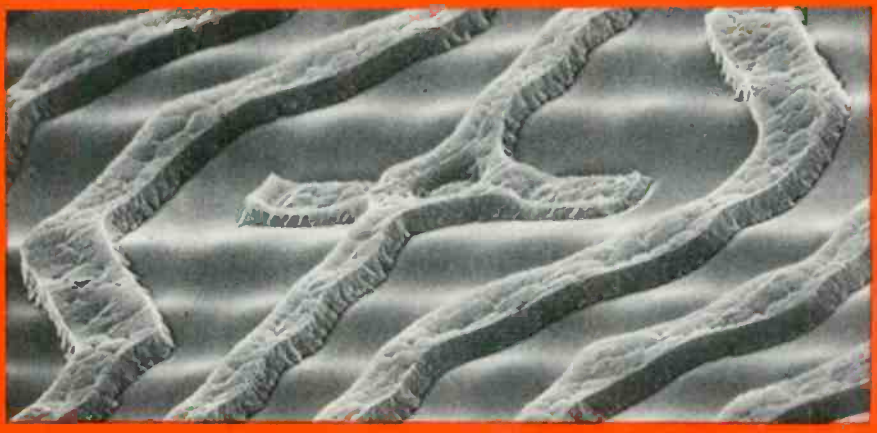
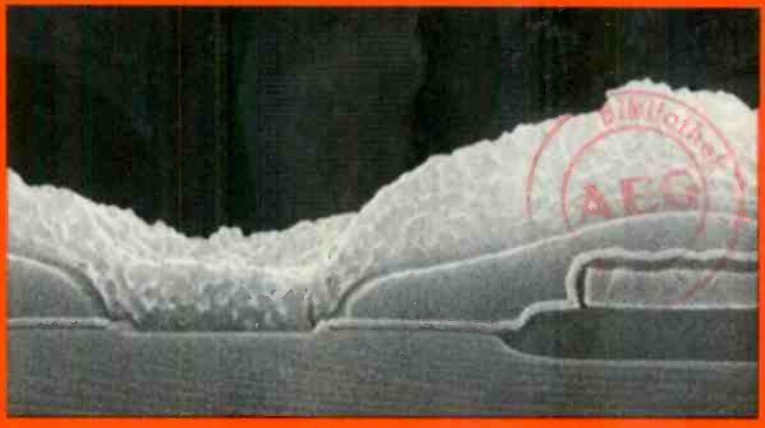
RCA Review





SEMs Showing BPSG Reflowed on CMOS 128K ROM

June 1985 Volume 46 No. 2

RCARCI46(2) 115–278 (1985)

Cover—Our cover photos are from the paper by Kern, Kurylo, and Tino on chemical vapor deposition of borophosphosilicate glass (BPSG). The top scanning electron micrograph (SEM) is a topographical view showing a via hole and aluminum interconnects over reflowed BPSG. The bottom SEM is a cleaved and demarcation-etched sectional view showing metal on top of BPSG and in a contact hole. This read-only memory (ROM) circuit has nominal feature sizes of 3 μm.

RCA Review, published quarterly in March, June, September, and December by RCA Laboratories Princeton, New Jersey 08540. Entered as second class matter July 3, 1950 under the Act of March 3, 1879. Second-class postage paid at Princeton, New Jersey, and at additional mailing offices. Effective January 1, 1983, subscription rates as follows: United States: one year $12.00, two years $21.00, three years $27.00; in other countries: one year $14.00, two years $24.50, three years $31.50. Single copies up to five years old $5.00

RCA Review

RCA Review (ISSN 0033-6831) is a technical journal published quarterly by RCA Laboratories in cooperation with the subsidiaries and divisions of RCA.

Contents

June 1985 Volume 46 Number 2 ISSN 0033-6831

Optimized Chemical Vapor Deposition of Borophosphosilicate Glass Films

Werner Kern and W. A. Kurylo
RCA Laboratories, Princeton, NJ 08540

C. J. Tino
RCA Aerospace and Defense Div., Somerville, NJ 08876

Abstract—The atmospheric-pressure chemical vapor deposition (APCVD) of borophosphosilicate glasses (BPSG) has been optimized for more efficient utilization of the reactants, which results in glass films with much fewer particle contaminants than could previously be obtained. These improvements have been achieved by lowering the deposition temperature and by increasing the oxygen/hydride ratio to minimize the particle-generating homogeneous gas phase nucleation and to maximize the heterogeneous CVD reaction that forms the desired glass films. Both laboratory-type and large-scale production reactors were used. The properties of BPSG films that are important in fusion-tapering applications of integrated circuits are not changed by the optimized CVD process.

New results concerning several additional aspects of BPSG technology are discussed and illustrated, including glass stability during processing, fusion in various ambients, and the application of isothermal rapid heating techniques for glass flow. In addition, a complete collection of references[1–75] on BPSG from 1955 through June 1985 is presented.

1. Introduction

The use of borophosphosilicate glass (BPSG) films in integrated circuit technology has found industrywide and very rapid acceptance in the last few years. The reason for this favorable acceptance is that thermal flow of BPSG glass films can be performed at reduced temperatures compared to conventional glass films. Lowest possible temperature during processing is desirable for any type of semiconductor device, but is essential for VLSI and VHSIC silicon inte-

grated circuits to minimize vertical and lateral diffusion of dopants and impurities from and into the shallow junction regions and to minimize the interdiffused junction width. It is also of critical importance for highly heat-sensitive radiation-hardened devices. When BPSG is used, realistic and practicable temperatures for the entire process do not exceed 900 to 950°C, a reduction of as much as 300°C compared to temperatures required when conventional phosphosilicate glass (PSG) is used.

Temperatures required for flow-tapering PSG containing the frequently used concentration range of 6 to 8 wt% P is 1050 to 1100°C. The addition of boric oxide (B_2O_3) to PSG by introduction of a suitable boron-containing reactant into the gas stream during chemical vapor deposition (CVD) yields films whose fusion temperatures are significantly lower (as low as 800°C) than those of PSG. The resulting BPSG glass films have several additional advantages over PSG that render them ideal not only for the fusion tapering of steps in the substrate, but also as interconductor insulator material for silicon-gate MOS integrated circuit applications and for overmetal passivation coatings.

This paper reports the results of research done at our laboratories to optimize the conditions of atmospheric-pressure CVD of BPSG films. Optimized conditions mean more efficient utilization of the hydride reactants, which translates into improved economy due to an increase of the reactant conversion yield to glass films or a corresponding increase in film deposition rate. At the same time, the particle density is decreased by minimizing the homogeneous reaction and maximizing the desirable heterogeneous reaction mechanism of the hydride oxidation. The interaction effects of diborane and phosphine under the new CVD conditions have also been studied and are shown to be considerably different from those observed under the conditions previously used.

Additional aspects of BPSG films that are discussed include stability in aqueous cleaning solutions used in processing, possible compositional changes during heat treatments, and the use of rapid high-intensity lamp heating for the fusion flow of BPSG.

The scientific and technical literature on BPSG has grown considerably in the last three years, and we are taking this opportunity to present a very complete (1955 up to June 1985) chronologically arranged collection of references.[1-75] These references, including U.S. and foreign patents, deal with many different aspects of BPSG, either as a primary topic or in the context of a related subject. References to both bulk glasses and films of BPSG have been included.

For a general review of the state-of-the-art of BPSG technology and its applications, the reader is referred to Ref. [64]. Two fundamental papers on the subject are those by Ramiller and Yau[13] and Kern and Schnable.[14]

2. Maximizing the Reaction Yield of Hydride to Glass Films

Atmospheric-pressure CVD of BPSG films from diborane, phosphine, silane, and oxygen at temperatures in the range of 410 to 440°C has been quite satisfactory for most production applications in the fabrication of integrated circuits. Thus, by using a continuous-production reactor under factory operating conditions, we have attained, for 800-nm-thick BPSG films, a thickness uniformity of ±1.5% over 100-mm wafers and ±4.2% from wafer to wafer. APCVD has the advantage that relatively high deposition rates (100 to 160 nm/min) can be easily obtained and that the process is fairly simple and readily controllable. Also, established and well-performing production reactors of various types and makes are available at relatively low cost.*[34] The lack of conformal coverage, which is typical for any APCVD process, is no disadvantage for flow-tapering application where the glass is fused and reshaped in subsequent processing.

The cooxidation of the nitrogen-diluted hydride mixture by the conventional CVD process at 430°C for BPSG is conducted with an oxygen-to-hydride ratio of 20:1,[13,14] the ratio normally used for the deposition of SiO_2 and PSG films at this temperature.[6,12,76] It has been well established that the deposition rates of SiO_2 and PSG films increase with temperature, but that the O_2/hydride ratio must be adjusted to the temperature of deposition to attain the maximal rate of film growth and the best dielectric qualities. For example, PSG films deposited from 5 vol% PH_3–95 vol% SiH_4 at 350°C require an O_2/hydride ratio of 10:1, whereas at 450°C a ratio of 20:1 is needed. The deposition rate increases by a factor of 1.6 when the temperature is increased from 350 to 450°C.[76]

We found that BPSG deposition does not follow the same characteristics. Model experiments were done with polished silicon wafers in an RCA rotary hot-plate bell jar reactor[77] as in earlier development work.[14] Fig. 1 shows the rate of film deposition for BPSG type A[14] (flowing at 900–925°C in steam) as a function of reactor-plate temperature for different O_2/hydride ratios. The same

* For example, Systems by Applied Materials, Inc.; Pacific-Western Systems, Inc.; Tempress (Unit of General Signal); or Watkins-Johnson Co.

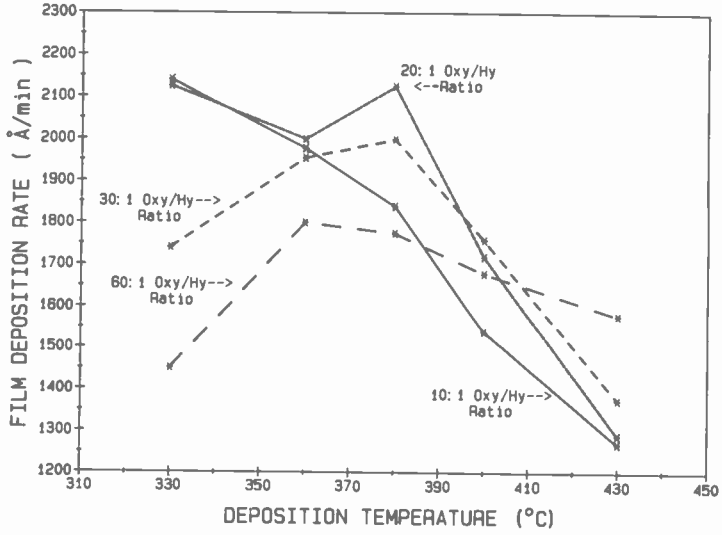


Fig. 1—BPSG type A deposition rate in the RCA rotary APCVD reactor as a function of plate temperature for various oxygen/hydride ratios.

data are plotted in Fig. 2 as temperature versus O_2/hydride ratio for the various deposition temperatures. Lowering the temperature from the conventional 430°C leads to substantially higher deposition rates that peak in the range of 360 to 380°C, depending on the O_2/hydride ratio, except for ratios of 10:1 and 20:1 for which the deposition rate continues to increase with lower temperature to at least 330°C. The largest increase is observed for the 10:1 and 20:1 ratios, the latter peaking at 380°C where the deposition rate corresponds to a 64% increase from that at 430°C. A ratio of 30:1 gives a rate increase of 43% for the same temperature interval. At the high-temperature portion (415 to 430°C), an increase in the O_2/hydride ratio increases the film deposition rate; below 415°C an increase in this ratio depresses the deposition rate.

The purpose of the above experiment was to test for large differences and to bracket the peak deposition rate in ranges of 20°C. Exact parameters were established by use of a large-scale CVD reactor, an AMS-2100 continuous in-line production system (Applied Materials, Inc.). This system features a large-area gas dispersion plate beneath which the substrate wafers pass on heated Inconel metal trays.[78] Even though the gas flow rates, geometry, design, and functioning of the large-scale reactor are entirely different from the small-scale rotary hot plate unit, both yield surprisingly

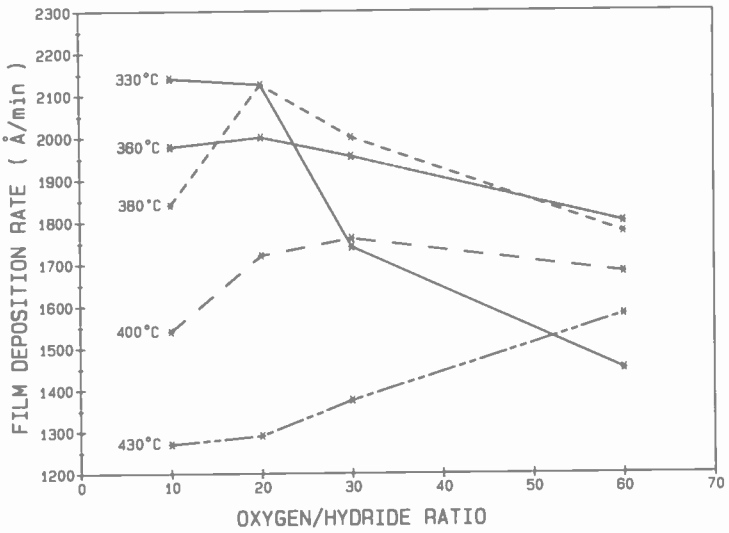


Fig. 2—Data from Fig. 1 plotted as a function of oxygen/hydride ratio.

similar results, at given temperatures, in terms of gas flow ratios and BPSG composition.[14]

The deposition temperatures stated throughout this paper refer to the temperature of the reactor plate on which the silicon wafers were placed during CVD. The plate in the RCA rotary reactor[77] is an aluminum alloy; that of the AMS-2100 continuous reactor is thermally oxidized Inconel® alloy.[78] The temperatures were measured after thermal equilibration of the CVD systems with thermocouples (pressed to the metal surface with glass fiberboard), with bimetallic surface thermometers, and with accurately calibrated spring-loaded thermocouple surface thermometers (Model 450-AKT Type K Digital Thermometer with 68101-K-SMP Surface Probe, Omega Engineering). The average values stated are believed to be correct to within ± 5°C of the true surface temperature. We have found that Inconel trays increase in temperature with extended use as wafer carriers (in a process that includes stripping of the deposited glass layers in diluted HF solution after each run) by as much as 50°C for reasons not entirely clear at this time.

The temperatures of 76-mm-diameter silicon wafers were measured in the same manner as described above. It is important that both the wafer and the plate be planar and free of insulator buildup to ensure close thermal contact for good heat transfer. To correlate the wafer temperature with the plate temperature, measurements of the silicon wafer temperature were taken in one-minute intervals

until the readings no longer changed. The wafer was then removed and the plate temperature was measured immediately afterwards in the same fashion. The temperature and its uniformity over the wafer were also tested by use of a noncontact optical instrument that uses an InSb photovoltaic infrared detector (AGA Thermovision® 782, AGA Infrared Systems, AB). The temperature scan is displayed on a color TV monitor screen. Thermal uniformity within 5°C was observed as long as the wafers were not bowed, as this would prevent good thermal contact. All the results showed the wafer temperature on the RCA reactor plate to be 10 to 20°C lower than the plate temperature; on the AMS-2100 reactor plate the wafer temperature measured by Thermovision was 5 to 10°C lower, whereas the other techniques indicated a 15 to 30°C difference.

Deposition tests were conducted in 10°C steps from 330 to 430°C after careful temperature equilibration of the reactor. Groups of twelve 75-mm-diameter polished silicon wafers were used for each run, with 16 to 20 "dummy" wafers positioned in the front and also in the rear of the test group moving through the reactor. BPSG type B^{14} (flowing at 800–825°C in steam) was used. Interferometric thickness measurements were taken on five positions (one in the center and one position six mm from wafer edge in each quadrant). The readings on all 12 wafers were closely similar. They were averaged and then plotted in terms of film deposition rates versus temperature (Figs. 3 to 5). Each data point thus represents the average of 60 thickness measurements.

Fig. 3 shows the increase of the film deposition rate as the temperature is lowered from 430 to 380°C for a fixed O_2/hydride ratio of 30:1. The increase in deposition rate for lowering the temperature from 430 to 380°C was 63%. No tests were made in this matrix at temperatures below 380°C; the shape of the graph indicates that not much additional rate increase, if any, is expected below that temperature.

A second matrix of tests was run in the temperature range extending from 400 to 330°C. An O_2/hydride ratio of 40:1 was used (by increasing the O_2 flow rate while keeping the hydride flow rate constant) because of evidence of decreased particle density (discussed in the next section). The same deposition and measurement conditions were used as in the previous matrix (Fig. 3). The data are shown in Fig. 4. Peaks in deposition rate are evident at 340°C and at 370°C, the latter being the maximum.

The data from both tests have been combined in Fig. 5 to facilitate direct comparison. The partial curve for the 30:1 ratio from 380 to 430°C has a shape similar to that of the curve for the 40:1 ratio;

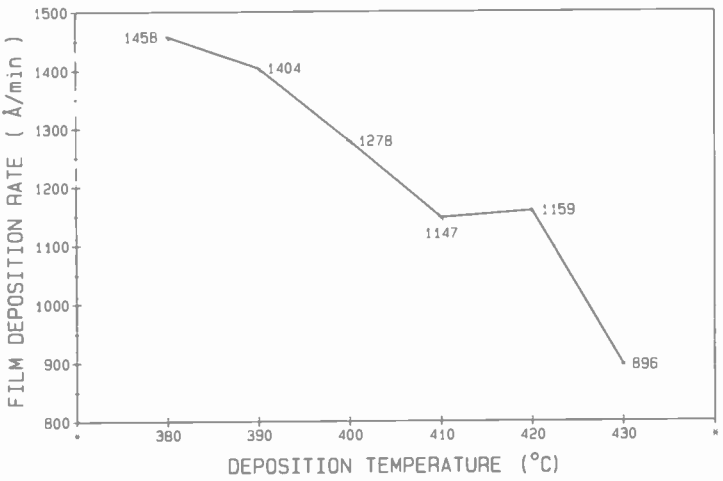


Fig. 3—BPSG type B deposition rate in a continuous APCVD in-line production reactor as a function of plate temperature for an oxygen/hydride ratio of 30:1.

the deposition rates are higher, as expected from the results in Figs. 1 and 2. A 53% increase in film deposition rate resulted when going from 430°C with a ratio of 30:1 to 370°C with a ratio of 40:1 which gives the maximum increase.

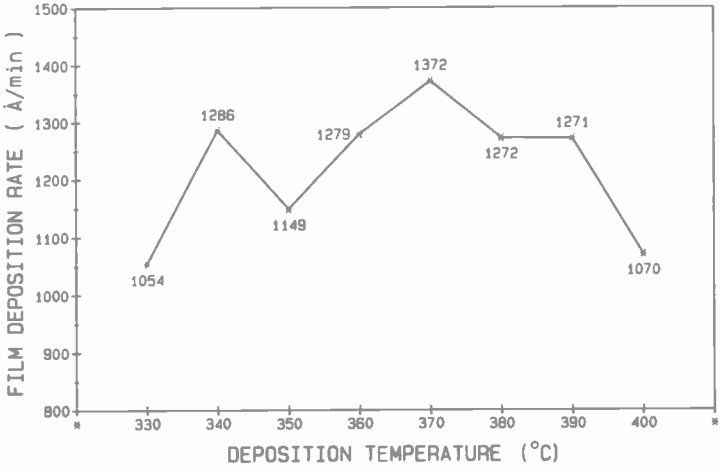


Fig. 4—BPSG type B deposition rate in a continuous APCVD in-line production reactor as a function of plate temperature for an oxygen/hydride ratio of 40:1.

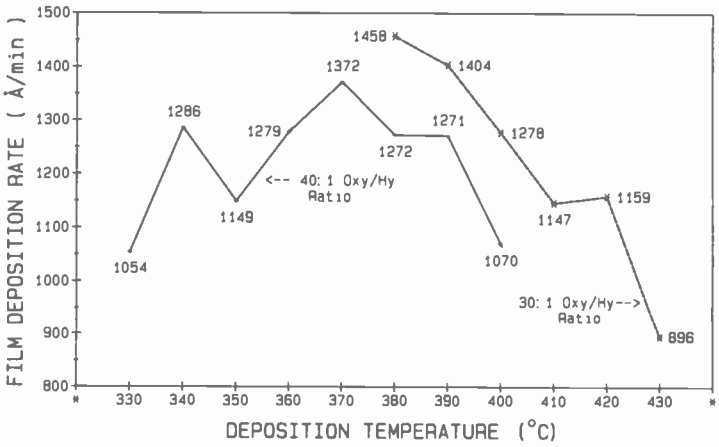


Fig. 5—Data from Figs. 3 and 4 are combined for comparison.

In general, the rate of BPSG film deposition increases significantly as the temperature is lowered by at least 50°C from the conventional 430°C. The increase is due in part to suppression of homogeneous gas phase nucleation, so that more reactant gas is available for glass-film formation by the heterogeneous reaction. Further lowering of the temperature tends to decrease the deposition rate, but the exact behavior depends on the O_2/hydride ratio and the CVD conditions and reactor type. The deposition rate below 400°C becomes depressed when the O_2/hydride ratio increases from 20:1 to 60:1. This oxidation-inhibiting effect occurs when a very large excess of oxygen is present in the reactant gas mixture, analogous to the oxidation of silane in the formation of SiO_2 films.[6]

3. Minimizing Particle Density

The numerical density and size of particles on the surface and in the layers of device wafers are major concerns in VLSI applications. These particles must be eliminated (down to less than the 0.3-μm size) or at least reduced to acceptable levels. Lowering the temperature of BPSG film deposition to the range of 340 to 370°C substantially reduces particle formation in the gas phase and the subsequent deposition on the reactor wall areas and on the critical wafer surface. The explanation is the same as for increased deposition rate at lower temperatures: there is a shift from the homogeneous reaction, which uses up reactant gas and nucleates particles, towards the heterogeneous reaction, which forms the desired glass film on the substrate surface.

Fig. 6 shows the average particle density per 76-mm-diameter wafer for BPSG type B deposited at a ratio of 40:1 as a function of temperature of deposition in a continuous APCVD reactor (AMS-2100). Each data point is the average of twelve 76-mm wafers. A Hamamatsu particle counter (Model C1515) used for the measurements was operated in the haze mode. Particles were measurable down to 0.3-μm size, and about 80% of all particles were in the range of 0.3 to 0.5 μm. The particle density decreased by 70% as the temperature was lowered from 430 to 370°C, then slowly decreased another 5% (total of 75%) down to 330°C.

Dornfest recently reported measurements based on laser light scattering from particle byproducts of the hydride gas phase reaction to form BPSG.[54] A laser beam was directed into the process chamber of an AMS-2100 reactor while film deposition was in progress, and light scattering was monitored with a photomultiplier tube. The data showed a decrease of the particle-forming gas phase or homogeneous reaction with decreasing temperature, accompanied by an increase of the film deposition rate. A plot of light intensity versus 1/K gave two linear curves typical for CVD reactions. The curves intersect at 350°C; the higher-temperature curve (350–420°C) represents the mass transport limiting regime, while the much more steeply sloped lower-temperature curve (approximately 260–350°C) is characteristic of the region limited by the surface reaction rate. These results corroborate our findings very well, even though an indirect method of detection of wafer contamination by particles was employed.

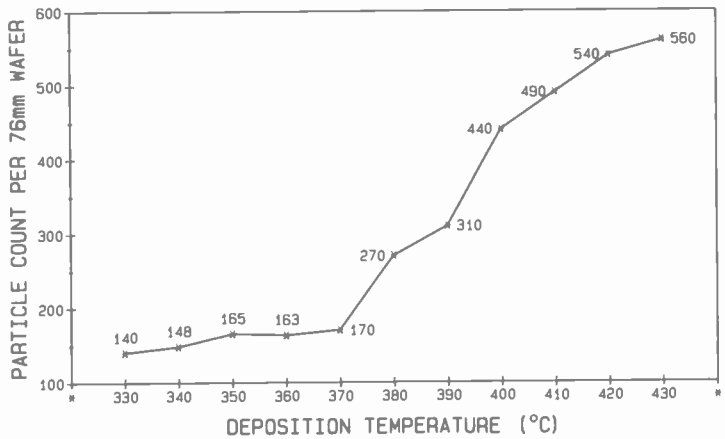


Fig. 6—Particle density per 76-mm wafer for BPSG type B as a function of plate temperature for deposition conditions as noted in Fig. 4.

We have found, furthermore, that an additional reduction of the particle density can be achieved by increasing the O_2/hydride ratio from the previous 20:1 to the range of 40:1 to 60:1. Fig. 7 shows the total number of particles down to 0.3-μm size per 76-mm wafer for BPSG type B deposited in the continuous-production reactor. Each bar represents the average of 48 wafers measured as described for Fig. 6. A substantial reduction of the particle density results as the O_2/hydride ratio is lowered, even at 430°C, the least desirable temperature condition in the test range. Combining a high O_2/hydrogen ratio with the lower temperature of 370°C is obviously the preferred CVD condition, as it yields maximal reactant conversion to glass and formation of a minimal number of particles.

Typical particle densities measured with the Hamamatsu particle counter are in the range of 5 to 15 particles per 100-mm wafer. Of these about 60% are in the size range of 0.3 to 0.5 μm, 15% between 0.5 and 1 μm, and 25% above 1 μm. Good agreement was obtained by darkfield scanning under the microscope, but better and worse results than these have been recorded. Quantitative particle measurements are notoriously difficult to repeat from one day to the next with high accuracy, and one should mainly compare measurements taken within a given set of samples processed under identical base conditions.

The optimal temperature and O_2/hydride ratio for BPSG production in a continuous APCVD reactor are 370°C and 40:1. Higher

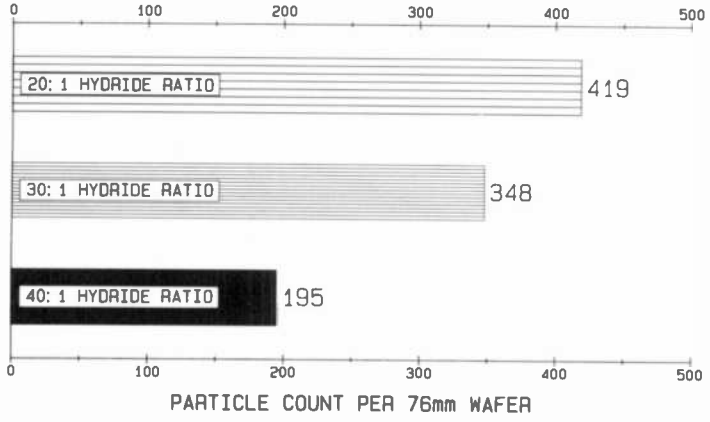


Fig. 7—Particle count per 76-mm wafer for BPSG type B deposited at 430°C in a continuous APCVD reactor as a function of oxygen/hydride ratio.

ratios suppress the deposition rate significantly without substantially lowering the particle density. A lower temperature of deposition can be used, but again the improvements in particle reduction are marginal. A very low temperature, e.g., 300°C, would require a lowering of the O_2/hydride ratio to the range of 10:1 to 20:1 to maintain a good reactant-to-glass-film yield. Also, the density and dielectric qualities of the glass films would decrease, limiting their applicability to fusion flow tapering.

The number of particles on the wafers is strongly related to the cleanliness of the deposition reactor and gas supply system. Optimal deposition conditions minimize the buildup of particulate matter on the reactor walls and gas distributor head by reducing the number of particles formed in the gas phase, but rigorous cleaning of the CVD system by specially developed techniques must be performed nevertheless to maintain consistently low particle counts on the wafers.

Although very high rates of film deposition can be obtained by optimized APCVD, we found that the rate should not exceed 120 nm/min to maintain low particle density and precise thickness control.

4. Film Composition as a Function of Optimized Deposition Parameters

The boron content in BPSG films deposited over the entire range of 330 to 430°C remains essentially constant, whereas the phosphorus concentration gradually increases with decreasing temperature by about 20%, similar to PSG deposited under comparable conditions.[76] Increasing the O_2/hydride ratio from 20:1 to 40:1 has no marked effect on the glass composition. Typical results obtained on examining these parameters are shown graphically in Figs. 8 to 11.

The boron concentration in BPSG can be readily monitored by use of dispersive[14] or Fourier-transform[35] infrared absorption spectroscopy. The absorbance ratio B-O/Si-O at maximum peak intensity is related to the boron concentration in the glass films, as shown previously.[14] Fig. 8 depicts this ratio as a function of processing temperature for BPSG type A films deposited at two O_2/hydride ratios in the RCA rotary reactor. The ratios of the as-deposited films vary considerably from each other and with temperature (for reasons other than composition). However, after a densification annealing treatment at 700°C (30 min, N_2) all film samples have the same ratio, indicating an identical boron concentration regardless of the O_2/hydride ratio or deposition temperature from 330 to 430°C.

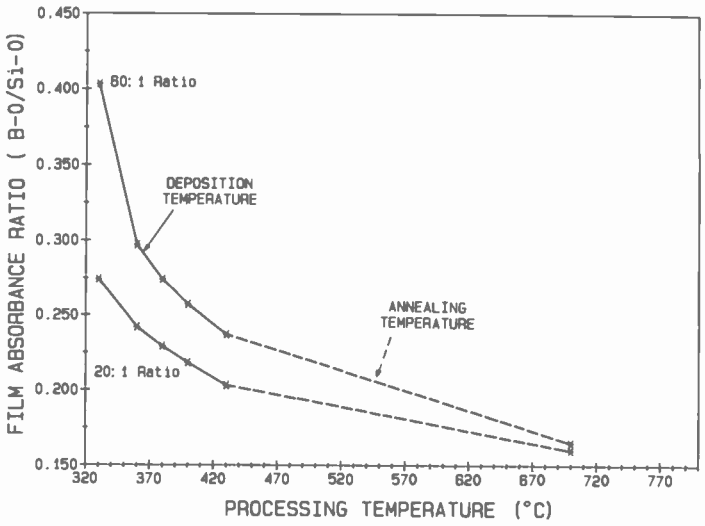


Fig. 8—Infrared absorbance ratio (B–O/Si–O) as a function of processing temperature for BPSG type A deposited in the RCA rotary reactor at oxygen/hydride ratios of 20:1 and 60:1. After the films deposited at temperatures of 330 to 420°C, respectively, were annealed at 700°C, the B–O/Si–O ratios were nearly equal, indicating the same boron concentration.

These techniques make it necessary to analyze the composition of BPSG after densification.

Figs. 9 and 10 are plots of absorbance ratio versus temperature of film deposition in an AMS-2100 continuous CVD reactor. Different O_2/hydride ratios, temperature ranges, and BPSG compositions were used for the tests in the two figures to cover a range of conditions. The B-O/Si-O ratios in all samples after densification (700°C, 30 min, N_2) were nearly level, indicating an essentially constant boron concentration for all conditions.

The samples for Fig. 9 were tested by etch-rate measurement at 25.0°C before and after densification (Fig. 11). Both buffered HF solution (BHF, 1 vol. HF 49%–6 vols. NH_4F 40%) and unbuffered HF solution (P-etch, 2 vols. HF 49%–1 vol. HNO_3 70%–60 vols. H_2O) were used. Unbuffered HF is a particularly sensitive, although nonspecific, etchant for BPSG composition tests because the etch rate increases with increases in both boron and phosphorus content. The etch rate of buffered HF decreases with increasing boron concentration in the glass.[5,13,14,79] Again it can be seen that for both types

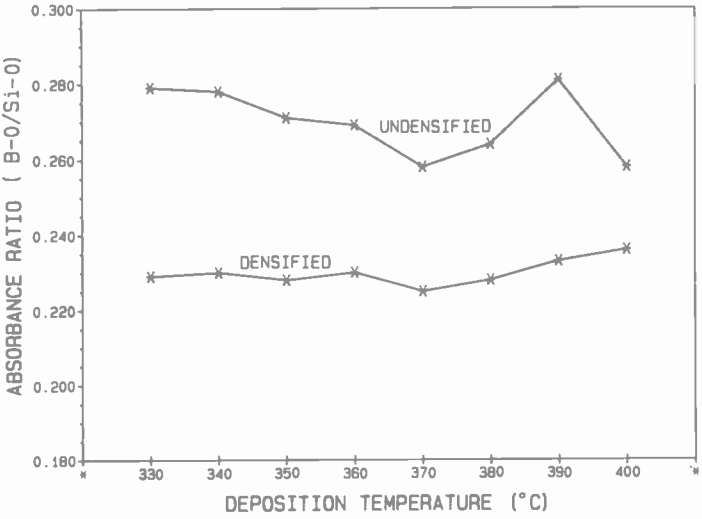


Fig. 9—Infrared absorbance ratio (B–O/Si–O) of undensified and densified BPSG type B films deposited in a continuous APCVD reactor as a function of deposition temperature at an oxygen/hydride ratio of 40:1.

of etchants, the etch rates of the samples are nearly level after densification, indicating no marked changes in film composition.

Phosphorus concentrations were measured by x-ray fluorescence

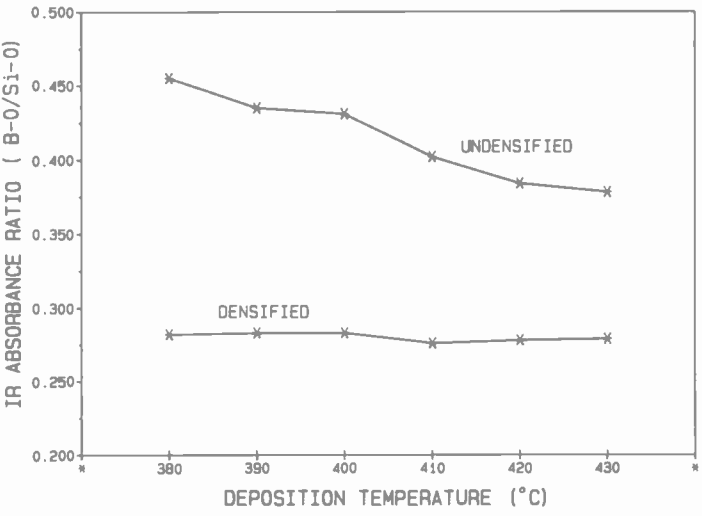


Fig. 10—Same as in Fig. 9, except for BPSG type BC[14] deposited with an oxygen/hydride ratio of 30:1.

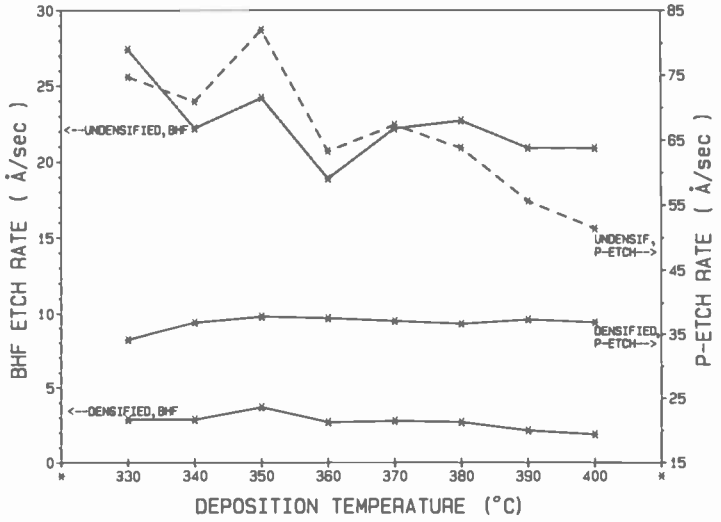


Fig. 11—Etch rates of BPSG samples from Fig. 9 at 25°C in two etchants. Nearly constant etch rates of the densified samples indicate an absence of marked changes in film composition.

analysis. The concentrations increased gradually with decreasing temperature of film deposition to a maximal value of approximately 20% for the lowest temperature (330°C), in agreement with results reported elsewhere.[54]

Some preliminary results from only a few samples were obtained in attempts to determine the oxidation states of the phosphorus in BPSG. Samples deposited at 370°C (RCA rotary reactor) with a very high ratio of O_2/hydride (60:1) were compared with samples deposited at 430°C with a 20:1 ratio. The results of wet-chemical microanalysis are listed in Table 1. They indicate that 23 to 24% of the total P is present as P_2O_3 in the 370°C/60:1 samples, whereas only 0 to 15% was detected in the 430°C/20:1 samples. No P_2O_3 was found after glass flow in a sample that originally had 0.3% P_2O_3. Clearly, more work is needed in this complex area[60] to ascertain the exact glass composition and its implications.

The concentration of boron and phosphorus through the thickness of the films is quite uniform. A typical profile obtained by secondary ion mass spectroscopy (SIMS) of a BPSG film deposited on silicon in a continuous CVD reactor (AMS-2100) is presented to demonstrate this uniformity (Fig. 12).

Another consideration concerns the inhibiting or synergistic effects of the hydrides on each other during cooxidation. We showed

Table 1—BPSG Compositions* of Samples Deposited at Low (370°C) and High (420°C) Temperatures with Different O_2/Hydride Ratios

BPSG No.	Temp. (°C)	O_2/Hydride Ratio	Total P (Wt%)	P as P_2O_5 (Wt%)	P as P_2O_3 (Wt%)	P_2O_3/ Total P	B (Wt%)
1	370	60:1	4.4	3.3	1.02	0.23	2.7
2	370	60:1	3.0	2.3	0.73	0.24	3.7
3	430	20:1	4.4	4.1	0.3	0.068	2.7
4	430	20:1	6.9	6.8	0.1	0.015	5.0
5	430	20:1	3.0	0.0	3.0	0.0	4.5
6	430	20:1	2.1	1.9	0.2	0.10	5.0
7	430	20:1	4.3	4.3	0.0	0.0	2.8

* Wet-chemical analyses were carried out by Balasz Analytical Laboratory, Mountain View, CA.

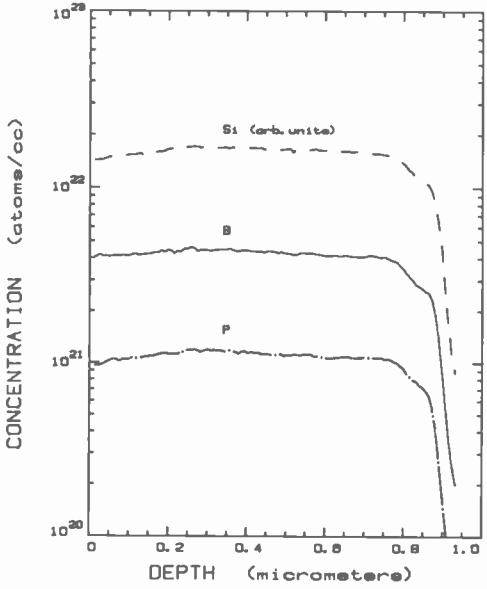


Fig. 12—SIMS profile analysis indicates constancy of boron and phosphorus concentrations through the film thickness. The BPSG was deposited in a continuous APCVD production reactor.

previously[14] that the rate of BPSG film formation differs from that expected on addition of the reactants needed to form the SiO_2, PSG, borosilicate glass (BSG), and borophosphate glass (BPG). Their rates of deposition were obtained by omitting one of the three hydride reactants at a time, but under otherwise comparable CVD conditions. At 430°C, with an O_2/hydride ratio of 20:1, SiO_2 and PSG had 1.5 to 2.1 times higher deposition rates than BPSG types A and B; the deposition rate of BSG was about 0.9 times lower, and that of BPG 0.3 times. These results indicated a strong mutual inhibition of the hydrides during cooxidation, the inhibiting effects of B_2H_6 on SiO_2 and PH_3 being especially strong.

Under the new, optimized CVD conditions at 370°C, with oxygen/hydride ratios of 40:1 and 60:1, the effects were quite different. The deposition rates of BSG were 1.05 to 1.16 times higher than those of BPSG, whereas those of SiO_2, PSG, and BPG were well below 0.3 times as low. These results indicate a very strong enhancement of SiH_4 and PH_3 oxidation by B_2H_6 that has not been reported before. The low deposition rates for SiO_2 and PSG films are readily predictable from the low temperature and high oxygen/hydride ratios.[6,14] While the deposition rate of PSG by itself under these con-

ditions is very low, the phosphorus concentration in the films is considerably higher (by some 40%) than that in the corresponding BPSG. This effect may be due to an oxidation enhancement by B_2H_6 that is greater for SiH_4 than for PH_3. The surprising finding is that these widely varying CVD reactions produce a BPSG of essentially the same elemental composition.

5. Stability of BPSG Films during Processing

Tests were also performed to determine the stability of BPSG films on exposure to high-temperature annealing treatments or immersion in corrosive chemical solutions after densification or fusion.

To test for possible loss of boron oxide or phosphorus oxides from the surface layers of BPSG during heat treatments, films of 800-nm thickness on silicon wafers were etched in P-etch at 25.0°C in 10-sec periods followed by thickness measurements by ellipsometry. The data should result in a linear plot if no changes in film composition occur.[79,80] Samples of BPSG type B were densified at 740°C in nitrogen or steam for 20 min (essentially equivalent to 30 min at 700°C), or fused at 850°C under the same conditions. Precision etch measurements of the critical upper portion of the films are shown in Figs. 13 and 14. No evidence of compositional changes was

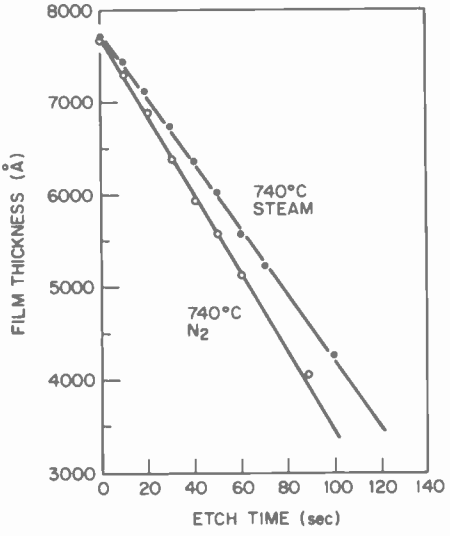


Fig. 13—Film thickness of BPSG type B densified at 740°C in steam or in nitrogen versus etch time in P-etch at 25.0°C. Linearity of curves indicates constant composition within film thickness.

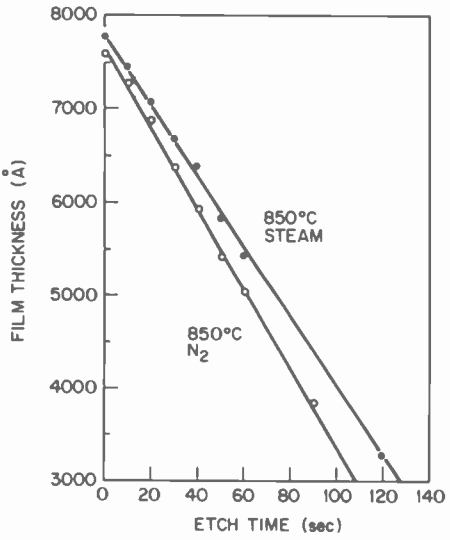


Fig. 14—Same as in Fig. 13, but samples were fused at 850°C.

found by this technique. In both cases, in densifying films, steam treatment was more effective than heating in nitrogen; this was as expected[81] and was also evidenced by the decreased etch rate represented by the slope of the curves (15.7% decrease with steam at 740°C, 13.5% decrease at 850°C).

More precise measurements were made with the same set of samples by use of SIMS profiling. The effect of a mixture of 1 vol. H_2O_2 30%–1 vol. H_2SO_4 98% at 120°C for 15 min followed by a rinse in deionized water at 23°C was also tested as an extreme example of a highly corrosive solution. The results are shown in Figs. 15 and 16, where the concentrations of boron and phosphorus are plotted as a function of layer depth. It can be seen that no depletion near the surface has taken place in the case of the samples heated for 20 min at 740°C in N_2 [Fig. 15(a)], even after the subsequent treatment in H_2O_2–H_2SO_4 [Fig. 15(b)]. Densification at 740°C in steam showed up to 30% depletion of boron in the first 20-nm top layer [Fig. 15(c)]. The corresponding samples fused at 850°C [Figs. 16(a), (b), (c)] show some depletion of the phosphorus and boron in the top 30 nm of the layer. The largest decrease occurred in the boron concentration of the sample fused in steam ambient, decreasing from an average value of 6.0×10^{21} boron atoms per cm^3 to 4.0×10^{21} at the surface. For practical application in flow processing, these relatively small changes have no serious effect.

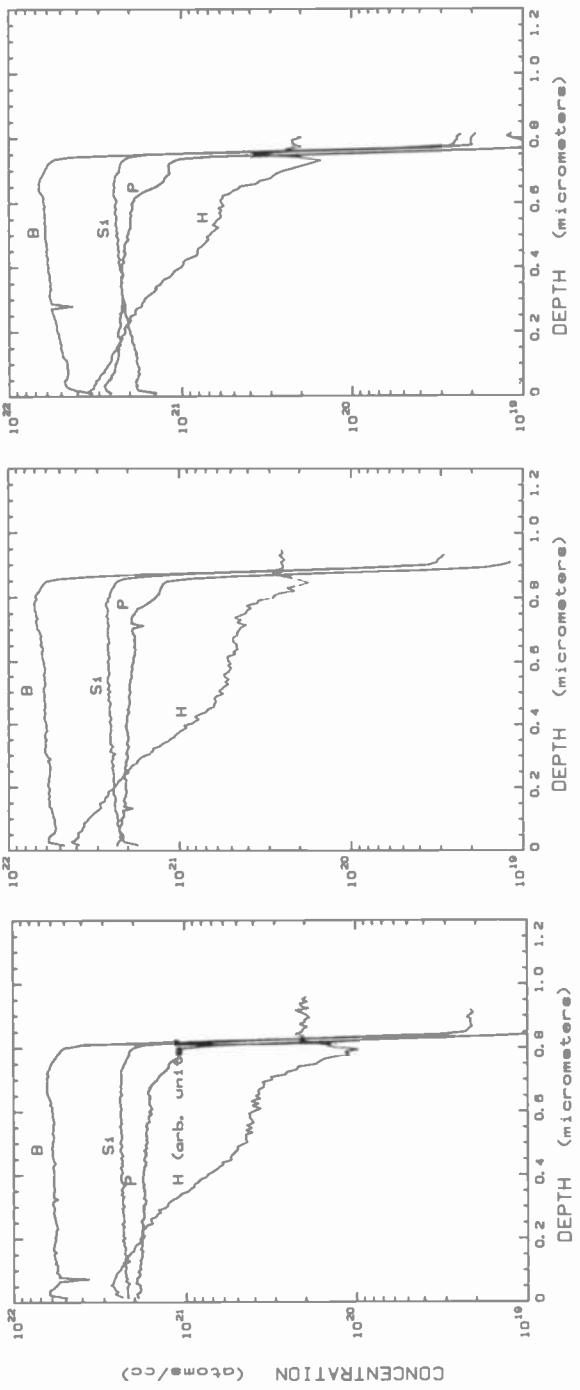


Fig. 15—SIMS profile of a BPSG film on silicon (a) after densification at 740°C in nitrogen, followed by (b) a 15-min immersion in a mixture of conc. H_2O_2–H_2SO_4 at 120°C and subsequent rinsing in deionized water. (c) shows SIMS profile after densification at 740°C in a steam ambient. (The Si and H concentrations are given in arbitrary units.)

