RCA Review

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

Contents

June 1978 Volume 39 Number 2 ISSN 0033-6831

Yield-Area Analysis: Part I—A Diagnostic Tool for Fundamental Integrated-Circuit Process Problems

W. E. Ham

RCA Laboratories, Princeton, N.J. 08540

Abstract—This paper considers the nature and value of the information contained in the distribution of good circuits across production IC wafers. It is shown that the actual production wafers are perhaps the only source of information available at reasonable effort and cost which reliably reflects the true yield-limiting factors existing in the process. Communication with these factors can be aided by analyzing the shape and orientation of the regions of the wafer wherein zero yield is obtained—thus, yield-area analysis. Fundamental divisions are made between parametric and point-defect limited yield. The key factor that allows the separation of these types is the quasi continuity of parametric values across wafers. It is suggested that any yield formula that assumes that point defects are the dominant yield-limiting factor must be multiplied by an area usage factor (AUF) which reflects the total fraction of available area being used for nonzero yield. On many wafers the AUF is less than 0.2, thereby making it one of the most important characteristic parameters of the process.

Yield-area analysis can be a very powerful tool to guide the improvement of existing processes, and the AUF should be considered a fundamental design parameter for new circuit types.

1. Introduction

During the course of manufacturing integrated circuits one finds many opportunities for destroying potentially good circuits on a wafer. These opportunities fall into three basic classes (1) ignorance, (2) neglect, and (3) inability to prevent. It is primarily with regard to the first class that this paper is addressed. If one knows that certain problems exist, then diagnostic efforts are redundant. If one does not exercise effective controls then he deserves the consequences of an uncontrolled process. On the other hand, if one is doing his best to control his process, it is essential

that he understand how to analyse all important available sources of information concerning the nature of the causes for nonworking circuits. One rich, yet frequently overlooked, source of information is the distribution of working circuits over the wafer. This paper attempts to provide a foundation for analysing these distributions and to suggest the causes of the nonideal distributions frequently observed on production wafers used for LSI circuits.

2. Preliminary Discussion

2.1 Point Defects versus Parametric Defects

When one has obtained a sizable set of wafer maps of working circuits from individual wafers produced from the process of interest he has the basic starting data. These maps may be obtained by any of several means as discussed later in this paper. If any circuits at all on a wafer work, one must assume that the gross features of the process were performed satisfactorily. Those wafers that have no working circuits may or may not have had gross processing mistakes. In any case, since no working circuits exist there is no data and these zero yield wafers will be ignored for the purposes of this paper. The failure of the non-working circuits to satisfy the testing criteria must have been caused by parametric variations that were incompatible with the circuit design or by local physical disruptions (commonly called point defects). Chemical nonuniformities may also exist but these will be reflected in an electrical or physical property and do not need to be directly considered. These causes for failure assume that the testing process itself reliably indicates working and nonworking circuits and does not suffer from mechanical problems, such as probing or electrical problems (for example, noise or transients).

For the purposes of this paper it is important to clearly define the terms "parameter" and "point defect". A parameter, as used here, is any measureable feature of the process that is quasi-continuous over an area of the wafer of the size of a circuit. Frequently the term parameter is used to mean electrically measured characteristics such as leakage current or sheet resistance. While these characteristics may be parameters in this paper, they are certainly not the only features of the process that are quasi continuous. Film thicknesses, line widths, and other physical or chemical properties are also considered as parameters for this paper if they are quasi continuous. Not all electrical properties qualify as parameters under the above definition. For example, the dielectric strength or leakage current of individual devices may not exhibit continuity and therefore not qualify as a parameter. Many electrical properties are, in fact, found to be quasi continuous, as recent detailed investigations by

the author has shown.[1,2] Thus with a few exceptions, electrical properties qualify as parameters.

Those features of the process that do not exhibit quasi continuity are termed here "point defects." If a leakage current is not quasi continuous it is defined as a point defect. This is not as ridiculous as it first sounds since the noncontinuity in the electrical current is very likely *caused* by a local physical disruption that would be typically termed a point defect. Therefore the terms parameter and point defect are retained since for most situations they have their traditional meanings.

In the absence of independent information (such as by visual inspection), one can consider that a differentiation between parametric problems and point-defect problems might be possible by examining the distribution of good and bad circuits on the wafer. It is extremely important that this distinction be made since the action taken to improve the situation will depend strongly on the dominant type of problem. One *must* recognize that for a given set of conditions either parametric properties or point defect properties can dominate the yield and that *a priori* assumptions are futile. Attempts to increase parametrically limited yield by decreasing defect densities or *vice-versa* are clearly antiproductive. The basic feature of parameters. as defined above, is their quasi continuous nature. Point defects by their very nature are not continuous with position. We will show that the distinction can often be made by examining a wafer yield map.

