the fields of plasma propulsion, plasma stability, microwave phenomena in gaseous and solid-state plasmas, solid-state physics, in the development of low-noise, high-efficiency IMPATT diodes, and the development of low loss high power PIN diodes. Most recently, he has worked on the development of high efficiency solar cells for high level solar concentration. His work in collaboration with Louis Napoli on a beam-plasma amplifier was cited by Industrial Research magazine for an I-R 100 award in 1963.

Dr. Swartz has received two RCA Achievement Awards and is a member of the American Physical Society and Sigma Xi.

Donald J. Tamutus attended St. Bonaventure University, University of Heidelberg (Heidelberg, Germany) and Rutgers University. After a tour in the U.S. Army, he joined RCA Laboratories in 1962, assigned to the Photomechanical Processing Lab. This group is responsible for fabricating printed circuit boards, etched masks, and all art work generation and photomask fabrication for integrated circuits. He has worked on a number of projects including solar cell connectors for various satellites, cryogenic memories in which he developed an electroform mask process, and video disc. In addition to the solar project, he is working on shadow mask development and flat panel TV. Mr. Tamutus is active in many civic functions in his community where he is an elected official and a former council president in New Jersey's ninth largest municipality.

Theodor M. Wagner received the Ing. grad. degree from the Oscar von Miller-Polytechnikum in Munich, Germany in 1953. He was employed with Rohde und Schwarz in circuit design of TV broadcast measuring equipment for six years. He was a group leader at Agfa-Gevaert's Lab of Physics, for six years prior to coming to the United States in 1964. In 1964 and 1965, he designed chroma processing circuits of color TV receivers with General Electric. In 1966 and 1967, his work included studies on vidicon tubes and the design of a high resolution industrial camera for medical and radar scope displays at General Electro-Dynamics Corp. in Garland, Texas.

He joined RCA Burbank in 1968. Mr. Wagner was a member of the team that received the 1969 David Sarnoff Outstanding Team Award in Engineering for the electro-optical design of the single-tube color camera. In 1972, Mr. Wagner transferred to the David Sarnoff Research Center at the RCA Laboratories in Princeton, N.J., where he is currently working on analyzing TV receiver colorimetry and optimizing measurement techniques.

He is member of the SMPTE and the SPSE.