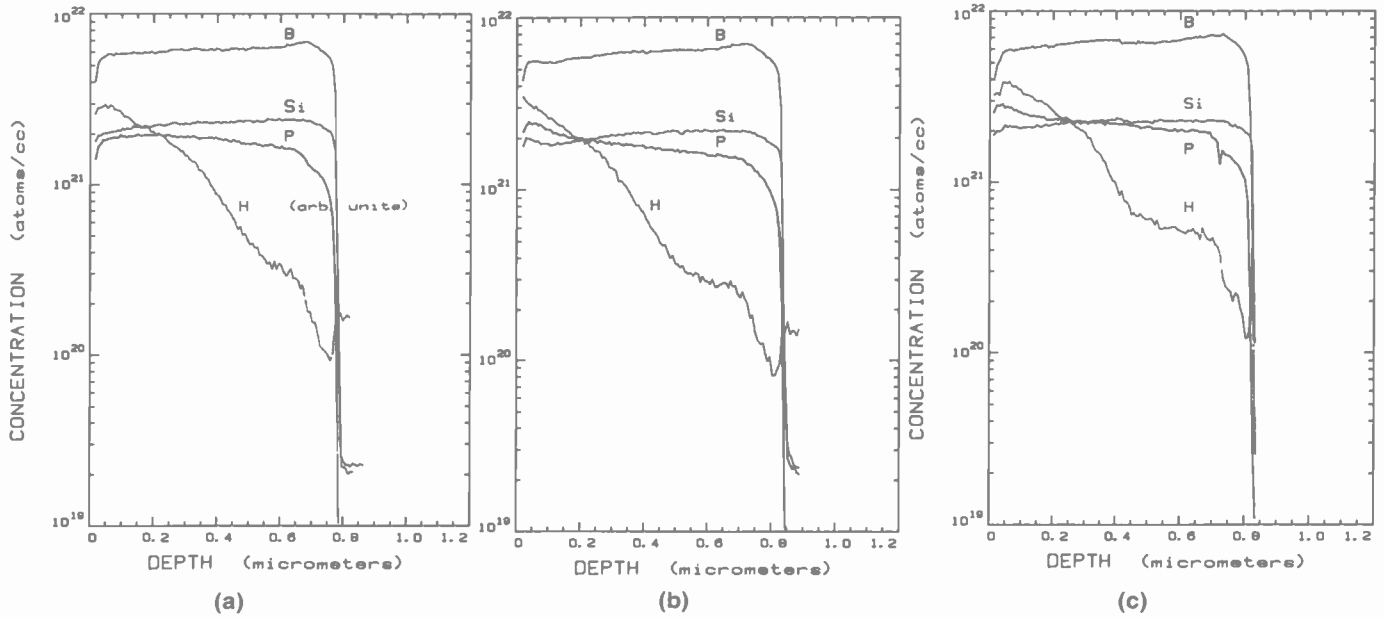


Fig. 16—SIMS profile of a BPSG film on silicon (a) after fusion at 850°C in nitrogen, followed by (b) a 15-min immersion in a mixture of conc. H_2O_2–H_sSO_4 at 120°C and subsequent rinsing in deionized water. (c) shows SIMS profile after fusion at 850°C in a steam ambient. (The Si and H concentrations are given in arbitrary units.)

It should be noted that the hydrogen concentration in these profiles is expressed in arbitrary units; the concentration is only relative, but it is comparable within the sample series. Before heating, the hydrogen concentration in the samples is uniform through the layer, as can be seen from the SIMS profile in Fig. 17, whereas after the heat treatment it decreases toward the silicon interface. (The decrease in the boron concentration near the surface is due to some boron loss during prolonged storage of the sample in room air.) The silicon traces are used for instrument calibration and should not be evaluated numerically. The sloping of the curves near the silicon–glass interface is due to electric charging effects, which are also reflected in the silicon curves. Sloping of the B and P concentrations at and near the surface is due to true depletion. The true surface of the samples is at the point where the curves begin, which (due to a displacement in the recording instrument) is at about 0.01 μm on the SIMS plot.

As-deposited BPSG films should never be exposed to corrosive or aqueous reagent solutions at elevated temperatures because boron oxide and phosphorus oxides would be leached out, similarly as in other vapor-deposited films of silicate glasses, such as PSG[76]. They

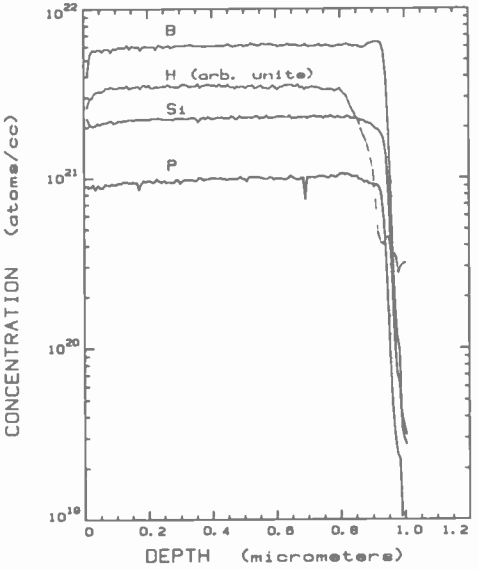


Fig. 17—SIMS profile of a BPSG film on silicon before heating (as-deposited at 430°C, 20:1 O_2/hydride ratio), showing a relatively uniform hydrogen concentration.

should be stored in a dry ambient until they can be densified or flowed.

An extreme example of a BPSG composition with a very high boron concentration was analyzed by SIMS profiling after the undensified sample had been exposed to humid ambient air for several weeks. Prior to analysis, the microcrystals of metaboric acid that had segregated on the surface were removed by rinsing in ethanol. The profile in Fig. 18 indicates depletion of boron in the top 30nm of the film by nearly an order of magnitude. No depletion of phosphorus occurred.

6. Fusion Properties of BPSG

We previously described[14] flow and reflow parameters of BPSG films and demonstrated the substantial decrease in temperature that can be achieved in a steam ambient instead of a dry ambient to produce the same degree of flow. Some additional information is presented in Figs. 19 and 20 where the glass flows obtained in steam, oxygen, and nitrogen are compared. Test samples of wet-chemically etched patterns of 800- to 900-nm-thick thermal SiO_2 on silicon were used. BPSG layers of 700- to 800-nm thickness were deposited and den-

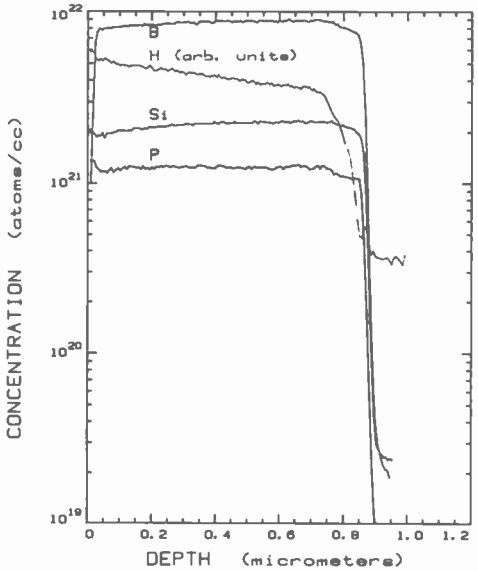


Fig. 18—SIMS profile of a BPSG film with a very high boron concentration on silicon after prolonged storage of the as-deposited sample in room air.

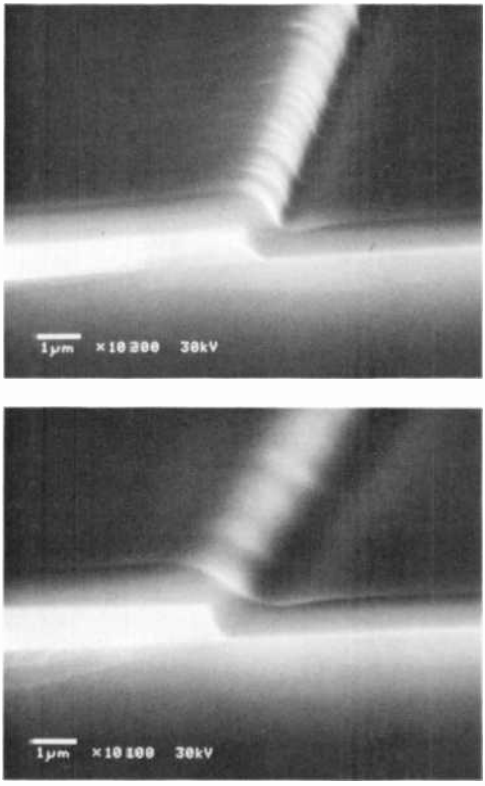


Fig. 19—SEM cross-sectional view showing BPSG deposited over SiO_2 patterns on oxidized Si, heated at 875°C for 30 min (a) in oxygen (slight fusion flow) and (b) in steam (strong fusion flow).

sified at 740°C in nitrogen for 20 min, then heated in a tube furnace at 875°C for 30 min in different ambients. SEM cross-section analysis showed marginal fusion flow in oxygen [Fig. 19(a)], but excellent flow in steam [Fig. 19(b)]. Comparison between oxygen [Fig. 20(a)] and nitrogen [Fig. 20(b)] ambient indicated no difference in the degree of glass flow.

The difference in glass flow obtainable in dry ambient at 875°C and at 950°C can be seen from the cross-section scanning electron micrographs presented in Fig. 21. Interconnect lines of oxidized polycrystalline silicon on thermally oxidized silicon were used as a test device. The surface was overcoated with Si_3N_4 by LPCVD at 800°C before deposition of a 880-nm-thick BPSG film type B. Fig. 21(a) shows the excellent taper contour resulting from flow and reflow at 875°C in oxygen for 20 min each. Fig. 21(b) shows semi-

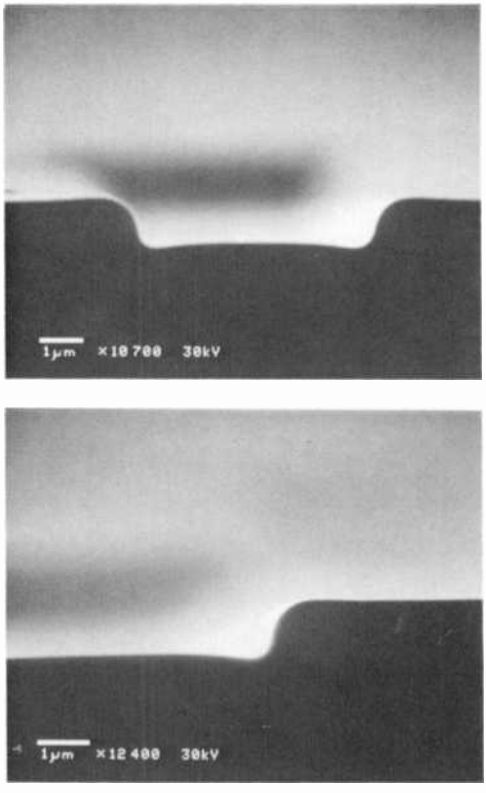
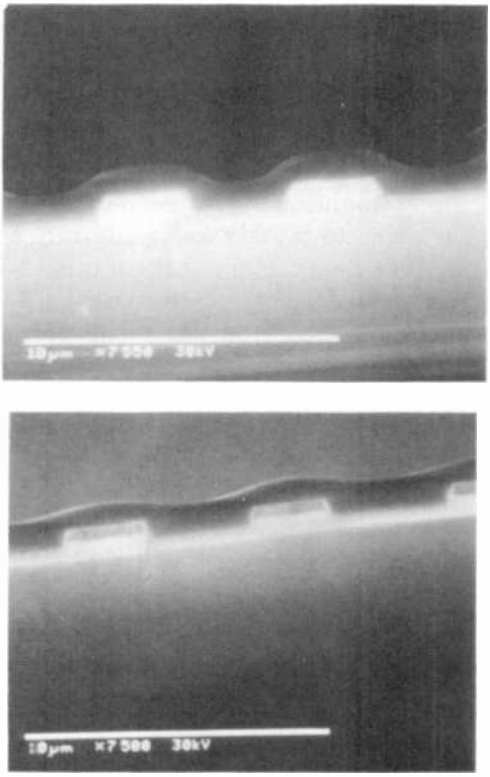


Fig. 20—SEM cross-sectional profile through a via showing a comparably slight flow tapering of BPSG heated at 875°C for 30 min (a) in oxygen and (b) in nitrogen.

planarization resulting from flow at 950°C in nitrogen for 25 min and reflow at 875°C in oxygen for 20 min (to taper plasma-etched via edges). These tests demonstrate the ease with which contours ranging from tapering to planarization can be achieved by relatively small differences in heat treatment. Instead of temperature or ambient changes, the BPSG composition can be varied to achieve the same results.

The time period used for heat treatment at a given flow temperature has a relatively weak effect on the degree of BPSG flow in comparison to changes in temperature. The application of rapid isothermal heating techniques for flow and reflow of BPSG appeared therefore especially promising for substantially reducing the heat exposure time of device wafers during glass fusion. Reduced heat



(a)

(b)

Fig. 21—BPSG flow taper resulting from (a) flow and reflow for 20 min each in oxygen at 875°C and (b) flow at 950°C for 25 min in nitrogen and reflow at 875°C for 20 min in oxygen, resulting in a semi-planarized topography.

exposure further decreases any dopant redistribution that might otherwise still occur.

Numerous investigations have been reported in recent years on the applications of short-duration heating techniques for VLSI,[47,82] several of which are concerned specifically with glass reflow.[43–45,55,83–85] PSG flow by rapid thermal annealing was found to be useful only for films containing high (8 wt%) phosphorus concentrations so that fusion could be achieved in 10 sec at 1000°C to minimize dopant redistribution; furthermore, higher temperatures led to the formation of voids in the glass.[43] Activation of As^+, B^+, and BF_2^+ implants in silicon was investigated by rapid isothermal annealing to determine conditions for minimum redistribution of

dopant atoms while attaining sufficient flow of BPSG for tapering, yielding acceptable results for future VLSI devices.[55]

Experiments conducted in our laboratory were aimed at correlating BPSG flow carried out conventionally in a tube furnace versus flow by isothermal rapid heating. A Heatpulse 210T Rapid Wafer Heating System (AG Associates)[85] was used in this work. This apparatus used a high-intensity, incoherent light produced by tungsten–halogen lamps to rapidly and uniformly heat a wafer to high temperatures. The heating chamber contains upper and lower banks of lamps and is enclosed by water-cooled reflective walls. A quartz isolation tube positions the wafer between the banks and uniformly diffuses the radiation from the lamps. The instrument was provided with an optional precision temperature control. The anneal cycle can be programmed. A wafer temperature of 1200°C can be attained in typically 10 sec.

Test wafers consisting of wet-chemically etched 1200-nm-thick SiO_2 stripes of 6-μm width and 7-μm spaces on thermally oxidized silicon were coated with 1000-nm BPSG of several compositions. Samples were annealed in the Heatpulse 210T system in flowing nitrogen in 50°C-steps from 900 to 1200°C. A standard ramp-up and ramp-down heating rate of 5% per sec of the peak-intensity power setting was used in the thermal cycle, with exposure times of 10 and 30 sec at constant peak temperature. For comparison, samples from the same wafers were heated in a quartz tube furnace in nitrogen ambient for 30 min at 900, 925, and 950°C, respectively. No special heat-up or cooling anneal was used. The heated samples were cleaved perpendicular to the line patterns and examined for the extent of glass fusion by SEM.

Two typical cross-section micrographs are presented in Fig. 22 to demonstrate the glass flows obtainable with the rapid isothermal heating technique. Fig. 23 shows the same degree of tapering resulting from conventional tube-furnace heating at 925°C in nitrogen for 30 min as is obtainable by rapid isothermal heating at 1100°C in nitrogen for just 30 sec (Fig. 22b). A correlation of optimal fusion-tapering conditions for the rapid heating and for the conventional tube-furnace heating was established from the results obtained. Essentially, fusion-equivalent temperatures for exposures in nitrogen at peak temperature for 30 sec in the Heatpulse 210T, and for 30 min in the tube furnace, are as follows: for BPSG type A, 1100°C pulse equals 925°C tube heating; for the intermediate BPSG type AB, 975°C pulse equals 875°C tube heating; and for BPSG type B, 900°C pulse equals 825°C tube heating. No differences were ap-

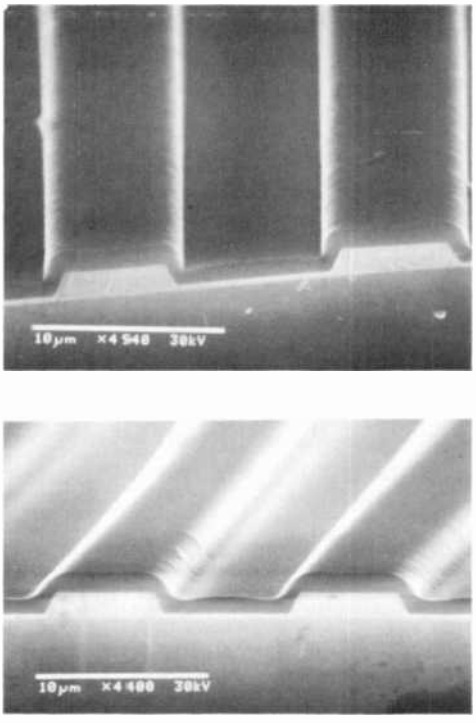


Fig. 22—BPSG type A over patterns of SiO_2 on silicon after (a) a 30-sec heat-pulse treatment at 1000°C in nitrogen, resulting in slight flow tapering, and (b) after 30 sec at 1100°C, resulting in strong flow tapering.

parent in the contour topography or morphology of the fused glass processed by the two techniques under fusion-equivalent conditions. Isothermal rapid heating thus achieves glass-fusion tapering in seconds instead of the minutes required by conventional tube-furnace techniques. Only a relatively small increase in temperature (75–175°C) is needed for 30-sec exposures of BPSG of typical compositions to achieve ideal glass-fusion tapering. As the fusion temperature decreases, a smaller increase in temperature suffices.

Several scanning electron micrographs of integrated circuits are presented in Figs. 24 to 26 to exemplify the contours of flowed BPSG in some actual device structures. Glass fusion of these devices was done by conventional tube-furnace techniques. Topographical and sectional views of a memory array area of a 128K ROM bulk CMOS circuit with nominally 3-μm feature sizes are shown in Figs. 24a and 24b, respectively. The contour-smoothing capability of flowed

BPSG is clearly evident in both micrographs. An example of a Schottky-barrier, infrared image sensor is shown in Fig. 25. This CCD circuit, similar to one described previously,[86] is constructed with two levels of polycrystalline silicon (minimum feature size of 3 μm) and uses flowed BPSG as the interlevel dielectric.[64] The smoothly tapered contours of the BPSG into the contact vias and over the double-polysilicon areas ensure excellent uniformity of the aluminum metallization. A much more severe topography, of a 2-μm CCD test structure, is shown in Fig. 26. Step coverage over depressions and elevations of about 1 μm and over interfeature spacings of as little as 0.5 μm is excellent, as can be seen from the aluminum interconnect lines crossing the polycrystalline conductor patterns. Further reduction of feature sizes and linewidths would eventually require semi-planarization by increasing the flow of BPSG, or by planarization of the topography by one of the several processes now available.[87]

7. Summary and Conclusions

We have shown that in the atmospheric-pressure chemical vapor deposition of BPSG, a lowering of the film-deposition temperature from the conventional 430°C to a 340–380°C range leads to a better utilization of the hydrides and thus a higher yield of BPSG. The increase in conversion yield is manifested by a greater-than-50% increase of the film deposition rate for the same quantity of reactants. Both the increase of the rate of film formation and the decrease of particle generation are largely due to suppression of ho-

Fig. 23—BPSG type A after 30-min tube-furnace heating at 925°C in nitrogen, resulting in strong flow tapering, equivalent to that seen in Fig. 22(b); the sectioned sample was given a brief etch to delineate the SiO_2 pattern.

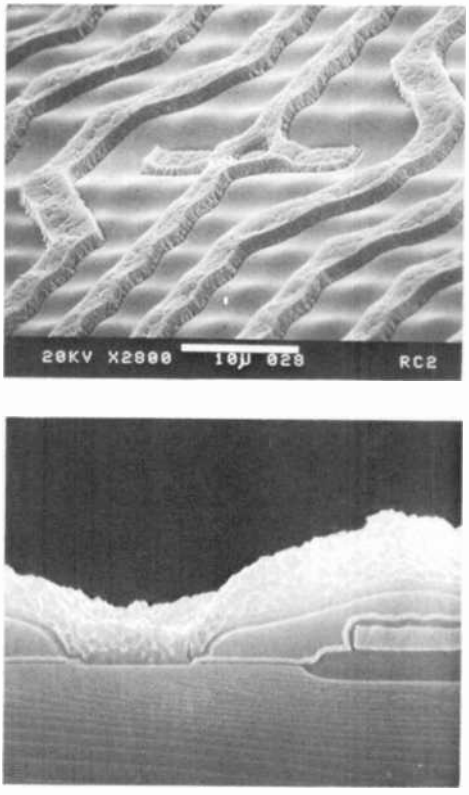


Fig. 24—Scanning electron micrographs of the memory array area of a CMOS 128K ROM: (a) topographical view showing a via and aluminum interconnects and (b) cleaved and demarcation-etched sectional view (approx. 6000×) showing metal on top of BPSG and in contact hole.

mogeneous gas phase nucleation and to enhancement of the heterogeneous CVD reaction; hence, more reactant is converted to glass and less to particles. We prefer to maintain the deposition rate at 100 to 120 nm/min by decreasing the reactant gas flow rates; this produces BPSG films of excellent quality, while allowing good control of the film deposition. In addition, the particle density in the BPSG films is substantially lowered by this change.

Increasing the oxygen-to-hydride ratio from the conventional 20:1 to the range of 40:1 to 60:1 suppresses particle generation still further, but results in some decrease of the rate of film formation. We have chosen a ratio of 40:1 and a nominal reactor-plate temperature of 370°C as a good compromise. The conditions hold for both rotary

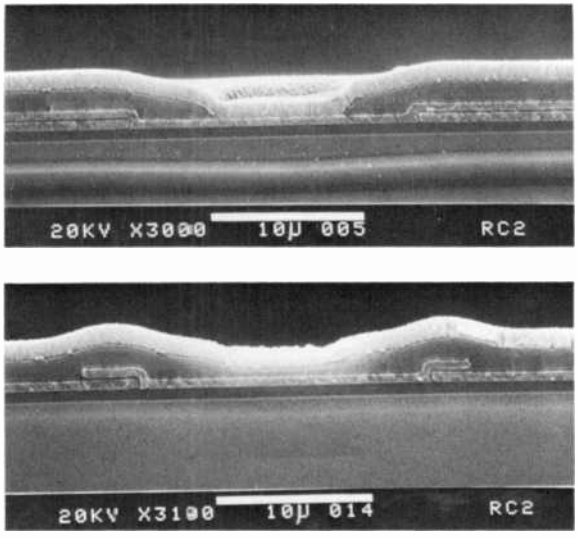


Fig. 25—Sectional scanning electron micrographs from the array perimeter of a Schottky-barrier infrared sensor, with BPSG as the interlevel dielectric. Photos show aluminum coverage over BPSG (a) into a contact hole and (b) over two double-polysilicon structures.

CVD reactors and large continous-production reactor systems, and are noncritical over a wide span in terms of film properties and BPSG composition. The optimized process has been successfully transferred to large-scale production.

The elemental composition of BPSG is not materially altered by these changes in the CVD process, even though the chemical interactions of SiH_4, B_2H_6, PH_3, and O_2 are quite different. There is some indication of an increase in the ratio of P_2O_3/P_2O_5 in the glass; more work is needed in this area. Other important properties, such as flow/reflow characteristics, etch rates, chemical stability, and infrared absorption spectra are not altered in BPSG produced by the optimized deposition process, as long as the samples are thermally densified before analysis.

The stability of BPSG films with respect to chemical composition during processing was investigated by film-profiling techniques. SIMS analysis showed essentially no changes in the boron and phosphorus concentrations after 20-min densification at 740°C in nitrogen, even after immersion for 15 min in a hot H_2O_2–H_2SO_4 mixture. Densification in steam at 740°C showed some decline of the boron concentration in the top 30 nm of the glass, corresponding to about 30% less boron on the surface. The largest change (33%) was

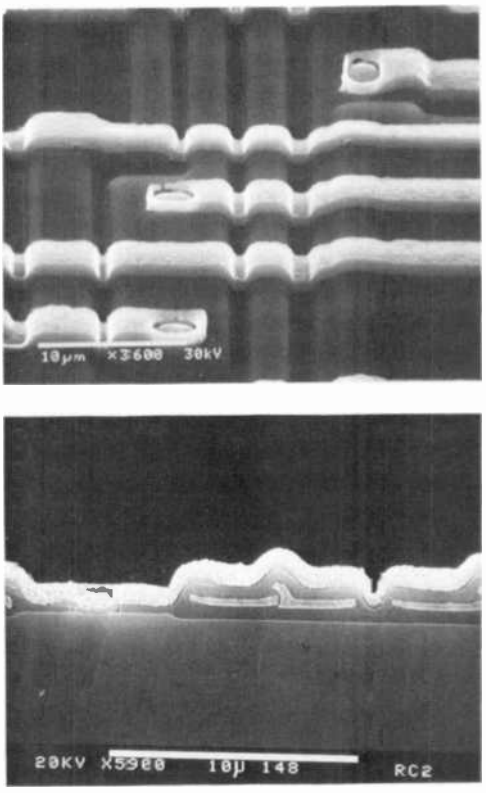


Fig. 26—Two views taken from a 2-μm-geometry CCD test circuit to illustrate metal coverage: (a) topographical view showing excellent fusion tapering with BPSG as the contoured interlevel dielectric over the polysilicon interconnects and (b) cleaved and demarcation-etched cross section showing smooth contours of flowed BPSG.

measured after fusion of the glass at 850°C in steam for 20 min; this produced a decrease of the boron concentration in the top 30-nm layer. For applications in flow-reflow of BPSG, these relatively small decreases in boron concentration have no noticeable effects. They also occur on prolonged storage of undensified samples in room air, especially at high relative humidity on samples of very high boron concentrations. We concluded that BPSG films of normally used compositions are sufficiently stable in practice, as long as they are densified or fused before one applies any wet-chemical treatments. Densification or fusion of as-deposited BPSG films may cause some depletion of boron (or, to a lesser extent, phosphorus) within the top 30 nm of the film thickness.

There is no difference in the degree of flow/reflow of BPSG in a nitrogen or oxygen atmosphere. Steam ambient, however, decreases the viscosity of the glass much more than dry ambients, resulting in a greatly enhanced fusion flow. The difference in equivalent BPSG flow (same taper angle) in dry versus steam ambient is estimated to be about 50°C. Semi-planarization of device topography can be readily achieved by either increasing the flow temperature or by increasing the boron (or phosphorus) concentration in the glass.

Rapid isothermal heating techniques, instead of conventional tube-furnace heating, can be used for glass fusion. An equivalent degree of fusion tapering can be attained in 30 sec of rapid heating versus 30 min of tube heating if the temperature during rapid heating is increased by 75 to 175°C.

Several topographical and sectional scanning electron micrographs of RCA integrated circuits are shown in Figs. 24 to 26 to exemplify typical fusion-tapering applications of BPSG in device technology.

Acknowledgments

The authors wish to thank the following individuals for their contributions to this work: Charles W. Magee and Ephraim M. Botnick for SIMS analyses, Gene Jordan (RCA Solid State Division, Palm Beach Gardens, Fla.) for coordination of production tests and process implementation, Ronald K. Smeltzer and Willard L. Lees, Jr., for SEMs of various integrated circuits, and Richard E. Berger for technical assistance. We are especially indebted to George L. Schnable for contributing to the reference collection, for his constructive comments throughout this work, and for critically reviewing the manuscript. We greatly appreciate the excellent editing of the manuscript by Ralph F. Ciafone and Eva Dukes. Unless otherwise indicated, all of the above-named persons are at RCA Laboratories, Princeton, N.J.

Portions of this work were presented at the Electrochemical Society Meeting in Toronto, Canada, in May 1985 (Ref. 56), and were summarized in the June 1985 issue of *Solid State Technology* (Ref. 64).

References:

[1] W. F. Horn and F. A. Hummel, "IX.-Notes on the System B_2O_3–SiO_2–P_2O_5: I. The BPO_4–SiO_2 Join," *J. Soc. Glass Technol.*, **39**(3), 113T–120T (1955).
[2] W. J. Englert and F. A. Hummel, "X.-Notes on the System B_2O_3–SiO_2–P_2O_5: II. Ternary System," *J. Soc. Glass Technol.*, **39**(3), 120T–127T (1955).

[3] W. Kern, "Silicate Glass Coating of Semiconductor Devices," U.S. Pat. 3,481,781, Dec. 2, 1969. (Patent refers to the synthesis of silicate glasses, including BPSG, by chemical vapor deposition from silane.)
[4] W. Kern and R. C. Heim, "Chemical Vapor Deposition of Silicate Glasses for Use with Silicon Devices, I. Deposition Techniques," *J. Electrochem. Soc.*, **117**, 562–568 (1970).
[5] *Ibid.*, II. Film Properties, *J. Electrochem. Soc.* **117**, 568–573 (1970).
[6] W. Kern and V. S. Ban, "Chemical Vapor Deposition of Inorganic Thin Films," in J. L. Vossen and W. Kern, Eds., *Thin Film Processes*, Academic Press, New York, 1978, pp. 257–331.
[7] W. F. Horn and F. A. Hummel, "Progress Report on the System BPO_4–SiO_2," *Trans. J. Brit. Ceram. Soc.*, **78**(4), 77–80 (1979). See also W. F. Horn and F. A. Hummel, "The System BPO_4–$AlPO_4$–SiO_2 at 1200°C," *Trans. J. Brit. Ceram. Soc.* **79**, 109–111 (1980).
[8] T. Edahiro et al., "Manufacture of Optical Fibers for Communication Use," Nippon Telegraph and Telephone Public Corp., Fujikura Cable Works, Ltd. Japan. Pat.: Japan Kokai Tokkyo Koho: JP 79134127 (79/10/18); Application: JP 7839611 (78/04/04). **CA:** 92(14)115246b.
[9] F. A. Hummel and W. F. Horn, "The Quaternary System B_2O_3–Al_2O_3–SiO_2–P_2O_5: I. Literature Review and Exploratory Data on the Ternary Subsystems," *J. Australian Ceramic Soc.* **17**(1), 25–32 (1981).
[10] W. F. Horn and F. A. Hummel, "The Quaternary System B_2O_3-Al_2O_3-SiO_2-P_2O_5: II. Exploratory Data," *J. Australian Ceramic Soc.* **17**(1) 33–36 (1981).
[11] K. A. Szwejda, "Formulation of the Dependence of Viscosity on the Chemical Constitution and Temperature for Glasses in the Silicon Dioxide–Phosphorus Pentoxide–Boron Oxide–Aluminum Oxide–Sodium Oxide System," Szilikatip. Szilikattud. Konf., (Elöadások), 13th; *Proc.*, Vol. **1**, 194–200 (1982). OMKDK, Budapest, Hungary; **CA:** 97(12)96998h.
[12] W. Kern, "Chemical Vapor Deposition of Inorganic Glass Films," *Semiconductor Int.* **5**(3), 89–103 (March 1982).
[13] C. Ramiller and L. Yau, "Borophosphosilicate Glass for Low Temperature Reflow," paper presented at SEMCON/WEST 82', May 25, 1982, San Mateo, CA; *Tech. Program Proc.*, pp. 29–37.
[14] W. Kern and G. L. Schnable, "Chemically Vapor-Deposited Borophosphosilicate Glasses for Silicon Device Applications," *RCA Rev.* **43**(3), 423–457 (1982).
[15] Toshiba Corp., "Taper Etching," Japan. Pat.: Japan Kokai Tokkyo Koho: JP 8283034 A2 (82/05/24); Application: JP 80159263 (80/11/12). **CA:** 97(20)173506a.
[16] L. Napoli, R. K. Smeltzer, R. Donnelly, and J. Yeh, "A Radiation-Hardened 256 × 4 Bulk CMOS RAM," *RCA Rev.* **43**(3), 458–463 (1982).
[17] L. S. Napoli, R. K. Smeltzer, J. L. Yeh, and W. F. Heagerty, "CMOS/SOS 4K RAMs Hardened to 100 Krads(Si)," *IEEE Trans. Nucl. Sci.* **NS-29**, 1707–1711 (1982).
[18] J. L. Yeh and R. K. Smeltzer, "A High Density, High Yield, Radiation Hardened, Buried Contact CMOS/SOS Technology," presented as a poster paper at IEEE Nuclear and Space Radiation Effects Conf., Las Vegas, NV, July 20–22, 1982.
[19] L. Napoli, R. Smeltzer, J.Yeh, and W.Heagerty, "A 200K Rad(Si) 150 nsec 5 Volt CMOS/SOS 4K RAM," paper presented at IEEE Nuclear and Space Radiation Effects Conf., Las Vegas, NV, July 20–22, 1982.
[20] W. C. Benzing, "Shrinking VLSI Dimensions Demand New Interconnection Materials," *Electronics* **55**(17), 116–119 (Aug. 25, 1982).
[21] D. W. Flatley and T. Hsu, "Process for Tapering Openings in Ternary Glass Coatings," U.S. Pat. 4,349,584, Sept. 14, 1982.
[22] H. Veloric, R. Denning, G. Schnable, and J.Yeh, "Reliability of Silicon-On-Sapphire Integrated Circuits," paper presented at the 1982 IEEE SOS/SOI Technology Workshop, Provincetown, MA, Oct. 5–7, 1982.
[23] G. I. Artamonova et al., "Low Melting Glass," USSR Pat.: SU 975620 A1 (82/11/23); Application: SU 3288719 (81/05/07); **CA:** 98(18) 148467b.
[24] R. G. Downing, R. F. Fleming, J. K. Langland, and D. H. Vincent, "Neutron Depth Profiling at the National Bureau of Standards," *Nucl. Instrum. Methods Phys. Res.* (Netherlands) **218**(1–3), 47–51 (Dec. 1983).
[25] C. W. Leung, H. Dawson, M. Blumenfeld, and D. Biondi, "Low Temperature Elevated Pressure Glass Flow/Reflow Process," U.S. Pat. 4,420,503, Dec. 13, 1983.

[26] A. C. Adams, "Dielectric and Polysilicon Film Deposition," Chap. 3, in S. M. Sze, Ed., *VLSI Technology*, McGraw-Hill, New York, 1983, pp. 93–129 (p. 124).
[27] I. Avigal, "Inter-Metal Dielectric and Passivation-Related Properties of Plasma BPSG," *Solid State Technol.* **26**(10), 217–224 (1983); see also "Errata," *Solid State Technol.*, **27**(2), 123, 139 (1984).
[28] G. L. Schnable and E. A. James, "Doped-Oxide Diffusion of Phosphorus Using Borophosphosilicate Glass," U.S. Pat. 4,433,088, Feb. 21, 1984.
[29] S. Shanfield and S. Bay, "Process Characterization of PSG and BPSG Plasma Deposition," *J. Electrochem. Soc.* **131,** 2202–2203 (1984).
[30] J. L. Vossen, Jr., "Technique for Planarization of Multilevel Metalization for Semiconductor Devices," *RCA Tech. Notes*, n1348, April 30, 1984.
[31] R. Kleppinger, "Laser-Based Defect Particle Counter Used in LSI Production," *RCA Engineer* **29**(2), 53–59 (March/April 1984).
[32] J. E. Tong, K. Schertenleib, and R. A. Carpio, "Process and Film Characterization of PECVD Borophosphosilicate Films for VLSI Applications," *Solid State Technol.* **27**(1), 161–170 (1984).
[33] Nippon Electric Co., Ltd., "Insulation of Semiconductor Wiring," Japan Pat.: Japan Kokai Tokkyo Koho: JP 8454242 A2, JP 5954242 (84/03/29); Application: JP82164452 (82/09/21). **CA:** 101(10)82685t.
[34] P. H. Singer, "Techniques of Low Pressure Chemical Vapor Deposition," *Semiconductor Int.* **7**(5), 72–77 (1984); see also "CVD Equipment," 1985 *Master Buying Guide*, *Semiconductor Int.* **7**(13), 82–89 (1984).
[35] K. Krishnan, "Simultaneous Determination of the Boron and Phosphorus Content in Silicate Glasses by FT-IR Spectroscopy," in *Semiconductor Processing*, STP 850, D. C. Gupta, Ed., ASTM, Nov. 1984, pp. 358–363.
[36] Toshiba Corp., "Passivation of Integrated-Circuit Wirings," Japan. Pat.: Japan Kokai Tokkyo Koho: JP 84181021 A2, JP 59181021 (84/10/15); Application: JP 8354310 (83/03/30). **CA:** 102(14)123978g. Toshiba Corp., "Integrated Circuits," Japan. Pat.: Japan Kokai Tokkyo Koho: JP 84181646 A2, JP 59181646 (84/10/16); Application: JP 8355988 (83/03/31). **CA:** 102(14)124090y.
[37] Toshiba Corp., "Insulation of Conductor Wiring," Japan. Pat.: Japan Kokai Tokkyo Koho: JP 84181648 A2. JP 59181648 (84/10/16); Application: JP 8356057 (83/03/31). **CA:** 102(14)124005z.
[38] W. Kern, "Borophosphosilicate Glass Films for VLSI Applications," *Microfabrication Technol. Rev. '84*, Annual Symp., New England Chapter, AVS; Nashua, NH, June 11–12, 1984.
[39] S. Gottesfeld and L. Gibbons, "Reliability Characterization of High-Speed CMOS Logic ICs," *RCA Rev.* 45(2), 179–193 (1984).
[40] NEC Corp., "Integrated Circuits," Japan. Pat.: Japan Kokai Tokkyo Koho: JP 84106133 A2, JP 59106133 (84/06/19); Application: JP 82216508 (82/12/09): **CA:** 101(24)220907v.
[41] Toshiba Corp., "Smooth-Surfaced MOS Integrated Circuits," Japan. Pat.: Japan Kokai Tokkyo Koho: JP 84126637 A2, JP 59126637 (84/07/21). **CA:** 102(4)38052f.
[42] P. K. Chu, "Quantitative Depth Profiling of Boron and Phosphorus in Borophosphosilicate Glass," *Chem. Phys.* (Springer Ser.) **36,** 332–335 (1984).
[43] J. S. Mercier, R. P. Beerkens, I. D. Calder, and H. M. Naguib, "Rapid Thermal Reflow of PSG Films," *Electrochem. Soc. Ext. Abstr.* **84-2,** 607–608 (1984).
[44] H. Matsui and M. Voshimaru, *Proc. SEMI Technol. Symp. '84*, pp. 1–37.
[45] M. Yoshimaru, H. Onoda, and H. Matsui, "Glass Flow of BPSG Film by Lamp Annealing," paper presented at the 1984 Fall Meeting, Japan Soc. Appl. Phys.; *Japan. Semiconductor Technol. News,* **3**(6), 42 (Dec. 1984).
[46] H. Veloric, M. P. Dugan, W. Morris, R. Denning, and G. L. Schnable, "Reliability of CMOS/SOS Integrated Circuits," *RCA Rev.* 45(2), 230–248 (1984).
[47] S. R. Wilson, W. M. Paulson, and R. B. Gregory, "Rapid Annealing Technology for Future VLSI," *Solid State Technol.* **28**(6), 185–190 (1985).
[48] B. Gorowitz, T. B. Gorczyca, and R. J. Saia, "Applications of Plasma Enhanced Chemical Vapor Deposition in VLSI," *Solid State Technol.* **28**(6), 197–203 (1985).
[49] G. Smolinsky and T. P. H. F. Wendling, "Measurements of Temperature Dependent Stress of Silicon Oxide Prepared by a Variety of CVD Methods," *J. Electrochem. Soc.*, **132,** 950–954 (1985).
[50] S. L. Chang, K. Y. Tsao, M. A. Meneshian and H. A. Waggener, "Spin-on BPSG and Its Application to VLSI," *Proc. Third Int. Symp. VLSI Sci. and Technol./1985*,

85-5, 231–236 (1985). W. M. Bullis and S. Broydo, Eds., The Electrochemical Soc., Pennington, NJ.
[51] R. A. Levy and K. Nassau, "Characterization of Borophosphosilicate and Germanophosphosilicate Glass for VLSI Applications," *J. Electrochem. Soc.* **132,** 409–415 (1985).
[52] P. K. Chu and S. L. Grube, "Quantitative Determination of Boron and Phosphorus in Borophosphosilicate Glass by Secondary Ion Mass Spectrometry," *Anal. Chem.* **57,** 1071–1074 (1985).
[53] H. Nozawa, J. Matsunaga, and K. Hashimoto, "Submicron CMOS Technology and Process Integration for VLSI Static Memories," *Proc. Third Int. Symp. VLSI Sci. and Technol./1985,* **85-5,** 59–70 (p. 69), (1985), W. M. Bullis and S. Broydo, Eds., The Electrochem Soc., Pennington, NJ.
[54] C. Dornfest, "The Effect of Reducing Deposition Temperature in an Atmospheric Pressure BPSG Process," *Electrochem. Soc. Ext. Abstr.* **85-1,** 347 (May 1985).
[55] H. Nishimura, Y. Suizu, and T. Tsujimaru, "A Rapid Isothermal Annealing for VLSI Interconnection Technology," *Electrochem. Soc. Ext. Abstr.* **84-2,** 760–761 (1984).
[56] W. Kern, "Recent Progress in BPSG Technology," Extended Recent Newspaper, presented at the Electrochemical Society Meeting, Toronto, Canada, May 1985. *J. Electrochem. Soc.* **132**(6), 222C, RNP754 (1985).
[57] "Reduced Particulates in BPSG Deposition," Applied Materials, Inc., Announcement, Santa Clara, CA, May 20, 1985.
[58] N. Armour, R. Kleppinger, R. Morey, and K.Pitts, "Statistical Process Control for LSI Manufacturing: What the Handbooks Don't Tell You," *RCA Engineer,* **30**(3), 44–53 (May/June 1985).
[59] R. L. Kleppinger, "Wafer Particle Reduction in LSI Production," *RCA Engineer,* **30**(3), 60–64 (May/June 1985).
[60] J. Houskova, K.-K. N. Ho, and M. K. Balazs, "Characterizing Plasma Phosphorus-Doped Oxides," *Semiconductor Int.,* **8**(5), 236–241 (1985).
[61] *Semiconductor Int.* **8**(5), 162–163 (Advertisement: Watkins-Johnson, "BPSG Progress Report") and 444 (Advertisement: Anicon Inc., "CVD System") (1985).
[62] T. Foster, G. Hoeye, and J. Goldman, "A Low Pressure BPSG Deposition Process," *J. Electrochem. Soc.* **132,** 505–507 (1985).
[63] K. Nassau, R. A. Levy, and D. L. Chadwick, "Modified Phosphosilicate Glasses for VLSI Applications," *J. Electrochem. Soc.* **132,** 409–415 (1985).
[64] W. Kern and R. K. Smeltzer, "Borophosphosilicate Glasses for Integrated Circuits," *Solid State Technol.* **28**(6), 171–179 (1985).
[65] W. Kern, "Deposited Dielectrics for VLSI," *Semiconductor Int.* **8**(7), 121–129 (July 1985).
[66] V. Teal, C. Ha, and C. Chowaniec, "BPSG as Interlevel Dielectric for Double-Level Metal Process," Second Int. IEEE VLSI Multilevel Interconnection Conf., Santa Clara, CA, June 24–26, 1985; *Conf. Proc.,* pp. 273–279.
[67] C. J. Dell'Oca, "Interconnect Technologies for VLSI," Second Int. IEEE VLSI Multilevel Interconnection Conf., Santa Clara, CA, June 24–26, 1985; *Conf. Proc.,* pp. 3–14.
[68] G. W. B. Ashwell and S. J. Wright, "The Reflow of Arsenosilicate Glass," Second Int. IEEE VLSI Multilevel Interconnection Conf., Santa Clara, CA, June 24–26, 1985; *Conf. Proc.,* pp. 285–291.
[69] W. Kern, "Inorganic Dielectric Systems," VLSI Multilevel Interconnection Short Course, Second Int. IEEE VLSI Multilevel Interconnection Conf., Santa Clara, CA, June 24–26, 1985; in *Visuals Booklet,* coordinated by T. E. Wade, pp. 213–320.
[70] K. F. Jensen and W. Kern, "Chemical Vapor Deposition for Solid-State Device Fabrication," Electrochemical Society Short Course, announcement for Las Vegas meeting of the Society in Oct. 1985; *J. Electrochem. Soc.,* **132**(7), 292C (July 1985).
[71] W. Kern, "Chemical Vapor Deposition for Electronics," American Vacuum Society Short Course, *Course Notebook,* Editions 1981 through 1985.
[72] F. S. Becker and D. Pawlik, "A New LPCVD Borophosphosilicate Glass Process Based on the Doped Deposition of TEOS-Oxide," submitted for presentation at the Electrochem. Soc. meeting, Las Vegas, NV (Oct. 1985).
[73] K. C. R. Chiu, W. R. Snow, and S. V. U. Dunton, "Modeling of BPSG Film Deposition," submitted for presentation at the Electrochem. Soc. meeting, Las Vegas, NV (Oct. 1985).

[74] M. Susa, Y. Hiroshima, K. Senda, T. Kuriyama, S. Matsumoto, S. Terakawa, and T. Takamura, "Borophosphosilicate Glass Flow for Solid State Imager Application," submitted to *J. Appl. Phys.*, May 1985.
[75] A. J. Learn and W. Baerg, "Growth of Borosilicate and Borophosphosilicate Films at Low Pressure and Temperature," paper submitted to *J. Thin Solid Films*, 1985; see also: "Full Spectrum CVD System," *Solid State Technol.*, **26**(12), 61–62 (1983).
[76] W. Kern, G. L. Schnable, and W. A. Fisher, "CVD Glass Films for Passivation of Silicon Devices: Preparation, Composition, and Stress Properties," *RCA Rev.*, **37**(1), 3–54 (1976).
[77] W. Kern, "Chemical Vapor Deposition for Glass Passivation of Integrated Circuits," *Solid State Technol.*, **18**(12), 25–33 (1975).
[78] W. C. Benzing, R. S. Rosler, and R. W. East, "A Production Reactor for Continuous Deposition of Silicon Dioxide," *Solid State Technol.*, **16**(11), 37–42 (1973).
[79] W. Kern, "Chemical Etching of Dielectrics," in H. G. Hughes and M. S. Rand, Eds., *Proc. Symp. Etching for Pattern Definition*, The Electrochem. Soc., Princeton, NJ, 1976, pp. 1–18d.
[80] W. Kern, "Analysis of Glass Passivation Layers on Integrated-Circuit Pellets by Precision Etching," *RCA Rev.*, **37**(1), 78–106 (1976).
[81] W. Kern, "Densification of Vapor-Deposited Phosphosilicate Glass Films," *RCA Rev.*, **37**(1), 55–77 (1976).
[82] D. F. Downey, C. J. Russo, and J. T. White, "Activation and Process Characteristics of Infrared Rapid Isothermal and Furnace Annealing Techniques," *Solid State Technol.*, **25**(9), 87–93 (1982).
[83] T. Hara, H. Suzuki, and M. Furukawa, "Reflow of PSG Layers by Halogen Lamp Short Duration Heating Technique," *Japan. J. Appl. Phys.*, **23**(7), Part 2, L452–L454 (July 1984).
[84] J. Kato and S. Iwamatsu, "Rapid Annealing Using Halogen Lamps," *J. Electrochem. Soc.*, **131**(5), 1145–1154 (1984).
[85] A. Gat and S. Shatas, "Introduction to Heatpulse Processing Technology," AG Associates (1982), pp. 1–14.
[86] W. F. Kosonocky and H. Elabd, "Schottky-Barrier Infrared Charge-Coupled Device Focal Plane Arrays," *Proc. SPIE*, **443**, 167–188 (1983).
[87] See, for example: *1985 Proc.*, Second Int. IEEE VLSI Multilevel Interconnection Conf., Santa Clara, CA, June 24–26, 1985 (IEEE Catalog No. 85CH2197-2); and Parts B and C, *VLSI Electronics—Microstructure Science*, Vol. 8, N. Einspruch and D. Brown, Eds., Academic Press, New York, 1985.

The Limitation of Short-Channel-Length N^+-Polysilicon-Gate CMOS ICs

S. T. Hsu
RCA Laboratories, Princeton, NJ 08540

Abstract—Submicrometer-channel-length n^+-polysilicon- or n^+-polycide-gate CMOS FETs exhibit relatively large subthreshold leakage current and small drive currents. An ideal gate material for CMOS VLSI should have a work function that is larger than the electron affinity energy of silicon by one half the forbidden-gap energy of silicon. Tungsten is one of the better gate materials for submicrometer-feature-size CMOS VLSI. Compared to an n^+-polysilicon-gate CMOS IC, a tungsten- or tungsten silicide-gate CMOS IC is expected to have higher speed, smaller leakage current, and a simpler fabrication process. This paper describes the reasons for the limitations on the use of n^+ polysilicon and n^+ polycide as gate materials for submicrometer-gate-length MOS ICs.

1. Introduction

Perhaps the most important contribution to MOS VLSI technology has been the development of the principles for scaling the physical dimensions of MOSFETs.[1] Scaling theory, which has been closely followed by scientists and engineers worldwide, was developed for n-channel MOSFETs but can be applied to p-channel MOSFETs by using a complementary material for the gate electrode. If n^+ polysilicon is used as the gate electrode of n-channel MOSFETs, p^+ polysilicon should be used as the gate electrode of p-channel devices.

Unfortunately, it is very difficult to prevent boron ions from penetrating through the gate oxide to the channel region of a p-channel MOSFET. Therefore, n^+ polysilicon is used exclusively for both n- and p-channel MOSFETs. For ICs with dimensions in the submi-

crometer range, however, it is not possible to fabricate p-channel MOSFETs with n^+ polysilicon gates using state-of-the-art technology. A new material is therefore needed for the gate electrode of submicrometer-feature-size CMOS ICs.

2. The Limitations of n^+ Polysilicon-Gate MOSFETs

N^+ polysilicon is an ideal gate material for a CMOS integrated circuit with relatively large physical dimensions. Its conductivity is on the order of 10 to 20 ohms/sq and the fabrication process is simple. However, it is not a good material for use as the gate electrode in ICs with channel lengths in the submicrometer range. The drive current of an n^+-polysilicon-gate n-channel MOSFET is relatively small and the junction capacitance is relatively large. It is virtually impossible to fabricate a low-leakage-current high-performance n^+ polysilicon-gate p-channel MOSFET with submicrometer dimensions.

2.1 N-Channel MOSFET

The energy-band diagram of the channel region of an n^+-polysilicon-gate n-channel MOSFET is shown in Fig. 1. Fig. 1a is the energy-

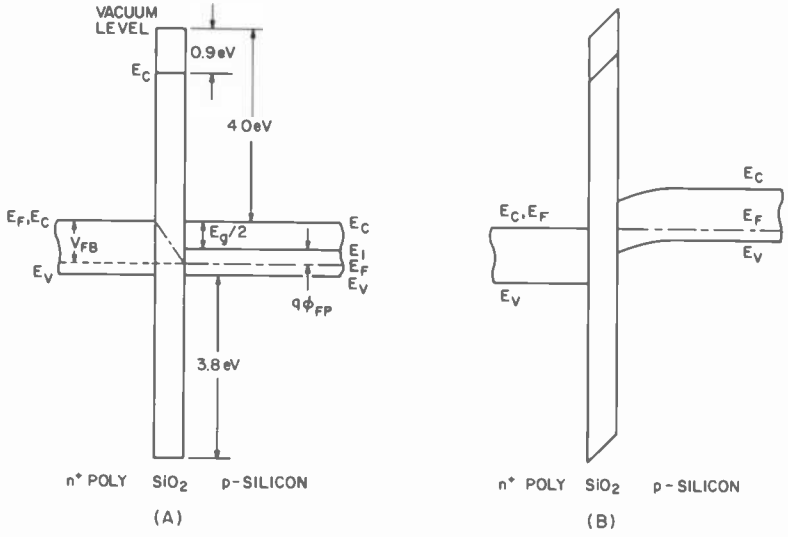


Fig. 1—Energy band diagrams of a n^+-polysilicon-gate n-channel MOSFETs (a) at flat-band condition and (b) at equilibrium condition.

band diagram of the channel region of the MOSFET at the flat-band condition and Fig. 1b is the energy-band diagram when the MOS structure is at the equilibrium condition. We have assumed that the forbidden gap of n^+ polysilicon is equal to the forbidden gap of single-crystal silicon. The flat-band voltage of this MOS structure is, therefore, equal to one half the bandgap plus the Fermi energy of the substrate and the effect of the fixed charge, Q_{ss}. The flat-band voltage of this structure is, therefore, approximately equal to twice the Fermi potential of the p-type silicon substrate. This large flat-band voltage implies that the substrate has to be heavily doped in order for the device to have a small drain-leakage current at small gate biases. From classical one-dimensional theory,[2] it can be shown that the drain current of a device made on a heavily doped substrate is smaller than that of a device made on a lightly doped substrate. An n^+-polysilicon-gate n-channel MOSFET, therefore, exhibits relatively small drain current and large junction capacitances.

When the gate bias voltage of a MOSFET is smaller than the threshold voltage of the device, the drain current of the MOSFET increases approximately with the exponential of the gate bias voltage. This subthreshold current is characterized by the subthreshold swing, S, defined as[3]

$$S = \ln 10 \frac{dV_G}{d(\ln I_D)}. \qquad [1]$$

For an enhancement-type MOSFET, S is given by[3]

$$S \simeq \frac{kT}{q} \ln 10 \left(1 + \frac{\epsilon_0 \epsilon_{Si}}{X_{de} C_{ox}} \right), \qquad [2]$$

where ϵ_{Si} is the dielectric constant of silicon, C_{ox} is the unit-area gate-oxide capacitance, and X_{de} is the width of the channel depletion region of enhancement-type MOSFETs.

According to scaling principles,[1] this subthreshold swing is practically independent of the scaling factor. Therefore, the subthreshold drain current of the MOSFET increases with the downward scaling of the threshold voltage. A submicrometer-gate-length n^+-polysilicon-gate n-channel MOSFET with properly scaled threshold voltage is expected to exhibit large subthreshold leakage current and is not an optimum high-performance low-leakage-current device.

2.2 P-Channel MOSFET

The energy-band diagram of an n^+-polysilicon-gate p-channel MOSFET is shown in Fig. 2. Fig. 2a is the energy-band diagram when the MOS structure is in the flat-band condition. Fig. 2b is the energy-band diagram when the MOS is in the equilibrium condition. The flat-band voltage is equal to the difference in potential between the edge of the conduction band and the Fermi potential of the n-type silicon substrate plus the effect of the fixed interface charge, Q_{ss}. Because of this negative flat-band voltage, the threshold voltage of an n^+-polysilicon-gate p-channel MOSFET is very large. A low-energy boron ion implantation is used to adjust the threshold voltage of a small-physical-dimension n^+-polysilicon-gate p-channel MOSFET. This boron ion implantation forms a thin p-type layer at the surface of the n-type silicon substrate (or n-well). A typical simulated impurity profile of an n^+-polysilicon-gate p-channel MOSFET is shown in Fig. 3. An additional deep phosphorus ion implantation was used to reduce the depth of the surface p-layer.[4]

For the sake of simplicity, we shall assume that the concentrations of donors and acceptors in the channel region are constant. Fig. 4 shows the space-charge distribution, the field intensity, and

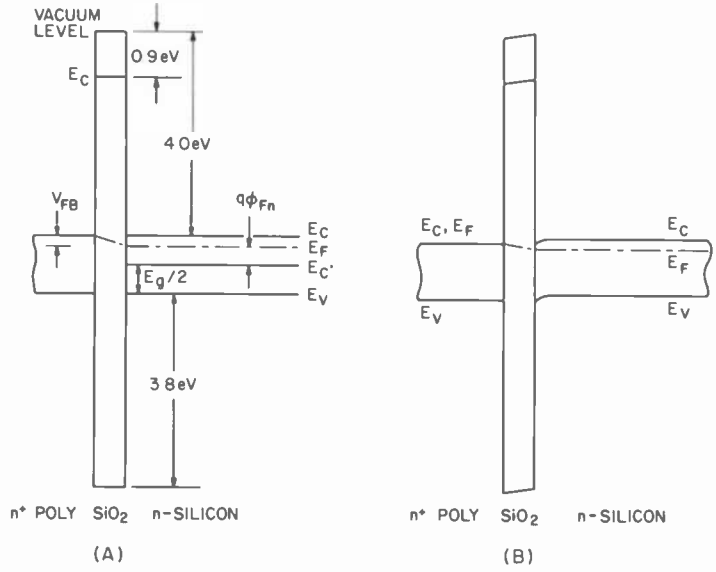


Fig. 2—Energy band diagrams of an n^+-polysilicon gate p-channel MOSFET (a) at flat-band condition and (b) at equilibrium condition.

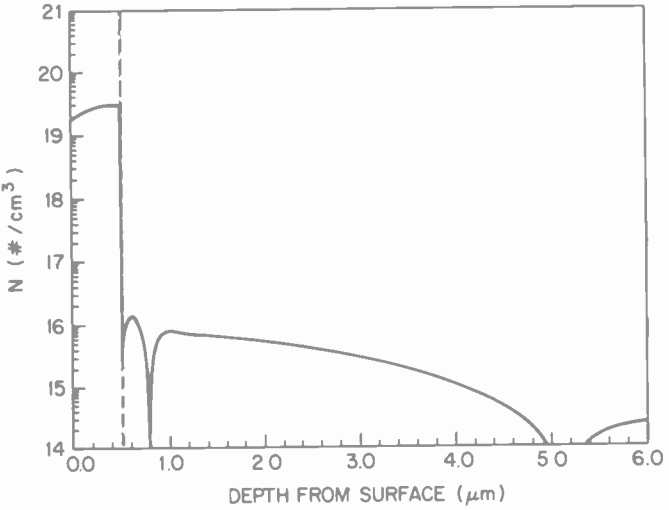


Fig. 3—Simulated impurity profile at the channel region of a p-channel MOSFET.

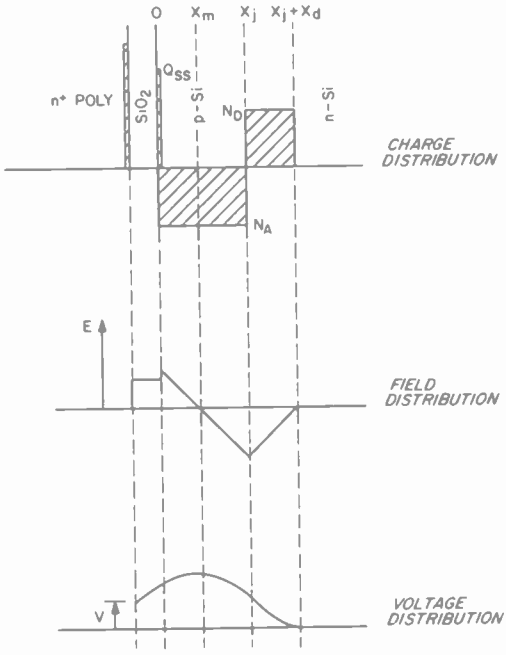


Fig. 4—Space-charge distribution, field intensity, and the potential profile at the channel region of an n^+-polysilicon-gate p-channel MOSFET.

the potential profile at the channel region of the MOSFET when the gate voltage is smaller than the threshold voltage of the device. The depth of the surface p-layer is equal to X_j; the width of the depletion region in the n-substrate is equal to X_d; and the acceptor and the donor densities are equal to N_A and N_D, respectively. Solving the Poisson's equation with the boundary conditions of zero voltage and zero field intensity at the substrate and with a gate bias voltage of V_G, the potential in the surface p-layer is given by

$$V_1(X) = \frac{qN_A}{2\epsilon_0\epsilon_{Si}}(X_j - X)^2 - \frac{qN_D}{2\epsilon_0\epsilon_{Si}} X_d(2X_j + X_d - 2_x). \quad [3]$$

The gate voltage, V_G, is given by

$$V_G = V_{FB} + \frac{qN_A}{C_{ox}} X_j + \frac{qN_A}{2\epsilon_0\epsilon_{Si}} X_j^2 - \frac{qN_D X_d}{C} - \frac{qN_D}{2\epsilon_0\epsilon_{Si}} X_d^2 + V_{Df}, \quad [4]$$

where V_{Df} is the diffusion potential of the p-n junction and

$$\frac{1}{C} = \frac{X_j}{\epsilon_0\epsilon_{Si}} + \frac{1}{C_{ox}}.$$

The minimum potential in the surface p-layer, V_{1m}, is

$$V_{1m} = -\frac{qN_D(N_A + N_D)}{2\epsilon_0\epsilon_{Si}N_A} X_d^2. \quad [5]$$

This minimum potential occurs at

$$X_m = X_j - \frac{N_D}{N_A} X_d. \quad [6]$$

Therefore, the net space charge between X_m and X_j is equal to the net space charge in X_d. This phenomenon is independent of the actual charge distribution in the surface p-layer and in the substrate.

The minimum potential barrier height at the p^+ source junction and the channel region is

$$\phi_B = \frac{E_g}{2q} + \phi_{Fn} + V_{1m}, \quad [7]$$

where E_g is the forbidden-gap energy of the semiconductor and ϕ_{Fn} is the Fermi potential of the n-type substrate. The minimum potential barrier, therefore, decreases with the doping concentration of the surface p-layer and with the increase of the gate-bias voltage.

The threshold voltage, V_T, of the device is defined as the gate

voltage required for the quasi Fermi potential at the minimum potential to be equal to the Fermi potential of the neutral p-surface layer. Under this condition, the minimum potential V_{1m} is equal to the diffusion potential of the p-n junction, V_{Df}, or

$$V_T = V_{FB} + \frac{qN_AX_j}{C_{ox}} + \frac{qN_A}{2\epsilon_0\epsilon_{Si}} X_j^2 - \frac{1}{C}\frac{2\epsilon_0\epsilon_{Si}qN_AN_DV_{Df}}{(N_A + N_D)} - \frac{N_AV_{Df}}{(N_A + N_D)} . \quad [8]$$

The subthreshold swing, S, of an n^+-polysilicon-gate p-channel MOSFET is, therefore, given by

$$S = \frac{kT}{q} \ln 10 \frac{N_AC_d}{(N_A + N_D)} \frac{(C_{ox} + C_{Si})}{C_{ox}C_{Si}} = \frac{kT}{q} \ln 10 \frac{N_A}{(N_A + N_D)} \left(1 + \frac{X_j}{X_d} + \frac{\epsilon_0\epsilon_{Si}}{X_dC_{ox}}\right) \quad [9]$$

For a short-channel-length device, the n-type substrate doping density, N_D, and the acceptor density at the surface p-layer, N_A, must be scaled upward to avoid punchthrough in the channel region. The depth of the surface p-layer, X_j, has to be scaled downward to obtain a proper threshold voltage for the device. The subthreshold swing of a short-channel n^+-polysilicon-gate MOSFET is larger for a p-channel type than for an n-channel type. From Eq. [9], it is obvious that the subthreshold swing of the device decreases with the decreasing surface p-layer depth. Consequently the surface p-layer must be very shallow in order for the device to have small subthreshold drain current. The depth of this surface p-layer can be reduced by an additional deep arsenic ion implantation as was proposed by Chiang et al.[4] However, it is very difficult to confine this surface p-layer to shallower than 0.1 μm in practical applications.

3. New Gate Material for Submicrometer CMOS Integrated Circuits

Optimum n- and p-channel device characteristics can be obtained simultaneously with a very simple fabrication process if the electrical properties of the gate material are symmetrical for both n- and p-type silicon. The electron affinity energy of silicon is equal to 4.15 eV.[5] If the interface charge density, Q_{ss}, is small, the ideal material for use as a gate electrode should have a work function of

4.7 eV. The work function of tungsten is approximately 4.6 eV.[6] Therefore, the Fermi level of tungsten is 0.1 eV above the center of the forbidden gap of silicon. This offset in flat-band voltage is compensated by the fixed charges at the silicon–silicon dioxide interface.

Since the work function of a refractory-metal silicide is approximately equal to that of the refractory metal itself, tungsten and tungsten disilicide are among the better gate materials for submicrometer-physical-dimension CMOS integrated circuits. Tungsten technology has been well developed, and both tungsten and tungsten disilicide can be deposited onto silicon and silicon oxide by the LPCVD process.[7] In addition tungsten can be selectively deposited onto bare silicon or bare polysilicon by the LPCVD process.[8,9]

The energy diagram of a tungsten-oxide-silicon system is shown in Fig. 5. The flat-band voltage of a tungsten-gate n-channel MOSFET is approximately 0.4 V smaller than that of an n^+-polysilicon-gate n-channel MOSFET. From classical one-dimensional theory, one can easily show that the drain current of a tungsten-gate n-channel MOSFET at $V_D = V_G = 5$ V and $V_T = 0.7$ V is approximately 20% larger than that of an n^+-polysilicon-gate n-channel MOSFET having the same threshold voltage and operated at the same bias condition. The flat-band voltage of a tungsten-gate p-channel MOSFET is equal to $V_{FN} - 0.1$ V plus the contribution of the fixed surface-state charge. Therefore, a surface p-layer is not required.

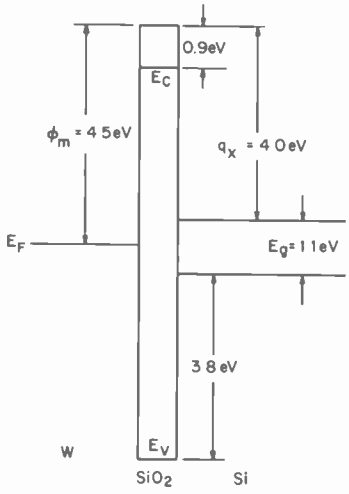


Fig. 5—Energy band diagram of a tungsten-gate MOSFET.

The simulated impurity profiles of a twin-tub tungsten-gate CMOS process are shown in Fig. 6. Fig. 6a shows the p-well impurity profile and 6b the doping profile of the n-well. The substrate is p-type silicon doped with $8 \times 10^{14}/cm^3$ acceptors. The depth of the n-well is equal to 4.2 μm. A deeper-well CMOS process is preferred for better latch-up immunity. There is no need for additional threshold-voltage adjustment. The doping concentration of both the n-well and p-well can be independently optimized. These doping concentrations are significantly smaller than those of n^+-polysilicon-gate devices. Both n- and p-channel MOSFETs are expected to

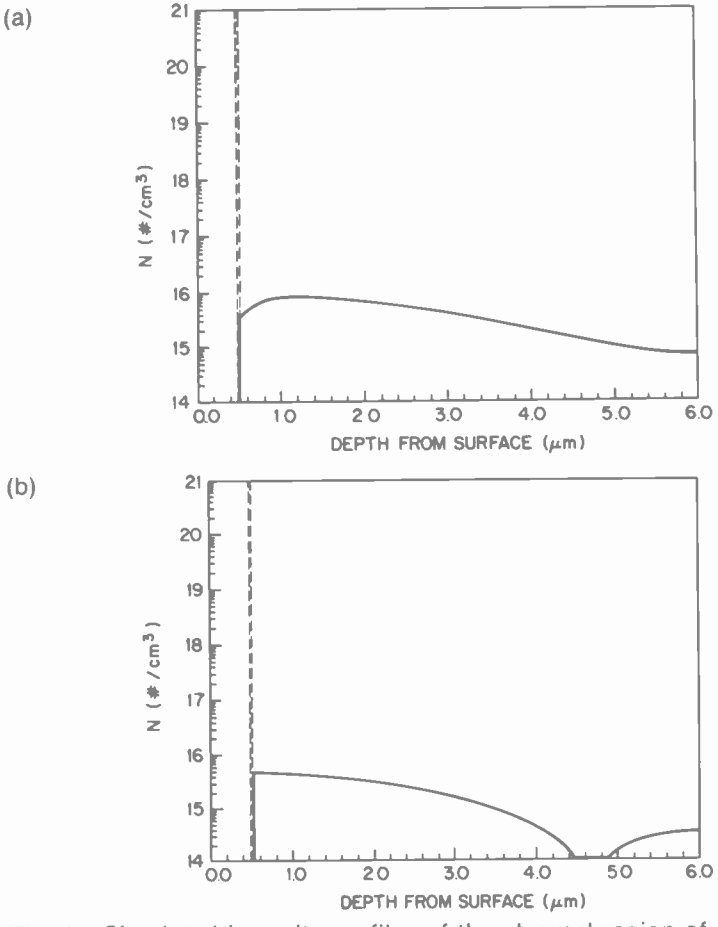


Fig. 6—Simulated impurity profiles of the channel region of a tungsten-gate CMOS FET: (a) p-well impurity profile and (b) n-well impurity profile.

have higher drive currents, lower junction capacitances, and smaller subthreshold swings.

5. Conclusion

It is virtually impossible to fabricate a high-performance submicrometer-size n^+-polysilicon-gate p-channel MOSFET with state-of-the-art technology. The n^+-polysilicon gate must be replaced by a new gate material having a work function that is larger than the electron affinity energy of silicon by one half the forbidden gap of the substrate. Tungsten meets this criterion and is probably one of the best materials for submicrometer-size CMOS silicon ICs. A tungsten- or tungsten silicide-gate CMOS IC is expected to have higher speed, lower leakage current, and a simpler fabrication process than n^+-polysilicon-gate CMOS ICs.

References:

[1] R. H. Dennard, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassous, and A. R. LeBlanc, "Design of Ion Implanted MOSFET's with Very Small Physical Dimensions," *IEEE Journal of Solid-State Circuits*, **SC-9**(5), p. 256 (Oct. 1974).

[2] A. S. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley & Sons, Inc., N.Y. (1967), Chapter 11.

[3] S. M. Sze, *Physics of Semiconductor Devices*, John Wiley & Sons, N.Y. (1981), p. 447.

[4] S. Y. Chiang, K. M. Cham, and R. D. Rung, "Optimization of Sub-micron P-channel FET Structure," *1983 IEDM Tech. Digest*, p. 534.

[5] B. E. Deal, E. H. Snow, and C. A. Mead, "Barrier Energies in Metal-Silicon Dioxide-Silicon Structures," *J. Phys. Chem. Solids*, **17**, p. 1873 (1966).

[6] *Handbook of Chemistry and Physics*, 64th Ed., 1983–1984, CRC Press. E-77.

[7] T. Moriya, S. Shima, Y. Hazuki, M. Chiba, and M. Kashiwagi, "A Planar Metalization Process—Its Application to Tri-Level Aluminum Interconnection," *1983 IEDM Tech. Digest*, p. 550.

[8] E. K. Broadbent and C. L. Ramiller, "Selective Low Pressure Chemical Vapor Deposition of Tungsten," *J. Electrochem. Soc.*, **131** (6), p. 1427 (1984).

[9] K. Y. Tsao and H. H. Busta, "Low Pressure Chemical Vapor Deposition of Tungsten on Polysilicon and Single-Crystal Silicon via the Silicon Reduction," *J. Electrochem. Soc.*, **131** (11), p. 2702 (1984).

Barrier-Height Measurements of Tantalum Silicide on Silicon

Sheng Teng Hsu and Jer-Shen Maa
RCA Laboratories, Princeton, NJ 08540

Abstract—The barrier heights of $TaSi_2$ to n-type and to p-type silicon were measured by the *C-V* method. A thin oxide layer (no thicker than 2 nm) between the $TaSi_2$ and n-type silicon was used to minimize the leakage current of the $TaSi_2$–n-type silicon diode. The measured barrier heights are 0.715 eV and 0.373 eV, respectively, for $TaSi_2$ to p-type and $TaSi_2$ to n-type silicon contacts.

1. Introduction

Refractory-metal silicide has been increasingly used in very-large-scale integrated (VLSI) circuits to enhance the conductivity of fine-geometry polysilicon lines and shallow n^+ and p^+ diffusion areas. As the physical dimensions of the device are scaled down, the width of the polysilicon gate becomes very narrow and the depth of the source/drain junction becomes very shallow. The resistivity of these narrow polysilicon lines and shallow n^+ and p^+ diffusion layers is very large. The depth of the source/drain junction is no more than 0.2 μm for a 1-μm-gate-length MOSFET. When aluminum is in direct contact to this shallow n^+ or p^+ layer, aluminum may spike through the junction. As a result, the leakage current of the junction can be very large and the junction breakdown voltage very small. A thin layer of a refractory metal or its silicide can be deposited onto the surface of the n^+ and the p^+ areas to reduce the resistivity of these shallow layers and to form a barrier for the aluminum-to-silicon contact. This contact barrier prevents the aluminum from spiking through the very shallow n^+ or p^+ source/drain layers. The refractory-metal contact barrier can either increase or decrease the resistance of the metal-to-silicon contact. The contact resistance depends on the doping concentration of silicon and the contact poten-

tial barriers at the refractory-metal silicide to silicon interface and at the refractory-metal silicide to aluminum interface. It is very important, therefore, to know the barrier height of the refractory-metal silicide to silicon contact.

In this paper we present measured barrier heights of $TaSi_2$–n-type silicon and $TaSi_2$–p-type silicon contacts. The sum of the measured barrier heights of $TaSi_2$ to n-type silicon and $TaSi_2$ to p-type silicon is approximately equal to the forbidden-gap potential of the silicon, as expected.

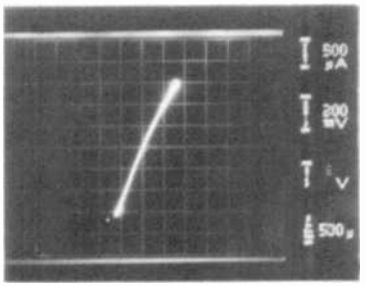
The capacitance-voltage (C-V) method was selected for the barrier-height measurements. This method requires only a capacitance meter which is available at any integrated circuit facility. The sample is an MOS structure with a very thin silicon dioxide layer. This oxide layer is used to reduce the leakage current of the contact so that accurate capacitance measurement can be made easily.

2. Sample Preparation

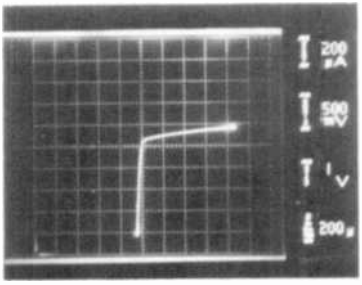
A 200-nm-thick silicon dioxide was thermally grown onto both n- and p-type silicon substrates. Oxide on the circular active area was defined by photolithographic and wet-etch processes. Tantalum and silicon were cosputtered onto the silicon wafer, then sintered at 900°C in argon ambient for 30 minutes to form tantalum disilicide.[1] The tantalum disilicide at the active area was protected with photoresist, and the silicide on the nonactive areas was removed with plasma etching. The diameter of the circular active area is equal to 1000 μm, and there is a 50-μm silicide overlap to the active area.

The current-voltage characteristics of $TaSi_2$ contacts to n-type silicon and to p-type silicon are shown in Figs. 1A and 1B, respectively. The wafers were dipped in 50:1 H_2O:HF solution for 30 seconds prior to the tantalum and silicon sputtering process. The $TaSi_2$ is, therefore, in direct contact with the silicon. The $TaSi_2$ to p-type silicon contact forms a good diode; the resistance at the $TaSi_2$ to n-type silicon contact is very small, which implies that the barrier height is small. The measurement of the contact capacitance is difficult. This conductive component of the contact impedance can be minimized by a very-thin-oxide MOS structure.

The RCA standard clean process[2,3] grows 1 to 2 nm of oxide onto the surface of bare silicon. We utilized this oxide for a $TaSi_2$–oxide–n-type silicon MOS structure. This thin oxide layer is transparent when the diode is forward biased, but it blocks the reverse-biased diode current. The typical current-voltage characteristic of a $TaSi_2$–



(A)



(B)

Fig. 1—*I-V* characteristics of $TaSi_2$ to silicon contacts: (A) $TaSi_2$–n-type-silicon contact and (B) $TaSi_2$–p-type-silicon contact.

SiO_2–n-type-silicon diode is shown in Fig. 2. The reverse diode current increases very slightly with the bias voltage.

3. Barrier Height Measurement

The barrier height of $TaSi_2$ to n-type silicon was measured on $TaSi_2$–oxide–silicon MOS diodes. The band diagram of a thin-oxide MOS diode is sketched in Fig. 3. Fig. 3A is the band diagram when the MOS is at the flat-band condition, and 3B is the energy diagram when the MOS diode is reverse biased. The substrate is at ground potential, and the voltage at the metal, V_m, is given by

$$V_m = \frac{qN_D}{2\epsilon_0\epsilon_{Si}} W^2 + \frac{qN_D W + Q_{ss}}{C_{ox}} \qquad [1]$$

$$= \frac{\epsilon_0\epsilon_{Si} qN_D}{2C^2} + \frac{qN_D W + Q_{ss}}{C_{ox}}.$$

Here, ϵ_{Si} is the dielectric constant of silicon, N_D is the doping con-

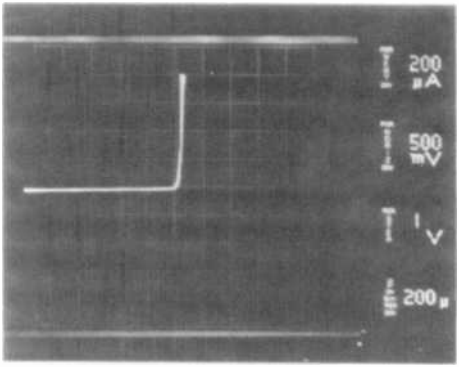


Fig. 2—*I-V* characteristic of a thin-oxide $TaSi_2$ MOS diode.

centration of the n-type substrate, W is the width of the space-charge region, Q_{ss} is the fixed-charge density at the silicon-oxide interface, and C_{ox} and $C = \epsilon_0\epsilon_{Si}/W$ are the unit-area oxide capacitance and the unit-area depletion layer capacitance, respectively. The capacitance of the thin oxide, C_{ox}, is larger than 1.7×10^{-6} F if the oxide is no thicker than 2 nm. The depletion layer capacitance,

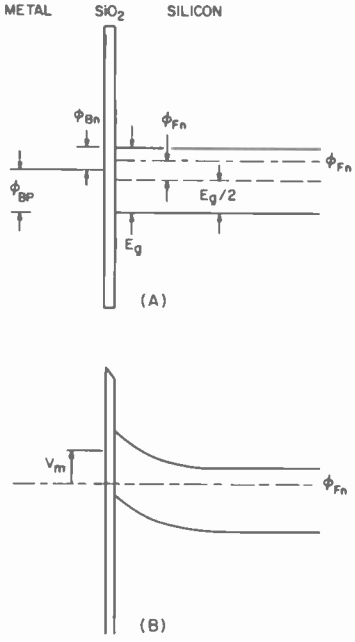


Fig. 3—Energy-band diagrams of thin-oxide n-type MOS diodes: (A) MOS diode at flat-band condition and (B) reverse-biased MOS diode.

C, is smaller than 1.0×10^{-7} F if the width of the depletion layer is larger than 0.1 μm. Consequently, the last term of Eq. [1] can be neglected:

$$V_m = \frac{\epsilon_0 \epsilon_{Si} q N_D}{2C^2}. \quad [2]$$

Differentiation of Eq. [2] with respect to V_m yields

$$\frac{d}{dV_m}\left(\frac{1}{C^2}\right) = \frac{2}{\epsilon_0 \epsilon_{Si} q N_D}. \quad [3]$$

A plot of $1/C^2$ versus V_m is a straight line. From the slope of this straight line, the doping concentration, N_D, and hence the Fermi potential, ϕ_{Fn}, of the silicon substrate can be calculated.

Eq. [3] can be rewritten

$$\frac{d}{dV_m}\left(\frac{1}{CA}\right)^2 = \frac{2}{\epsilon_0\, \epsilon_{si} q N_D A^2}, \quad [4]$$

where A is the area of the thin-oxide MOS diode and CA is the capacitance of the thin-oxide MOS diode. If we extrapolate the $1/C^2$ versus V_m curve to intersect the V_m axis and call the voltage at the intersection V_{mo}, the capacitance C at V_{mo} is equal to infinity and the MOS structure is at the flat-band condition. From Fig. 3A it is obvious that the barrier height, ϕ_{Bn}, is given by

$$\phi_{Bn} = \frac{\epsilon_g}{2q} + |V_{mo}| - |\phi_{Fn}|, \quad [5]$$

where ϵ_g is the energy gap of the silicon and ϕ_{Fn} is the Fermi energy of the n-type silicon substrate.

The same method can be used to measure the barrier height of refractory metal or silicide on a p-type silicon substrate, ϕ_{Bp}. The sum of ϕ_{Bn} and ϕ_{Bp} should be equal to the bandgap potential, ϵ_g/q, of silicon.

Fig. 4 presents the measured data of a typical $TaSi_2$-to-p-type-silicon contact. The I-V characteristic of this diode is shown in Fig. 1B. The curve in Fig. 4 is the diode capacitance, and the points on the straight line are $1/C^2$ as calculated from the measured capacitance. The fit is quite good. The slope of this line is equal to 1.59×10^{20}/(V-cm^2-F^2). Therefore, the substrate doping concentration and the Fermi potential of the silicon are equal to 1.28×10^{15}/cm^3 and 0.286 V, respectively. The line intersects the voltage axis at 0.46 V; the barrier height, ϕ_{Bp}, therefore, is 0.715 V.

Fig. 5 presents the measured data of a typical $TaSi_2$ on an n-type

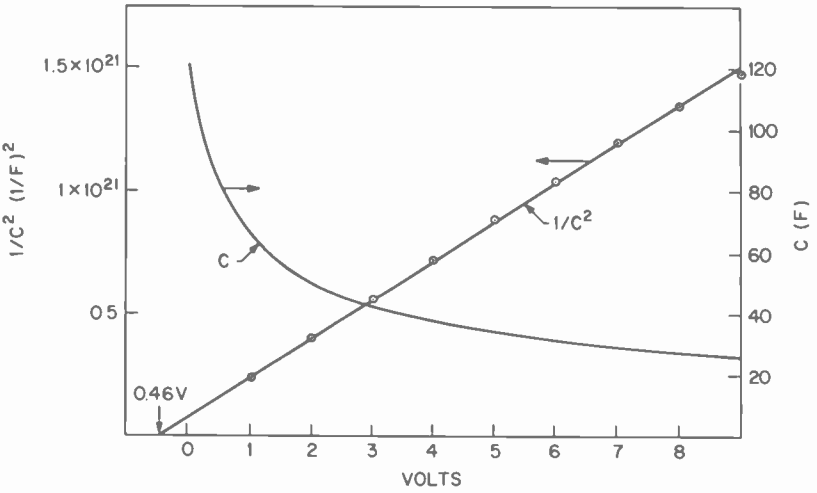


Fig. 4—Barrier-height measurement of a $TaSi_2$-to-p-type silicon contact.

silicon thin-oxide MOS diode whose *I-V* curve is shown in Fig. 2. The curve shows the measured diode capacitance and the points on the straight line are obtained by calculating $1/C^2$ from capacitance measurements. Again, the fit is very good. This straight line intersects the voltage axis at 0.08 V. From the slope of the line, the doping concentration of the n-type silicon substrate is calculated to be $4.15 \times 10^{14}/cm^3$. The Fermi potential of the silicon substrate,

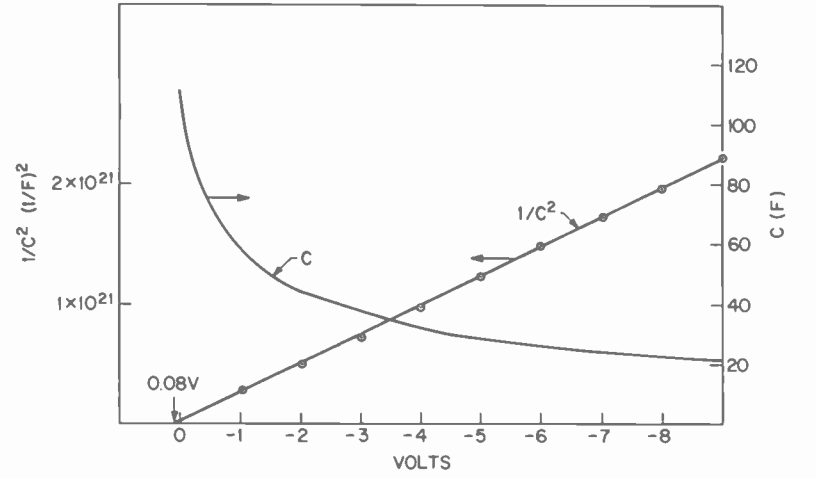


Fig. 5—Barrier-height measurement of a $TaSi_2$-to-n-type-silicon thin-oxide MOS diode.

therefore, is 0.257 V. The barrier height of the $TaSi_2$-to-n-type-silicon contact, ϕ_{Bn}, is 0.373 V. The sum of ϕ_{Bp} and ϕ_{Bn} is 1.088 V, which is approximately the forbidden gap of silicon.

4. Conclusion

The barrier heights of $TaSi_2$ to n-type silicon and to p-type silicon were measured by the C-V method. The barrier height of the $TaSi_2$-to-p-type-silicon contact is found to be 0.715 eV. The $TaSi_2$-to-n-type-silicon contact barrier height is 0.373 eV, as compared to the published data of 0.59 eV.[4] Since the sum of our measured barrier heights of $TaSi_2$ to n-type and $TaSi_2$ to p-type silicon is equal to the forbidden gap of silicon, our measurement is basically accurate. The discrepancy between the published data and our measured result is due, possibly, to the difference in the silicidization process.

References:

[1] J.-S. Maa, C. W. Magee, and J. J. O'Neill, "Phosphorus Out-Diffusion From Double-Layered Tantalum Silicide/Polycrystalline Silicon Structure," *J. Vac. Sci. Technol.*, **B1** (1), p. 1, Jan.-Mar. 1983.
[2] W. Kern and D. Puotinen, "Cleaning Solutions Based on Hydrogen Peroxide for Use in Silicon Semiconductor Technology," *RCA Rev.*, **31** (2), p. 187, June 1970; S. Shwartzman, A. Mayer, and W. Kern, "Megasonic Particle Removal from Solid-State Wafers," *RCA Rev.*, **46** (1), p. 81, March 1985.
[3] A. D. Weiss, "Wafer Cleaning Update," *Semiconductor International*, April 1984, p. 82.
[4] S. M. Sze, *Physics of Semiconductor Devices*, 2nd Ed., John Wiley & Sons, New York (1981), Chap. 5, p. 292.

Preparing Master Grooves for Lapping Discs Used in Micromachining of VideoDisc Styli

W. R. Roach and I. Gorog
RCA Laboratories, Princeton, NJ 08540

Abstract—The tip of a long playing VideoDisc stylus has a shape resembling a keel of microscopic dimensions. The machining of such a keel shape on the faceted tip of a diamond log requires a lapping disc having an approximately 4.5-μm deep, 2.4-μm-wide groove of minimal taper. The method of generating a photoresist master having a groove of this geometry by laser exposure and development is described. Details of a laser recording system for the production of a uniform lapping groove over a 3-inch-wide band on a 12-inch-diameter disc and methods for evaluating the groove profile are presented. A general theory of the exposure/development process is used to generate a family of groove profiles that traces the evolution of the groove shape during photoresist development. This theory provides a qualitative framework for establishing optimum exposure/development conditions.

1. Introduction

The RCA VideoDisc recording is played using a diamond stylus whose tip is precisely shaped so as to ride smoothly in the 140-degree V-shaped groove. The trailing vertical surface of the stylus carries an electrode that capacitively senses the presence of signal elements on the vinyl disc. The stylus extends the full width of the groove so that the signal-to-noise performance on playback is maximized. The stylus might thus be configured as shown in Fig. 1. However, after extensive use, such a stylus would wear causing the width of the tip to broaden, which would eventually result in the configuration shown in Fig. 2. During playback such a stylus could pick up unwanted signals from the adjacent grooves. It could also

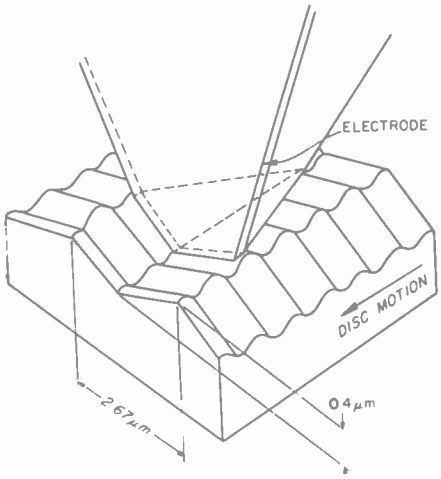


Fig. 1—View of faceted stylus in RCA VideoDisc groove.

lose the wanted signal because minor groove-pitch nonuniformities could cause the tip to not conform precisely to this extended profile for many of the grooves on the disc. It was suggested by E. O. Keizer[1] that both problems could be avoided by shaping the stylus tip as indicated in Fig. 3. With such a configuration, the stylus tip width will not vary during life, and thus it will consistently ride in and sense signals from a single groove.

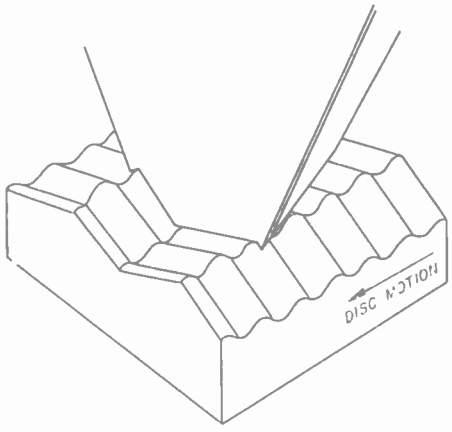


Fig. 2—Effect of wear on improperly shaped stylus showing electrode overlap to adjacent grooves.

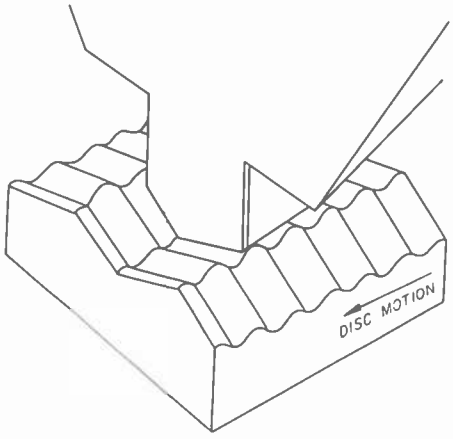


Fig. 3—View of ideal keel-lapped stylus tip riding in VideoDisc groove.

The production process for a Video-Disc diamond stylus with a tip geometry approaching that of Fig. 3 includes a machining operation that shapes the electroded and faceted tip by lapping it into a spiral groove of precisely shaped profile. The required profile of the keel-lapping groove is relatively narrow (2.4 μm wide) and deep (4.5 μm deep). Because of the large aspect ratio, both the production of the original groove and its replication to form the final lapping discs are relatively difficult. As indicated in Fig. 3, the ideal keel-lapped stylus would have vertical walls, but in practice it is impractical, if not impossible, to replicate lapping discs unless the walls have some taper. Taper is necessary because, during replication of the lapping disc, a reentrant structure will form an interlocking system with the next generation part, and the two cannot be separated without damaging them. However, too much taper would reduce the effectiveness of the resultant keel-lapped shape. A good compromise is to accept a 2.5° outward tilt of the groove walls, as indicated in Fig. 4. The groove must also be symmetrical to produce a keel whose prow is properly centered for optimum tracking performance. The groove pitch on the master recording used to make the lapping discs should be constant so that the lapping station can be set up with a constant pitch drive. The groove spacing should be large enough to insure that the shoulders of the fully keel-lapped tip are flat. The maximum number of turns per disc is desirable, consistent with the shoulder-width requirement, so that the largest possible number of stylii can be lapped on a given disc. This requirement is adequately met by a 9-μm groove period.

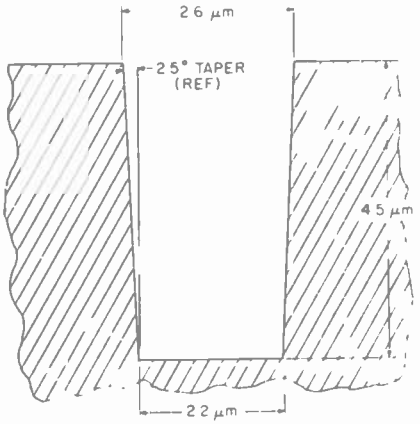


Fig. 4—Groove profile required for keel-lapping master.

The keel-lapping process is typically followed by a very short lap of the stylus on another spiral-grooved disc on which the grooves have the same V-shaped profile as the production VideoDiscs (140° included angle) to insure that no wear-in time is necessary to reach the maximum electrical output from a new stylus.

This paper outlines the method used to generate the master part used in the production of lapping discs, provides details of the optical recording system used to generate grooves with a uniform profile over a 3-inch-wide band of approximately one-foot outside diameter, describes the special measurement techniques that have been developed to evaluate the resultant groove shape, and discusses a theoretical model used to gain insight into the optical recording/development process.

2. Keel-Lap Mastering

The technique used to generate the metal keel-lap master consists of the following steps:

(a) coating a large-diameter copper-plated aluminum substrate with a uniform layer of photoresist,
(b) baking the resist to drive off the solvent,
(c) exposing a spiral track with a focused laser beam,
(d) developing the resist to remove exposed photoresist down to the substrate,
(e) washing off the developer and drying the resist,

(f) coating the deep relief surface with a thin conductive layer, and (g) electroplating the nickel to form the metal master.

Of these steps, this paper discusses in detail only (c) keel-lap recording and (d) photoresist developing. These two aspects are not only central to the production of the keel-lap master but also have wider applicability in the production of any system where deep and narrow grooves of controlled profile are required on a large diameter substrate.

There is one other crucial step in the process of obtaining a suitable master, and that is to evaluate the geometry of the grooves actually produced and, using this information as feedback, to improve the choice of recording/developing conditions for subsequent recordings. Because of the dimensions involved, only a scanning electron microscope (SEM) has the resolution necessary to characterize the groove profile adequately. Unfortunately, all processes that we are aware of to prepare samples for accurate SEM evaluation of the deep-grooved recording profile are destructive of the part involved. Also, when the original nickel master is pulled off the substrate, the parting takes place in the resist, resulting in damage to the relief pattern. Thus, only one electroform nickel part (metal master) can be made directly from the photoresist-covered substrate. If we want to evaluate a possibly useful recording (one that might actually be utilized for the production of lapping discs) we must generate from the master a part that can be sacrificed. In our case we are forced to generate from the production master a mold, which is then used to make several stampers. Of these, the first and second stampers are reserved for further fanout, since they are the best. The third and fourth stampers are used to press test lapping discs. Therefore, the fifth stamper is the first available for groove-profile evaluation. This multiple replication procedure in itself takes several days, and the groove-profile evaluation several more. Even then we have gained information on only one set of recording/developing conditions.

To reduce our dependence on this lengthy full-fanout feedback, we have devised an alternative method involving the production of a group of six identically coated substrates and the recording of as many as 20 separate test bands on one or two of these substrates, followed by immediate groove profile evaluation of thin electroform metal masters. Information from these test recordings can be used to establish conditions for recording the remaining members of the set, provided that this set of substrates has resist layers that are matched in thickness to within 0.2 μm. By this method we can make

two test recordings, evaluate their groove shapes, and record the remaining four members of the set as production masters, all within approximately three weeks. The four production masters can then be fanned out and evaluated at leisure while another complete set is in process.

Processing steps that cause modifications of the laser-recorded groove profile have also been developed, but these are beyond the scope of this paper.

3. Keel-Lap Recording Requirements

The recording of the keel-lapping grooves involves laser exposure of a spiral track in a photoresist layer that is coated onto a flat, 1/2-inch thick, 14-inch diameter, copper-plated, aluminum substrate. The thickness of the photoresist is approximately 4.5 μm with a uniformity of ±0.1 μm over the band from 2.8- to 5.8-inch radius. The resist used is a positive-developing RCA proprietary compound. The RCA formulation is used to insure complete quality control over the composition. The resist thickness must be maintained within the limits specified, since the exposure/development process produces a groove whose depth is the resist thickness and whose profile is a complicated function of the resist thickness, beam power, beam shape, focus position, substrate reflectivity, and development parameters.

The function of the keel-lap recorder then is to provide a laser beam that is focused in the photoresist at a fixed height above the copper–photoresist interface and scanned along a spiral track on the substrate. This height must be controlled to better than ±6% of the film thickness to insure a uniform groove shape. To maintain as many parameters constant as possible during the recording of production masters, these recordings are typically made at a constant laser power with a constant linear velocity of the substrate under the beam. The recording time was chosen to be ½ hour. To scan the approximately 500 cm^2 area during this time, the required linear velocity of the beam over the substrate surface is 300 cm/sec. It is fortunate that at this scan velocity, a moderate power He-Cd laser causes a large increase in the solubility of the exposed regions of the photoresist in the developer without simultaneously causing significant bleaching at the 4416 Å wavelength. It is essential to maintain a strong absorption in the resist throughout the exposure process if we are to obtain the optimum groove shape.

4. Keel-Lap Recorder

The system that has been used for recording masters of a uniformity adequate for mass production of keel-lapped styli is similar to that described by A. H. Firester et al[2] for optically reading VideoDisc substrates. The keel-lapping recorder is shown in Fig. 5. Every RCA stylus that has been sold has a keel lapped tip whose profile is traceable to an optical master generated on this machine.

4.1 Beam Forming Optics

The major components of the optical system necessary for the generation of the keel-lap masters are shown in Fig. 6. The 1-mm diameter beam from a 20-mW helium-cadmium laser is directed through a polarizer, quarter-wave plate, and a beamsplitter. Distortions in the beam profile introduced by these elements are reduced by passing the beam through a $\times 5$ microscope objective and a 25-μm pinhole. To produce our smallest diameter beam in the photoresist we next insert a 150-mm focal-length lens positioned a focal length away from the pinhole to recollimate the beam. Next a beam-limiting aperture (7.6 mm in diameter in this case) is used

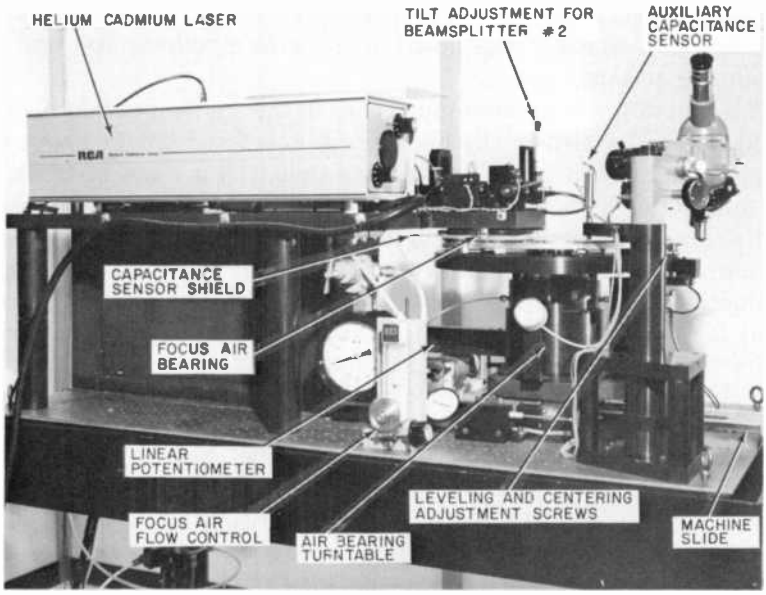


Fig. 5—View of optical keel-lap recorder.

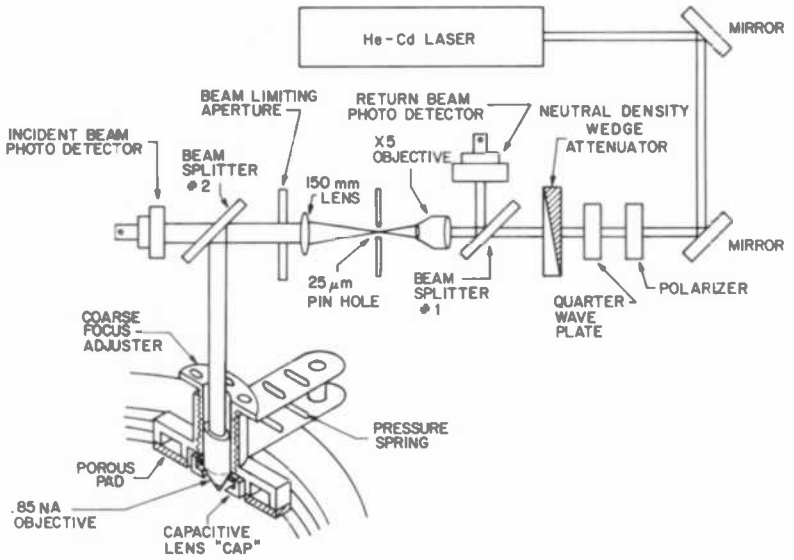


Fig. 6—Schematic of keel-lap-recorder optics.

to block all but the central portion of the Airy ring pattern[3] generated by diffraction from the pinhole.

With the Airy rings thus removed, the beam profile is approximately Gaussian. For a Gaussian profile, the power distribution $I(\rho)$ as a function of radius ρ is

$$I(\rho) = I(O) \exp(-\rho^2/w_p^2), \quad [1]$$

where $I(O)$ is the power on the optical axis and w_p is the Gaussian (power) parameter. The diameter of the beam at which $I(\rho)$ is just $1/2$ of $I(O)$ (defined as the full width at half maximum, FWHM) is

$$\text{FWHM} = 2w_p \sqrt{-\ln(0.5)} = 1.665\, w_p. \quad [2]$$

For the cylindrical Gaussian beam, this FWHM diameter also contains half the total power in the beam.

The collimated apertured beam (3.4-mm FWHM) is reflected off the beam splitter and passes through the final focusing objective (a $\times 50$, 0.85 NA, Leitz microscope objective with an input aperture of 8 mm) to produce a focused beam in the photoresist. When the objective is thus operated unfilled, with a small-diameter approximately Gaussian beam as input, it produces in the photoresist a focused beam that is also very nearly Gaussian. Under these conditions of illumination, although the spot diameter is larger than that attainable under uniform illumination of the focusing objec-

tive, the rings in the diffraction pattern are attenuated (in power) by a factor of approximately 100 as compared to those produced by a uniformly illuminated aperture.[4]

A measurement of the spot diameter at focus was made using the knife-edge scan technique described by Firester et al.[2] The result of this scan is shown in Fig. 7, where the downward displacement along the S-shaped curve represents increasing light intensity passed by the knife edge as it is removed from the beam. The horizontal axis is time, which can be related to distance by using the auxiliary sinusoidal signal, which is also shown. This auxiliary signal is the output from a He-Ne laser-illuminated Michelson interferometer that has one end mirror mechanically driven by the knife-edge motion. One complete cycle of the sinewave corresponds to one-half-wavelength (0.3164 μm) displacement of the knife edge.

For the exposure of the grooves to be uniform around the disc, the objective must be maintained at a fixed height (to within ± 1700 Å) above the photoresist surface while the substrate is rotated and translated under it at a scan velocity of several hundred cm/sec. To maintain focus to this accuracy, the final focusing objective is supported by a focus air bearing[2] consisting of two porous pads epoxied into aluminum cups. A 60-psi supply of filtered air is bled through a needle valve into the resulting chambers. This air exits downward through the pads and causes the structure supporting the lens to be lifted against a spring pressure and to ride a few mils above the resist surface. Very fine control of the air flow through these pads permits maintenance of the focus position to the required tolerance while the substrate is being rotated under the beam.

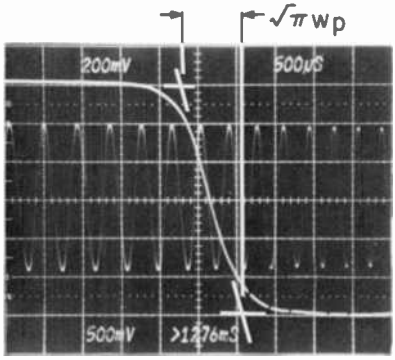


Fig. 7—Knife-edge spot size measurement. Downward deflection indicates increased light intensity. Sinusoidal interferometer peak spacing is 0.3164-μm knife-edge displacement.

4.2 Capacitance Sensor

To monitor the vertical position of the lens relative to the substrate, the final focusing objective was fitted with an aluminum cap. The flat ring-shaped bottom surface of the cap typically rides ~75 μm above the substrate surface, and the capacitance between the two surfaces (typically ~20 pF) varies as the reciprocal of the distance between them. Since the groove profile in the 4.5-μm thick resist must be substantially constant, the focal-plane position in the resist must not shift during the recording by more than ±0.3 μm. Refraction in the resist (see Sec. 4) requires that the lens height be maintained to within ±0.17 μm, corresponding to changes in lens-cap-to-substrate capacitance of ±0.05 pF. Since the stray capacitance associated with the lens cap, leads, etc., is ~25 pF, we require a system that will accurately measure changes of ±0.05 pF out of 45 pF. A capacitance bridge feeding a digital voltmeter provides a measure of the position to better than this precision, but such a system provides no absolute indication of where the lens focal plane is located. To establish the absolute scale for the capacitance sensor, we employ the optical-focus sensor described below.

4.3 Optical-Focus Sensor

The optical-focus sensor utilizes the recording beam after it has been reflected off the substrate. This return beam passes backward through the focusing objective, beam-splitter No. 2, beam-limiting aperture, collimating lens, pinhole, and microscope objective to beam-splitter No. 1 and the return-beam photodetector. The primary path of the laser beam is through the photoresist, followed by a reflection off the photoresist–substrate interface and back through the photoresist. The return-beam power as a function of focal position has one local maximum when the laser beam is focused at the photoresist-substrate surface. Another local maximum occurs in the return beam when the final objective has been raised sufficiently that the beam waist occurs at the upper surface of the resist. In this case, a small portion of the incident beam is reflected off the upper resist surface and is recollimated by the final focusing objective and passed backward along the optical path.

The positions of these two peaks on a return-beam-power-versus-capacitance plot establish the absolute scale for the capacitance sensor. Once the capacitance values corresponding to both top-surface focus and bottom-surface focus are established, an intermediate position can be maintained by controlling the capacitance reading

at the appropriate intermediate capacitance value. If the resist thickness of the production substrate is the same as that of the test substrate, the focus for the production recording can be made by setting the system up to have the same capacitance as that observed to give the best grooves on the test substrate. The response of the return-beam power to changes in lens height, as measured by the capacitance sensor, is shown in Fig. 8 for the case of a bare copper substrate. Note the well defined focus position.

4.4 Optical Behavior of Photoresist

When through-focus measurements of return-beam power versus capacitance are made on a photoresist-covered substrate, care must be taken to account for the possibility of bleaching the resist during the measurements. The results of a set of through-focus measure-

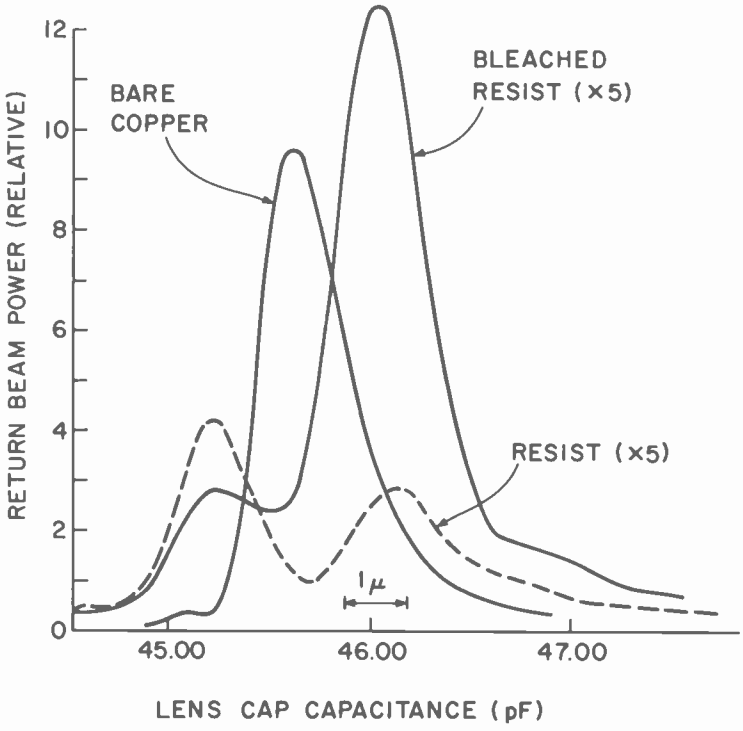


Fig. 8—Response of focus sensor versus lens cap capacitance. For approximate calibration, length of arrow represents 1-μm change in lens cap elevation.

ments of return-beam power are also shown in Fig. 8. Note, however, that the photoresist data is displayed with the vertical scale magnified by a factor of 5. In these measurements, the power in the focused laser beam was fixed at approximately 0.5 mW and the linear velocity of the substrate under the beam was varied. The bleached condition was obtained by allowing the beam to write over the same circular track at 325 cm/sec for several minutes. This bleached circular track was further probed by making scans of the return beam versus focal position at various surface velocities. Down to 2 cm/sec, these all produced essentially identical traces. Resist physical damage can occur for very slow scans (less than 10 cm/sec) for an input power of only 2 mW.

Note in Fig. 8 the two distinct peaks corresponding to focus at the air–photoresist interface and at the photoresist–copper interface. The vertical displacement of the lens necessary to shift the focal plane from the top surface to the photoresist–copper interface can be calculated as follows. The photoresist index of refraction n(resist) is given[5] as a function of the free-space wavelength of light, λ_0, by

$$n(\text{resist}) = 1.539 + \frac{0.00832}{\lambda_0^2 - 0.036}, \qquad [3]$$

which for $\lambda_0 = 0.4416$ μm is $n = 1.645$. Rays exiting the lens at an angle i will be refracted by the resist surface and traverse the film at an angle r. If unrefracted rays incident at an angle i meet at a distance H_i below the air-resist surface, the refraction causes the rays to be brought to focus at a distance H_r below the resist surface,

$$H_r = H_i (\tan i/\tan r). \qquad [4]$$

Our apertured beam, upon passing through the lens, has marginal rays (those intercepting the local FWHM circumference) traveling at an angle $i = 22°$. From Snell's law, the angle r in the photoresist is found to be 13°. From Fig. 8, the observed value of H_i is found to be 2.35 μm. From Eq. [4], we find $H_r = 4.05$ μm, which is in reasonable agreement with the 4.9-μm thickness of the film that was more accurately determined by analyzing the interference fringes in its reflectivity-versus-wavelength spectrum, where the wavelength dependence of the index is assumed to be given by Eq. [3].

The magnitude of the peaks seen in Fig. 8 provides information about the absorption properties of the photoresist during exposure. We note that the bare-copper data of Fig. 8 provides an absolute reflectivity reference. The normal amplitude reflection coefficient

at the boundary between material of (complex) index of refraction n_1 and that of (complex) index of refraction n_2 is given by

$$S = (n_2 - n_1)/(n_2 + n_1). \quad [5]$$

In the case of copper at a wavelength of 4416 Å, the complex index is found from Heavens[6] to be

$$n(\mathrm{Cu}) = 0.87 + 2.2j. \quad [6]$$

The power reflection coefficient R at the boundaries of interest can be calculated directly as the absolute square of S determined from Eq. [3], with the results that R(air-Cu) = 58.4%, R(air-resist) = 5.9%, and R(resist-Cu) = 48.7%. From the measured relative peak heights of Fig. 8, we can calculate the power absorption coefficient α_P defined by

$$P(z) = P(O) \exp(-\alpha_P z), \quad [7]$$

where P(O) is the power entering the film, $P(z)$ is the transmitted power, and z is the path length through the film. The results of these calculations are that α_p(fully bleached) = 0.11/μm, α_p(unexposed) = 0.23/μm, and α_p(record) = 0.13/μm.

The result for the fully bleached case is consistent with the data of Kaplan and Meyerhofer,[7] adjusted to compensate for the differences in concentration of sensitizer in the films. Our value of the absorption constant for the unexposed case is significantly higher than their value, which for 4416 Å was essentially unchanged from the bleached case. For the unexposed, resist, the 4416 Å line lies on a very strong absorption edge which can be shifted in wavelength by processing. Their film was of pure sensitizer, while our resist has only 12% sensitizer, so such a disagreement is not surprising. The decrease of the reflectivity by the upper surface upon bleaching is reproducable and too large to be accounted for by changes in the published bulk properties (index of refraction and absorption) with bleaching.

4.5 Laser Power and Alignment

The laser power to the resist surface is controlled by the position of a motor-driven neutral-density wedge attenuator and the incident power is monitored by a photodetector. During test recordings, power and focus are varied in order to insure that the proper conditions for good grooves are obtained somewhere on the recording. Since we also require that the average slope of the groove walls be

vertical, care must be taken to insure that the incident laser beam is perpendicular to the photoresist. Beam splitter No. 2 can be rotated and translated to accomplish this final adjustment. Unfortunately, because the beam that is reflected off the substrate can be reflected again by other optical surfaces in the system, it is difficult to duplicate the exposing beam geometry unless the substrate is actually in place. When it is in place, we can no longer measure the light distribution. Thus, to see that the final tilt adjustment has been made properly, we must make a series of test recordings into the photoresist while varying the translation and tilt of beam splitter No. 2 and then evaluate the resulting grooves to establish optimum settings for these controls. Once a convenient developer and development time have been chosen, the development is always done with a constant developer composition and for a fixed time.

Perhaps the most frustrating aspect of the recording/developing process is its nonreproducability. We have attempted to minimize this nonreproducibility by coating all substrates of a set in a single morning and exposing and developing them under nominally identical conditions. Despite this attempt to maintain constant coating conditions, exposure conditions, and development conditions, the development rate still varies significantly from recording to recording. Thus, to stand a reasonable chance of transferring a given groove profile observed on a test recording to a production substrate of the same set, several production recordings must be made.

The He-Cd laser is polarized, and this aspect is utilized to prevent feedback of power into the laser cavity, which would cause unwanted oscillations in the laser power. The isolation of the laser is accomplished by inserting a polarizer and a quarter-wave plate in the laser beam before it enters the variable attenuator. The polarizer is oriented to pass the incident laser light; the quarter-wave plate converts this to circularly polarized light on the first pass; and then the return beam is converted back to linearly polarized light whose plane of polarization has been rotated by 90 degrees with respect to the original polarization direction. This reflected beam is then blocked by the polarizer.

4.6 Recorder Mechanics

The recorder turntable consists of a 15-inch-diameter aluminum plate mounted on a spindle supported by an air bearing. The spindle also incorporates the rotor for a dc motor. The motor housing is mounted on a machine slide equipped with a 10-turns-per-inch (TPI) lead screw. This screw is driven by a dc motor through a 100:1

speed-reducing worm gear. A microscope with an auxiliary capacitance sensor rigidly attached is swung over the substrate when the turntable slide is in the load position. The top surface of the substrate is brought up to a fixed height, and the surface is leveled by adjusting 6 vertical 100 TPI leveling screws until the capacitance sensor indicates that, upon slow rotation of the substrate, the surface is everywhere at the proper height to better than 2.5 μm vertical runout. Centering is accomplished by adjusting horizontal 100 TPI screws located on three of the leveling pads. The centering accuracy is measured by slowly rotating the substrate while viewing, through the microscope, a single circular groove that was mechanically cut into the substrate surface after this surface was machined flat.

The turntable rotational and translational drive motors are both controlled by a voltage that is proportional to the inverse of the radius. The required control signal is derived by supplying a linear potentiometer with a fixed voltage, coupling the potentiometer slider to the radial motion of the turntable, and using the output from the slider to drive an operational amplifier. The op-amp output is applied to the denominator input terminal of an analog divider, and a fixed reference voltage is applied to the numerator input terminal. By this method, we can obtain a control voltage that can generate constant linear velocity and constant pitch spiral to an accuracy of a few percent.

5. Photoresist Development

After exposure, the substrate is washed with freon and spin-dried to remove any traces of grease or other airborn contamination. The substrate is mounted face-up on a slowly rotating turntable, and a rubber collar is placed around the edge to form a shallow pan. Approximately 600 cm^3 of developeer is poured onto the surface, filling the pan to a depth of 1 cm.

The speed of the process of development depends on the developer used and on how much laser energy the photoresist has been exposed to locally. Initial tests indicated that contact of the photoresist with full strength Shipley AZ developer for 180 seconds would produce grooves of approximately 2.4 μm width down to the substrate when the resist is previously exposed locally to somewhat less than the full available focused laser power at a scan velocity of 300 cm/sec. Thus, these development conditions (Shipley AZ developer, 180 seconds) were established as our standard, and fine changes in groove width are accomplished by changing the laser power slightly.

After development, the substrate is washed with deionized water and baked.

6. Groove-Profile Evaluation

As indicated earlier, because of the dimensions of the grooves involved, the detailed experimental evaluation of the exposure/development process can only be made using a scanning electron microscope (SEM). As this instrument requires a conductive sample, the photoresist relief pattern must first be converted into a metal replica. This is accomplished by evaporating a 1200 Å thick gold layer over the entire surface of the test substrate and then electroplating a 0.005-inch-thick layer of nickel onto this gold film to form a thin master. The metal master is easily separated from the substrate at the resist layer, and acetone can be used to remove the residual photoresist from the master. The SEM can easily resolve features much smaller than the dimensions of the replicated grooves, but because the groove walls are smooth in the tangential direction, a fiducial viewing line must be provided to enable us to visualize the groove profile. Two methods of generating such a line, which corresponds to the intersection of a radial plane (normal to the disc surface) with the replica surface, are described in the subsections below.

6.1 Contamination Line Profiles

The simpler method, used for quick surveys of test recordings, involves punching a 0.5-inch-diameter disc from the master, mounting it in the SEM, pumping the system down to $\sim 10^{-4}$ Torr pressure, and making multiple scans across a groove along a single path with a small-diameter electron beam. This SEM beam locally heats the sample, and contaminants in the vacuum system collect on the sample surface along this line. When the SEM scan is shifted back into the raster-scan viewing mode, the contamination line stands out in sharp contrast to the rest of the surface. The sample can now be rotated almost 90° so as to cause the plane of the contamination line to be presented almost perpendicular to the beam. In this orientation, the contamination line represents a profile of the groove that can be photographed and measured.

Fig. 9 gives results of this procedure for grooves generated using a constant laser power under four different focus conditions: Fig. 9a is for laser focus (beam waist) 0.6 μm above the photoresist-copper interface; Fig. 9b corresponds to laser focus approximately 1.9 μm

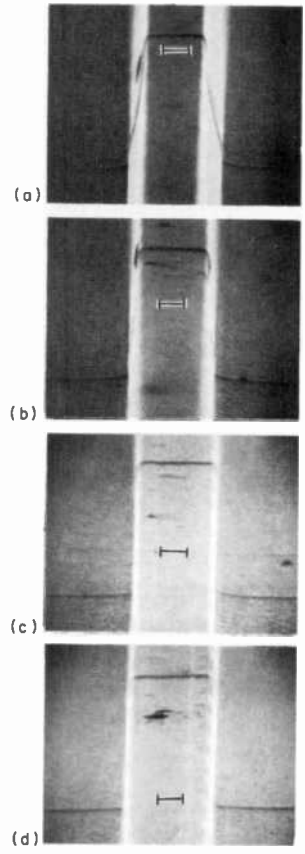


Fig. 9—Contamination-line views of replica of grooves generated by a focused beam with focal plane at the positions indicated by the black horizontal line. The length of the black horizontal line corresponds to the diameter of the incident beam waist which is 0.91 μm (FWHM). The depth of the groove in each case is 4.7 μm.

away from that surface; Fig. 9c to focus 3.1 μm up from the copper; and Fig. 9d to focus at 4.4 μm up, i.e., almost at the top surface. The head-on contamination line photograph of Fig. 9a for the focus near the bottom surface of the photoresist shows a strongly tapered nonreentrant groove profile. In Figs. 9c and 9d, the contamination line disappears along the sidewalls, and we can only guess what the corresponding groove profiles are.

We can, however, generate a contamination line groove profile in these cases by tilting the sample about a tangent to a groove so that the bottom, top, and one face of each groove are exposed to the

contaminating beam. An SEM photograph for that face is then taken by tilting the sample about both a radial and a tangential axis. Figs. 10a, 10b, 10c, and 10d were made in this manner to display the profile of the face of the groove nearest the substrate center. This complete procedure was repeated again to generate the photographs of Figs. 11a, 11b, 11c, and 11d, which show the face nearest the outside radius. From such a set of photographs, it is possible to obtain information that is adequate to establish where the optimum focus position is. However, because it is extremely difficult to set up angles in the SEM precisely, both for the contamination line scans and for the subsequent viewing, and because there are inherent distortions in the SEM scan, the technique described above is inadequate to establish groove tilt to the 1° reso-

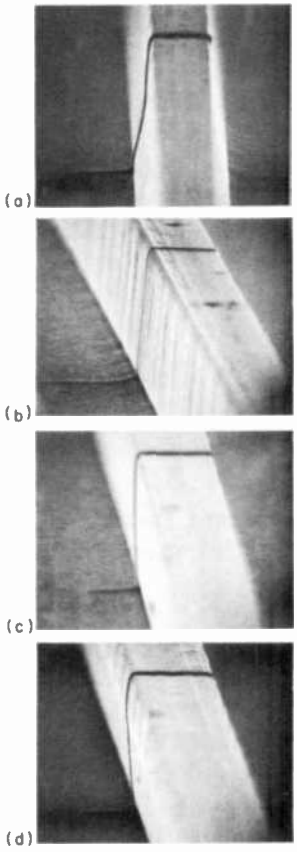


Fig. 10—Contamination line view of side of groove face nearest center of disc for grooves corresponding to those of Fig. 9.

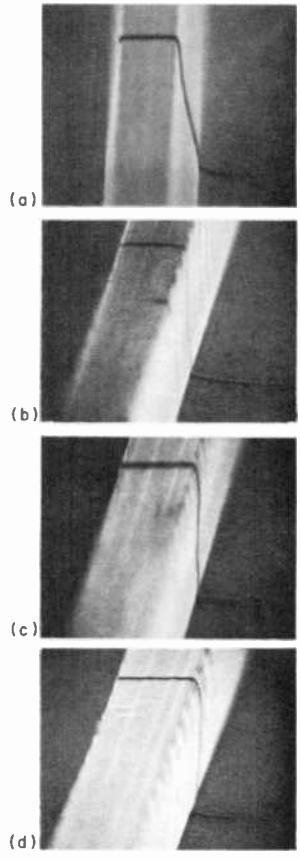


Fig. 11—Contamination line view of side of groove face nearest outer radius for grooves corresponding to those of Fig. 9.

lution required. To circumvent these limitations, the method described below for precise evaluation of the groove profile was developed.

6.2 Cross-Sectioned Samples

This second method of groove profile evaluation consists of cutting a small radial strip from the gold-covered nickel master, electroplating a few mils of nickel onto this gold layer, and then standing this strip on edge with the radius horizontal. The sample is potted in epoxy and allowed to cure. It is sawed, again along a radius, and the sawed surface is polished to present an edge view of the grooves. The polished surface is then etched with a gold etching solution

that does not attack the nickel. The polished surface is thus transformed into a deep relief structure in which the groove profile is easily seen using an optical microscope and can be accurately measured using a scanning electron microscope.

In Fig. 12 we see explicitly the groove profiles corresponding to the four focus conditions described for Figs. 9, 10, and 11. Note that a single head-on view now gives us the complete profile, even for the case of re-entrant grooves. A study of the re-entrant grooves on the test master is useful to establish focus and groove tilt, but a metal part containing grooves of this shape could not be replicated without damage during separation. Notice also the hole clearly vis-

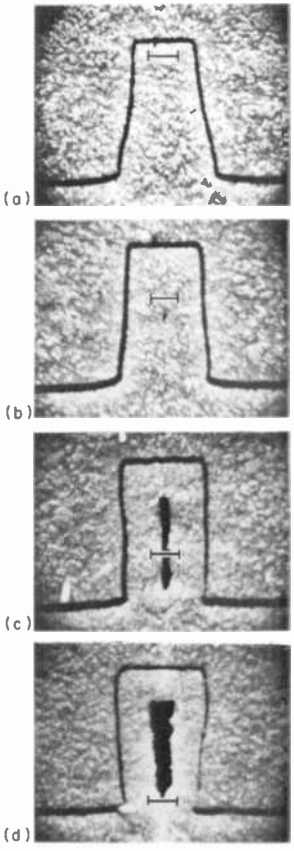


Fig. 12—Sectioned profiles of the grooved replica described in Fig. 9. Again, the position of each horizontal line corresponds to the focal plane of the incident laser beam and its length corresponds to the diameter of the beam waist of 0.91 μm (FWHM).

ible in the center of Figs. 12d and 12c and just visible in 12b. For a re-entrant structure, a conformal electrodeposition will produce such a void. In the real case, the deposition usually occurs more rapidly at sharp edges (that are sticking into the bath), so that deep and narrow vertical walled structures, or even slightly tapered ones, can also be expected to generate these voids.

Figs. 9–12 show one major effect of the change in focus for one particular laser power, collimating-lens focal length, photoresist thickness, and substrate reflectivity. An experimental determination of an large number of groove shapes for various possible combinations of the above variables would obviously be extremely tedious. For this reason and in order to more fully understand the exposure/development process, we have developed a theoretical model to explore the range of parameters relevant to keel-lap mastering. The purpose of this analysis is to provide some guidance for experimental efforts. The basic physical/chemical aspects of deep resist recording are not sufficiently understood to allow the use of the model developed here for the precise specification of the process parameters that produce a desired relief pattern, but it does yield useful information about how the recording and development parameters affect the shape of the relief surface.

7. Theoretical Model

The etch model used as the basis for our calculations assumes (a) that the exposure is via a Gaussian beam scanned uniformly across the surface, (b) that local etching proceeds normally to the etch boundary, and (c) that the evolution of the boundary can be adequately described by the numerical integration of a set of difference equations. Each of these aspects is discussed in some detail below.

7.1 Laser Exposure Pattern

The incident beam inside the resist layer is assumed to be described in regions away from the focal plane by the more complete Gaussian expression

$$E_I(\rho, z) = A_I \frac{w_0}{w(z)} \exp\left(\frac{-\rho^2}{w^2(z)} + j\,\phi(\rho, z)\right). \qquad [8]$$

Here $E_I(\rho,z)$ is the field amplitude described in terms of a cylindrical coordinate system, $\rho = \sqrt{x^2 + y^2}$; A_I is the amplitude normalization constant; $\phi(\rho,z)$ is the phase function; w_o is the beam waist;

$$w(z) = w_0 \sqrt{1 + (z/z_0)^2}; \quad [9]$$

z is the distance from the waist along the beam direction;

$$z_0 = \pi w_0^2/\lambda; \quad [10]$$

and λ is the wavelength of the exposing light in the photoresist. The phase function $\phi(\rho,z)$ can be shown to be[8]

$$\phi(\rho,z) = k\left(z + \frac{\rho^2}{2z[1 + (z_0/z)^2]}\right) - \arctan\left(\frac{z}{z_0}\right), \quad [11]$$

where $k = 2\pi/\lambda$. The positions $z = 0$ and $z = z_a$ are, respectively, the positions of focus and the air-resist interface. The effect of absorption of the resist can be accounted for by an amplitude absorption coefficient α_A. The field amplitude is thus reduced as it progresses through the resist according to

$$A_I = A_0 \exp[-\alpha_A (z - z_a)]. \quad [12]$$

We now assume that the resist is coated onto a partially reflecting substrate located at $z = z_R$ and having a (complex) reflection coefficient S. The reflected field configuration is simply obtained by making the reflection transformation substitution where z goes to $z' = 2z_R - z$ and noting that $A_R = A_I(z')S$. We then have, for the reflected field,

$$E_R(\rho, z) = E_I(\rho, z')\, S. \quad [13]$$

For the typical recording arrangement, if we assume that the substrate is moved uniformly across the beam in the y direction, the total local exposure $D(x,z)$ is found from

$$D(x, z) = \eta \int_{-\infty}^{\infty} dy\, |E_I + E_R|^2, \quad [14]$$

where η is a constant proportional to the dielectric constant of the resist and is inversely proportional to the velocity. If the integration indicated in Eq. [14] is performed, the result can be conveniently written as the sum of three terms D_I, D_R, and $D_{I,R}$ due, respectively, to the incident, reflected, and interfering fields. Thus

$$D(x, z) = D_I + D_R + D_{I,R}, \quad [15]$$

where

$$D_I(x, z) = \eta \sqrt{\frac{\pi}{2}} A_i^2 \frac{w_0^2}{w(z)} \exp\left(-2\alpha_A (z - z_a) - \frac{2x^2}{w^2(z)}\right), \quad [16]$$

$$D_R(x, z) = |S|^2 D_i(x, z'), \quad [17]$$

$$D_{I,R}(x, z) = \eta G(z, z') \frac{2\sqrt{\pi} \exp(-x^2 g_1)}{(g_1^2 + k^2 g_2^2)^{1/4}} \cos\{\zeta\}, \quad [18]$$

with the auxiliary expressions

$$G(z, z') = |S| A_I^2 \frac{w_0^2}{w(z)w(z')} \exp(-2\alpha_A (z_R - z_a)), \quad [19]$$

$$g_1 = \frac{1}{w^2(z)} + \frac{1}{w^2(z')}, \quad [20]$$

$$g_2 = \left(\frac{z}{w^2(z)} - \frac{z'}{w^2(z')}\right) \frac{w_0^2}{2z_0^2}, \quad [21]$$

$$\zeta = k(2z - 2z_R + x^2 g_2) + \delta - \Phi, \quad [22]$$

$$\delta = \arctan\left(\frac{kg_1}{g_2}\right) + \arctan(z/z_0) - \arctan(z'/z_0). \quad [23]$$

Here ϕ is the phase factor of the amplitude reflection coefficient S.

7.2 Local Etching

We assume that the local etching proceeds normal to the surface contour. The local etch rate (for given resist and developer formulations) is a function $F(x,z)$ of only the local total exposure $D(x,z)$. The etch velocity, $\mathrm{V}(x,z)$ is then a vector which is locally of the form

$$\mathrm{V}(x, z) = F(x, z)\, \mathrm{n}(x, z), \quad [24]$$

where $\mathrm{n}(x,z)$ is the unit vector normal to the surface. The most commonly employed models for etching assume one of the three following forms:

$$F1(x, z) = C_1 + C_2 D(x,z) \quad [25]$$

$$F2(x, z) = C_1 \exp(C_2 D(x, z)) \quad [26]$$

$$F3(x, z) = C_1 + C_2 D^{\gamma}(x, z). \quad [27]$$

It is convenient to refer to the functions $F1$, $F2$, and $F3$ respectively, as the linear-with-erosion, exponential, and power-law model etch rates. In Eqs. [25] through [27], C_1 describes the erosion rate, i.e., the etch rate at zero exposure, and C_2 is the exposure normalization constant.

7.3 Numerical Integration Procedure

The evolution of the resist–developer interface contour is traced by converting Eq. [24] to a set of difference equations that are numerically integrated. To do this, we divide the contour at time T into a number of segments whose end points are labeled by the integer N. The positions of the N points are chosen such that each of the N points is (straight line) equidistant from its nearest neighbors. The normal to the contour at point N is taken to be the normal to the line joining points $N - 1$ and $N + 1$. The difference equation form of Eq. [24] is then used to determine the position of each of the end points at time ΔT later. The new set of points is used to define a smooth curve using a standard spline routine. The spline routine requires a knowledge of the slopes of the curve at the end points, and in most real situations this can be done by selecting the end points in regions of negligible exposure where the slope is the initial slope of the air–resist interface. The procedure for dividing this new contour into the same number of equal segments is then repeated, and the motion of the new end points during the next time interval ΔT is again calculated from the difference equations.

7.4 Results of Computer Calculations

Extensive computer calculations were done using the linear, linear-with-erosion, and exponential models of photoresist development. For the linear model, $C_1 = 0$ and the single constant C_2 was chosen to obtain the observed development rate at the center of the beam of 3.0 μm/minute. In the case of the linear-with-erosion and exponential models, the constants C_1 were chosen to give the observed erosion rate of 0.05 μm/minute. The constant C_2 in the exponential model was then adjusted to obtain the observed etch rate at the center of the beam. The results of some of these calculations are displayed as Figs. 13, 14, 15, and 16.

The calculations for Fig. 13 assume zero absorption in the photoresist layer and an amplitude reflection coefficient at the photoresist–copper boundary of 0.7. Fig. 13 illustrates, using the linear model without absorption, the effects of beam-waist size and focal position on the evolution of the groove. The heavy horizontal line in each figure is located in the focal plane and the vertical bar at its end indicates the position of the half-power beam-waist radius (FWHM/2) for the incident beam. In going from top to bottom of

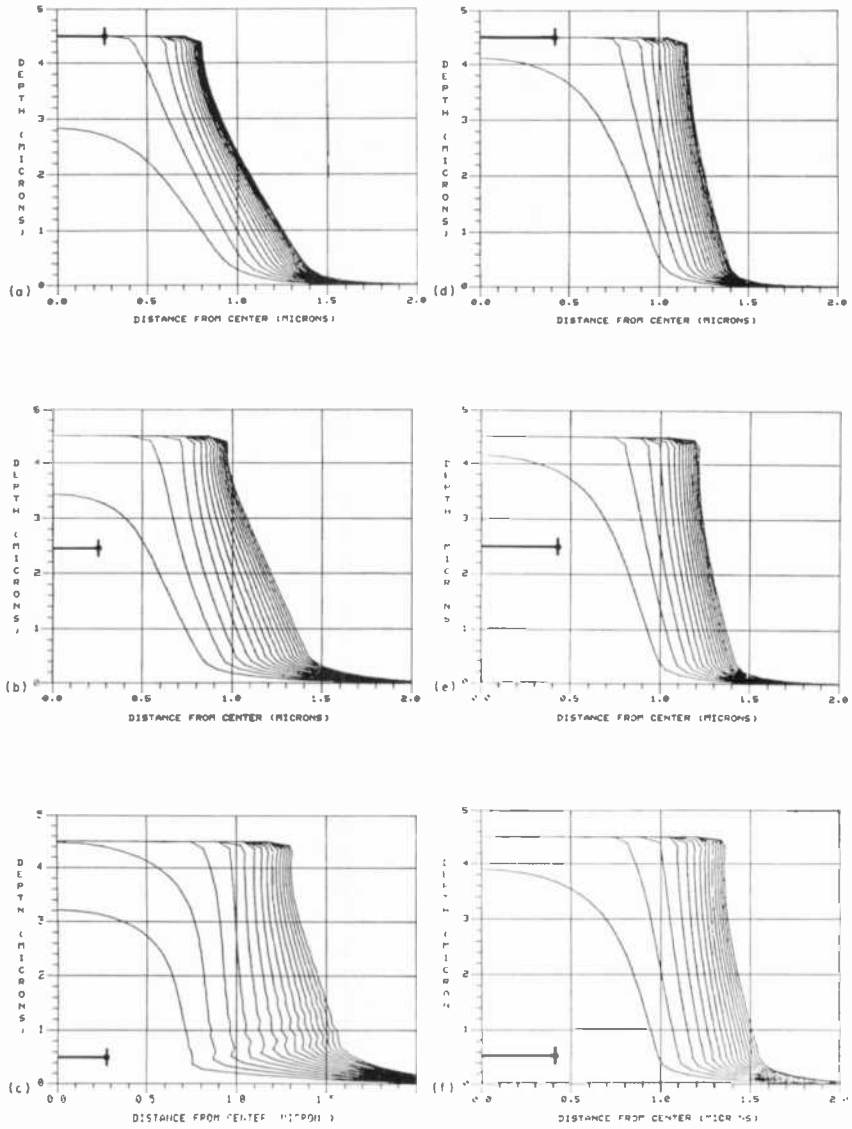


Fig. 13—Effect of focal position and spot size on the evolution of the grooves during development. The computer calculations used the linear-model, zero resist absorption, and an interface amplitude reflectivity of 0.7. Times shown are 1 to 8 minutes at 0.5 minute intervals. The vertical position of the dark horizontal line indicates the input-beam focal plane while the horizontal position of the end marker indicates the input-beam waist radius (FWHM/2). Note the foreshortened vertical scale.

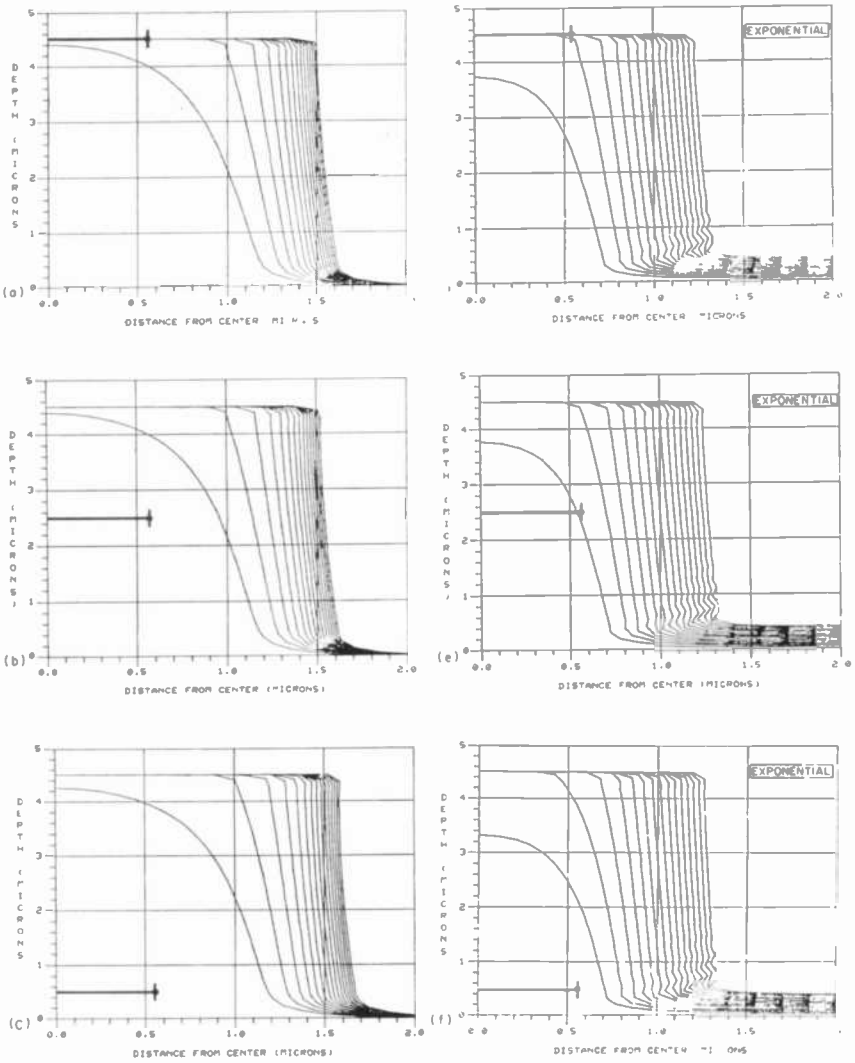


Fig. 14—Effect of focal position and model on the evolution of the grooves. Computer calculation uses linear model for (a), (b), and (c) and the exponential model for (d), (e), and (f). Zero resist absorption and an interface amplitude reflectivity of 0.7 are assumed. Times shown are 1 to 8 minutes at 0.5 minute intervals. The vertical position of the dark horizontal line indicates the input-beam focal plane, while the horizontal location of the end marker indicates beam waist radius (FWHM/2). Figs. 14 (a), (b), and (c) can also be considered as an extension of Fig. 13 to a larger beam radius. Note the foreshortened vertical scale.

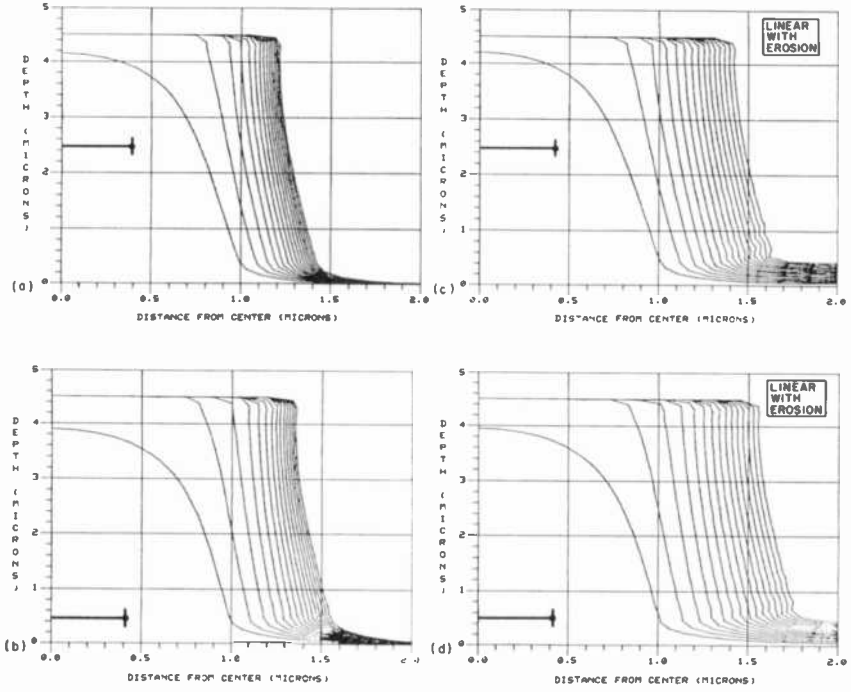


Fig. 15—Effect of focal position and erosion on the evolution of the grooves. Computer calculation uses linear model for (a), (b), and (c) and the linear-with-absorption model for (d), (e), and (f). Zero resist absorption and interface reflectivity of 0.7 are assumed. Times shown are 1 to 8 minutes at 0.5 minute intervals. The vertical position of the dark horizontal line indicates the input-beam focal plane, while the horizontal position of the end marker indicates the input-beam waist radius (FWHM/2). Note the foreshortened vertical scale.

Fig. 13, we notice the strong effect of changing focus on the groove shape, which is consistent with our experimental observations.

In Figs. 14a, 14b, and 14c, we extend the linear model to a larger spot size. Figs. 14d, 14e, and 14f present the results of calculations using the exponential model. Aside from reducing the overall groove height, the use of the exponential model calculations are not significantly different than those using the linear model. For designing optimum recording conditions, however, the linear model is probably more appropriate, because the exponential model is likely to over-estimate the effect of photoresist nonlinearities.

Similarly, Fig. 15 shows a side-by-side comparison of the results

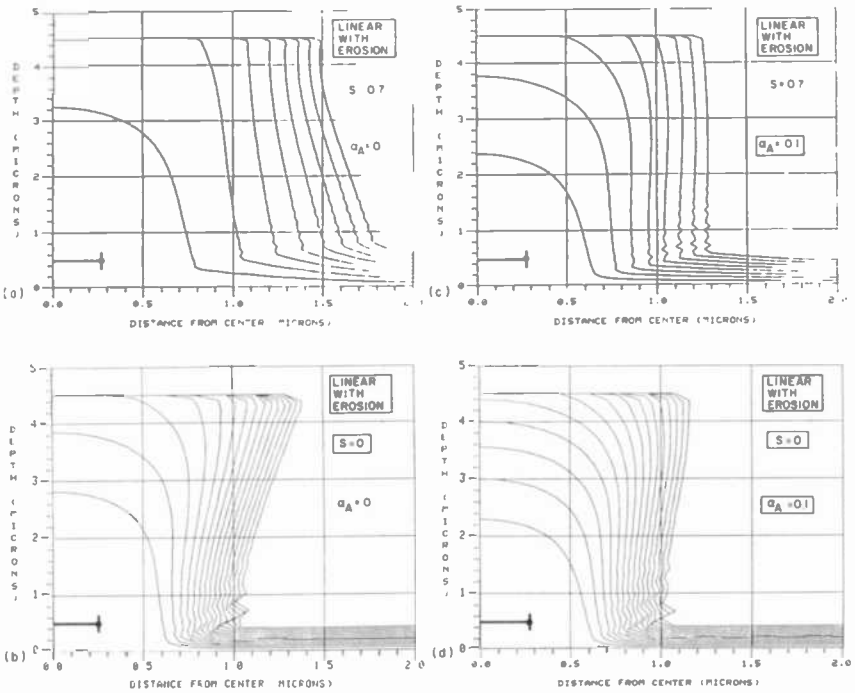


Fig. 16—Effect of changes in resist absorption and/or resist–copper interface reflection on the evolution of the grooves. The calculations were made with (a) $S = 0.7$, $\alpha_A = 0$; (b) $S = 0$, $\alpha_A = 0$; (c) $S = 0.7$, $\alpha_A = 0.1$; and (d) $S = 0$, $\alpha_A = 0.1$. Times shown for (a) and (c) are 1 to 8 minutes at 1.0 minute intervals; for (b) and (d) are 1 to 8 minutes at 0.5 minute intervals. The vertical position of the dark horizontal line indicates the input-beam focal plane, while the horizontal position of the end marker indicates the input-beam waist radius (FWHM/2). Note the foreshortened vertical scale.

of model calculations, this time for the linear model with and without erosion. Again, with the trivial exception of the reduction of the overall groove height by the linear-with-erosion model, the particular model used does not significantly effect the qualitative behavior of the groove evolution during development.

In Fig. 16 we show explicitly how absorption by the photoresist and reflectivity at the photoresist-copper interface affect the evolution of the groove profile. The beam diameter and focal position are that of Fig. 13c, which is reproduced as Fig. 16a. The effects of changing the absorption, the interface reflectivity, or both are dramatically demonstrated.

None of these calculations include the interference term of Eq. [18] explicitly, but since it is an oscillatory function whose average value is zero, the effect of neglecting this term would only be noticed if the calculations were carried out on a very small scale. Since we have often noticed the strong interference pattern that is etched into the resist walls, as can be seen in Figs. 9b, 10b, and 11b, some computer calculations seemed justified. To include such effects, however, the spatial grid of the numerical calculations must be made much finer, and when this is done the computer cost goes up rapidly. Thus only small development times were explored, and these did show the evolution of fringes spaced at approximately one half the exposure wavelength (in the resist).

Thus we see that a fine balance must be maintained between many contributing factors to enable us to generate grooves of the required profile. Note also from the theoretical curves that not only the width but also the shape is critically affected by development time as well as by the other parameters discussed.

Acknowledgments

The authors wish to acknowledge the special efforts of the following members of the Laboratory staff. The original idea for machining the stylus tips in the shape of a keel is due to E. O. Keizer. The original keel-lap recorder was assembled and early recordings were made by A. H. Firester, J. P. Russell, and M. E. Heller, with the assistance of C. B. Carroll in the mechanical design and construction. W. C. Henderson assisted with or made hundreds of test recordings. K. F. Etzold and D. W. Thompson have contributed significantly in many areas of the keel-lap effort including the development of the sectioning technique. A. Sabo has admirably handled all of the resist processing. All of our experimental views of the grooves are the result of painstaking effort by the SEM analysis group: J. H. Reisner, W. L. Lees, Jr., W. R. Campbell, 3rd, and R. E. McCoy. P. V. Valembois and E. J. Holub provided the finely machined substrates. Numerous discussions with G. Kaganowicz concerning specifications of the optimum lapping-groove profiles and with H. Scheible on resist processing techniques are gratefully acknowledged. M. Kaplan and D. Meyerhofer provided all resist-related data and were helpful with numerous discussions relating to the physical model of resist etching. Much of the computational algorithm is the result of suggestions made by R. W. Klopfenstein. System programming assistance was provided by A. W. Jessup, Jr.

References:

[1] E. O. Keizer, Keel-Tipped Stylus for VideoDisc System, U.S. Patent No. 4,162,510, July 24, 1979.
[2] A. H. Firester, C. B. Carroll, I. Gorog, M. E. Heller, J. P. Russell, and W. C. Stewart, "Optical Readout of the RCA VideoDisc," *RCA Rev.*, 39(3), p. 392, Sept. 1978.
[3] M. Born and E. Wolf, *Principles of Optics*, Fifth Edition, Pergammon Press, (1975), p. 395.
[4] P. Jacquinot and M. B. Roizen-Dossier, "Apodization," *Progress in Optics*, John Wiley & Sons, N.Y. (1963), p. 31.
[5] M. Kaplan, private communication.
[6] O. S. Heavens, *Optical Properties of Thin Solid Films*, Dover Publications, Inc., New York (1965), p. 200.
[7] M. Kaplan and D. Meyerhofer, "Response of Diazoquinone Resists to Optical and Electron-Beams Exposure," *RCA Rev.*, 40(2), p. 166, June 1979.
[8] H. Kogelnik and T. Li, "Laser Beams and Resonators," *Proc. IEEE*, 54, p. 1312, (1966).

Measurement of the Thermal Resistance of Packaged Laser Diodes

J. J. Hughes, D. B. Gilbert, and F. Z. Hawrylo
RCA Laboratories, Princeton, NJ 08540

Abstract—This paper describes a technique for measuring the thermal resistance associated with the components of a packaged laser diode: device chip, die-attachment material, intermediates (such as BeO substrate), and case-header. Results are presented for both normal (p-down) and inverted (p-up) device chip mounting to case-header. The relationship between the thermal resistance of a packaged device and the distance from the p-n junction to various components of the laser diode assembly is analyzed. The effects of high-power aging on thermal resistance are reviewed for laser diodes with lifetimes of more than one year.

Introduction

An evaluation of the thermal behavior of a laser diode mounted on a heat sink is essential for laser-performance and reliability optimization. A study was made to identify the thermal resistance associated with the chip, die-attachment material, and L-mount (case-header). The laser diodes used in the study are constricted double-heterojunction large-optical-cavity (CDH-LOC) devices.[1] These devices are high-power, single-mode lasers with 10-μm-wide contact stripes and cavity lengths of 290 μm.

Thermal resistance, θ_{JC}, in these devices is defined as the increase in the forward-biased p-n junction temperature, ΔT_j, due to power dissipation, P_{DISS}, i.e.,

$$\theta_{JC} = \Delta T_j / P_{DISS},$$

where the subscript JC denotes the junction-to-case thermal resistance.

Temperature is sensed by using the typical forward-voltage-versus-junction-temperature relationship (Fig. 1) for a forward-

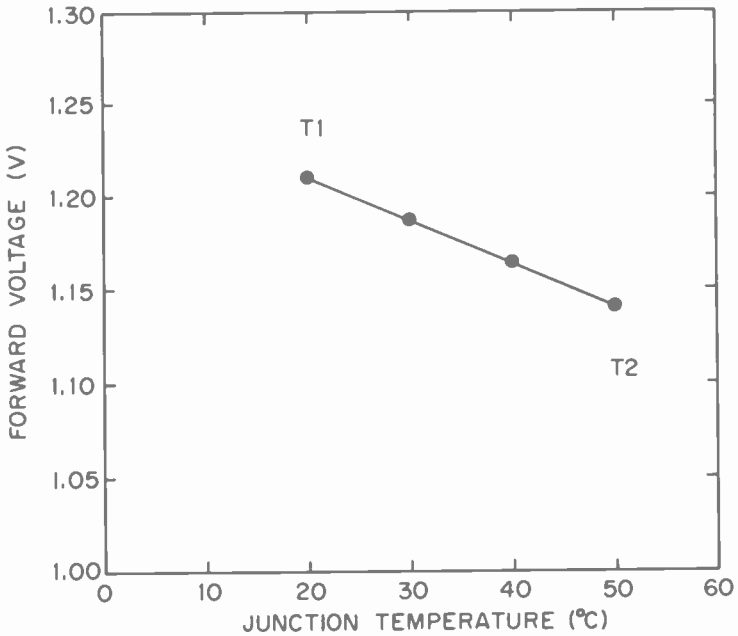


Fig. 1—Typical relationship between forward voltage and junction temperature for a forward-biased diode driven by a small, constant measurement current.

biased diode driven by a constant measurement current. The measurement current (1 mA) overcomes surface leakage effects without heating the diode. The forward voltage (V_F) at 1 mA is measured over a temperature range of 30°C for each diode. The difference between V_{F1} measured at a temperature of T_1 and V_{F2} measured at T_2 is directly proportional to the junction-temperature increase as shown in Fig. 1[2]:

$$\Delta T_j = |K(V_{F1} - V_{F2})|.$$

Here K is the reciprocal of the slope of the V_F/T_j curve with units expressed in degree Celsius per millivolt.

Fig. 2 is a schematic representation of a packaged laser diode, and Fig. 3 shows an electrical circuit analog of the heat flow for each component. Each section is modeled by a parallel R-C circuit having a thermal time constant t_θ equal to the product of the heat capacity C_θ and thermal resistance θ. The heat propagating from the junction encounters the thermal impedance of each discrete section while propagating to the case ambient. A steady state is reached when the heat capacities of the elements no longer affect

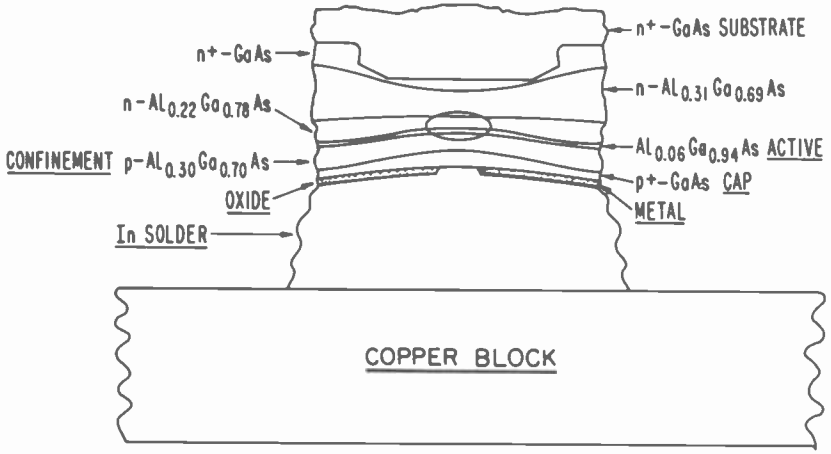


Fig. 2—Schematic representation of packaged laser diode.

the flow of heat. The sum of the thermal impedances is equal to the thermal resistance θ_{JA} from the junction through the series packaging elements to ambient air.

The variation in transient thermal behavior of the heat flow from

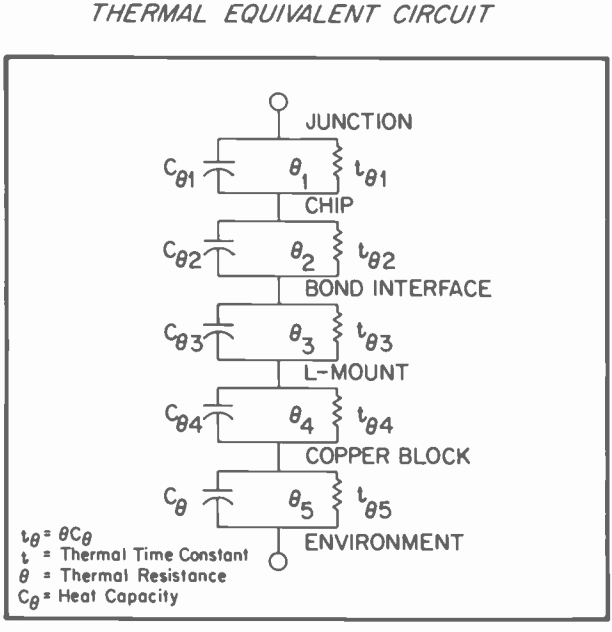


Fig. 3—Electrical circuit analog of thermal impedance for packaged laser diode.

a high temperature (junction) to a low temperature (case) can be used to measure the integrity of the individual series bonds between sections. The large differences (orders of magnitude) in time constants between sections allow use of electrically driven transient measurement techniques.

In this circuit analogy, both radiative and convective heat flow and conductive heat flow through the bond wires are ignored. These probably represent less than 10% of the total heat-flow path in our devices.

Measurement Technique

The use of electrical-transient measurements to evaluate die-bond integrity has been implemented in an instrument appropriately named a Die Attachment Evaluator.* The instrument employs a three-step procedure to measure the differences in the temperature-sensitive parameter V_F.

(1) A measurement current of ~1 mA forward current is applied to the diode. The forward voltage V_{F1} is measured.
(2) A current pulse with controlled duration and amplitude is then used to heat the diode repetitively.
(3) At the end of the heating pulse, the forward voltage V_{F2} at 1 mA is again measured.
(4) The duration of the heating pulses is then changed (in the range from 10 μs to 10 s), and the process (steps 1 through 3) is repeated.

The difference between the voltages $V_{F1} - V_{F2}$ is proportional to the change in junction temperature rise ΔT_j. The thermal resistance θ_{JX} is

$$\theta_{JX} = \Delta T_j / P_{DISS} = \frac{K|V_{F1} - V_{F2}|}{P_{DISS}},$$

where the P_{DISS} is the product of the heating current and the junction voltage during the heating pulse. The subscript X denotes a distance from the junction determined by the duration t_θ of the heating pulse. For example, if the heating pulse duration is equal to the sum of the time constants $t_{\theta 1} + t_{\theta 2}$ (shown in Fig. 3), then θ_{JX} represents the thermal resistance of the device plus the die-bond interface material.

* Sage Enterprises, Mountain View, CA, Model D.A.E. 100.

Measurement Data

Both the cw threshold current and K-factor of each diode are measured before employing the electrical-transient measurement technique. The threshold current of the device under test is used to set the heating-pulse amplitude mentioned in step 2 of the measurement procedure. The heating-pulse duration t_θ can be manually set or automatically swept through each pulse duration.

Fig. 4 shows a plot of the thermal resistance values for 42 data points from 10-μs to 10-s pulsed duration. The data were recorded on a diode mounted on a gold-plated, oxygen-free-high-conductivity (OFHC) Cu L-mount as shown in the photograph of Fig. 5. The following observations can be made before identifying the location of the series of components in the thermal path:

(1) The curves show a thermal relationship for the device under test versus time.

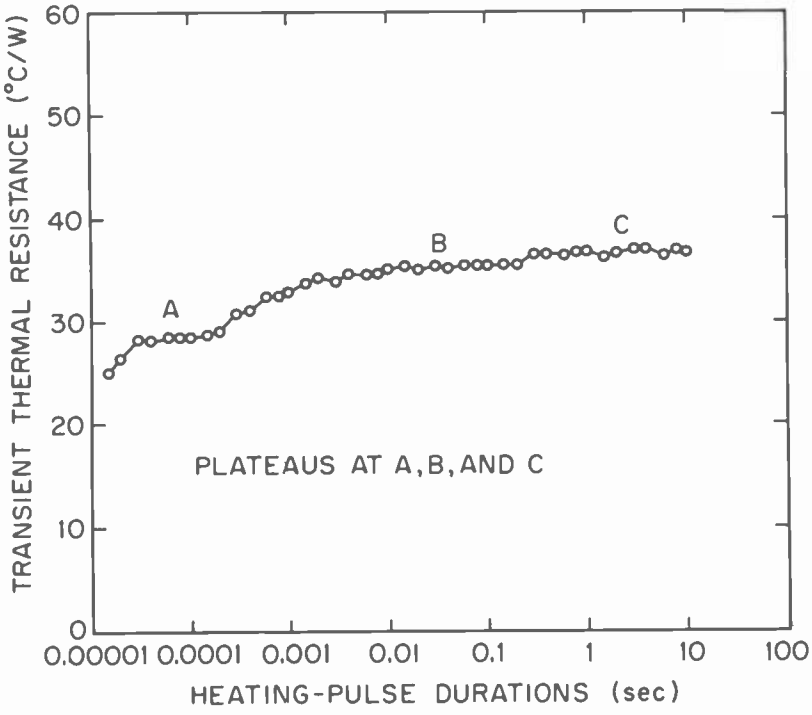


Fig. 4—Thermal resistance data curve.

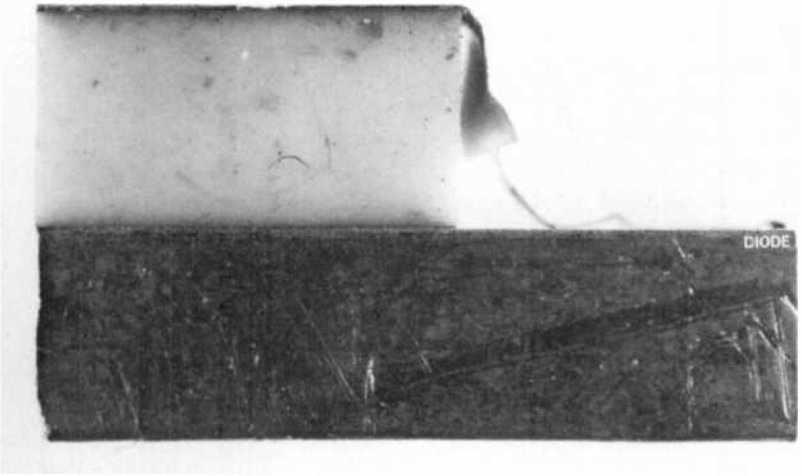


Fig. 5—L-mount package.

(2) There are three plateaus (labeled A, B, C in Fig. 4), with the first plateau being much shorter in length than the other two.
(3) The thermal time constant for the packaged device is approximately 1.0 s.

Fig. 6 shows data taken on a device having three different combinations of thermal path lengths. This measurement technique is used to identify the component of the packaged device associated with each plateau in the thermal transient data curve (Fig. 4). Curve A in Fig. 6 (initial data) was taken on a packaged device consisting of a diode-chip, indium solder, L-mount, and heat-sink copper block. Curve B was generated with the same components except that the L-mount was purposely misaligned in the main heat sink (copper block) to create a poor thermal path. Note that curves A and B are nearly identical out to the third plateau (2 s). Curve B, however, does not reach steady state due to the poor thermal path caused by the misalignment of the L-mount and copper block. In identifying the third plateau with the copper block we can conclude the following:

(1) The ambient-air plateau is not detected by our measurements and therefore the length of the third plateau associated with the copper block has not been determined.
(2) The second plateau is associated with the thermal resistance of the L-mount since it is in intimate contact with the copper block.

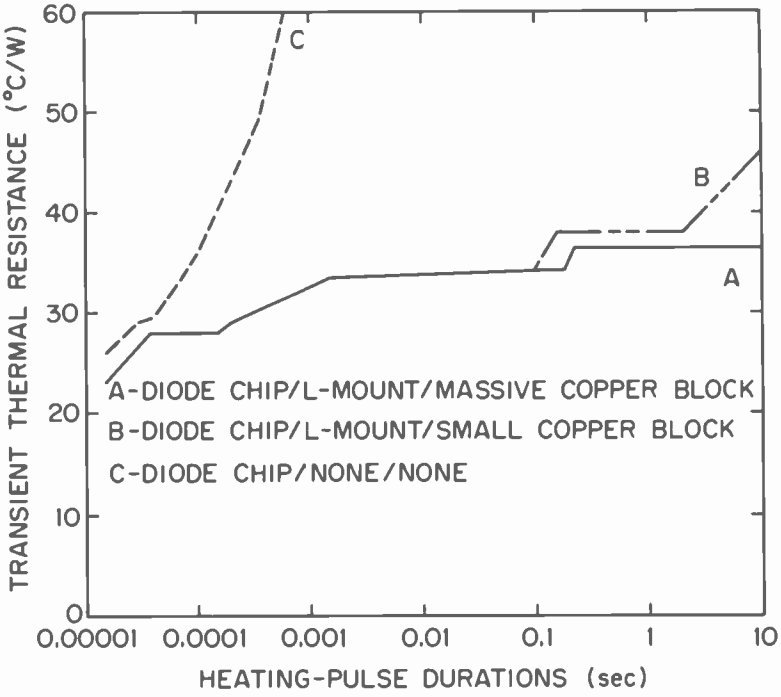


Fig. 6—Transient thermal resistance for three different combinations of thermal path lengths.

In curve C, both the L-mount and copper block are removed from the thermal path and the device is biased between two 1-mil wires without any heat-sink components. The indium solder remaining on the chip, when removed from the L-mount, is considered to be a heat spreader from the chip directly to air. Curve C approaches its plateau at 3×10^{-5} s before thermal run-away of the junction temperature. Therefore, the association with the first plateau is related to the p metallization and indium solder. Mounting defects in the indium solder (voids, etc.) would result in a shorter and higher plateau, the length of which depends on the magnitude of the bond defect.

Fig. 7 shows the relationship between the thermal resistance of a packaged device and the distance from the junction to various components of the laser-diode assembly, assuming the heat flow is approximately one dimensional. The correlation is approximate, but it avoids complicated equations relating the pulse duration to a linear distance scale; it requires that we know the physical dimensions of each layer shown in Fig. 2. This model assumes that heat

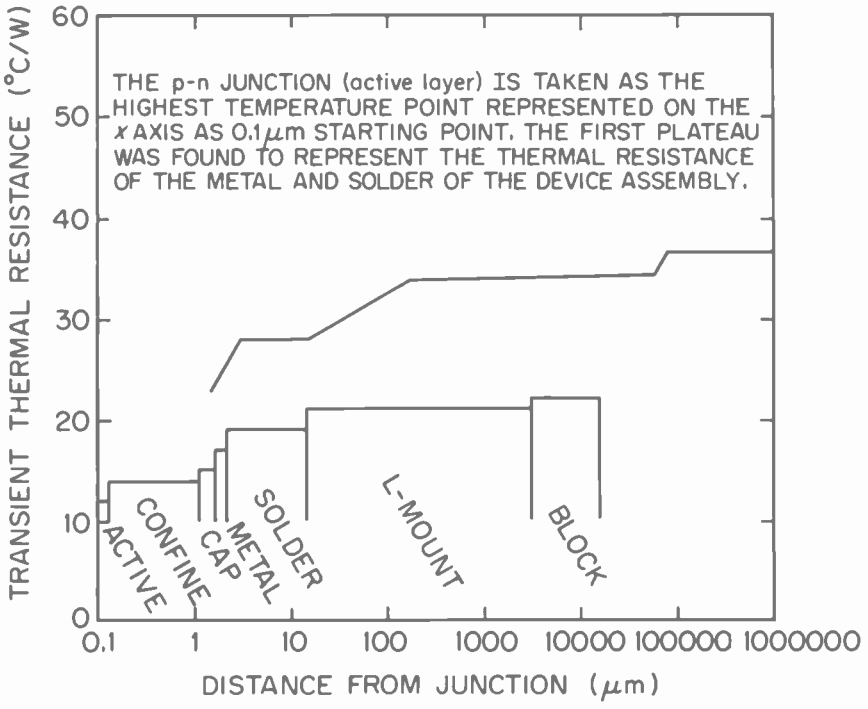


Fig. 7—Transient thermal resistance versus linear distance from junction to heat sink.

generation starts in the active layer (x axis = 0.1 μm) and heat flows straight through without spreading. The length of each layer is graphically shown along the x axis of Fig. 7 with the thermal resistance curve from Fig. 6. Although the appropriate thermal constants of each material are different, we can see that the plateaus in this graph occur at the expected locations. One observes good agreement between each plateau and the physical location of the components (indium and L-mount) identified in our study. This crude modeling is effective up to the beginning of the plateau associated with the L-mount. It fails to correlate with either the length of the L-mount or the copper blocks, because heat spreading in these components is not equivalent to that in a one-dimensional model. However, our analysis can be used to identify the components along the first and second plateaus of the data curve.

Theoretical thermal resistance values have been reported by Manning[3] on laser diodes with similar parameters to those in this study; they are 12°C/W for the diode chip, 6–11°C/W for the indium solder bond (depending on the thickness), and 7°C/W for the copper

heat sink. In our analysis, the first plateau of the data curve represents the total values (18–23°C/W) of the chip (12°C/W), metallization, and indium solder (6–11°C/W). The second plateau includes the additional 7°C/W of the L-mount (inserted in the copper heat sink) for a total of 25–30°C/W for the packaged device.

We also examined the thermal effects of a beryllium oxide (BeO) intermediate, mounted between the laser diode and the L-mount. This structure has the following advantages in the packaging process:

(1) It avoids stress-induced degradation because of the matching thermal expansions between GaAs and BeO.
(2) The intermediate substrate (BeO chip) makes laser diode alignment for optical-fiber coupling easier.

The penalty for using this low-thermal-conductivity material is an additional thermal resistance of approximately 11°C/W in the thermal path from junction to case.[4]

Fig. 8 shows the position and length of the BeO substrate (6 mils

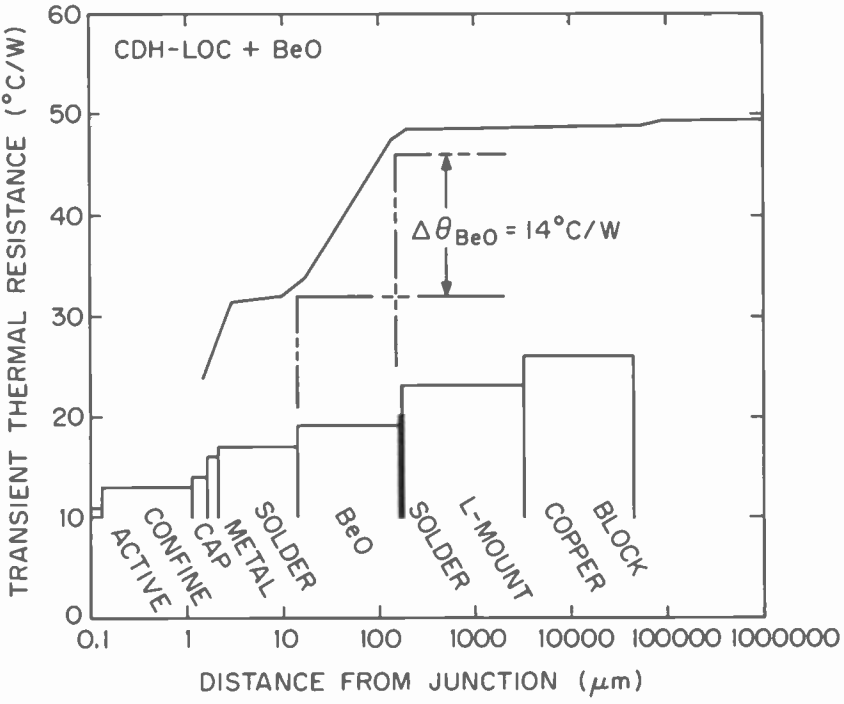


Fig. 8—Transient thermal resistance of packaged device mounted on an intermediate substrate (BeO).

thick) in the thermal path from junction to copper block. There is an increase of approximately 14°C/W from the first to the second plateau due to the addition of the BeO substrate. Note that this increase in thermal resistance is located over the BeO block shown on the x axis of the graph.

P-up Mounted Devices

The electrical-transient measurement curve for a p-up mounted device is similar to the data curve shown in Fig. 8. The GaAs substrate of the device chip behaves as an intermediate between the active region and the case-header, but has 2–3 times higher thermal resistance than BeO. For low-power or pulse-operated devices, this thermal penalty is outweighed by the following advantages:

(1) Optical obstructions and reflections caused by the die-attachment material and L-mount are minimized, because the emitting layer is removed from the mounting area.
(2) Solder voids on the p-side of the device are avoided, since only an electrical connection is required on the p-surface stripe contact.
(3) Alignment requirements for making a planar connection between the optical-cavity output area and the edge of the case-header are reduced to more reasonable tolerances.

Fig. 9 shows the measurement curves for p-up (inverted) and p-down (normal) mounted devices. Note that the first plateaus of both curves have approximately the same thermal impedance. The increase in thermal resistance caused by the inverted mounting is observed between the first and second plateaus, as was the case for the BeO intermediate. This is in agreement with our previous discussion of electrical-transient behavior. At pulse durations less than 10 μs, the heat dissipated within the wafer and the junction (active layer) through the p-side thermal path remains the same regardless of the orientation used in mounting the device. In effect, the junction-to-case-header thermal path for p-up mounted devices is from the active region toward the p-surface metallization and thick indium solder used to attach the connecting wire. The heat flows up to the indium solder, which acts as a heat spreader, and down through the device over a larger area to the L-mount.[5]

Thermal Resistance Versus High-Power Aging

Fig. 10 shows the thermal-resistance curves for a laser diode aged at 20 mW, 30°C. Each progressive curve (0, 4000, and 12,000 hours)

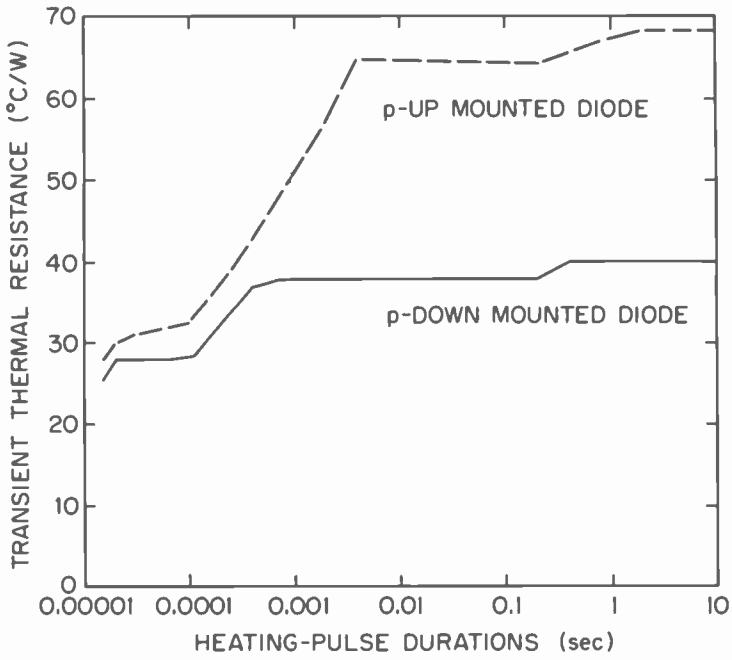


Fig. 9—Transient thermal resistance of p-up and p-down diodes in a packaged device.

shows an increase in the thermal resistance value of the first plateau, which correlates with the p-side Au metallization and die-attachment material (In) of the device. The indium solder reacts with the Au metallization to form an intermetallic compound of high thermal resistance. After 12,000 hours of continuous operation the first plateau has changed about eighty percent. A more drastic example can be seen in Fig. 11 for a device that failed after 5000 hours. Curve B is the 12,000-hour curve from Fig. 10, repeated for comparison. It is believed that, in addition to formation of an intermetallic compound (Au–In), the migration of In solder into the Au metallization leaves voids under the diode. These voids lead to the formation of hot spots in the active volume of the laser. The hot spots enhance bulk degradation of the AlGaAs active layer which, in turn, leads to failures.

Curves similar to those resulting from the formation of an intermetallic compound can be found in initial tests of laser diodes because of poor fabrication processes. Here high thermal resistance occurs at $t = 0$ h rather than after thousands of hours. The time to failure for such devices is of the order of 10–100 hours. Early fail-

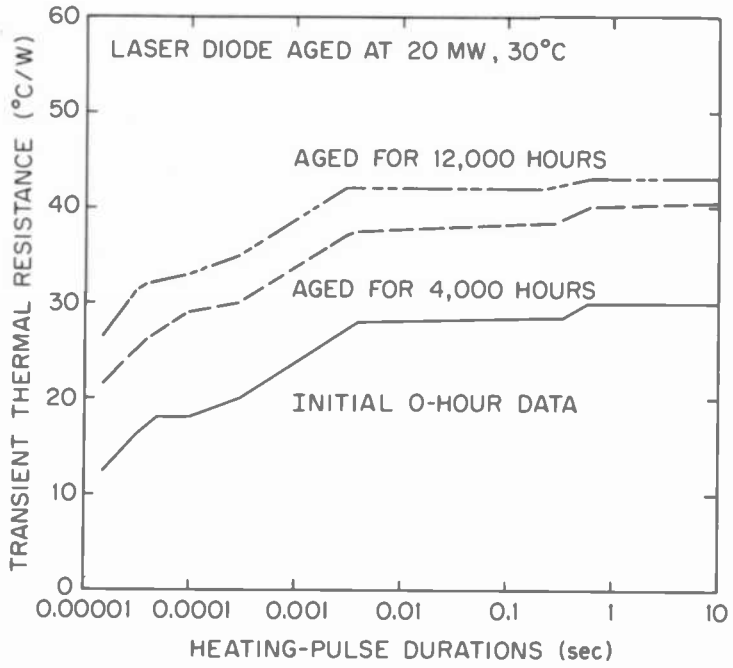


Fig. 10—Transient thermal resistance versus high-power aging.

ures of this type are characterized by poor adhesion of the metallization or oxide coating on the p^+-GaAs cap layer shown in Fig. 2.

Discussion of Data

Our study has shown an accurate method of tracking the thermal characteristics of a packaged device as a function of *time* or *distance*. The thermal-resistance measurements obtained by use of the Die Attachment Evaluator were compared to a null technique used to determine thermal resistance from the longitudinal-mode wavelength, as described by Paoli.[6] In that measurement, use is made of the increase in wavelength caused by cw heating of the active region and the reduction in heat-sink temperature necessary to return the wavelength to its value at the same pulsed current. Typical data for a CDH-LOC device showed a 2.43 Å/°C change in the longitudinal-mode wavelength for a temperature rise of 30°C (20–50°C). A shift in wavelength of 17 Å was found between pulsed and cw operation for an input power of 125 mW at 20°C heat-sink temperature. The thermal resistance value obtained by use of the null tech-

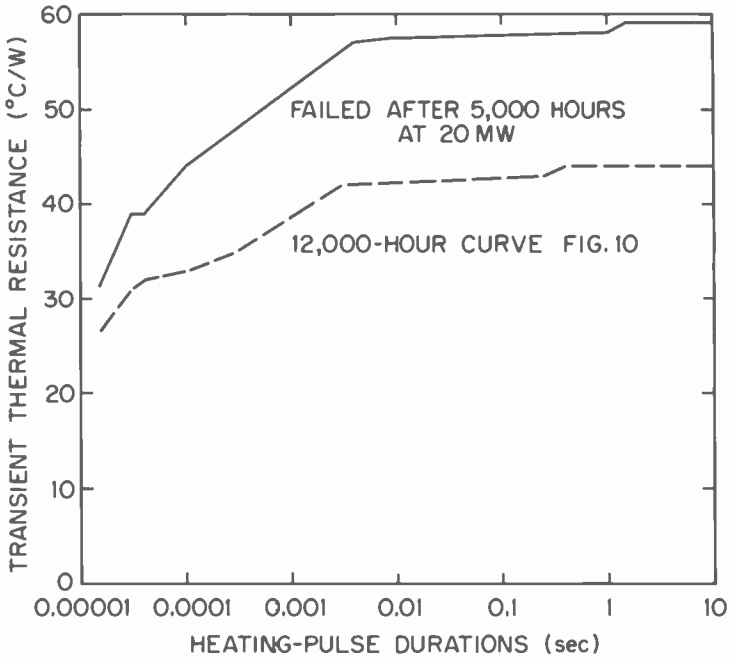


Fig. 11—Transient thermal resistance curves comparing a failed and an aged device.

nique was approximately 55°C/W. The thermal resistances θ_{JC} found by the null technique (50–57°C/W) and thermal-transient measurements (35–45°C/W) are in fair agreement.

Summary

Our analysis has been particularly useful in determining the integrity of the die-bond material and the thermal-resistance increase attributed to intermediates used to match the linear expansion coefficient of GaAs. Although the duration of the measurement pulse (10 μs) is too long to detect the junction-to-chip thermal resistance, the value of the first plateau puts a limit on the chip contribution in the thermal path from junction to die-bond material. For example, consistently high values of thermal resistance at initial test in the first plateau for devices from the *same* wafer with indium-solder or p-up mounting indicate fabrication problems. On the other hand, consistently high values of thermal resistance for a particular die-attachment material as compared to a consistently lower reference value obtained with indium as a die-bond material indicate a poor bond or damage to the chip during the bonding process.

The increase in thermal resistance attributed to an intermediate between the diode and case-header is detected between the first and second plateaus of the thermal transient curve. Also, when a large sample population is used, the difference in the thermal-resistance value attributed to the die-attachment material can be determined by measuring the difference between the first and second plateaus.

Acknowledgments

The authors wish to express their appreciation to M. G. Harvey and N. S. Klein for their valuable technical support. The authors also acknowledge the comments and technical review of this paper by D. B. Carlin and D. Redfield.

References:

[1] Botez, D., "Constricted Double-Heterojunction AlGaAs Diode Lasers—Structures and Electro-Optical Characteristics," *IEEE J. Quantum Electron.*, **QE-17**(12), 2290–2309 (1981).
[2] Siegal, B., "Electrical Transients Simplify LED Junction-Temperature Measurements," *Electro-Optical Systems Design*, Kiver Publications, Inc., Nov. 1981, pp. 47–49.
[3] Manning, J., "Thermal Impedance of Diode Lasers: Comparison of Experimental Methods and a Theoretical Model," *J. Appl. Phys.*, **52**(5), 3179–3184 (May 1981).
[4] Duda, E., Carballes, J., and Apruzzese, J., "Thermal Resistance and Temperature Distribution in Double-Heterostructure Lasers: Calculations and Experimental Results," *IEEE J. Quantum Electron.*, **QE-15**(8), 812–816 (1979).
[5] Joyce, W. B., Dixon, R. W., "Thermal Resistance of Heterostructure Lasers," *J. Appl. Phys.*, **46**(2), 855–862 (Feb. 1975).
[6] Paoli, T. L., "New Techniques for Measuring the Thermal Impedance of Junction Lasers," *IEEE J. Quantum Electron.*, **QE-11**(7), 498 (July 1975).

A Comparison of the Flux-Corrected Transport Method with Other Low-Dispersive Schemes for the Two-Dimensional Linear Advection Equation

Harvey P. Miller
RCA Government Communications Systems Div.,
Camden, NJ 08102

Abstract—The flux-corrected transport (FCT) method developed for compressible, nonlinear flows is compared with other low-dispersive, explicit, finite-difference schemes for the two-dimensional linear advection of a 2-D Gaussian initial temperature distribution of varying half-width. The flow field is specified a priori as consisting of a slowly varying, oscillating, uniform x component of velocity and a constant y component of velocity. The artificial numerical effects of diffusion (dissipation), dispersion, and anisotropy are discussed, and numerical results for both one-dimensional and two-dimensional linear advection problems are presented. Also discussed are applications of the theory to thermal dispersion of surface discharges of once-through condenser cooling water in estuaries and at coastal sites and to direct jet impingement for cooling high-heat-flux electronics components.

1. Introduction

The flux-corrected transport (FCT) method for significantly reducing numerical dispersion in convection-dominated, compressible, nonlinear flows has been developed by Boris et al.[1–3] Miller et al[4] simplified the FCT method for solving the linear problem of thermal dispersion* in a convection-dominated, two-dimensional,

* Thermal dispersion refers to the process of forced convection-thermal diffusion, that is, combined advection-diffusion.

incompressible flow field for the special case of a slowly varying oscillating component of velocity in the x direction and a constant component of velocity in the y direction. An explicit FCT method for these simplified flow conditions has been compared by Miller[5] with various other low-dispersive, explicit, finite-difference schemes for the one-dimensional example problem of the incompressible forced convection of an initially square-wave distribution of the transported fluid property. The simplified FCT method has also been compared with methods by Fromm,[6] Leith,[7] and Godunov.[8] It is clear from this numerical comparison that the FCT method is the best choice for eliminating numerical dispersion for a square test wave.

A better comparison of the FCT method for a square test wave, however, would be with comparable convective approximations, such as Fromm's monotonic scheme by VanLeer,[9] which is also formally second-order accurate. Forester[10] provides an extensive comparison for the one-dimensional linear advection problem using a square wave and a sharply peaked Gaussian distribution of half-width $2\Delta x$ and a Courant number of 0.1. Zalesak[11] presents a new algorithm for implementing the critical flux-limiting stage in multidimensions without resorting to time-splitting; the clipping phenomenon associated with the original FCT flux-limiter is also eliminated or alleviated. Note, that Zalesak also used a Gaussian distribution of half-width $2\Delta x$ with a Courant number of 0.1 to illustrate the improvement of the new flux-limiter, which results in reduced clipping as compared with the original flux-limiter of Boris et al. Thompson[12] notes that a further reduction in clipping is possible in one dimension by projecting the provisional solution values to form peaks between grid points and including these peaks on either side of the point in question in the determination of the extrema.[11] Zalesak[13,14] presents the results for the linear advection of a square test wave using various low-order and high-order schemes for the FCT algorithms, as well as for tests of linear advection using FCT algorithms for the sharply peaked Gaussian distribution of half-width $2\Delta x$ used by Forester. A noted increase in accuracy is observed as the order of the spatial derivatives is increased.

The computational noise, i.e., short-wavelength spatial oscillations occurring near steep gradients, can be controlled by filters that smooth out the oscillations (referred to as dispersion "wiggles" by Gresho[15]) by selectively applying large local dissipation, i.e., artificial diffusion. The Shuman filter was used by Vliegenthart[16], but he applied too much diffusion for realistic use. The switched Shuman filter was applied in a hybrid scheme by Harten and Zwas[17]

and by Harten and Tal-ezer[18] in second- and fourth-order schemes of the Crank-Nicholson type to control oscillations occurring with shocks. In Ref. [12], Thompson summarizes current research on the use of filters.[10,19–22]

The FCT method with filtering was applied by Weber et al[23] using the original flux-limiter, and the FCT method employing time-splitting was used by Anderson.[24] VanLeer[25] gives a flux-limiter that is more compact than that used in the original FCT algorithm and compares it to the FCT limiter for convection of a triangular wave. The clipping phenomenon characteristic of monotonicity-preserving algorithms[26] is still observed, but the VanLeer limiter does give a somewhat smoother solution. The monotonic Fromm scheme introduced by VanLeer[9] uses Fromm's second-order composite scheme,[6] which is formed as an average of schemes with lagging and leading phase error (so-called method of zero average phase error), and is made monotonic by including a flux-limiter composed of nonlinear feedback terms. As noted by Thompson, the diffusion is naturally increased, but dispersion oscillations (or wiggles) are virtually eliminated in the monotonic form. VanLeer[26] determined that the minimum artificial diffusion for second-order methods corresponds to that in the Lax-Wendroff method,[27] which reduces to Leith's method[7] for linear advection. VanLeer also determined the minimum artificial diffusion required for monotonicity. This corresponds to the first-order Godunov method.

After a review of the above-mentioned studies, it was decided to compare the simplified FCT method in two dimensions using time-splitting with the original method by Fromm,[6] Leith's method,[7] and Godunov's method.[8] The initial distribution is a two-dimensional Gaussian with half-widths of $2\Delta x$, $4\Delta x$, and $8\Delta x$ (note, $\Delta x = \Delta y$). The advection field consists of a uniform, slowly varying, oscillating x component of velocity, and a constant y component of velocity. In Sec. 3 of this paper, Fromm's scheme is extended to this case of an oscillating x-component of velocity, and the two-dimensional solution is obtained for all the explicit schemes by time-splitting. In Sec. 4, the four explicit finite-difference schemes are then compared (1) qualitatively in terms of dissipation (or diffusion) and dispersion (wiggles) of the resulting convected, two-dimensional, Gaussian distributions of varying half-width and (2) quantitatively in terms of the L_1-norm for measuring error in different-size regions in the regular cartesian grid system.

2. Governing Equation for Thermal Dispersion

The incompressible, two-dimensional dispersion of heat in a tur-

bulent uniformly varying flow field can be expressed in cartesian coordinates as[4]

$$\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} = B_x\frac{\partial^2 T}{\partial x^2} + B_y\frac{\partial^2 T}{\partial y^2} + S. \quad [1]$$

The left side of Eq. [1] represents the convective change upon T, and the right side represents turbulent eddy diffusion and a source term. Definitions of symbols in Eq. [1] and throughout the paper are given in the Appendix. With appropriate initial and boundary conditions, Eq. [1] describes the thermal dispersion in a particular domain of interest. For a slowly varying, uniform velocity field (u,v), Eq. [1] need not be cast into the conservative form,[28] although Boris et al developed the FCT method for a system of nonlinear, compressible conservation equations. Note, the source term S does not affect numerical stability and convergence in thermal dispersion models, and, therefore, it will not be considered here.

For problems where convection dominates eddy transport, Eq. [1] reduces to the linear hyperbolic equation[29]

$$\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} = 0, \quad [2]$$

where, for a uniform flow field, the incompressible form of the continuity equation is identically satisfied. The momentum equations need not be solved simultaneously, provided the temporal variations of u and v are known a priori. For the convection-dominated linear thermal-dispersion equation presented here

$$u(t) = u_{max} \sin(\omega t) \quad [3]$$

$$v = c_1. \quad [4]$$

Hence, the linear advection problem is to be solved numerically for a slowly varying, oscillating x component of velocity and a constant y component of velocity. The thermal dispersion of the two-dimensional Gaussian initial-temperature distribution is given by

$$T(x, y, t = 0) = T_\infty + \Delta T \exp\{-r^2/L'^2\}, \quad [5]$$

where

$$r^2 = (x - x_0)^2 + (y - y_0)^2 \quad [6]$$

$$T(x_0, y_0, t = 0) = T_{max} \quad [7]$$

$$\Delta T = T_{max} - T_\infty. \quad [8]$$

The boundary conditions for the regular cartesian grid system used

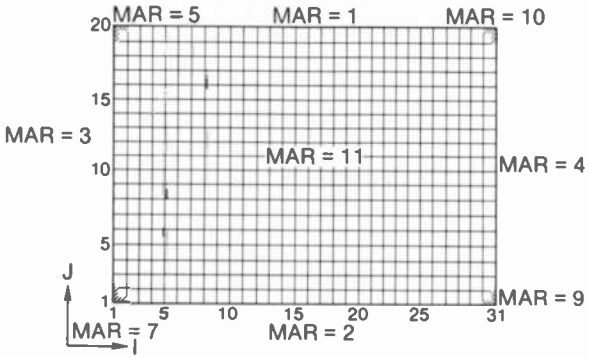


Fig. 1—Grid system and MAR node values.

are assumed to be the adiabatic thermal condition on both x and y boundaries.

3. Method of Solution

A two-dimensional regular cartesian grid system in (x,y) is established, where $\Delta x = \Delta y$. Fig. 1 illustrates the regular rectangular grid system, where a node-numbering matrix, MAR(I,J), is used for computational convenience in selecting differencing schemes for interior nodes, corners, and x and y boundaries.[30] The corner and x and y boundaries identified by MAR(I,J) = 1, 2, 3, 4, 5, 7, 9, 10 are considered adiabatic. Following Roache,[31] the adiabatic lateral wall (or open boundary) conditions are specified as zero-gradient, i.e., on x and y boundaries,

$$T_w = T_{w+1}. \qquad [9]$$

Eq. [2] has been solved using forward-time explicit finite-difference schemes following Boris et al,[2] Fromm,[6] Leith,[7] and Godunov.[8] The dominant transport process, upon which the integration time-step, Δt, is based for an explicit scheme is

$$\Delta t = \text{MIN}(\Delta t_{cx}, \Delta t_{cy}) \qquad [10]$$

where

$$\Delta t_{cx} \leqslant \frac{1}{2} \frac{\Delta x}{u_{max}} \qquad [11]$$

$$\Delta t_{cy} \leqslant \frac{1}{2} \frac{\Delta y}{v_{max}} \qquad [12]$$

3.1 The Simplified FCT Method in Two Dimensions

The explicit FCT method[2] for convection-dominated thermal dispersion can be simplified for incompressible, two-dimensional flow for a slowly varying uniform velocity (specified a priori) by using time-splitting in two steps.[4] For the x cycle, using the Boris et al notation,

$$F^n_{I+1/2} = \frac{1}{8}(T^n_{I+1} - T^n_I) \quad [13]$$

$$F^d_I = T^n_I + (f^n_{I+1/2} - f^n_{I-1/2}) \quad [14]$$

$$\epsilon^{\pm}_I = \frac{1}{2} \pm u^n_I (\Delta t/\Delta x) \quad [15]$$

$$Q^+_I = \epsilon^-_I / (\epsilon^+_{I+1} + \epsilon^-_I) \quad [16]$$

$$Q^-_{I+1} = 1 - Q^+_I \quad [17]$$

$$F^{td}_I = 4(Q^+_I)^2 f^n_{I+1/2} - 4(Q^-_I)^2 f^n_{I-1/2} + Q^+_I (T^n_I) + Q^-_I (T^n_I) \quad [18]$$

$$\delta F^t_I = F^{td}_I - F^d_I \quad [19]$$

$$\delta F^{td}_{I+1/2} = F^{td}_{I+1} - F^{td}_I \quad [20]$$

$$f^t_{I+1/2} = f^n_{I+1/2} + \frac{1}{8}(\delta F^t_{I+1} - \delta F^t_I) \quad [21]$$

$$f^c_{I+1/2} = \mathrm{SGN}(f^t_{I+1/2}) \,\mathrm{MAX}\{0, \mathrm{MIN}[\mathrm{SGN}(f^t_{I+1/2})\, \delta F^{td}_{I-1/2}, \mathrm{ABS}(f^t_{I+1/2}), \mathrm{SGN}(f^t_{I+1/2})\, \delta F^{td}_{I+3/2}]\} \quad [22]$$

$$T^{n+1}_I = F^{td}_I - (f^c_{I+1/2} - f^c_{I-1/2}) \quad [23]$$

The operator $L[T^n_I, u^n_I, \Delta x, \Delta t]$ is introduced to denote Eqs. [13] through [23], which advances $\{T^n_I\}$ by a time step Δt on a grid spacing Δx. Then for the x cycle,

$$T^{n+1}_I = L[T^n_I, u^n_I, \Delta x, \Delta t]. \quad [24]$$

The x half cycle and the x whole cycle have been reduced to a one-step x cycle by considering that $v_x{}^o = v_x{}^{hx} = u$ over a time step, Δt, for a uniform slowly varying x component of velocity. Similarly, by considering $v_y{}^x = v_y{}^{hy} = v$ for a constant or uniform slowly varying y component of velocity, the y half cycle and the y whole step[2] have been reduced to a one-step y whole step as follows:

$$T^{n+1}_{I,J} = L[T^{n+1}_I, v^n_J, \Delta y, \Delta t]. \quad [25]$$

3.2 Fromm's Method in Two Dimensions

The explicit Fromm method of zero average phase error[6] for a uniform flow field can be extended to solve the two-dimensional linear advection problem with an oscillating x component of velocity as follows:

$$\alpha^n_{I,J} = \alpha^n = \frac{u^n \, \Delta t}{\Delta x} \qquad [26]$$

$$\beta^n_{I,J} = \beta^n = \frac{v^n \, \Delta t}{\Delta y}. \qquad [27]$$

Then, for $\alpha > 0$,

$$\begin{aligned} \tilde{T}_{I,J} = {} & T^n_{I,J} + \frac{\alpha^n}{4}(T^n_{I-1,J} - T^n_{I+1,J} + T^n_{I-2,J} - T^n_{I,J}) \\ & + \frac{(\alpha^n)^2}{4}(T^n_{I-1,J} - 2T^n_{I,J} + T^n_{I+1,J}) \\ & + \frac{(\alpha^n)^2 - 2\alpha^n}{4}(T^n_{I-2,J} - 2T^n_{I-1,J} + T^n_{I,J}), \end{aligned} \qquad [28]$$

and for $\alpha < 0$,

$$\begin{aligned} \tilde{T}_{I,J} = {} & T^n_{I,J} + \frac{\alpha^n}{4}(T^n_{I-1,J} - T^n_{I+1,J} + T^n_{I,J} - T^n_{I+2,J}) \\ & + \frac{(\alpha^n)^2}{4}(T^n_{I-1,J} - 2T^n_{I,J} + T^n_{I+1,J}) \\ & + \frac{(\alpha^n)^2 + 2\alpha^n}{4}(T^n_{I,J} - 2T^n_{I+1,J} + T^n_{I+2,J}). \end{aligned} \qquad [29]$$

Then, for $\beta > 0$,

$$\begin{aligned} T^{n+1}_{I,J} = {} & \tilde{T}_{I,J} + \frac{\beta^n}{4}(\tilde{T}_{I,J-1} - \tilde{T}_{I,J+1} + \tilde{T}_{I,J-2} - \tilde{T}_{I,J}) \\ & + \frac{(\beta^n)^2}{4}(\tilde{T}_{I,J-1} - 2\tilde{T}_{I,J} + \tilde{T}_{I,J+1}) \\ & + \frac{(\beta^n)^2 - 2\beta^n}{4}(\tilde{T}_{I,J-2} - 2\tilde{T}_{I,J-1} + \tilde{T}_{I,J}). \end{aligned} \qquad [30]$$

The tilde in Eqs. [28], [29], and [30] indicates fractional time-step values resulting from the oscillating x component of velocity, which

causes convection in the x direction. Then, by conventional time-splitting, convection in the y direction for $v = c_1$ (>0) is given by Eq. [30] for advancing the solution over one time step, Δt. Strang[32,33] gives other time-splitting methods, but there appeared to be no need in this study to implement these methods.

3.3 Leith's Method in Two Dimensions

The Lax-Wendroff method[27] for reducing numerical dispersion reduces to Leith's method[7] for the linear advection Eq. [2] in two dimensions for a uniform slowly varying velocity field (u,v). Leith's method with time-splitting is given as

$$\tilde{T}_{I,J} = T^n_{I,J} - \frac{\alpha^n}{2}(T^n_{I+1,J} - T^n_{I-1,J}) + \frac{(\alpha^n)^2}{2}(T^n_{I+1,J} - 2T^n_{I,J} + T^n_{I-1,J}) \quad [31]$$

and,

$$T^{n+1}_{I,J} = \tilde{T}_{I,J} - \frac{\beta^n}{2}(\tilde{T}_{I,J+1} - \tilde{T}_{I,J-1}) + \frac{(\beta^n)^2}{2}(\tilde{T}_{I,J+1} - 2\tilde{T}_{I,J} + \tilde{T}_{I,J-1}). \quad [32]$$

Leith's method is a one-step, two-level, second-order accurate method. The one-dimensional version has been given by Noh and Protter.[34] For convection in the x direction, the Courant number $\alpha^n = u^n\Delta t/\Delta x$ can be regarded as an interpolation parameter.[31] For $\alpha^n = \alpha = 1$, no grid interpolation is required and the exact solution would obtain. The numerical stability restriction NORM(α) ≤ 1 is required to ensure interpolation between adjacent grid points, $(I + 1)$ and $(I - 1)$, rather than extrapolation. Also, for $\beta^n = \beta = v^n\Delta t/\Delta y = c_1\Delta t/\Delta y$, which is less than unity for this study, the y component of convection satisfies the numerical stability criterion.

3.4 Godunov's Method in Two Dimensions

The Godunov double-sweep method[8,35] is quite similar in form to the two-step Lax-Wendroff method except that $\Delta t/(2\Delta x)$ is replaced

by $\Delta t/\Delta x$, and $\Delta t/(2\Delta y)$ is replaced by $\Delta t/\Delta y$. For a uniform-velocity field, Godunov's method with conventional time-splitting is given as

$$\tilde{T}_{I,J} = T^n_{I,J} - \frac{\alpha^n}{2}(T^n_{I+1,J} - T^n_{I-1,J}) + (\alpha^n)^2 (T^n_{I+1,J} - 2T^n_{I,J} + T^n_{I-1,J}) \quad [33]$$

and,

$$T^{n+1}_{I,J} = \tilde{T}_{I,J} - \frac{\beta^n}{2}(\tilde{T}_{I,J+1} - \tilde{T}_{I,J-1}) + (\beta^n)^2 (\tilde{T}_{I,J+1} - 2\tilde{T}_{I,J} + \tilde{T}_{I,J-1}). \quad [34]$$

This method is a first-order explicit finite-difference scheme in Δt, but as will be illustrated in a numerical example of the two-dimensional linear advection of a two-dimensional Gaussian initial temperature distribution of varying half-width ($2\Delta x$, $4\Delta x$, $8\Delta x$), Godunov's scheme does possess better dispersion-reducing properties than Leith's scheme. However, this is shown to be offset by the additional artificial diffusion added, which causes significant dissipation (or diffusion) in the region of steep gradients of the two-dimensional Gaussian distribution.

4. Results and Discussion

Miller et al[4] constructed a numerical example using the case of one-dimensional thermal dispersion with a specified, uniform, slowly varying, oscillating x component of velocity for the incompressible forced convection of an initially square-test-wave distribution of temperature. In the study, an explicit FCT method in one dimension was compared with a conventional upwind differencing method. Significant dissipation (or diffusion), of the numerical solution is shown for the upwind differencing scheme in one dimension, whereas this severe dissipation was not evidenced by the explicit FCT method.[4]

In another numerical study, Miller[5] compared the explicit FCT method in one dimension, for the same square test wave, with the low-dispersive methods of Fromm,[6] Leith,[7] and Godunov.[8] The results of that numerical example are shown in Fig. 2 for the forced convection of an initially square test wave for the transported fluid temperature. The values of grid size, time step, turbulent eddy thermal diffusivity, oscillating x-component of velocity,[36] and initial

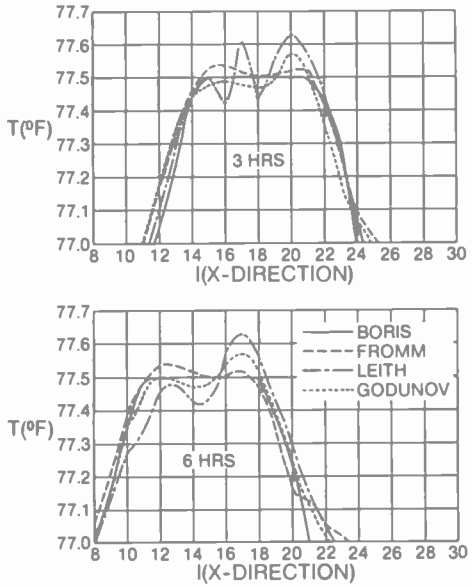


Fig. 2—One-dimensional comparison of FCT method with other low-dispersive schemes (Δt = 6 min).

temperature profile for this one-dimensional linear advection problem are given as*

Δx = 2025 ft (617.2 m)
Δt = 360 sec
B_x = 6 ft^2/sec (0.56 m^2/sec)
u = 1.0 sinωt, ft/sec (0.3 sinωt, m/sec) in which, $\omega = 2\pi f = 2\pi/12$ (rads/hr)
$T(x, t = 0)$ = 77.5°F (25.3°C) for $17 \leq I \leq 26$
$T(x, t = 0)$ = 77.0°F (25°C) elsewhere.

As can be seen in Fig. 2, only the explicit FCT method eliminates the dispersive effects exhibited by the three other low-dispersive methods (Fromm, Leith, and Godunov). Fromm's method[6] yields the least dispersion of the three; Godunov's method[8] is next in minimizing dispersion; and the most dispersive is Leith's method.[7] Again, Forester[10] indicates that VanLeer's monotonic Fromm method[9] would be a strong competitor with the explicit FCT method

* These values of current and eddy diffusivity were taken from an existing field data base for a convection-dominated coastal site.[36]

for the one-dimensional linear advection of an initially square test wave.

The distortion of the numerical solutions in the neighborhood of the end points is shown in Fig. 2. A similar distortion was indicated by Boris et al[1] for a square test wave. Sod[29] points out that there is a minimal numerical diffusion across the contact discontinuity with each time step, causing the contact discontinuity to be smeared in the course of the calculation. With the square-wave distribution, the clipping phenomenon that plagues most versions of FCT is of minor significance.

In this paper, the simplified FCT method in two dimensions, using conventional time-splitting for the linear advection Eq. [2] for incompressible flow, is compared with the original method by Fromm, Leith's method, and Godunov's method for predicting the linear advection of a two-dimensional Gaussian initial temperature distribution of varying half-width ($2\Delta x$, $4\Delta x$, $8\Delta x$). VanLeer's monotonic Fromm method[9] was not included, since results showed that the original Fromm scheme without flux-limiting was a good choice for the linear advection problem for steep gradients and exhibited negligible oscillations (shown by the isotherm contours in Figs. 3–8). Furthermore, as pointed out by Thompson,[12] VanLeer's flux-limiter increased diffusion of steep gradients, although dispersion effects were virtually eliminated. Zalesak's[11] fully multidimensional FCT would have improved the clipping exhibited by the simplified FCT method presented in this study. However, presenting these numerical results indicates that time-splitting the Book et al[2] transport operator for this linear advection problem does not result in the computational difficulties experienced by Zalesak.[11,37]

The important distinction to be noted in this linear advection problem is that each of the terms of the continuity equation are zero for this case. Therefore, the numerical difficulty of generating nonphysical temperatures (severe undershoot and overshoot of temperatures,[38]), which were traced by Zalesak to the false compressibility of the incompressible fluid when seen in each of the splitting directions (i.e., $\partial u/\partial x \neq 0$ and $\partial v/\partial y \neq 0$, which yields nonvanishing divergence when seen in the x and y splitting directions) did not result. For this linear advection problem,

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} = 0,$$

and, hence $\nabla \cdot \mathbf{V} = 0$ is satisfied in each direction. This is an important distinction and was the reason for using the time-split version of FCT for this study. A further comparison can be made with

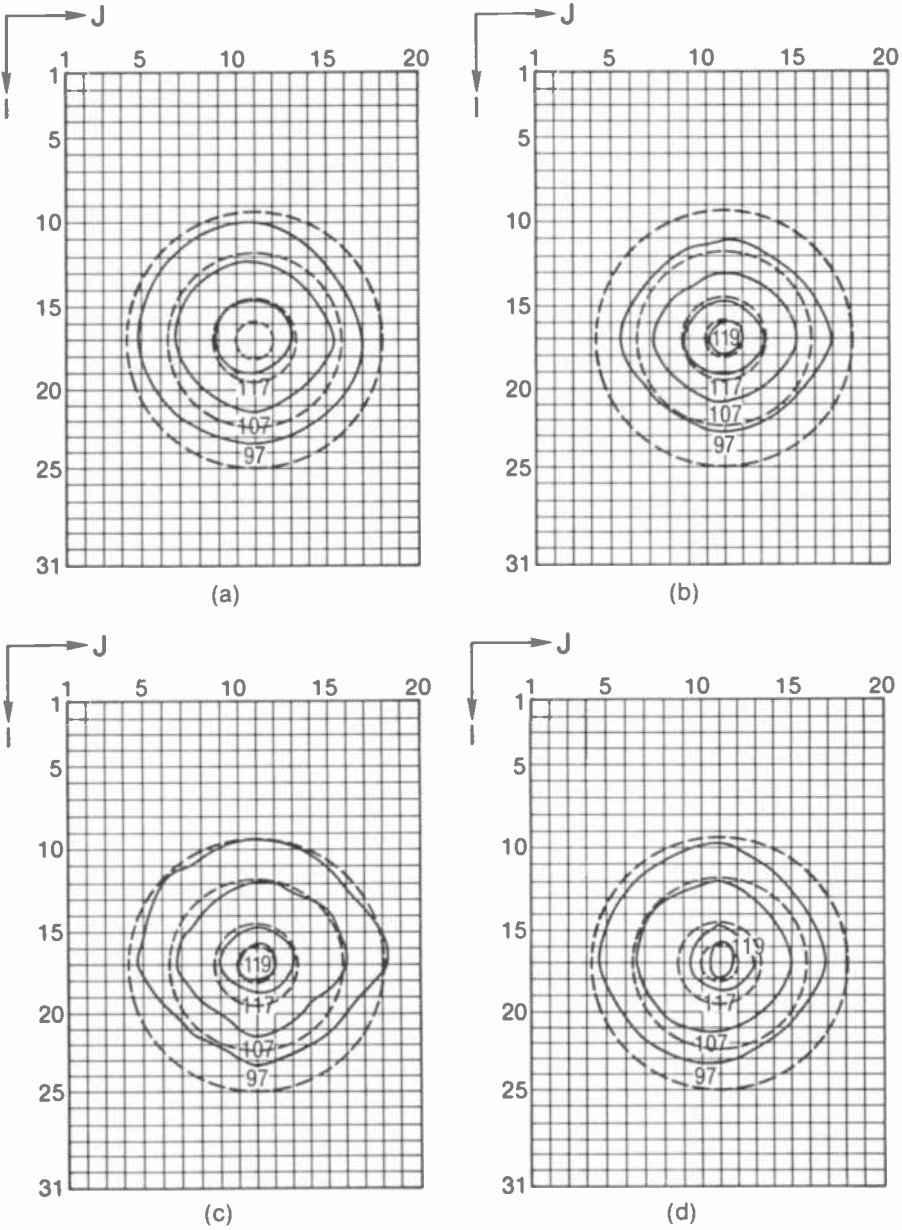


Fig. 3—Isotherms, for half-width of $8\Delta x$ at $t = 6$ hours, for (a) Boris 2D; (b) Fromm 2D; (c) Leith 2D; and (d) Godunov 2D. Dashed contours are shown for the exact solution at $T = 119$, 117, 107, and 97°F and $\Delta t = 6$ min.

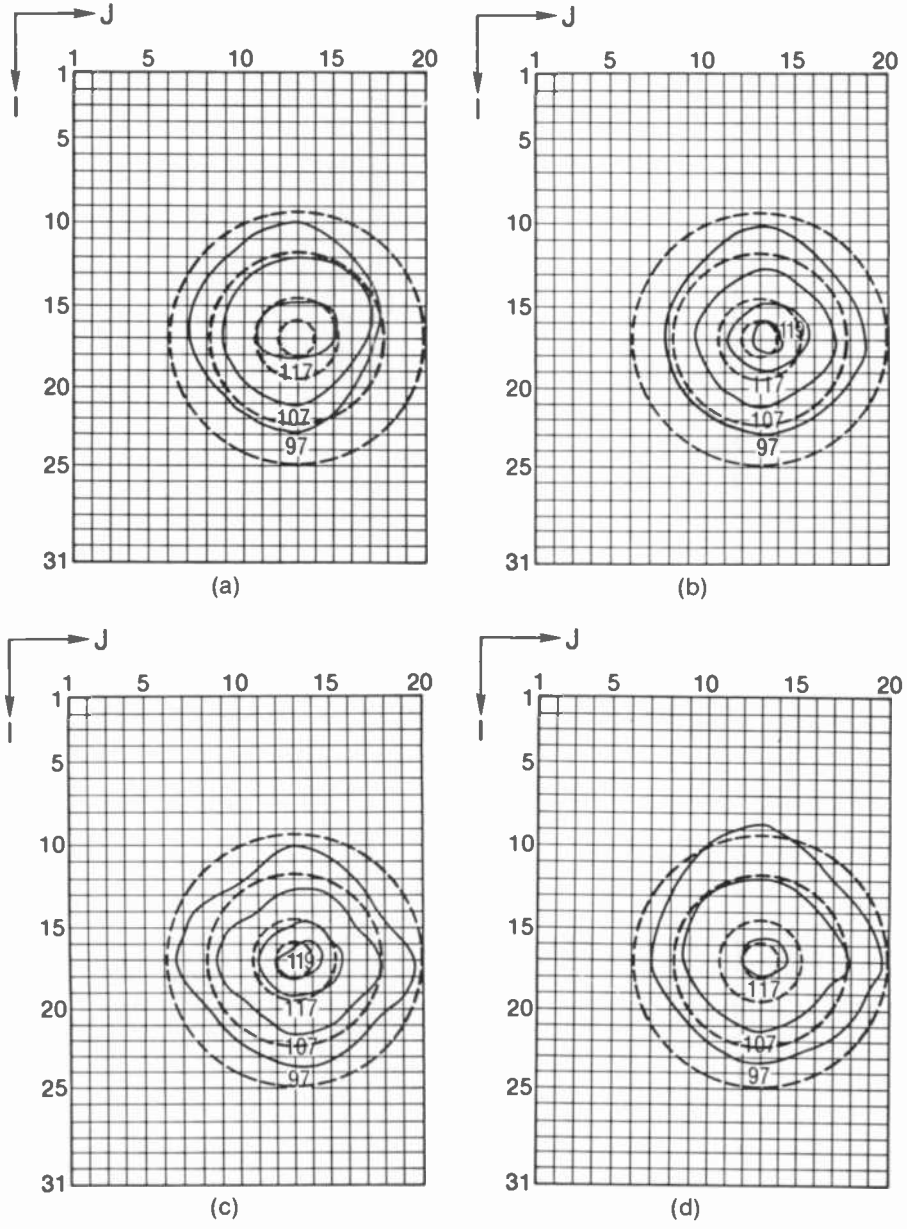


Fig. 4—Isotherms, for half-width of $8\Delta x$ at $t = 18$ hours, for (a) Boris 2D; (b) Fromm 2D; (c) Leith 2D; and (d) Godunov 2D.

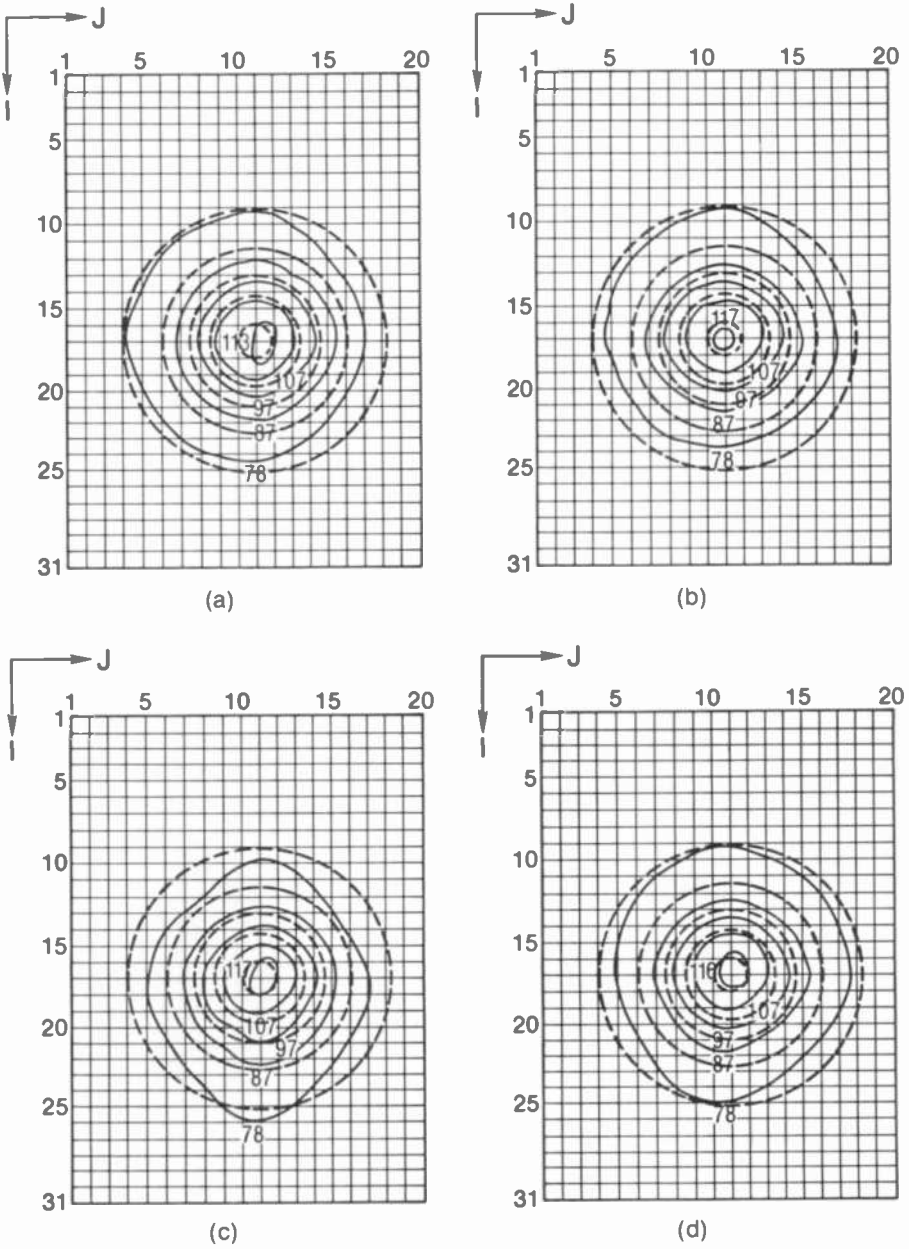


Fig. 5—Isotherms, for half-width of $4\Delta x$ at $t = 6$ hours, for (a) Boris 2D; (b) Fromm 2D; (c) Leith 2D; and (d) Godunov 2D. Dashed contours are shown for the exact solution at $T = 117, 107, 97, 87$, and 78°F and $\Delta t = 6$ min.

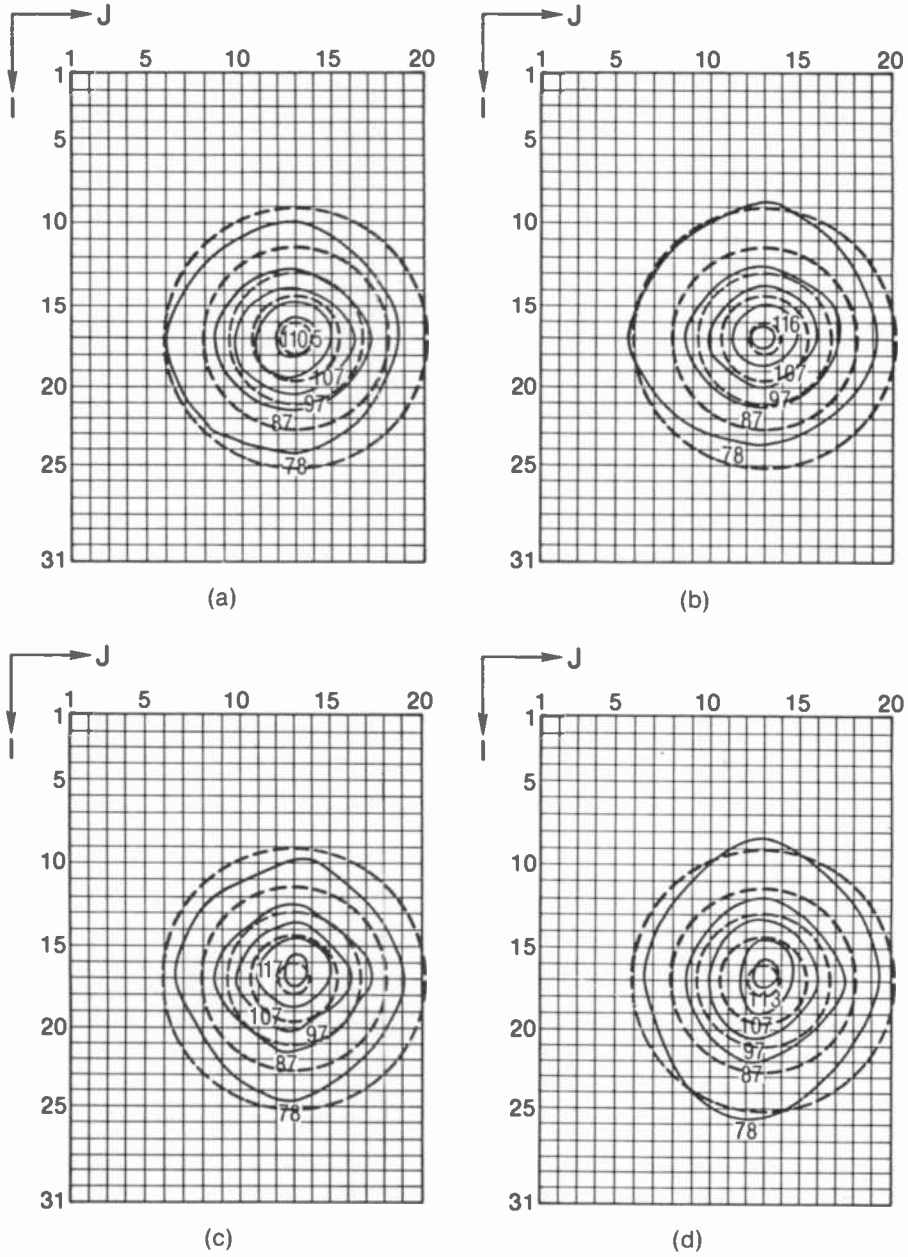


Fig. 6—Isotherms, for half-width of $4\Delta x$ at $t = 18$ hours, for (a) Boris 2D; (b) Fromm 2D; (c) Leith 2D; and (d) Godunov 2D.

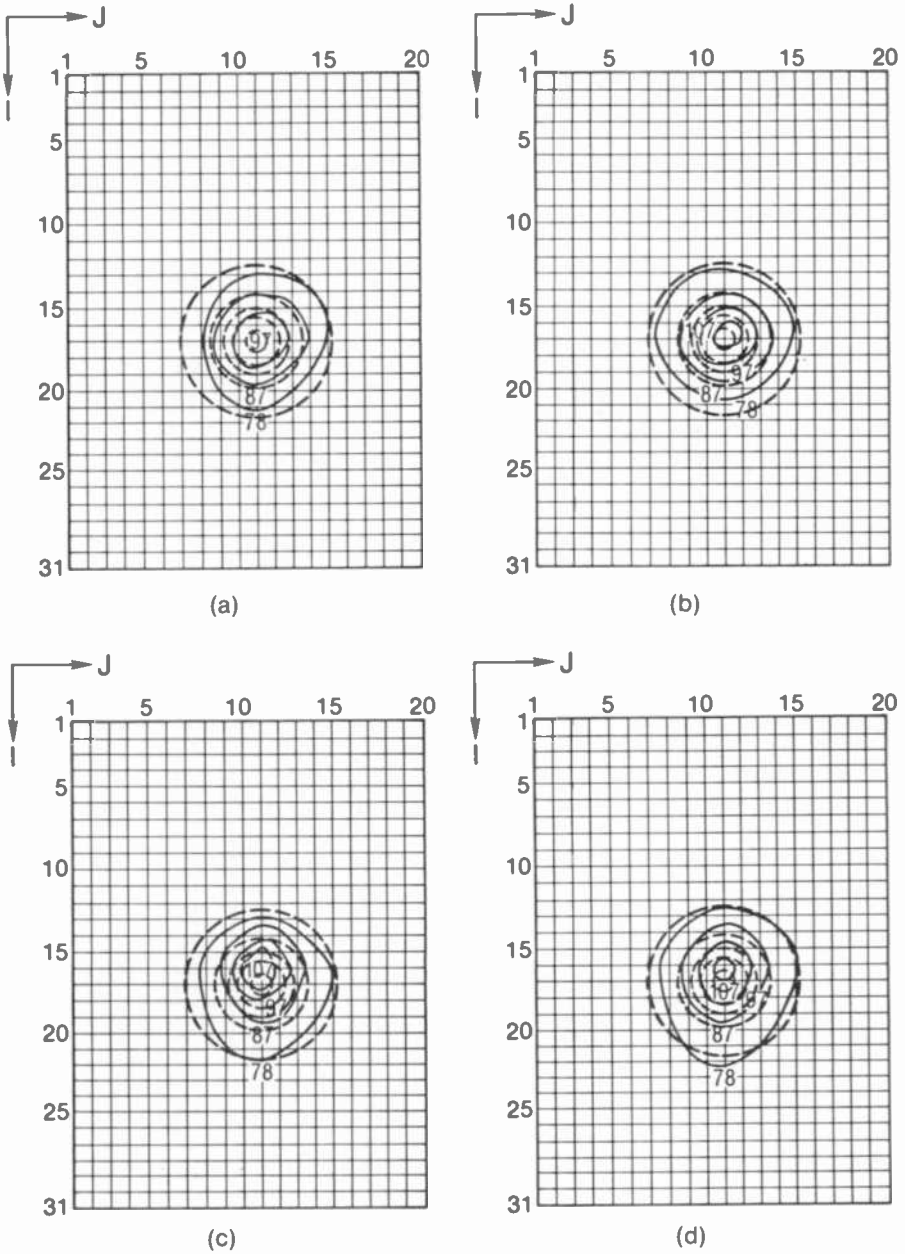


Fig. 7—Isotherms, for half-width of $2\Delta x$ at $t = 6$ hours, for (a) Boris 2D; (b) Fromm 2D; (c) Leith 2D; and (d) Godunov 2D. Dashed contours are shown for the exact solution at $T = 117$, 107, 97, 87, and 78°F and $\Delta t = 6$ min.

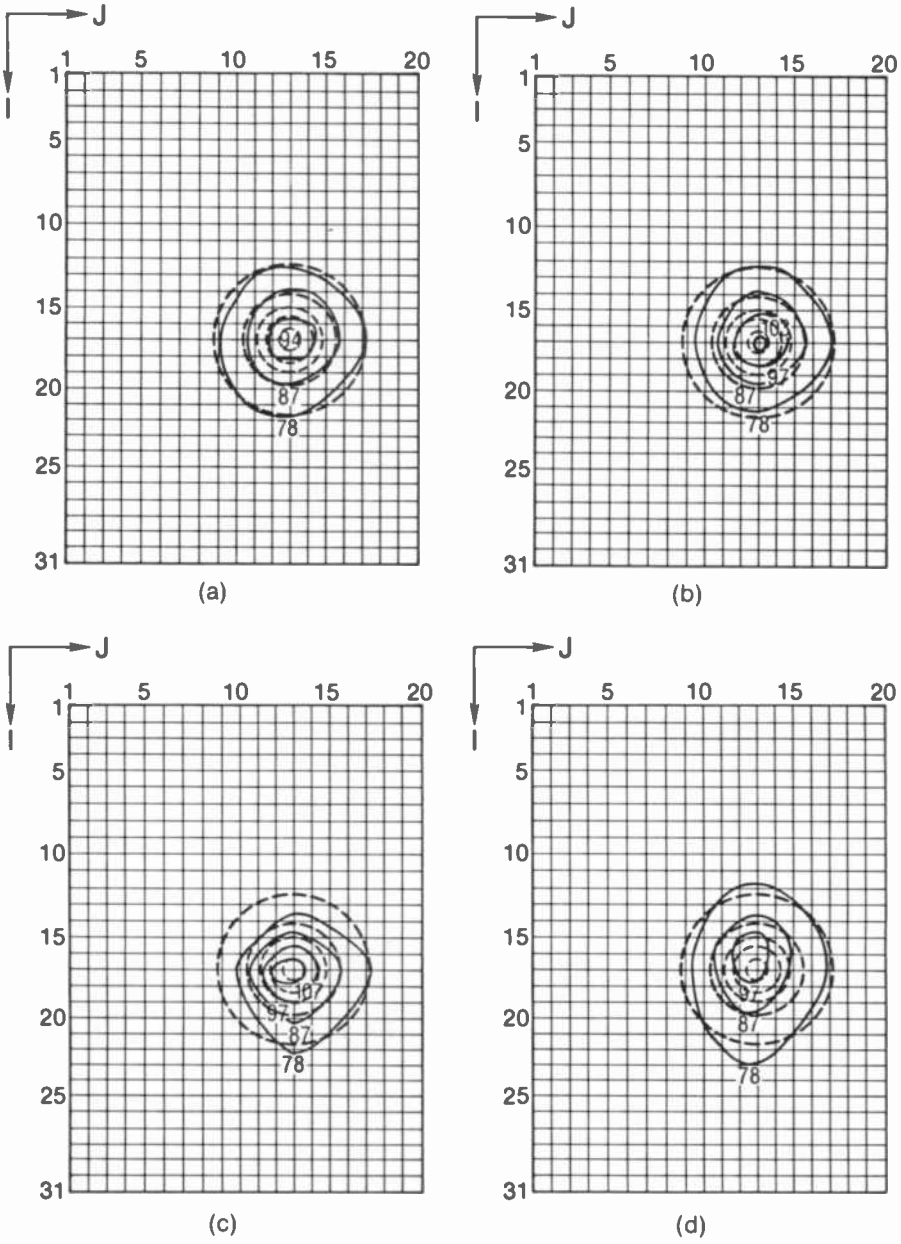


Fig. 8—Isotherms, for half-width of $2\Delta x$ at t = 18 hours, for (a) Boris 2D; (b) Fromm 2D; (c) Leith 2D; and (d) Godunov 2D.

Leith's method and Godunov's method to indicate[12] minimum numerical diffusion required for second-order methods (viz., Lax-Wendroff or Leith) and the minimum numerical diffusion required for monotonicity (viz., the first-order Godunov method). Total variation diminishing (TVD) schemes, as introduced by Harten[39] and recently reviewed by Sweby,[40] when extended to two dimensions may be reduced to first-order accuracy.[41] Therefore, there appeared no need to include them in this comparison.

For linear advection of the two-dimensional Gaussian initial temperature distribution, the same values of grid size, time step, and x-velocity component are used as in the previous example for linear advection of an initial square wave of temperature. In addition, a uniform rectangular grid is considered (i.e., $\Delta x = \Delta y$) and the y component of velocity is included,

$v = 0.1$ ft/sec (0.0305 m/sec).

Most oscillating flow regions in a typical convection-dominated coastal site[36] would have a cosinusoidal component of current normal to the shoreline corresponding to a sinusoidal component of current parallel to the shoreline.[42] However, this example was run for a constant normal component of current to indicate the maximum distance convected in the y direction of the thermal parcel. The two-dimensional Gaussian temperature distribution corresponds to a surface, or near-surface, thermal discharge for once-through condenser cooling[43] for the following values,

$T_\infty = 77.0$°F (25.0°C)

$\Delta T = 43.0$°F (23.9°C).

The initial distribution is located at $I = 10$ and $J = 10$ in the uniform rectangular grid system, with $T_{max} = 120$°F (48.9°C), i.e., the 2-D Gaussian distribution is initially centered at (x_0, y_0).

Figs. 3–8 illustrate the isotherms at $t = 6$ hours and $t = 18$ hours* for the two-dimensional, incompressible linear advection of a 2-D Gaussian initial distribution of temperature. Each figure gives the isotherms for the four different methods (Boris, Fromm, Leith, and Godunov). For a half-width of $8\Delta x$ (Figs. 3 and 4), the Boris et al scheme indicates that the 119°F (48.3°C) isotherm is gone due to clipping, and the 117°F (47.2°C) and 107°F (41.7°C) isotherms compare reasonably well with the exact solution (dashed contours). The

* The results are not shown at $t = 12$ hours and $t = 24$ hours, since dispersive effects are virtually cancelled due to the complete reversal of flow at these times.

97°F (36.1°C) isotherm is somewhat distorted by the right open boundary condition. The Fromm scheme exhibits fine results for the 119°F and 117°F isotherms, but the 107°F and 97°F isotherms have been considerably distorted by the open boundary conditions. The Leith scheme exhibits numerical oscillations or dispersion ripples (wiggles) for the 119°F, 117°F, 107°F, and 97°F isotherms. The Godunov scheme indicates that the 119°F isotherm is completely diffused at $t = 18$ hours, but the 117°F, 107°F and 97°F isotherms exhibit significantly less dispersion ripples than computed by Leith's scheme, as was expected.

For a steeper 2-D Gaussian with half-width of $4\Delta x$ (Figs. 5 and 6), the Boris et al scheme clipped to 109°F (42.8°C) after $t = 24$ hours, which is worse than computed by Godunov's scheme. The 107°F, 97°F and 87°F (30.6°C) isotherms are fairly good, and the 78°F (25.6°C) isotherm is significantly distorted by the right open (adiabatic) boundary condition. Fromm's scheme indicated that the 117°F isotherm diffused away after $t = 15$ hours and at $t = 24$ hours yielded a maximum temperature of 116°F (46.7°C). The 117°F, 107°F, and 97°F isotherms are in fair agreement with the exact solution as predicted by Fromm's scheme. However, the 87°F and 78°F isotherms were indeed distorted by the boundary conditions. Obviously, for actual applications, a larger grid system would have been employed to solve this boundary problem for the broader Gaussians. Leith's scheme indicates that the 117°F isotherm is fine, but the 107°F and 97°F isotherms and particularly the 87°F isotherm exhibit dispersion ripples. The 78°F isotherm is again distorted by the boundary conditions. The Godunov scheme indicated that the 117°F isotherm diffused very fast, worse than for Fromm's scheme. The 107°F, 97°F, 87°F, and 78°F isotherms are reasonably accurate and indicate significantly less dispersion than for Leith's scheme.

The steepest 2-D Gaussian distribution of half-width $2\Delta x$ (Figs. 7 and 8) indicates that the Boris et al scheme immediately clipped (or diffused away) the 117°F and 107°F isotherms. The 97°F isotherm was clipped at $t = 9$ hours. The 87°F and 78°F isotherms are rather accurate, particularly at $t = 18$ hours. Fromm's scheme indicates that 117°F isotherm is immediately diffused and the 107°F isotherm diffused at $t = 9$ hours. The 97°F, 87°F, and 78°F isotherms are reasonably accurate. Leith's scheme indicates that the 117°F isotherm diffused immediately, and the 107°F, 97°F, 87°F, and 78°F isotherms exhibit dispersion ripples, again due to dispersive phase error. Godunov's scheme indicates also, that the 117°F isotherm diffused almost immediately and, similar to Fromm's scheme, the 107°F isotherm diffused at $t = 9$ hours. The 97°F, 87°F, and 78°F

isotherms indicate some elongation in the x direction, particularly the 78°F isotherm. Several computer runs were made setting the y component of current equal to zero and resulted in short-wavelength anisotropy[44] in the vicinity of the steep gradients. Therefore, the study could not ascertain, without further varying the y component of current, whether the elongation of the isotherms in the x direction is due to diffusion and/or anisotropy.

A more quantitative measure of error would be to compare these four advection schemes with the exact solution using the L_1-norm. In discrete form, the l_1-norm for this problem is given by

$$l_1 = \frac{\sum_I \sum_J \text{ABS}[T(I,J) - T_{EX}(I,J)]\Delta x \Delta y}{\sum_I \sum_J T_{EX}(I,J)\Delta x \Delta y}.$$

Tables 1 and 2 present the l_1-norms for the Boris et al, Fromm, Leith, and Godunov schemes for the 2-D Gaussian half-widths $8\Delta x$, $4\Delta x$, and $2\Delta x$, i.e., increasingly steeper gradients of temperature. Table 1 gives results at $t = 6$ hours and Table 2 at $t = 18$ hours.

The various regions corresponding to l_{12}, l_{13}, l_{14} and l_{15} are given in the Appendix. It can be readily observed that Fromm's scheme is the best choice at $t = 6$ hours, and although Leith's scheme is comparable for $8\Delta x$ and $4\Delta x$ half-widths at $t = 18$ hours, Fromm's scheme again excels for the steepest Gaussian distribution, i.e., a half-width of $2\Delta x$. The l_1-norms show that the Boris et al scheme compares rather well with Fromm's scheme for the steepest 2-D Gaussian distribution, except in the smaller regions near the center of the Gaussian distribution. The Boris et al scheme and Godunov's scheme are quite close, as would be expected since both methods possess strong diffusion in order to ensure monotonicity. Fromm's scheme exhibited negligible oscillations for this linear advection problem, whereas Leith's scheme did not.

Although these numerical experiments have been performed to study the 2-D advection of hypothetical thermal discharges (plumes) for the case of once-through cooling of the condenser in fossil and nuclear power plants,[43] the 2-D Gaussian distributions of varying half-width also physically represent the thermal profile upon direct jet impingement on high-flux, discrete, distributed, heat sources in modern electronics equipment.[45] According to Miller and Wolfe,[46] power transistors found in radio amplifiers and power supplies often dissipate over 50 watts each in physical areas of less than one-half-inch square (0.000322 m^2). Scaling of the numerical example given

here for a geophysical flow to that of an effective cooling technique utilizing direct jet impingement of high-flux electronics components would typically result in the following model parameters:

$\Delta x = \Delta y = 0.05$ ft (0.0152 m)
$u_{max} = 3$ ft/sec (0.915 m/sec)
$v = 0$
$T = 1.5$ sec; $\omega = 2\pi/T = 4.18$ rads/sec
$\Delta t \cong 3$ msec

5. Conclusions

Numerical experiments have been performed using four different low-dispersion schemes for the two-dimensional, incompressible linear advection of 2-D Gaussian distributions of temperature of varying half-width. The flow field is specified a priori as consisting of a slowly varying, uniform, oscillating x component of velocity and a constant y component of velocity. Based upon an extensive review of the literature for low-dispersive advection schemes, the methods by Boris et al, Fromm, Leith, and Godunov were selected for these numerical studies.

It is shown that time-splitting the Boris et al one-dimensional FCT operator does not result in nonphysical temperatures for a uniformly oscillating flow field, although for steep gradients the clipping phenomenon may not be acceptable in small regions near the center of the 2-D Gaussian distribution. Fromm's scheme has been extended to an oscillating x component of velocity and is a good choice for this particular problem. Leith's scheme, although relatively accurate, does produce undesirable dispersion ripples. Godunov's scheme, although relatively oscillation free for this problem, provides excessive diffusion of the solution and, generally, results in the worst measure of error of the four schemes.

All four advection schemes exhibit transportive and reasonably conservative properties (based upon a review of Tables 1 and 2). Except for Leith's scheme, the results obtained indicate fairly accurate preservation of monotonicity of the solutions for temperature. The open boundary conditions did cause some distortion of the broader Gaussian distributions (i.e., half-widths of $8\Delta x$ and $4\Delta x$), although this did not occur for the steepest gradient (half-width of $2\Delta x$). Use of a larger grid system would be expected to alleviate this problem.

The artificial numerical effects of diffusion, dispersion, and an-

Table 1—l_1-norms at $t = 6$ hours

Code	L'	l_{12}	l_{13}	l_{14}	l_{15}
Boris 2D	$8\Delta x$	0.0079	0.0048	0.0062	0.0088
	$4\Delta x$	0.0085	0.0152	0.0224	0.0321
	$2\Delta x$	0.0114	0.0298	0.0616	0.0868
Fromm 2D	$8\Delta x$	0.0030	0.0016	0.0008	0.0007
	$4\Delta x$	0.0037	0.0058	0.0074	0.0090
	$2\Delta x$	0.0071	0.0163	0.0298	0.0421
Leith 2D	$8\Delta x$	0.0031	0.0016	0.0011	0.0007
	$4\Delta x$	0.0068	0.0124	0.0146	0.0114
	$2\Delta x$	0.0146	0.0305	0.0552	0.0621
Godunov 2D	$8\Delta x$	0.0047	0.0041	0.0052	0.0056
	$4\Delta x$	0.0086	0.0137	0.0185	0.0216
	$2\Delta x$	0.0125	0.0281	0.0499	0.0661

isotropy have been discussed. The proper choice based upon the l_1-norms would be Fromm's scheme, but the Boris et al scheme performed rather well for this 2-D linear problem, except near the center of the 2-D Gaussian for a half-width of $2\Delta x$.

These low-dispersive 2-D advection schemes have been applied (1) to thermal dispersion (considering forced convection only) of a surface discharge located offshore at a convection-dominated coastal site, and (2) to an electronics cooling technique. Note, that these 2-D advection methods can also be applied to three-dimensional time-dependent incompressible flows where it is necessary to consider steep gradients of some transported fluid property.[47–49] This appli-

Table 2—l_1-norms at $t = 18$ hours

Code	L'	l_{12}	l_{13}	l_{14}	l_{15}
Boris 2D	$8\Delta x$	0.0250	0.0122	0.0094	0.0160
	$4\Delta x$	0.0146	0.0238	0.0293	0.0523
	$2\Delta x$	0.0152	0.0394	0.0727	0.1302
Fromm 2D	$8\Delta x$	0.0067	0.0024	0.0022	0.0023
	$4\Delta x$	0.0083	0.0111	0.0157	0.0229
	$2\Delta x$	0.0131	0.0302	0.0479	0.0810
Leith 2D	$8\Delta x$	0.0063	0.0026	0.0013	0.0007
	$4\Delta x$	0.0080	0.0129	0.0151	0.0122
	$2\Delta x$	0.0157	0.0336	0.0602	0.0705
Godunov 2D	$8\Delta x$	0.0112	0.0119	0.0149	0.0158
	$4\Delta x$	0.0178	0.0246	0.0396	0.0512
	$2\Delta x$	0.0172	0.0401	0.0636	0.1071

cation is particularly apt if horizontal diffusion is negligible[50] and vertical transport processes do not dominate, e.g., in relatively deep estuary or coastal geophysical flows where the horizontal advection currents are strong.

A numerical study of direct jet-impingement as a viable cooling technique for high-flux electronics heat sources is difficult because of the multiplicity of discrete, distributed, high-flux components. However, valuable information can be acquired by applying the results given here to one-dimensional and two-dimensional linear advection of the jet thermal profile.

Acknowledgments

All of the computations were made on a Perkin-Elmer OS/32 computer at RCA Government Communications Systems, Camden, NJ. Thanks are accorded to Dr. Steven T. Zalesak of the Naval Research Laboratory for his helpful suggestions during the course of the development of the numerical experiments.

Appendix—Nomenclature

B_x = Turbulent eddy transport coefficient in x direction, m^2/sec
B_y = Turbulent eddy transport coefficient in y direction, m^2/sec
c_1 = Constant value of y component of velocity, m/sec
f = Frequency of tidal cycle, 1/sec
l_1 = Discrete form of (continuous) L_1-norm
l_{12} = l_1-norm for region of size $14\Delta x$ by $14\Delta y$
l_{13} = l_1-norm for region of size $8\Delta x$ by $8\Delta y$
l_{14} = l_1-norm for region of size $4\Delta x$ by $4\Delta y$
l_{15} = l_1-norm for region of size $2\Delta x$ by $2\Delta y$
L' = Half-width of 2-D Gaussian distribution
S = Source term in convection-diffusion equation, °C/sec
T = Temperature, °C
t = Time, sec
u = x component of velocity, m/sec
v = y component of velocity, m/sec
Δt = Time step, sec
Δx = Grid size in x direction, m
Δy = Grid size in y direction, m
ω = Angular frequency of variation of x component of velocity, rads/sec

Subscripts

cx = Convection in x direction
cy = Convection in y direction
EX = Exact value
I = Grid index in x direction
J = Grid index in y direction
max = Maximum value of quantity
w = Boundary value of quantity
∞ = Ambient condition

Superscripts

c = Flux corrected
d = Diffused
n = Time level
t = Transported
td = Transported and diffused

References:

[1] J. P. Boris and D. L. Book, "Flux-Corrected Transport I: SHASTA, a Fluid Transport Algorithm that Works," *J. Computational Physics*, **11**, p. 38 (1973).
[2] D. L. Book, J. P. Boris, and K. Hain, "Flux-Corrected Transport II: Generalizations of the Method," *J. Computational Physics*, **18**, p. 248 (1975).
[3] J. P. Boris and D. L. Book, "Flux-Corrected Transport III: Minimal-Error FCT Algorithms," *J. Computational Physics*, **20**, p. 397 (1976).
[4] H. P. Miller and K. C. S. Tong, "An Ad-hoc Method for Convection Dominated Flows," ASME Paper 81-FE-15 (1981).
[5] H. P. Miller, "A Comparison of Explicit Finite Difference Schemes for Incompressible Convection Dominated Flows," SIAM 1983 Fall Meeting, Norfolk, Virginia, Nov. 1983.
[6] J. E. Fromm, "A Method for Reducing Dispersion in Convective Difference Schemes," *J. Computational Physics*, **3**, p. 176 (1968).
[7] C. E. Leith, "Numerical Simulation of the Earth's Atmosphere," *Methods in Computational Physics*, **4**, p. 1 (1965).
[8] S. K. Godunov, A. W. Zabrodin, and G. P. Prokopov, "A Difference Scheme for Two-Dimensional Unsteady Problems of Gas Dynamics and Computation of Flow with a Detached Shock Wave," *USSR J. Computational Mathematics and Mathematical Physics*, p. 1020 (1961).
[9] B. VanLeer, "Towards the Ultimate Conservative Difference Scheme. II. Monotonicity and Conservation Combined in a Second-Order Scheme," *J. Computational Physics*, **14**, p. 361 (1974).
[10] C. K. Forester, "Higher Order Monotonic Convective Difference Schemes," *J. Computational Physics*, **23**, p. 1 (1977).
[11] S. T. Zalesak, "Fully Multidimensional Flux-Corrected Transport Algorithms for Fluids," *J. Computational Physics*, **31**, p. 335 (1979).
[12] J. F. Thompson, "Convection Schemes for Use with Curvilinear Coordinate Systems—A Survey," Miscellaneous Paper E-84-4, U.S. Army Corps of Engineers Final Report, June 1984.

[13] S. T. Zalesak, "Very High Order and Pseudospectral Flux-Corrected Transport (FCT) Algorithms for Conservation Laws," *Advances in Computer Methods for Partial Differential Equations IV,* R. Vichnevetsky and R. S. Stepleman (Editors), p. 126 (1981).
[14] S. T. Zalesak, "A Physical Interpretation of the Richtmyer Two-Step Lax-Wendroff Scheme, and Its Generalization to Higher Spatial Order," *Advances in Computer Methods for Partial Differential Equations, V,* R. Vichnevetsky and R. S. Stepleman (Editors), p. 491 (1984).
[15] P. M. Gresho and R. L. Lee, "Don't Suppress the Wiggles—They're Telling You Something," *Computers and Fluids,* **9,** p. 223 (1981).
[16] A. C. Vliegenthart, "The Shuman Filtering Operator and the Numerical Computation of Shock Waves," *J. Engineering Mathematics,* **4,** p. 341 (1970).
[17] A. Harten and G. Zwas "Switched Numerical Shuman Filters for Shock Calculations," *J. Engineering Mathematics,* **6,** p. 207 (1972).
[18] A. Harten and H. Tal-Ezer, "On a Fourth Order Implicit Finite Difference Scheme for Hyperbolic Conservation Laws. II. Five-Point Schemes," *J. Computational Physics,* **41,** p. 329 (1981).
[19] D. N. Holla and P. C. Jain, "Implicit Dissipative Schemes for Solving Systems of Conservation Laws," *J. Engineering Mathematics,* **13,** p. 257 (1979).
[20] K. Srinivas, J. Gururaja, and K. K. Prasad, "An Assessment of the Quality of Selected Finite Difference Schemes for Time Dependent Compressible Flows," *J. Computational Physics,* **20,** p. 140 (1976).
[21] K. Srinivas and J. Gururaja, "An Improved Form of the Artificial Diffusion Parameter-X," *J. Computational Physics,* **31,** p. 289 (1979).
[22] A. Harten and G. Zwas, "Self-adjusting Hybrid Schemes for Shock Computations," *J. Computational Physics,* **9,** p. 568 (1972).
[23] W. J. Weber, J. P. Boris, and J. H. Gardner, "ALFVEN—A Two-Dimensional Code based on SHASTA, Solving the Radiative, Diffusive MHD Equations," *Physics Communications,* **16,** p. 243 (1979).
[24] D. V. Anderson, "Axisymmetric Multifluid Simulation of High Beta Plasmas with Anisotropic Transport Using a Moving Flux Coordinate Grid," *J. Computational Physics,* **27,** p. 246 (1975).
[25] B. VanLeer, "Towards the Ultimate Conservative Difference Scheme. IV. A New Approach to Numerical Convection," *J. Computational Physics,* **23,** p. 276 (1977).
[26] B. VanLeer, "Towards the Ultimate Conservative Difference Scheme. I. The Quest of Monotonicity," in *Lecture Notes in Physics,* Vol. 18, p. 163, Springer, Berlin (1973).
[27] P. D. Lax and B. Wendroff, "Difference Schemes with High Order of Accuracy for Solving Hyperbolic Equations," *Communications on Pure and Applied Mathematics,* **17,** p. 381 (1964).
[28] A. Arakawa, "Computational Design of Long-Term Numerical Integration of the Equations of Fluid Motion. I. Two-Dimensional Incompressible Flow," *J. Computational Physics,* **1,** p. 119 (1966).
[29] G. A. Sod, "A Survey of Several Finite Difference Methods for Systems of Nonlinear Hyperbolic Conservation Laws," *J. Computational Physics,* **27,** p. 1 (1978).
[30] S. Sengupta, S. S. Lee, and H. P. Miller, "A Three-Dimensional Free-Surface Numerical Model for Transport Processes in Biscayne Bay," *Computing Methods in Geophysical Mechanics,* ASME AMD, **25,** p. 169 (1977).
[31] P. J. Roache, *Computational Fluid Dynamics,* Hermosa Publishers, Albuquerque, New Mexico (1976).
[32] G. Strang, "Accurate Partial Difference Methods, I. Linear Cauchy Problems," *Arch. Rational Mech. Anal.,* **12,** p. 392 (1963).
[33] G. Strang, "On the Construction and Comparison of Difference Schemes," *SIAM Journal of Numerical Analysis,* **5,** p. 506 (1968).
[34] W. F. Noh and M. H. Protter, "Difference Methods and the Equations of Hydrodynamics," *J. Mathematics and Mechanics,* **12,** p. 149 (1963).
[35] M. Holt, *Numerical Methods in Fluid Dynamics,* Second Edition, Springer-Verlag Publishers, Berlin (1984).
[36] Raytheon Company Final Report, "Thermal Dispersion Modeling of the Proposed Norco Nuclear Power Station," Aug. 1974.

[37] D. L. Book, *Finite-Difference Techniques for Vectorized Fluid Dynamics Calculations*, Springer-Verlag Publishers, Berlin (1981).
[38] S. T. Zalesak, private communication.
[39] A. Harten, "High Resolution Schemes for Hyperbolic Conservation Laws," *J. Computational Physics*, **49**, p. 357 (1983).
[40] P. K. Sweby, "High Resolution Schemes Using Flux Limiters for Hyperbolic Conservation Laws," *SIAM Journal of Numerical Analysis*, **21**, p. 995 (1984).
[41] J. B. Goodman and R. J. LeVeque, "On the Accuracy of Stable Schemes for 2D Scalar Conservation Laws," Courant Mathematics and Computing Laboratory, New York University, DOE/ER/03077-203, Nov. 1983.
[42] H. P. Miller and K. C. S. Tong, "2-D Convection Dominated Thermal Dispersion in Coastal Waters," in *Waste Heat Utilization and Management*, S. Sengupta and S. S. Lee, Editors, Hemisphere Publishers, p. 183 (1983).
[43] S. Glasstone, and W. H. Jordan, *Nuclear Power and Its Environmental Effects*, American Nuclear Society Publication (1980).
[44] R. Vichnevetsky and J. B. Bowles, *Fourier Analysis of Numerical Approximations of Hyperbolic Equations*, SIAM Studies in Applied Mathematics, p. 115 (1982).
[45] H. P. Miller, private communication.
[46] H. P. Miller and D B. Wolfe, "Computer Aided Thermal Analysis and Design of Complex Electronics Equipment," ASME Paper 85-DE-E-1, presented at Eastern Design Engineering Conference, Boston, MA, June 1985.
[47] S. Sengupta, S. S. Lee, and H. P. Miller, "Three-Dimensional Numerical Investigations of Tide and Wind-Induced Transport Processes in Biscayne Bay," *Sea Grant Technical Bulletin*, No. 39 (1978).
[48] S. Sengupta, H. P. Miller, and S. S. Lee, "Numerical Simulation of Dissolved Constituent Transport in Biscayne Bay," *International Journal of Modelling and Simulation*, **2**, No. 3 (1982).
[49] H. P. Miller, "Three-Dimensional Free-Surface Suspended Particles Transport in the South Biscayne Bay, Florida," *International Journal for Numerical Methods in Fluids*, **4**, No. 10, p. 901 (1984).
[50] H. P. Miller, "Numerical Three-Dimensional Free Surface Circulation Model for the South Biscayne Bay, Florida," *Applied Mathematical Modelling*, **8**, No. 5, p. 313 (1984).

Optoelectronic Performance Issues in Fiber-Optic Communications*

D. J. Channin
RCA Laboratories, Princeton, NJ 08540

Abstract—This paper presents a survey of technical issues concerning the specification and use of optoelectronic devices and components in fiber-optic communications systems. It is intended to bring together current understanding of the performance of fiber-optic systems using relatively straightforward models for device and transmission characteristics. Where possible, formulas and graphical presentations of the phenomena are presented. Representative observations of significant effects are offered as guides to the magnitude and impact of the phenomena.

Introduction

This paper presents a survey of technical issues concerning the specification and use of optoelectronic devices and components in fiber-optic communications. It is divided into five sections. Sec. 1 discusses overall system characterization and introduces the basic issues of system power budget and signal fidelity. Sec. 2 gives the characteristics of isolated optoelectronic devices that have important impacts on system performance. Interactive phenomena that involve both device and transmission parameters are discussed in Sec. 3. Sec. 4 presents some issues concerning modulation of light for signal transmission. The last section shows how the performance of various kinds of fiber-optic communication systems are influenced by the considerations discussed in the prior sections.

1. Overall System Characterization

1.1 Optical Power Budget

The most straightforward characterization of a fiber-optic commu-

* Presented at SPIE Int. Conf., San Diego, CA, Aug. 23–24, 1984. (*Proc. SPIE–Int. Soc. Opt. Eng.*, Vol. 512, *Fiber Optic Communication Technology* (1984), pp. 84–100.)

nication system is by the optical signal power at points between the emitter and the receiver. A specific amount of signal power is coupled into the optical fiber and suffers loss during transmission through the fiber, connectors, and possibly multiport couplers. For proper operation, the power incident on the detector of the receiver must be greater than the receiver's sensitivity, yet less than the maximum set by the receiver's dynamic range. Usually the optical power is given, in decibel-milliwatts (dBm), as

$$P \text{ (in dBm)} = 10 \log \frac{P \text{ (in mW)}}{1 \text{ mW}}, \qquad [1]$$

so that multiplicative connector and coupler losses are added together algebraically, and exponential fiber losses are expressed as products of distance times loss (in dB/km).

Accurate prediction and measurement of an optical power budget can be hindered by inadequate specification of the modal distribution of the guided light and variabilities in connector loss and port-to-port coupler efficiency. Consequently, the power budget for a practical communication link should contain appropriate excess power margin for these and other uncertainties in specified parameters.

1.2 Signal Fidelity

Transmission through a fiber-optic communication system can produce distortions in the modulated waveform of the optical signal. Such distortions may be produced by the optoelectronic devices themselves or by interaction between characteristics of the optoelectronic devices and the fiber-optic transmission media. These two sources of signal distortion are discussed in Secs. 3 and 4, respectively. Broad characterization of signal fidelity for the system involves the following considerations:

(a) Bandwidth capabilities of sources, receivers, and fiber-optic transmission lines can limit the maximum rate of information transfer and the signal format.
(b) Noise may be introduced by the emitter or by the propagation characteristics of the fiber, as well as by optical or electronic phenomena at the receiver.
(c) Distortion due to nonlinearities in the source or transmission medium may be significant, especially in analog signal transmission.

In the most straightforward digital system designs, the band-

widths of optical transmitter, receiver, and fiber link are specified to give insignificant distortion of the signal at the appropriate data rate and format. Receiver sensitivity is limited by noise due to well-characterized electronic processes in the detector or input amplifier. If nonlinear distortion has negligible impact, signal fidelity ceases to become an explicit issue and the capabilities of the system are defined solely by the optical power budget.

In many cases, particularly those involving advanced components to achieve maximum transmission range, very high data rates, or analog signal transmission, this straightforward approach is not adequate, and more complex issues of optoelectronic characteristics and system interactions must be analyzed. Inadequate bandwidth or signal distortion can often be compensated for by electrically matched equalization. This requires additional signal strength, leading to an optical *power penalty* that must be added to the power budget. Predistortion of the electrical drive currents of the optical sources may also be used to compensate for linearity or nonsymmetrical distortions of the optical output.

2. Characteristics of Isolated Optoelectronic Elements

2.1 Bandwidth of Optical Sources

The frequency response of LED emitters is often characterized by a simple transfer function of the form[1]

$$H(f) = \frac{1}{1 + i\dfrac{f}{f_c}}. \qquad [2]$$

f_c indicates the cut-off frequency, which is typically 50-100 MHz for surface-emitting devices and up to 200 MHz for edge-emitting diodes. Though properly used only for small-signal (linear) modulation, such a transfer function is generally adequate for describing pulse response so long as the device is maintained in forward bias.

Laser diodes at low modulation levels have a frequency response that is represented by the following transfer function[2]:

$$H(f) = \frac{1}{1 - \left(\dfrac{f}{f_o}\right)^2 + if\Gamma(f)}. \qquad [3]$$

Here the resonance frequency f_o is a function of the bias current I_b and the lasing threshold current I_{th}, as well as of internal param-

eters τ_e and τ_p which represent the carrier and photon lifetimes in the laser:

$$f_o^2 = \frac{1}{(2\pi)^2 \tau_e \tau_p} \left(\frac{I_b}{I_{th}} - 1 \right). \quad [4]$$

Typically f_o is in the range 1–3 GHz, though special designs operating well above threshold can extend the frequency range to at least 10 GHz. The damping parameter Γ in Eq. [3] is the result of various internal processes within the laser, including optical gain saturation, carrier diffusion, and spontaneous emission. The transfer function in Eq. [3] may be multiplied by an additional transfer function incorporating effects of internal diode capacitance, device resistance, and external parasitic reactance. Fig. 1 shows the relative small-signal modulation characteristics of LED and laser sources for fiber optics.

Pulse modulation of laser diodes is not well represented by a linear transfer function such as Eq. [3], since the response of the device involves strong nonlinearities in the coupled-rate equations for photons and carriers that describe its internal processes. These nonlinearities are especially important when the device is modulated through the lasing threshold and is therefore switched between the lasing and nonlasing states.

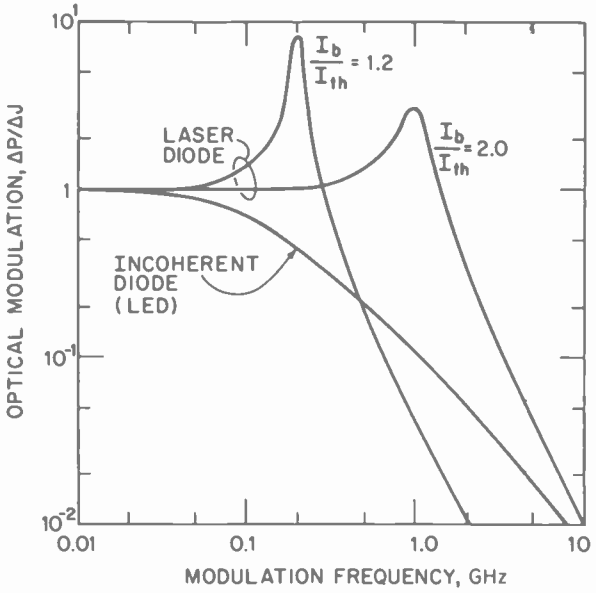


Fig. 1—Frequency response of LED and laser diode emitters.

Numerical integration of the laser rate equations provides a useful representation of the device response to high-frequency digital modulation. Signal distortion effects include rise and fall time, ringing and transient phenomena, delay in optical pulse response, and intersymbol distortion. All these effects are strongly influenced by the operating point of the laser, especially when the modulation involves switching through the lasing threshold. Generally, fast rise and fall times are achieved by switching through threshold, but transients, delay, and intersymbol distortion are minimized with above-threshold modulation. Fig. 2 shows results of a laser simulation model[3] incorporating lateral carrier diffusion, spontaneous emission, and optical gain saturation to represent the laser-pulse response as the bias current is varied from below to above the lasing threshold. The same model is used to simulate digital data trains in order to determine the effect of laser-pulse response for transmission at various data rates and signal formats.

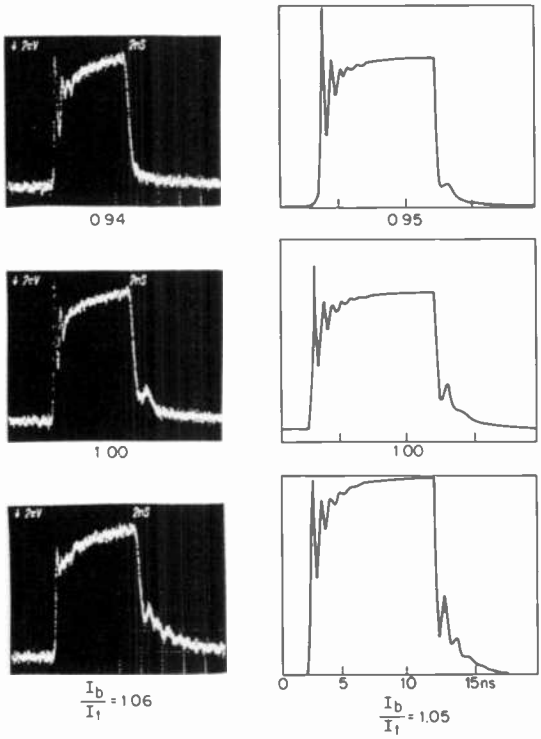


Fig. 2—Pulse response observed (left) and computed (right) for laser diode bias current I_b below, at, and above threshold current I_t.

2.2 Bandwidth Characteristics of Optical Fibers

Though our focus is on optoelectronic issues, a full system analysis must also consider the bandwidth characteristics of the optical fiber medium. Two kinds of dispersion limit this bandwidth. Modal dispersion occurs in multimode fiber because of unequal propagation velocities among the guided optical modes. Optical or chromatic dispersion is a refractive property of the glass itself whose effect on bandwidth depends on the optical spectrum of the source. For that reason this effect will be discussed in Sec. 2 on interactive phenomena.

The impact of modal dispersion on signal propagation can be described by the transfer function[1]

$$H(f) = \frac{1}{1 + i\sqrt{3}\,\frac{zf}{D}}, \qquad [5]$$

where z is the propagation distance and D is a distance-bandwidth product characterizing the fiber. For graded index fibers, D can be in the range of 500 to 1500 GHz-km. However, if the modes of such a fiber are not uniformly excited by the source, the use of such a simple transfer function may be inaccurate.

2.3 Noise Characteristics of Optical Sources

Optical signal sources may also generate optical noise that, in some cases, will degrade the performance of a fiber-optic communication system. Such problems are associated primarily with laser diodes, since LEDs are incoherent emitters that do not generate significant noise in their optical output. Laser diodes can generate significant noise, with considerable variability in magnitude and in spectral distribution. We are concerned here with amplitude noise having a frequency spectrum that falls in the passband of typical communication systems.

The noise intensity from an optical source is important in relation to the intensity of the modulated optical signal. If we let mS_o represent this signal, where S_o is the peak optical intensity and m the modulation index, $<|\Delta S(f)|^2>$ represents the noise power spectral density at frequency f. The relative intensity noise is defined as[4]

$$N_R(f) = \frac{<|\Delta S(f)|^2>}{m^2\, S_o{}^2}\, 2\Delta f, \qquad [6]$$

where Δf is the transmission bandwidth. Since the detector photocurrent is proportional to the optical intensity, this definition is equivalent to the inverse of the power signal-to-noise ratio (SNR) at the photodetector.

The value of $N_R(f)$ depends strongly on the operating point of the laser, and for a typical device is highest when the laser is biased at or very near threshold. When operated at a typical bias level of about 1.2 times threshold, $N_R(f)$ at $f = 5$ MHz can be expected to be in the range of -120 to -140 dB (referenced to 1 Hz bandwidth).[5] Both gain-guided and index-guided lasers operate with similar noise within this broad range. The spectrum of N_R is generally peaked at the resonance frequency given by Eq. [4]. These observations apply to lasers in the absence of any optical feedback and driving transmission and detection systems that are insensitive to optical spectral fluctuations.

Laser diodes sometimes show a strong enhancement of their normal noise output that also results in a narrowing of the noise spectrum. This phenomenon, when very intense, is described as self-oscillations or self-pulsing. Many mechanisms have been proposed for self-oscillations. Some of these mechanisms are often associated with degradation processes that also lead to rapid failure of the diode.[6] Fig. 3 shows the noise power spectrum of a particular laser that exhibits a transition from a normal broad noise spectrum at lower bias currents to a sharp self-oscillation resonance as the bias current increases.

Another laser noise phenomenon can occur at particular current levels at which the optical wavelength switches between two longitudinal nodes. The resulting intensity noise shows an approximately $1/f$ spectrum. Fig. 4 shows a typical example of this type of noise.

2.4 Nonlinear Distortion

Analog signal applications generally impose severe requirements on linearity as well as on signal-to-noise ratio. Both LED and laser diodes can be used in such applications only if proper account is taken of the nonlinear transfer characteristics of the devices. Without special device selection or appropriate circuit compensation, present optoelectronic sources generally impose too much distortion to allow high-quality AM television transmission, for example.

LED emitters, while frequently showing significant nonlinearity

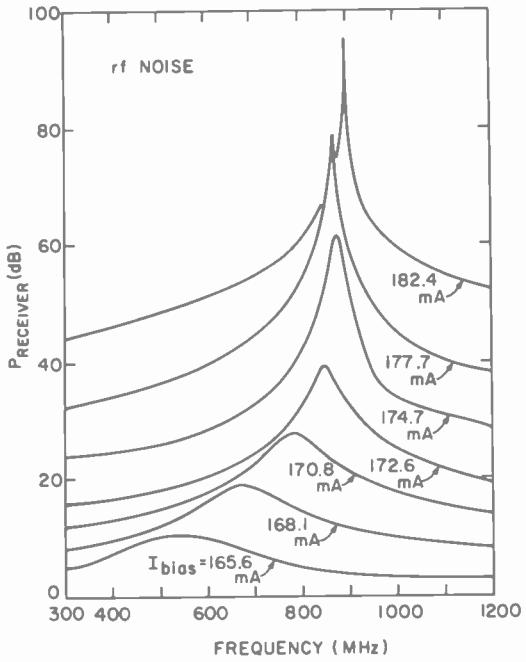


Fig. 3—Noise spectra of laser diode at various bias currents, showing transition to self-oscillation at higher currents.

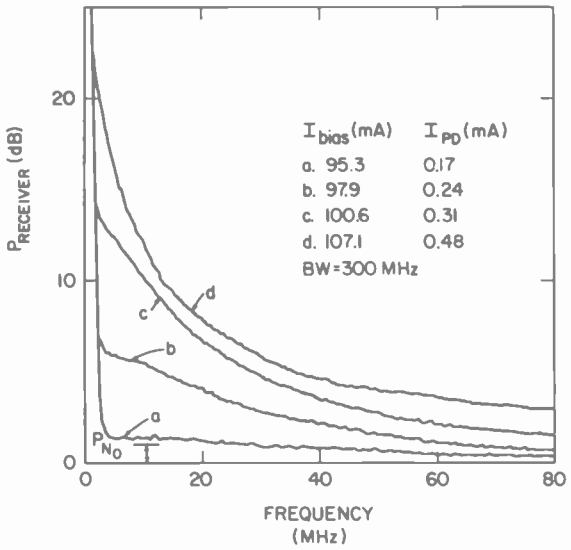


Fig. 4—Noise spectra of laser diodes at various bias currents and corresponding detector photodiode currents (I_{pd}), at a noise bandwidth of 300 MHz, showing amplitude noise associated with longitudinal mode-hopping.

in their optical power/current transfer characteristics, lack a sharp threshold that imposes critical requirements on operating point control, as is the case for laser diodes. Their transfer characteristics are at least potentially stable over the device lifetime. Electronic compensation or negative feedback techniques can be used to linearize the device operation.

Laser diodes have an inherent nonlinearity associated with their threshold. Operation at large modulation depth ($m \geq 1$) will always cause distortion. Linearity at useful modulation requires selection of devices with stable spatial mode structure ("kink-free") and linear transfer characteristics. Third-order intermodulation signals as much as 80-dB down from the fundamentals have been reported at 1-mW output for low-power monomode BH (buried heterojunction) 1.3-μm lasers, and 60-dB intermodulation for 0.85-μm monomode CDH-LOC (constricted double-heterojunction, large optical cavity) at 6-mW modulated output.[7]

Fig. 5 indicates harmonic distortion characteristics of high-power

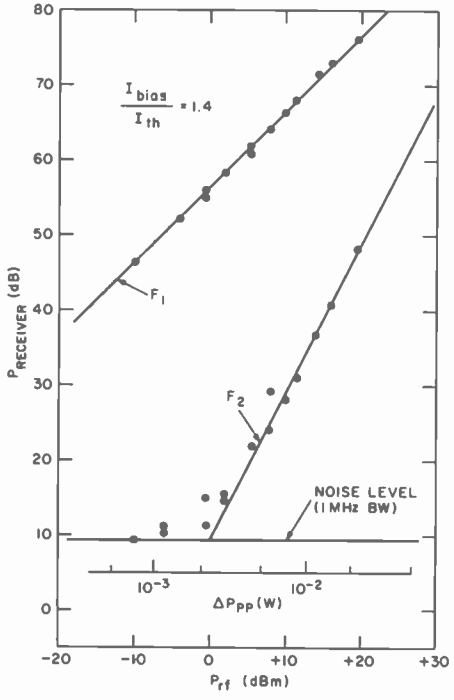


Fig. 5—Analog modulation capabilities of laser diode, with fundamental receiver power (F_1) and second harmonic power (F_2) shown versus modulation power P_{rf} (into 50 Ω) applied to laser (modulation frequency 200 MHz), with 1-MHz noise bandwidth (BW).

CDH-LOC lasers in a format similar to that commonly used to show the capabilities of electronic amplifiers. The signal fundamental frequency F_1 and second harmonic F_2 are plotted against input rf power. A subsidiary scale ΔP_{pp} shows the magnitude of the peak-to-peak modulated optical power generated by the laser. The laser noise floor is also shown. System capability in terms of SNR, second harmonic distortion, rf drive required (into 50-Ω load), and available modulated optical power can be obtained directly from the graph.

2.5 Receiver Noise

Many fiber-optic digital communication links are constrained in range by the limits set on receiver sensitivity by internally generated noise. This noise may be electronic in nature or it may be due to shot noise or avalanche multiplication noise in the photodetector. Unlike most of the other phenomena discussed in this paper, receiver noise is accurately characterized by readily determined device and circuit parameters. A number of detailed analyses and review papers describe the characterization of receivers for optical communication.[8,9] Some fundamental concepts and useful results are as follows.

Digital optical receiver sensitivity is usually determined as the modulated optical power required to achieve a specified bit error rate (BER). Communication theory provides relationships between BER and SNR for various digital formats and noise statistics. For most purposes, the widely used relationships for Gaussian noise statistics

$$P(E) = \frac{1}{\sqrt{2\pi}} \frac{1}{Q} e^{-Q^2/2}, \qquad [7]$$

$$Q^2 = \text{SNR (for a bipolar signal)} \qquad [8]$$

are adequate. For BER = 10^{-9}, $Q = 6.0$, making this error rate a common benchmark for specifying sensitivity. If a simple p-i-n photodiode detector (with unity gain) is used, the sensitivity is given by the optical signal power:

$$P = \frac{h\nu}{\eta q} Q \langle i^2 \rangle_c^{1/2}. \qquad [9]$$

Here $h\nu$ is the photon energy, q the electronic charge, η the detector quantum efficiency, and $\langle i^2 \rangle_c^{1/2}$ is the root-mean-square noise cur-

rent that is superimposed on the signal photocurrent at the input of the front-end amplifiers.

Receiver design to minimize the noise current for both p-i-n and avalanche photodiodes is described in the cited papers. A central consideration in such design is achieving minimum noise while maintaining adequate signal bandwidth and dynamic range. The impact of this consideration on receiver performance is shown in Fig. 6, which illustrates three input amplifier adjustments. In case (a), the bandwidth and noise are optimized. Case (b) illustrates excess bandwidth resulting in higher noise and reduced sensitivity. Case (c) shows inadequate bandwidth causing data drop-outs for some data patterns in the pseudorandom test sequence, even at adequate optical power for high SNR.

The relationship between noise and sensitivity for a potential long-wavelength p-i-n detector receiver design is shown in Fig. 7. The receiver uses an FET preamplifier combined with two bipolar stages in a transimpedance circuit. Gain and transimpedance feedback resistance were treated as variables to optimize sensitivity at

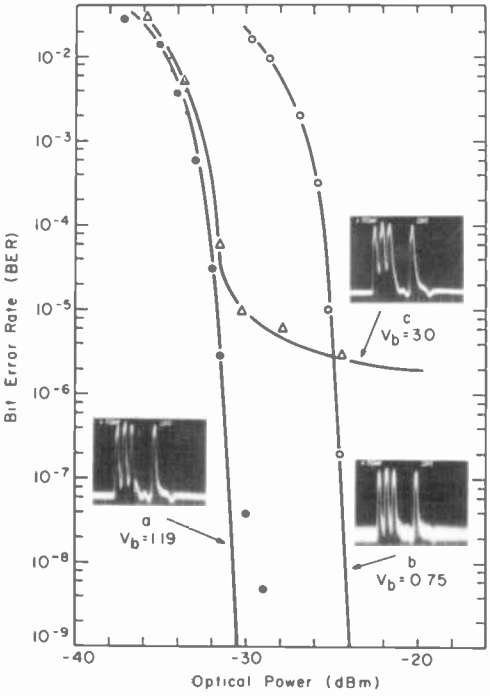


Fig. 6—Bit-error-rate characteristics of optical receivers: (a) optimum characteristics, (b) excess noise, and (c) inadequate bandwidth.

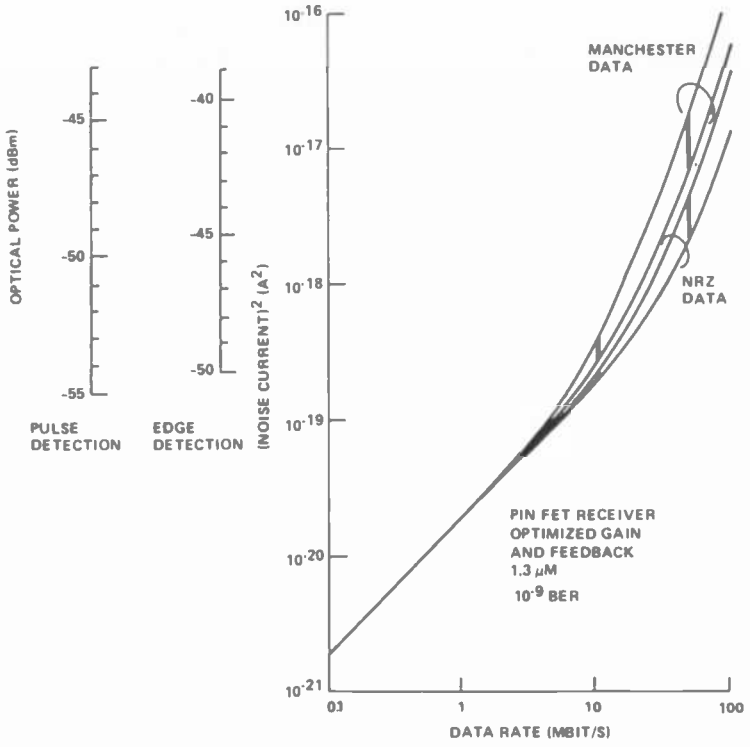


Fig. 7—Noise and sensitivity characteristics of long-wavelength PIN-FET transimpedance optical receiver. Gain and feedback are optimized as functions of data rate. (Manchester = biphase; NRZ = non-return-to-zero.)

any data rate. Noise current and corresponding optical sensitivity (for BER = 10^{-9}) are shown as functions of data rate. The different data formats and digital detection schemes will be discussed in Sec. 4. Here we note the increase in noise and reduction in sensitivity as the data rate increases. At the lower data rates, detector shot noise dominates and is independent of data format. At the high data rates, thermal noise at the transimpedance feedback resistor dominates. The indicated variance in noise for each of the two data formats represents the influence of manufacturing variance in detector capacitance. Differences in noise for the different data formats result from the different transimpedance resistances required to match the differing bandwidth requirements of the two formats.

3. Interactive Phenomena

Many phenomena that can limit the performance of fiber-optic sys-

tems involve the interaction of the characteristics of the optoelectronic source with those of the fiber-optic medium. This section deals with these issues.

3.1 Optical Dispersion

The bandwidth limits set by optical dispersion in the glass fiber may be described by the following transfer function[10]:

$$H(f) = \frac{1}{(1 + i\eta)^{1/2}} \exp\left[- \frac{\eta^2(\lambda - \lambda_o)^2}{(1 + \eta^2)2\sigma^2} (1 - i\eta) \right] \quad [10]$$

where $\eta = \dfrac{2\pi z f S \sigma^2}{c\lambda^2}$. [11]

The source is assumed to have a Gaussian spectral envelope with RMS spectral width σ at center wavelength λ. The parameter S is a material constant characterizing the group delay τ in the fiber ($S \approx 0.05$):

$$\frac{d\tau}{d\lambda} = S \frac{\lambda - \lambda_o}{c\lambda^2}, \quad [12]$$

where λ_o is the wavelength of minimum optical dispersion when the pulse broadening goes to zero. Eq. [11] gives the following expression for the distance-bandwidth product:

$$Z_B f_B = \frac{1}{2\pi} \frac{c}{S} \left(\frac{\lambda}{\sigma} \right)^2 \eta. \quad [13]$$

$|H(f)|$ is down by 3 dB from its low-frequency limit. Z_B and f_B are defined as the range and bandwidth as determined by Eq. [13]. η is the root of the following transcendental equation obtained from Eq. [10]:

$$1 + \eta^2 = 16 \exp\left[- 2 \frac{\eta^2 (\lambda - \lambda_o)^2}{(1 + \eta^2)\sigma} \right]. \quad [14]$$

Fig. 8 shows solutions of Eqs. [13] and [14] for LED and superluminescent emitters operating in the long-wavelength transmission bands. With typical linewidths of 600Å for edge-emitting LEDs the distance-bandwidth product can be equivalent to the modal dispersion (about 1 GHz-km) if the wavelength of the spectral emission maximum coincides with λ_o. This analysis is conservative since it neglects preferential transmission of the wavelengths near λ_o due to the minimum fiber absorption near this wavelength.[11]

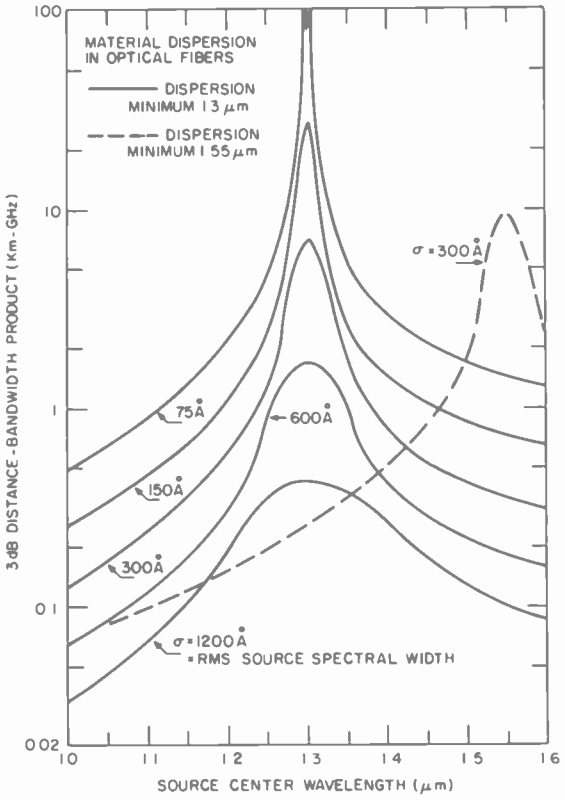


Fig. 8—Bandwidth characteristics of optical fibers using LED and superradiant emitters, considering only chromatic dispersion.

A similar analysis using Eqs. [13] and [14] can be performed for transmission in single-mode fiber using narrow-spectral-width laser sources, though the assumption of a Gaussian spectral distribution becomes weak in this case. However, analysis of partition noise in single-mode systems produces a similar analytical result and a more stringent limit on the achievable distance-bandwidth product. Numerical results are presented in the next section.

3.2 Optical Partition Noise

This noise can be considered a dynamic analog to the wavelength dispersion discussed above. It occurs because the spectral distribution of a multimode laser diode incorporates fluctuations and mode hopping that distribute power among the modes on a subnanosecond time scale. The spectral envelope is a time-averaged result

of the actual fluctuations. Optical dispersion in the fiber converts these spectral fluctuations into intensity fluctuations in the received signal, which are referred to as optical partition noise.

Related sources of partition noise include wavelength dependence in the fiber loss, splice or connector transmission, and photodetector quantum efficiency. Polarization-dependent attenuation in connectors can also contribute to partition noise. However, the discussion here is restricted to noise associated with the fiber-optical dispersion.

The signal-to-noise ratio (SNR) associated with partition noise is given by the relationship[12]

$$\frac{1}{(SNR)^2} = \frac{1}{2}(\pi f)^4 [A_1{}^4\sigma^4 + 48A^4\sigma^8 + 42A_1{}^2A_2{}^2\sigma^6], \quad [15]$$

where

$$A_1 = \frac{S(\lambda - \lambda_o)}{c\lambda^2} Z, \; A_2 = \frac{S}{2c\lambda^2} Z. \quad [16]$$

The other parameters and variables are those of Eqs. [10], [11], and [12]. The SNR can be related to the bit error rate using Eqs. [7] and [8]. The distance-bandwidth product is then given by the expression

$$Z_B f_B = \frac{2^{1/4}}{\sqrt{\pi Q}} \frac{c}{S}\left(\frac{\lambda}{\sigma}\right)^2 \left[3 + 21\left(\frac{\lambda}{\sigma}\right)^2\left(1 - \frac{\lambda_o}{\lambda}\right)^2 + \left(\frac{\lambda}{\sigma}\right)^4\left(1 + \frac{\lambda_o}{\lambda}\right)^4\right]^{-1/4}. \quad [17]$$

Eq. [17] may be compared with Eqs. [13] and [14].

Fig. 9 shows distance-bandwidth products (for BER = 10^{-9}) using laser diodes with differing source optical spectral widths. The fiber dispersion minimum, λ_o, is 1.3 μm. This figure shows the impact on system capability of both the spectral width of the laser and of displacing the center of the source spectrum from the fiber dispersion minimum. The three values of σ correspond to different levels of laser technology. A multimode laser can have a spectral width of about 2 nm. A mode stabilized laser (e.g., a distributed feedback laser) that is constrained to operate in a single longitudinal mode under modulation has a spectral width of about 0.1 nm. Finally, σ = 0.01 nm indicates a potential limit to laser linewidth control in which the linewidth is comparable to the modulation bandwidth of about 1 GHz. Current production technology is capable of σ = 2

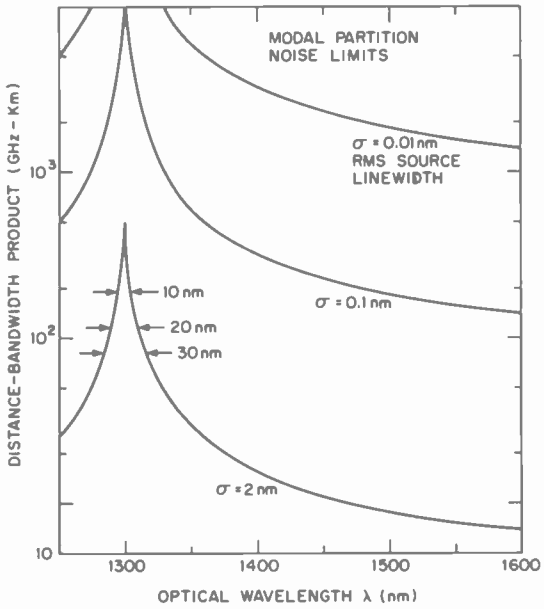


Fig. 9—Bandwidth characteristics of optical fibers using laser-diode emitters, considering only modal partition noise due to chromatic dispersion.

nm, while $\sigma = 0.1$ nm represents laboratory capability that is rapidly being pushed toward production.

Considering the case $\sigma = 1$ nm, we see that high performance requires that the center of the laser optical spectra be closely matched to the dispersion minimum of the fiber. The arrows indicate tolerances that must be achieved at various levels of gain-bandwidth performance. A reasonable benchmark for single-mode fiber systems at present technology limits is about 100 GHz-km. By comparison, 1 GHz-km is the corresponding limit for multimode fiber systems.

Using a mode-stabilized laser with $\sigma = 0.1$ nm gives dramatic improvement in distance-bandwidth product when the laser spectrum and fiber dispersion are well matched. We will see in Sec. 5 that such performance can not be translated into great range because of the fiber loss. The important advantage of the mode-stabilized laser is that it can operate in the transmission window at 1.55 μm without requiring that the optical dispersion be shifted out to this longer wavelength. Doping the glass fiber to achieve this shift compromises the extremely low loss (~0.2 dB/km) that can otherwise be achieved at 1.55 μm.

The third case, $\sigma = 0.01$ nm, suggests the possible limiting performance of coherent optical transmission technology. This area is presently strictly a research activity, but may ultimately be the standard for long-range fiber-optic communication.

Evidently, the ability to control laser-diode spectral characteristics is critical for high-performance transmission systems. Unstabilized lasers that operate single mode when unmodulated show significant partition noise when operated at large modulation depths. Fig. 10 shows test results on such a laser. A spectrometer set to pass the principal longitudinal mode simulated the dispersive fiber or other wavelength-selective transmission medium. Modulation at greater than 75% created partition noise that limited achievable BER. The same device was then mode-stabilized by an external diffraction grating. With such stabilization, the BER was not affected by modulation even when driven through threshold ($m = 1.40$).

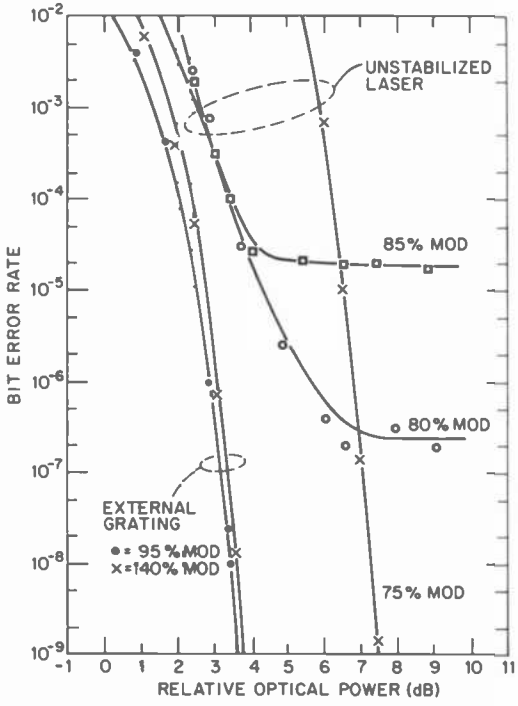


Fig. 10—Bit-error-rate characteristics of a laser in unstabilized and stabilized operation, with varying modulation depth at 150 Mbit/s. An optical spectrometer simulates chromatic fiber dispersion.

3.3 Modal Noise and Distortion

Transmission of coherent light in multimode fibers can yield noise and distortion due to interference between the waves occupying the different fiber modes. This interference is similar to the speckle patterns produced by laser beams reflecting off rough surfaces. In the first instance, this "noise" is simply a modulation of the amplitude distribution of the fiber-guided light. Intensity noise results if the guided light must pass splices, connectors, or other elements that are inhomogeneous in their transmission of the guided light. Distortion as well as noise results when the optical phase of the guided light is varied by modulation.

Quantitative characterization and specification of modal noise is difficult, since it depends on the degree of excitation of perhaps hundreds of guided modes. Analysis has not been carried to the point where predictions of system performance in terms of measured device characteristics can easily be shown. In the following we will discuss only the principal factors controlling the impact of modal noise on system performance.

(a) Optical Coherence

Interference between modes of a multimode fiber require that the coherence time of the light source, τ_c, be greater than the differential propagation delay, $\Delta\tau_m$, of these modes.[13] If, under this condition, the light signal encounters a connector, splice, or other source of modal-selective transmission, it is susceptible to modal noise. Coherence time is related to the optical spectral width of the source by the formula

$$\tau_c = \frac{\lambda^2}{c\sigma}. \qquad [18]$$

Optical coherence times of a single longitudinal mode are of the order of 5×10^{-8} seconds for index-guided lasers and 1×10^{-10} seconds for gain-guided lasers.[14] These times correspond to coherence lengths within the fiber of 9 meters and 2 cm, respectively. A gain-guided laser operating with a broad multimode spectrum has a coherence length that is still further reduced. Thus, modal noise can easily be introduced at typical connector spacings if index-guided lasers are used, but gain-guided lasers are relatively immune to this phenomenon. The coherence of index-guided lasers can be reduced by modulation through the lasing threshold. High-frequency (~1 GHz) tone modulation superimposed onto a lower-fre-

quency signal modulation has been shown effective in reducing modal noise and distortion. Such modulation reduces the coherence of the lasing modes to that of typical gain-guided lasers.

(b) Connector Loss

Reducing connector loss also reduces modal noise effects. An approximate expression gives the SNR in terms of the average (over all the modes) coupling loss ξ as[15]

$$SNR = N \frac{\xi}{1 - \xi}. \qquad [19]$$

Here N is the total number of guided modes and may be on the order of 200. This indicates that a 3-dB coupling loss may give a 23-dB SNR, while a coupling loss of 0.8 dB increases the SNR to 30 dB. This difference can have a significant impact on communication system performance.

(c) Noise Frequency Spectrum

Modal noise frequently appears in a sporadic manner driven by mechanical fluctuations of the fiber-optic cables. As such, the frequencies are low and may be accommodated by the automatic gain control (AGC) loop on a digital receiver. The effective dynamic range is thereby reduced, however.

High-frequency modal noise and distortion are driven by modulation of the laser itself, and appear in the signal bandwidth. These are serious considerations in analog systems, and in severe situations may affect digital transmission as well.

3.4 Optical Feedback

Laser diodes are generally unstable to light reflected into their output facet from external elements. This feedback will produce noise and distortion when the amplitude of the feedback light is comparable to the spontaneous emission coupled into the lasing mode by internal processes. Reflections may affect both the optical spectrum and the intensity of the laser output.

For light reflected far from the laser, so that the transit time is much greater than the coherence time of the laser, individual modes of the external cavity thus formed may mix or mode-lock with the lasing modes, giving noise that usually exhibits a spectrum extending up to the resonance relaxation frequency (Eq. [4]). The frac-

tion of the output intensity required to cause this instability is estimated by Petermann[4] to be 10^{-8} to 10^{-6} for single-longitudinal-mode lasers and 10^{-5} to 10^{-3} for gain-guided multilongitudinal-mode lasers. Noise and distortion can degrade both digital and analog performance, especially in single-mode fibers which are more effectively coupled to the laser output.

Light reflected close to the laser, where the coherence time exceeds the transit time, can have a strong effect on the lasing wavelength. This is because the extended cavity due to the reflection may lock with one of many modes of the laser cavity. Very small changes in the standing-wave conditions can cause mode-hopping and instability. Intensity noise with spectral peaks at multiples of the round-trip time of the external cavity are seen as well. Pulse distortion occurs when round-trip times are comparable with pulse widths, a typical situation with 1-meter fiber pigtails and data rates of the order of 200 MHz.

4. Signal Modulation

The performance and required characteristics of optoelectronic components depends on the nature of the modulation applied to the optical carrier. In this section we review the basic properties of digital and analog modulation approaches in terms of optoelectronic and communication system performance.

4.1 Digital Data Format Characteristics

Many important aspects of digital data formats can be seen by inspection of their power spectra when random information is transmitted. Fig. 11 shows power spectra for three important binary formats: NRZ (non-return-to-zero), biphase (Manchester), and delay modulation.[16] In these formats, binary information is represented by, respectively, the signal level over the length of the bit interval, the sense of a signal transition *within* the bit interval, and the presence or absence of a signal transition within the bit interval.

The effect of the choice of format on the power spectra can be seen in the required transmission bandwidth, the duty-cycle variations in the transmitted signal, and the ease of synchronization of a receiver clock oscillator from the data stream. NRZ data has a low bandwidth requirement, with most of the signal power at frequencies below the bit rate. This format imposes on the system the requirement of a wide range of duty-cycle capability, as indicated by the large dc component of the power spectrum. It also has low spec-

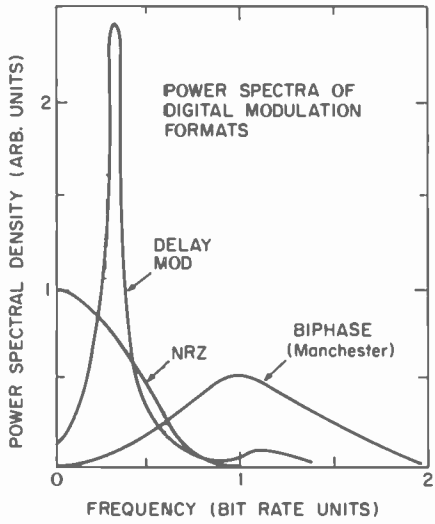


Fig. 11—Signal power spectra of three digital modulation formats.

tral content at the bit frequency, making it unsuited for clock synchronization. Biphase Manchester data has excellent clock recovery power at the bit frequency, and no dc signal component. It requires a high bandwidth, however, extending out to twice the bit frequency. Delay modulation represents something of a compromise, with limited bandwidth, low (but nonzero) dc signal component, and available (but weak) clock signal at the bit frequency.

4.2 Optoelectronic Transceiver Approaches

Modulation considerations will constrain the design choices available for fiber-optic transmitters and receivers. Some key issues are introduced below.

(a) Digital Signal Detection

Two means of detecting the digital signal from the received and amplified optical signal are widely used. Pulse detection uses the integrated signal power over the timing interval to drive an electronic switch into one of two binary states. Edge detection drives the binary switch by a signal derived by differentiation of the transitions (edges) of the pulses making up the signal.

Edge detection is useful because of its insensitivity to the duty cycle, or dc component, of the data signal. It lends itself to general

purpose receivers in which the data rate and format are unspecified and expected to vary over a wide range. However, it makes much less efficient use of the power available in the signal to obtain a maximum SNR. As a result, a penalty of about 4 dB in increased optical signal over that of a pulse detector is required in an edge-detection receiver to obtain a given BER.[17] Consequently, high-performance fiber-optic systems use pulse detection.

(b) Multirate Digital Receivers

Some applications call for operation over a range of data rates. In such cases, both the upper and lower limits of the receiver passband are important. The upper limit is set by the greatest data rate allowed. The noise bandwidth set by this upper limit then determines the sensitivity for *all* data rates. The lower limit is set by the minimum data rate required. Technically, the ability to operate at very low rates is set by the $1/f$ frequency spectra of electronic devices. GaAs FET devices used in high-bandwidth preamplifiers can have significant low-frequency noise. In addition to this factor, efficient electronic design constrains the low-frequency capability of broadband ac-coupled amplifiers. DC-coupled amplifiers are subject to drift and are poorly suited for high-sensitivity optical receivers.

(c) Analog Transceivers

As discussed in Sec. 2, the nonlinear characteristics of optoelectronic emitters impose restrictions on performance for demanding analog applications, such as high-quality video transmission. A widely used approach for such transmission is to convert the AM signal to frequency modulation of a carrier frequency oscillator. This FM signal then modulates the optical intensity. After detection at the receiver, the resulting FM signal is demodulated to recover the analog information.

FM transmission is inherently resistant to nonlinear distortion, since the result of such distortion is to create harmonics outside the frequency band containing the information. In addition, use of wide frequency deviation in the modulation gives enhanced SNR in the demodulated signal over the carrier-to-noise ratio (CNR) of the detected signal. As a result, high-quality video can be transmitted over an analog fiber-optic channel that would severely degrade the signal if AM were used.

Fig. 12 shows test results and analytical predictions for video SNR of a studio-quality TV signal transmitted over a fiber-optic

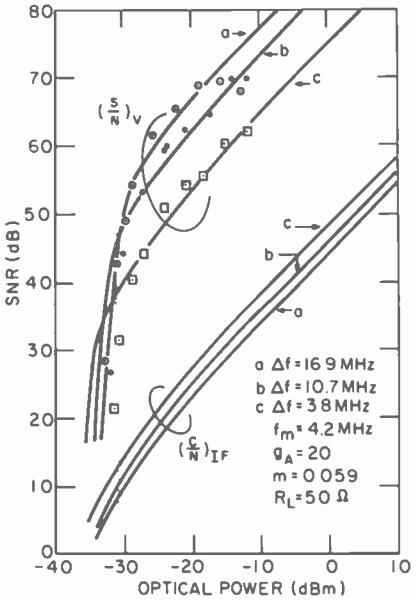


Fig. 12—Performance of a 5-km analog fiber-optic link using FM for studio-quality TV transmission. Video SNR and calculated CNR (at 70 MHz i-f) are shown versus received optical power.

link on a 70-MHz FM carrier. Several FM frequency deviations, Δf, were used. For comparison, the CNR of the i-f amplifier prior to FM demodulation is shown. The high quality of the FM transmission over a noisy and nonlinear analog channel is evident.

(d) Local Area Network (LAN) Transceivers

Both transmitters and receivers for fiber-optic network systems require special approaches to accommodate the transmission requirements of these applications. Special problems are often encountered in the following areas.

Burst Transmission

In contrast to point-to-point data links, fiber-optic networks operate with packets of data rather than continuous data transmission. Fig. 13 shows a typical packet structure for RCA's 200 Mbit/s network.[18] During the preamble, the transmitter must stabilize its output, the receiver must stabilize to the power level of the packet, and the clock recovery circuit must synchronize to the data frequency and phase. The time required for these processes lowers the network

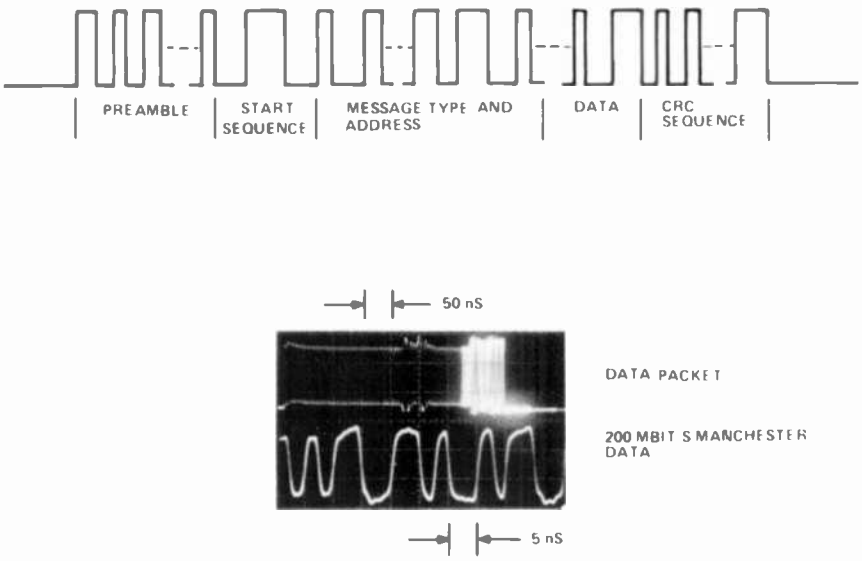


Fig. 13—Data packet structure for 200 Mbit/s fiber-optic LAN.

efficiency. Networks operating at hundreds of Mbit/s data rates require stabilization within several hundred nanoseconds or less. By contrast, point-to-point links are continuously stabilized with closed-loop circuits having millisecond response times.

Laser operation in high-speed networks is complicated by the need to keep nontransmitting terminals from generating background light. The combined background light from many transmitters on a common bus will degrade receiver performance. One solution is to modulate from complete-off rather than from threshold, as is done conventionally. However, at high speeds the resulting turn-on delay causes pulse distortion and intersymbol interference. Another approach is to store in each terminal the operating parameters for the lasers in that terminal and switch rapidly to these stored values when operation starts. Since these parameters change with time, they must be updated periodically and means must be provided to do this without seriously degrading network operation.

Dynamic Range

Multiterminal network receivers must accommodate a wide range of signal amplitudes and must switch rapidly between levels of different packets and the interpacket absence of light. High sensitivity and stability are best obtained using ac-coupled amplifier chains

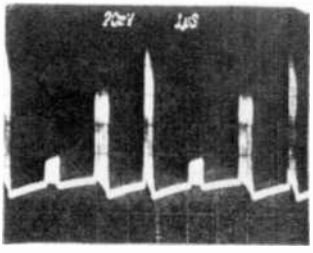
which, however, produce considerable baseline wander between packets. The performance expected of a network receiver is indicated in Fig. 14, which shows the optical-line signal with large interpacket dynamic range and baseline wander. The properly detected digital output for this line signal is shown, along with an expanded picture of one packet. Stabilization within a 50 nanosecond interval at the start of the preamble is seen.

Repeater Delay

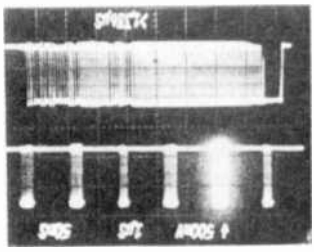
Fully regenerative bus receivers can impose significant total propagation delay on the transmitted signal. In network systems the accumulated delay from multiple repeaters can seriously reduce efficiency. Operational protocols for some networks impose requirements on delays to accommodate packet-collision detection and avoidance procedures. Network-compatible transceivers must meet the requirements of the appropriate protocol.

5. System Capability

In this final section we look at various ways of expressing the requirements and capabilities of fiber-optic communication systems. These capabilities result from the optoelectronic-component issues and fiber-optic transmission characteristics discussed in the previous sections. We look especially at the trade-offs between different optoelectronic devices and specifications and how they affect system performance.



Bus Line Signal
200 MBIT/s
> 7.8 dB Dynamic
Range



Receiver Output
Token Packets
3-Node Sequence

Fig. 14—Data packet sequences for 200 Mbit/s fiber-optic LAN. Left trace is for optical line signal; right, for bus receiver output.

5.1 Laser Versus LED

Fig. 15 shows the performance capabilities of two approaches to multimode digital transmission. One approach uses a laser diode and avalanche photodetector (APD) operating at 0.84 μm. The alternative is a 1.3-μm LED emitter and a PIN-FET receiver. The latter system shows considerably greater range at low data rates, despite the greater power output and receiver sensitivity of the laser-APD system. The low loss of the 1.3-μm fiber accounts for the improved performance. However, at high modulation frequencies the modal dispersion of the fiber dominates the performance of both systems and leads to nearly equivalent performance.

5.2 Long-Wavelength Transmission

Fig. 16 illustrates the transmission capabilities of long-wavelength digital fiber-optic systems using different wavelengths, detector technologies, and fiber distance-bandwidth products. The straight lines give performance limits for various choices of distance-bandwidth products. These products, in turn, are related to laser and fiber spectral characteristics by partition noise and chro-

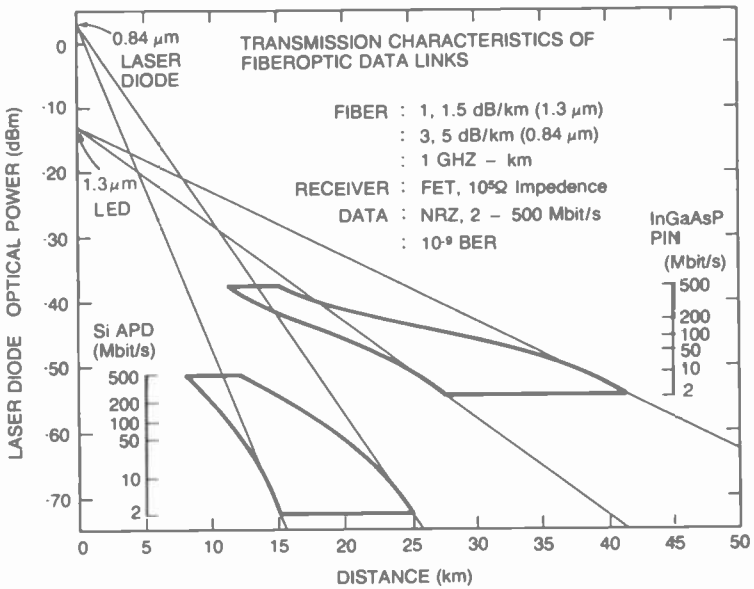


Fig. 15—Performance capabilities of typical 0.85-μm laser-based and 1.3-μm LED-based digital links.

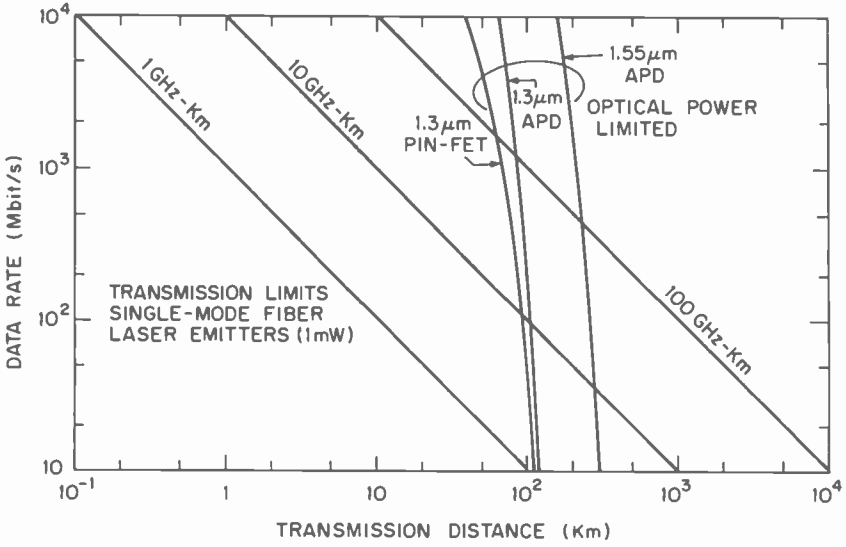


Fig. 16—Performance capabilities of long-wavelength single-mode fiber-optic links. Straight lines give partition-noise limits from chromatic dispersion (see Fig. 9). Optical-power-limited capabilities reflect receiver sensitivity and fiber loss.

matic dispersion, as shown in Figs. 8 and 9. The curved lines in Fig. 16 represent receiver sensitivities limited by detector or amplifier noise. Depending on the bandwidth and receiver characteristics, the systems may be either bandwidth limited or sensitivity (optical power) limited. These two limits are equivalent, for example, with 100 GHz-km bandwidths and an APD-based receiver operating at 1.55 μm at a data rate of approximately 500 Mbit/s.

5.3 Multirequirement Applications

When an application requires a wide range of link lengths and data rates, the optoelectronic transceiver performance specification must take into account the dispersal of system requirements. Since performance compromises and high cost result from demands for large ranges of capability in a single transceiver, the optimum solution may be a set, or family, of components that have capabilities matched to the distribution of system requirements.

Fig. 17 shows a solution to this problem for Air Force ground-based digital link requirements.[19] The links derive from diverse systems and applications. Each block on the histogram represents

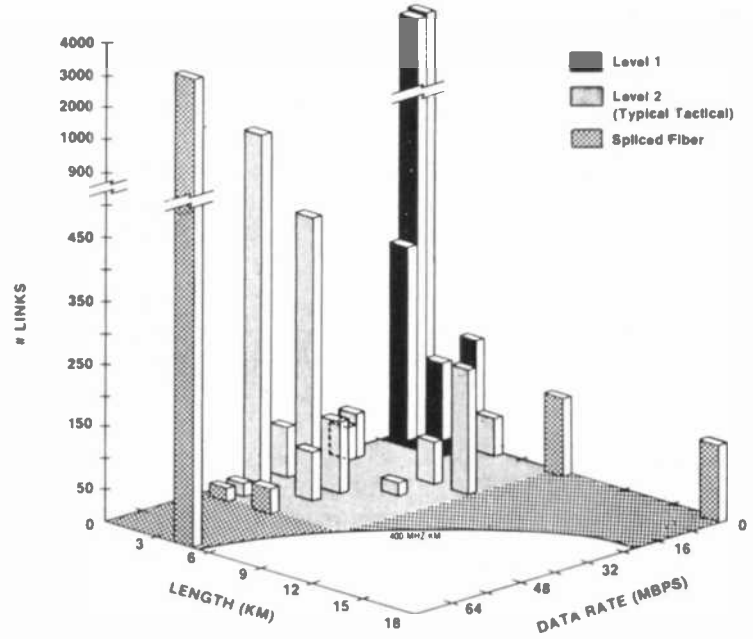


Fig. 17—Distribution of link-length and data-rate requirements for Air Force ground-based systems.

a maximum range requirement and a specific data rate. The height of each block represents the projected number of individual links for each application. Fiber losses are those associated with operation at 1.3 μm with multiple demountable connectors (one for each kilometer) on typical tactical cable.

Two classes of transceivers, designated levels 1 and 2, were specified to meet the requirements of these diverse applications. Level 1 represents limited performance at low cost designed to capture the large group of applications at the far corner of the histogram. Level 2 has greater performance to capture the large majority of applications using tactical cable with demountable connectors at 1 km intervals. Going to spliced fiber to remove connector loss allows performance to reach the bandwidth limit of the fiber and captures all applications. Additional effectiveness was gained by recognizing that the high-data-rate applications were in fact digitized video, which could be transmitted more efficiently as frequency modulation of an optical pulse train. This reduced the pulse-rate requirement to 50 Mbit/s for all applications.

5.4 Local Area Networks

The capabilities of network systems are often limited by distribution loss of the light among multiple terminals. A star coupler offers the most efficient way to accomplish this distribution. Fig. 18 shows the capability of a star-coupled network to accommodate various numbers of terminals at different data rates using different optoelectronic emitters. At a low rate (10 MHz), an LED emitter and PIN photodiode receiver can accommodate on the order of 100 terminals. This drops to ten terminals at 100 Mbit/s unless the LED is replaced by a laser. With rates of 1000 Mbit/s, it is necessary to use high-power laser devices in order to accommodate a significant number of terminals.

Summary

The performance of fiber-optic communication systems depends on two basic issues, optical power budget and signal fidelity. Optoelectronic device and component characteristics that affect these issues have been reviewed. Where possible, analytical descriptions and quantitative examples that relate directly to system performance

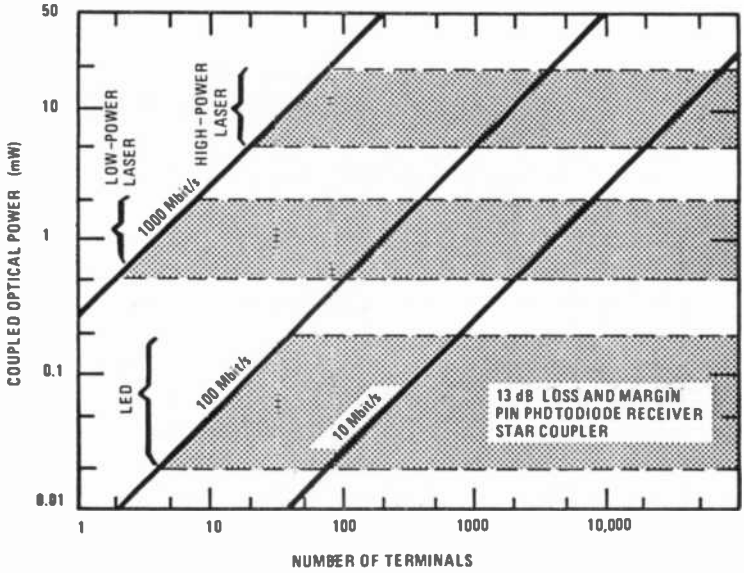


Fig. 18—Capabilities of star-architecture LAN to support multiple terminals at varying data rates and emitter optical power.

are given. In addition to performance, issues such as optoelectronic packaging, interfacing and control requirements, and reliability are also essential to system development, as are economic and standardization issues.

References:

[1] D. Gloge, A. Albanese, C. A. Burrus, E. L. Chinnock, J. A. Copeland, T. P. Lee, Tingye Li, and K. Ogawa, "High-Speed Digital Lightwave Communication Using LEDs and PIN Photodiodes at 1.3 μm," *Bell Syst. Tech. J.*, **59**, p. 1365, Oct. 1980.
[2] H. Kressel and J. K. Butler, *Semiconductor Lasers and Heterojunction LEDs*, Academic Press, New York (1977).
[3] D. J. Channin, D. Redfield, and D. Botez, "Effect of Injection-Current Confinement on Modulation of CDH-LOC AlGaAs Laser Diodes," 9th Intl. Semiconductor Laser Conf., Rio de Janeiro (1984).
[4] Klaus Petermann and Gunther Arnold, "Noise and Distortion Characteristics of Semiconductor Lasers in Optical Fiber Communication Systems," *IEEE J. Quantum Electron.*, **QE-18**, p. 543, April 1982.
[5] Kristian E. Stubkjaer and Martin B. Small, "Noise Properties of Semiconductor Lasers Due to Optical Feedback," *IEEE J. Quantum Electron.*, **QE-20**, p. 472, May 1984.
[6] D. J. Channin, M. Ettenberg and H. Kressel, "Self-Sustained Oscillations in (AlGa)-As Oxide-Defined Stripe Lasers," *J. Appl. Physics*, **50**, p. 6700, Nov. 1979.
[7] Hans Melchior, "Analog Fiber-Optic Systems," Digest of Conf. on Optical Fiber Communication, New Orleans, Jan. 1984.
[8] R. G. Smith and S. D. Personick, "Receiver Design for Optical Fiber Communication Systems," in *Semiconductor Devices for Optical Communications*, H. Kressel, Ed., Springer-Verlag, Berlin (1980).
[9] Tran Van Muoi, "Receiver Design for High-Speed Optical-Fiber Systems," *J. Lightwave Tech.*, **LT-2**, p. 243, June 1984.
[10] D. Gloge, K. Ogawa, and L. G. Cohen, "Baseband Characteristics of Long-Wavelength LED Systems," *Electron. Lett.*, **16**, p. 366, May 1980.
[11] Leonard G. Cohen and Stan Lumish, "Effects of Water Absorption Peaks on Transmission Characteristics of LED-Based Lightwave Systems Operating Near 1.3 μm Wavelength," *IEEE J. Quantum Electron.*, **QE-17**, p. 1270, July 1981.
[12] Kinichiro Ogana, "Analysis of Mode Partition Noise in Laser Transmission Systems," *IEEE J. Quantum Electron.*, **QE-18**, p. 849, May 1982.
[13] R. E. Epworth, "The Phenomenon of Modal Noise in Analog and Digital Optical Fiber Systems," *Proc. 4th European Conf. on Opt. Comm.*, Genoa, Italy, p. 492, Sept. 1978.
[14] Wolfgang Elsasser, Ernst O. Gobel, and Jurgen Kuhl, "Coherence Properties of Gain- and Index-Guided Semiconductor Lasers," *IEEE J. Quantum Electron.*, **QE-19**, p. 981, June 1983.
[15] K. O. Hill, Y. Tremblay, and B. S. Kawasaki, *Opt. Lett.*, **5**, 270, 1980.
[16] D. J. Morris, *Pulse Code Formats for Fiber Optical Data Communications*, Marcel Dekker, Inc., New York (1984).
[17] Vincent L. Mirtich, "A 20-Mband Full Duplex Fiber Optic Data Link Using Fiber Optic Active Components," Motorola, Inc., *Application Note AN-794*.
[18] D. J. Channin, J. B. Sergi, David R. Patterson, Stefan A. Siegel, and J. P. Viola, "Fiber-Optic 200-Mbit Local Area Network," *Digest of Conf. on Optical Fiber Communication*, p. 94, New Orleans, Jan. 1984.
[19] Technical Report (Interim) F30602-83-C-0142, Rome Air Development Center, Griffiss AFB, New York 13441 (1984).

Patents Issued to RCA Inventors—First Quarter 1985

January

T. N. Altman Binary Drive Circuitry for Matrix-Addressed Liquid Crystal Display (4,496,219)
F. Aschwanden Digital Sync Separator (4,491,870)
F. S. Bernard, S. M. Eliscu, and E. P. Batterman Apparatus for Generating Scaled Weighting Coefficients for Sampled Data Filters (4,494,214)
P. J. Callen and P. R. Pierce Satellite Dual Bus Power System (4,494,063)
J. M. Cartwright, Jr. Programmable Logic Gates and Networks (4,495,427)
R. J. D'Amato Arc Suppression Structure for an Electron Gun (4,491,764)
M. S. Deiss Remote Controlled Receiver With Provisions for Automatically Programming a Channel Skip List (4,495,654)
W. A. Dischert Video Disc Player Having Carriage Drive Mechanism (4,493,070)
L. Faraone Method of Making Semiconductor Device With Multi-Levels of Polycrystalline Silicon Conductors (4,494,301)
T. J. Forquer and H. Li Multiprocessor-Memory Data Transfer Network (4,491,915)
W. E. Ham Process for Forming an Improved Silicon-on-Sapphire Device (4,496,418)
J. M. Hammer and C. C. Neil Microtranslator and Microtranslator Assembly (4,495,704)
S. Harada and S. Tosima Method for Characterizing Solder Compositions (4,491,412)
F. Z. Hawrylo Method of Soldering a Light Emitting Device to a Substrate (4,491,264)
M. V. Hoover FET Negative Resistance Circuits (4,491,807)
K. C. Kelleher, N. J. Kiser, and T. J. Christopher DC Motor Servo System (4,494,052)
R. F. Keller Magnetron Filament Having a Quadrilateral Cross-Section (4,494,034)
J. G. Lea and D. J. Carlson Variable Frequency U.H.F. Local Oscillator for a Television Receiver (4,494,081)
P. A. Levine Compensation Against Field Shading in Video from Field-Transfer CCD Imagers (4,496,982)
R. B. Lovick and W. L. Firestone Television Descrambler With Security Plug Having Folded Flexible Printed Circuit Board Providing Tier TAG Memory (4,494,143)
D. W. Luz and D. H. Willis Single Controllable Switch, Push-Pull Inverter for a Television Receiver Ferroresonant Power Supply (4,492,900)
S. V. Naimpally and J. C. Tallant, 2nd Wideband Kinescope Driver Amplifier (4,494,146)
J. J. Piascinski, R. H. Axelrod, and J. Mount Method for Combined Baking-Out and Panel-Seating of a Partially-Assembled CRT (4,493,668)
W. R. Roach Stylus Manufacturing Apparatus and Method (4,490,945)
A. N. Schmitz and R. N. Clark Audio-Visual Diplexed Television Transmitter in Which the Aural Signal can be Multiplexed Without Switching (4,491,871)
G. E. Theriault Terminated Switch (4,492,937)
A. E. White, D. E. Coleman, and S. R. Miller Statistical Teaching Apparatus (4,493,651)

February

F. Aschwanden Slow Genlock Circuit (4,498,103)
A. S. Baran Stencilling a Unique Machine-Readable Marking on Each of a Plurality of Workpieces (4,497,848).
O. H. Bismarck Coincident Pulse Cancelling Circuit (4,502,014)
T. V. Bolger FIR Chrominance Bandpass Sampled Data Filter With Internal Decimation (4,500,912)
R. M. Bunting and A. Acampora Apparatus for Frame-to-Frame Comb Filtering Composite TV Signal (4,498,100)
P. J. Coyle Protective Cartridge for Disc Record (4,499,996)

D. D. Crawshaw Field-Transfer CCD Imagers With Reference-Black-Level Generation Capability (4,498,105)
C. E. Deyer System for Converting the Frequency of a Pulse Train to a Binary Number (4,499,588)
A. R. Dholakia Pickup Cartridge Having Imposed Stylus Flylead (4,502,135)
R. A. Dischert and R. J. Topper Television Gamma Corrector With Symmetrical Response about Black-Level (4,499,494)
W. E. Duvall and M. P. Hwang Television Receiver Standby Power Supply (4,500,923)
P. D. Filliman Signal Sampling Network With Reduced Offset Error (4,502,079)
C. J. Gentile, M. L. Hagge, and R. C. Croes Level Shift Interface Circuit (4,501,978)
K. J. Hamalainen and G. A. Reitmeier Television Signal Standards Conversion Using a Helical Scan VTR (4,500,930)
L. A. Harwood Hue Control System (4,500,910)
G. C. Hermeling, Jr. and M. W. Muterspaugh Double Conversion Tuner for Broadcast and Cable Television Channels (4,499,602)
W. Hinn and H. R. Fecht Noise Suppressing Interface Circuit in a Kinescope Bias Control System (4,502,073)
L. L. Jastrzebski and J. Lagowski Method to Determine the Crystalline Properties of an Interface of Two Materials by an Optical Technique (4,498,772)
L. B. Johnston System for Segmentally Refreshing the Stored Electron Gun Drive Voltages of a Flat Panel Display Device (4,500,815)
P. A. Levine CCD Imager With Improved Low Light Level Response (4,499,497)
H. E. McCandless Multibeam Electron Gun With Composite Electrode Having Plurality of Separate Metal Plates (4,500,808)
R. K. McNeely Stylus Arm Insertion Apparatus (4,501,063)
E. A. Miller Distributor Tube for CVD Reactor (4,499,853)
W. S. Pike Diode Simulator Circuit (4,500,798)
D. H. Pritchard Apparatus for Frame-to-Frame Comb Filtering Composite TV Signal (4,498,099)
G. A. Reitmeier Digital Television Signal Processing System (4,502,074)
A. Schwarzmann Method and Apparatus for Neutron Radiation Monitoring (4,498,007)
A. Schwarzmann Folded Dipole Radiating Element (4,498,085)
B. W. Siryj and A. G. Lazzery Automatic Handling Mechanism for an Optical Disc Enclosed in a Protective Cartridge (4,502,133)
P. D. Southgate, D. W. Fairbanks, R. B. Davis, and J. P. Beltz Automatic Stripe Width Reader (4,498,779)
S. A. Steckler and A. R. Balaban Digital Television Receivers (4,502,078)
R. L. Turner Video Disc Package (4,499,995)
L. K. White Method for Detecting Distance Deviations to a Photoresist Surface in an Optical Printer (4,498,775)
D. J. Wierschke Method for the Manufacture of Record Stampers (4,500.393)
J. P. Wilbur, Jr. and M. Vanrenssen Shadow Mask Washer/Spring Welding Apparatus (4,500,767)
T. Yamazaki, M. Faltas, and P. P. Webb Silicon Photodiode with N-Type Control Layer (4,499,483)

March

D. F. Battson Reducing Grain in Multi-Phase-Clocked CCD Imagers (4,507,684)
S. L. Bendell Color TV Camera With Four-Port Prism (4,507,679)
E. A. Brauer Stylus Cartridge Having Stylus Arm Restraint (4,504,941)
L. A. Christopher and D. L. Sprague Digital Filter Overflow Sensor (4,507,725)
A. G. Dingwall and V. Zazzu Flash A/D Converter Having Reduced Input Loading (4,507,649)
R. A. Dischert and J. J. Williams, Jr. Digital Matrixing System (4,507,676)
F. C. Easter and R. H. Aires Switching DC-to-DC Converters (4,504,896)
A. M. Goodman Surface Preparation for Determining Diffusion Length by the Surface Photovoltage Method (4,507,334)
R. N. Hurst, Jr. Independent Fleshtone Contours (4,506,293)
K. C. Kelleher Linear Pulse Width to Current Converter for Brushless DC Motors (4,507,591)
K. H. Knop Diffractive Color Separation Filter (4,506,949)

H. G. Lewis, Jr. Color Television Receiver With a Digital Processing System that Develops Digital Driver Signals for a Picture Tube (4,503,454)
H. G. Lewis, Jr. CCD Delay Line System for Translating an Analog Signal (4,506,288)
H. G. Lewis, Jr. Television Receiver With Digital Signal Processing Having a Digital-to-Analog Converter Control Capability (4,506,291)
P. T. Lin Mechanical Scriber for Semiconductor Devices (4,502,225)
J. T. McCann, L. W. Dobbins, and C. W. Reno High Efficiency Precision Laser Beam Switch (4,507,773)
S. Osaka and M. Sagawa Venetian Blind Construction (4,503,900)
K. J. Phillips and T. G. Tracy Control of Nutation in a Spacecraft (4,504,032)
J. J. Prusak, B. E. Lock, and J. H. Thorn Lubricant Detector and Measuring Device (4,503,806)
D. Raychaudhuri Time Division Multiple Access Communication Systems (4,504,946)
G. A. Reitmeier and F. Aschwanden Digital Phase Comparator Circuit Producing Sign and Magnitude Outputs (4,506,175)
D. J. Sauer Dynamic CCD Input Source Pulse Generating Circuit (4,503,550)
G. L. Schnable and C. P. Wu Method of Forming Semiconductor Contacts by Implanting Ions of Neutral Species at the Interfacial Region (4,502,206)
M. H. Wardell, Jr. Method for Casting a Base Directly on an Electron Tube (4,504,430)
H. R. Warren Color Moviola for Two-Tract VTR (4,504,869)
W. F. Wedam and W. E. Babcock Deflection Circuit With Linearity Correction (4,503,367)
A. E. Widmer and R. Fehlmann LPCVD Deposition of Tantalum Silicide (4,504,521)
C. M. Wine Analog Signal Comparator Using Digital Circuitry (4,503,465)
J. R. Young and B. Vanbreemen Projection Television Receiver With Unitary Optics Frame (4,506,295)
J. Zelez Diamond-Like Film and Process for Producing Same (4,504,519)

AUTHORS

Donald J. Channin received his BS degree from Case Western Reserve University in 1964 and his PhD degree from Cornell University in 1970. His PhD thesis was based on high-frequency phonon studies in solids at cryogenic temperatures, in which he used thermal pulse and microwave techniques. In 1970, Dr. Channin joined RCA Laboratories, Princeton, NJ, as a Member, Technical Staff. There he worked first on acousto-optic laser light deflection and modulation. Subsequent work on optical waveguides led to the development of several electro-optic waveguide devices and the analysis of scattering phenomena. His work on liquid crystals has included the development of an integrated-circuit inspection process and a new technique for scanned displays. During 1977–78 Dr. Channin spent a year at the University of São Paulo, at São Carlos, Brazil, as a Visiting Professor. His research there involved acousto-optic effects in liquid crystals. Since his return to RCA Laboratories he has been doing research on modulation and high-frequency effects in injection lasers and has been responsible for developing components and systems for optical communications. He has participated in efforts to establish standards in the field of fiber optics and is chairman of a JEDEC committee on fiber-optic transducers. In 1983, he was appointed Research Leader in the Optoelectronic Devices and Systems Research group.

Dr. Channin received an RCA Laboratories Outstanding Achievement Award in 1973 and 1975. He has published many technical papers and holds nine U.S. patents. He is a Senior member of the IEEE and a member of the American Physical Society, the Optical Society of America, Sigma Xi, and Tau Beta Pi.

Dean B. Gilbert graduated from RCA Institutes in 1970 with an Associate degree in Electronics Technology. He then joined RCA Laboratories, Princeton, NJ, as an Associate Member of the Technical Staff. He has been involved with testing and development work done in the Semi-conductor Laser group. In particular, Mr. Gilbert has been working on the characterization of single heterostructure lasers, cw double-heterostructure lasers, 1.3-μm InGaAsP lasers, single-mode constricted double-heterostructure (CDH) lasers, and the CDH large-optical-cavity lasers (CDH-LOC). He has also assisted in designing the automated laser-characterization system (ALCS).

Mr. Gilbert has coauthored six papers in the laser field and is a member of the IEEE.

Istvan Gorog received the BSc (1961), MSc (1962), and PhD (1964) degrees in Electrical Engineering from the University of California at Berkeley. In 1964, he joined RCA Laboratories Princeton, NJ, as a Member of the Technical Staff. His research activities have included lasers and their applications; holography; optical communications; the investigation of prerecorded-video recording and playback techniques; laser deflection and modulations; photochromic, cathodochromic, and electrochromic devices; and a wide variety of display systems. He has also been interested in the analysis of image transmission and display in terms of the quantitative descriptors of human visual perception. This work included the initiation of an experimental psychophysics program at RCA Laboratories. He also initiated and was responsible for all optical reading, recording, and instrumentation efforts related to the development of the RCA VideoDisc. During 1968 he was on leave from RCA as a National Science Foundation Post-doctoral Fellow at the Laboratori Nazionali di Frascati and the European Space Research Institute in Frascati, Italy. There he worked on the problems of production of high-temperature plasmas by laser irradiation of solid targets and scattering of laser radiation by collective plasma fluctuations.

In 1983, Dr. Gorog was appointed Director of the Manufacturing Technology Research Laboratory at RCA Laboratories. He directs research and development in the areas of automation, process control, and applied physics, in support of RCA's manufacturing operations.

Dr. Gorog has published and presented over 60 technical papers, has been granted 21 U.S. patents, and has several patent applications pending both in this country and abroad. He is a member of the American Physical Society and of Eta Kappa Nu.

Frank Z. Hawrylo joined RCA Laboratories, Princeton, NJ, in 1960, and has since been involved in many diverse projects dealing with III-V crystal growth and device fabrication. In 1966, he was promoted to Technical Staff Associate and placed in charge of a support research laboratory in III-V technology. He then participated in the fabrication of the first diode laser made by liquid-phase epitaxy. He also made important contributions to the method of growing high-quality AlGaAs lasers by liquid-phase epitaxy, which led to the development of close-confinement structures and lasers operating cw at room temperature. More recently he has been involved with the liquid-phase preparation of InGaAsP heterojunction lasers, LED devices used for fiber-optic communications, and the machine bonding of III-V semiconductor devices.

In 1978 Mr. Hawrylo was promoted to Associate Member, Technical Staff. In 1979 he received the BS degree in Material Sciences from Thomas A. Edison College and was presented the "Edison Centennial Award." In 1980 he was promoted to Member of the Technical Staff.

Mr. Hawrylo holds 23 U.S. patents and is author or coauthor of approximately 40 technical publications. He has received an IR-100 Award (1969), two RCA Laboratories Outstanding Achievement Awards (1976 and 1981), and NASA Technical Awards (1976, 1977, 1980, and 1982). He is a member of the American Physical Society and a Senior member of the IEEE.

John J. Hughes received the BSEE degree from Thomas Edison College in 1980. From 1950 to 1961, he worked in the microwave tube department of ITT Laboratories. From 1958 to 1961, assigned to Associated Testing Laboratories as a representative of ITT, he monitored and evaluated the environmental testing of microwave tubes. In 1962, he joined the RCA Electron Tube Division, Harrison, NJ, as a production foreman in the microwave tube division. In 1963, he joined RCA Laboratories, Princeton, NJ, as an Associate Member of the Technical Staff. He has been doing research on solid-state microwave devices, microstrip technology, solar generators, energy management systems, and laser semiconductor devices.

Mr. Hughes holds seven U.S. patents and has written more than 16 papers. He is a member of the IEEE.

Sheng T. Hsu received the BS degree in Electrical Engineering from National Taiwan University in 1958; the MSEE degree from National Chiao-Tung University, Taiwan, in 1960; and the PhD degree in Electrical Engineering from the University of Minnesota in 1966. From 1966 to 1970, Dr. Hsu was with Fairchild Semiconductor Research and Development Laboratory as a Member of the Technical Staff. From 1970 to 1972, he was an Assistant Professor in the Department of Electrical Engineering, University of Manitoba. In 1972 he became a Member, Technical Staff, at RCA Laboratories, Princeton, NJ, working on semiconductor devices and integrated-circuit technologies. In 1985, Dr. Hsu was appointed a Fellow of the Technical Staff. He holds 24 U.S. patents in integrated-circuit technology and has published numerous papers in professional journals.

Dr. Hsu is a member of the IEEE.

Werner Kern received his education in chemistry in Switzerland and the United States. His 1947 research thesis on the discovery of fluorescing carcinogenic hydrocarbons in soil established an important new branch of environmental science. He was a research chemist with Hoffmann-La Roche in Switzerland and New Jersey. Later he joined the Nuclear Corporation of America and became Chief Chemist directing research in radiation chemistry. In 1959 Mr. Kern joined RCA Electronic Components and Devices to investigate semiconductor processes by radiochemical methods, was head scientist on several research projects, and was put in charge of radiological safety. In 1964 he transferred to RCA Laboratories, Princeton, NJ, as a Member, Technical Staff; he was promoted to Senior Member in 1983 and Fellow of the Technical Staff in 1985. Mr. Kern has been working in semiconductor process research, specializing in the characterization and device applications of dielectric films, chemical vapor deposition, and etching processes. From 1974 to 1979 he was project scientist for government-sponsored research contracts on glass passivation, dielectric defects, and silicon solar cells. He has been a technical consultant on silicon device processing and chemical safety to RCA Laboratories and the RCA product divisions for the past 20 years.

In 1983 and 1984 Mr. Kern was Chairman of the Electrochemical Soci-

ety's Dielectrics and Insulation Division and a member of the Board of Directors of this Society. He is a member of the American Chemical Society, the American Vacuum Society, and the Research Honorary Society of Sigma Xi. He is listed in *American Men and Women of Science*, *Who's Who in Technology Today*, and *Who's Who in the East*, is the author or coauthor of over 80 scientific publications, holds 10 U.S. patents plus several pending, has presented numerous technical lectures and seminars, and has organized and chaired scientific symposia.

He received the Callinan Award of the Electrochemical Society in 1971 for his pioneering research in chemical vapor deposition. Mr. Kern received an RCA Laboratories Outstanding Achievement Award in 1966 for his work in integrated-circuit process research; another in 1973 for his contributions to the glass passivation of silicon device structures; and a third in 1983 for developing a ternary glass (CVD BPSG) for low-temperature fusion. He was also awarded a 1981 NASA certificate of recognition. He coedited and coauthored *Thin Film Processes*, published by Academic Press in 1978, and has been Course Lecturer on Chemical Vapor Deposition for the American Vacuum Society since 1981.

William A. Kurylo received a BS degree in 1972 from East Stroudsburg State College and a diploma from Data Processing Repair Institute in 1974. He joined RCA Laboratories, Princeton, NJ, in 1978 as a Research Technician in the Integrated Circuit Technology Group. He was promoted to Technical Associate and Senior Technical Associate in 1980 and 1985, respectively. His work has centered around the deposition of improved dielectric films for silicon devices by chemical vapor deposition on both laboratory and production scales for several R&D programs. He has also carried out physical and chemical measurements of these films using infrared spectroscopy, ellipsometry, SEM, and chemical-etch-rate techniques.

Jer-shen Maa received a BS degree in Metallurgy in 1968 from National Cheng-Kung University, Taiwan, and a PhD degree in Materials Science in 1974 from the University of Minnesota. From 1975 to 1979, he served as an Associate Professor at National Tsing Hua University, Taiwan, and briefly as a temporary staff member of the Thin Film Technology group, RCA Laboratories, Princeton, NJ. In 1980, Dr. Maa joined RCA's Solid State Division, Somerville, NJ, as a Member, Technical Staff. In 1981, he returned to RCA Laboratories. His research activity has been concerned primarily with thin-film materials, including thin-film nucleation and growth, various deposition technologies, thin-film reactions, diffusion and oxidation, materials characterization, and dry-etching technology. In recent years, he has been engaged in research and development on refractory metal silicides and plasma-etching technology for integrated electronics application.

Dr. Maa has published 20 technical papers and has one patent pending. He is a member of the American Vacuum Society and the Electrochemical Society.

Harvey P. Miller received his BS in Mechanical Engineering from Rensselaer Polytechnic Institute, MS in Mechanical Engineering from Massachusetts Institute of Technology, M. Phil. in Plasma Physics from Columbia University, and is completing his Ph.D. in Applied Physics from Columbia University in the fall of 1985. He joined RCA Government Communications Systems in 1984 as a Senior Member of the Engineering Staff. Mr. Miller has been responsibile for computer-aided thermal design involving computational heat transfer and fluid dynamics. He has developed a front ended computer-aided-design general purpose heat-transfer program for use on the GCS Auto-trol engineering workstation to allow interactive graphics for 3-D model generation and transient thermal analysis.

Mr. Miller is a member of the American Society of Mechanical Engineers, American Physical Society, Society for Industrial and Applied Mathematics, American Nuclear Society, and Tau Beta Pi and Pi Tau Sigma Honor Societies. Mr. Miller was a Raytheon Fellow from 1966 to 1967 while attending MIT, and has been elected to Who's Who in Frontiers of Science and Technology for the past two years. He has numerous technical publications in Computational Fluid Dynamics and Heat Transfer, and is a registered Professional Engineer in New Jersey and Pennsylvania.

W. Ronald Roach received the BE degree in Physics from Cornell University in 1961, and the MS and PhD degrees in Physics from the University of Illinois, Urbana, in 1963 and 1966, respectively. His doctoral thesis involved measuring the heat capacity of dilute solutions of helium-3 in helium-4 over a temperature range of 0.010-0.3 K. In 1966 Dr. Roach joined RCA Laboratories, Princeton, NJ, as a Member, Technical Staff, to do optical research. He built a 2-photon spectrometer, using it to study the band structure of insulators; investigated mode locking in Nd^{3+}-doped YAG lasers and the design and fabrication of electro-optic crystal modulators for producing a train of picosecond pulses; and studied the optical properties of films of VO_2, demonstrating the use of these films as a high-contrast erasable holographic storage medium. He also developed photoconductively controlled deformable mirror light valves that were coupled to a laser scanner to give a real-time, large-area TV display.

In 1975 Dr. Roach began working on optical instrumentation for the manufacture and inspection of VideoDisc parts. He was responsible for the precision laser recording of deep, narrow grooves into photoresist. He developed a diffraction system that reproducibly measures the 80-Å-deep audio signals on VideoDisc surfaces to ± 1Å. Later he worked on a system for laser scribing of the three layers of amorphous-silicon solar cells to produce series-connected high-voltage panels. In 1982 Dr. Roach was promoted to Senior Member of the Technical Staff and, in 1983, was appointed Head of the VideoDisc Playback Research group, in the VideoDisc Systems Research Laboratory.

For contributions to a team effort that developed the laser defect detector Dr. Roach received an RCA Laboratories Outstanding Achievement Award in 1975. He received a second such award in 1976 for developing

a diffraction spectrometer for the optical measurement of the groove and signal geometries of the VideoDisc, and a third one in 1980 for his contributions to the micromachining of VideoDisc styli.

Dr. Roach holds 12 U.S. patents and has authored or coauthored 14 technical papers. He is a member of Sigma Xi and the American Physical Society.

Christopher J. Tino began his career in the U.S. Navy as a Fire Control Technician Petty Officer in charge of a 3-dimensional air search radar system. He has been employed by the Olivetti Corporation in their Advanced Electronic Laboratories, Applied Materials, Inc., and National Semiconductor. During this time, he served as a research technician, field service engineer and instructor, and as a consultant on chemical vapor deposition (CVD) and epitaxial production line operations.

Since joining RCA Solid State Technology Center, Mr. Tino has implemented a computer controlled maintenance management system for SSTC and has participated in several task forces on CVD for the Solid State Division. He is presently assigned to the CVD section and is involved in characterizing and implementing several thin film deposition processes.