It is, of course, possible that the *density* of defects may be systematically different at different positions on the wafer surface.[3,4] For example, improper heating or cooling cycles can preferentially induce defects in bulk silicon near the edges.[5] Edges of wafers can preferentially induce defects in the masks which are subsequently printed on the next wafer. Wafers are handled near the edges by processing fixtures. We therefore are not particularly interested in the edges of the wafers, since they will nearly always be relatively heavily defected. Away from the edges, however, there is little reason to expect preferential locations or quasi continuous gradation of point-defect densities. If such a distribution were present, the density itself would qualify as a parameter. Therefore, the primary distinction between parametric yield limitations and point-defect limitations lies in the fact that as one moves continuously in the same direction across the wafer surface a dividing line will exist, beyond which the parametric values (because of their quasi continuous nature) do not allow the circuit to function or perform within acceptable bounds. Random-defect-limited yield, on the other hand, should exhibit no such dividing line.

Within the area of the wafer where circuit failure is not determined by parametric distributions, point defects are necessarily (almost by

definition) responsible for nonfunctioning circuits. Similarly joint responsibility might be claimed where parameters are out of specification.

2.2 Deliniation of Zero Yield Areas

If areas of appreciable size exist where there are *no* good circuits, specifically not within a "handling distance" of the wafer edge, and the distribution of good circuits is reasonably uniform over the part of the wafer where good circuits do exist, then there is a very good chance that the circuits in the "bad" area are caused by parametric problems. Thus, the nature of the information contained in the yield map is that of a locus (or loci) that divides the wafer into the regions of zero yield and nonzero yield. The shape and position of these regions can provide substantial information concerning the causes of the zero-yield portions, while the density of good circuits in the nonzero-yield portions contains the information concerning the point-defect distributions.

At the time of this writing, we have not yet produced a specific algorithm for drawing these dividing lines. It is worthwhile to comment on the criteria used to produce the lines drawn on the wafers in the figures to follow. They were not drawn entirely arbitrarily, but they are somewhat subjective. A rigorous method for drawing these lines might be useful but also might mask long-range trends that are readily visible to the eye but difficult to incorporate into a computer program.

The mean separation between good circuits is first noted. The locus or (loci) should not pass any closer to a good circuit than half of this mean separation. Areas confined only to the wafer edge are deemphasized, since these areas are expected to have near zero yield. Continuation of a loci from a non-edge region of the wafer into the edge region is allowed. One simply looks for areas of the wafer that have no good circuits within an area whose diameter is approximately twice the mean separation between good circuits. If such an area exists, the locus is drawn in the region between the nearest good circuits and the zero-yield portion at a distance of 0.5–1.0 times the mean separation from the good circuits and approximately parallel to the line intersecting the centers of two neighboring good circuits. This line is not precisely defined by this method, but large differences between the loci drawn by different people would not be expected as examination of the wafer maps to follow will show. This method uses the ability of the human brain to perform very sophisticated statistical analyses on the highly statistically distributed location of good circuits in the nonzero-yield portions.

2.3 The Area Usage Factor

It is apparent that the area usage of the wafers *modulates* the point defect limited yield formulas and as such can be more important than traditional considerations, such as chip area A and average defect density D. In *general* one can write:

$$\text{Yield} = (\text{AUF}) \times (\text{Results of whatever point defect yield formula is appropriate}). \quad [1]$$

where AUF is an area usage factor. It is simply defined as the area of the wafer used for yield, A_y, divided by the total area of the wafer A_w. It may also conveniently be defined in terms of the handling area A_H.

$$\text{AUF} = \frac{A_y}{A_w - A_H}. \quad [2]$$

We note that the AUF is *not dependent on chip size*, and is the absolute maximum yield obtainable with *no* loss from point defects. The AUF may be readily determined by examining wafers from several typical lots and is a valid and very important characteristic number for a particular set of design rules, processing conditions, and testing criteria. It may be very sensitive to small changes in these features, as will be shown in an example to be discussed later.

A measure of the effects of defects in the nonzero-yield areas can be obtained if one can assume that a constant, though spatially random, density of defects exists on the same wafer. In this case the yield for a given wafer is given by:

$$Y = (\text{AUF}) f(A,D)$$

where D is the defect density. For the particular circuit design, processing conditions, and testing conditions being used, one can determine D in principle if f is known. As a crude approximation for f a simple exponential can be used to arrive at a first order estimate of D. For example, for wafer i,

$$Y_i = (\text{AUF})_i \exp(-D_i A),$$

which implies that

$$AD_i = -\ln (Y_i/\text{AUF}_i)$$

It is of course necessary to know A or D_i for some specific wafer to obtain an absolute measure of D_i. In general, only the product AD_i is determinable. Furthermore D_i defined this way is a conglomerate of all mask levels and processing steps used to make the circuit and represents an *effective* defect density. The main value of this exercise is to determine the *spread* in the effective defect density being produced by the process.

The ratio of AD_i between different wafers is a measure of the reproducibility of the defect density and does not depend on A. This ratio is useful when discussing the need for, and methods to be used for, better processing conditions and process controls. It reflects a characteristic of the process which probably actually depends on point-defect densities.

3. Experimental

3.1 Data Collection

Except in special cases where a dominant problem exists that is easily visible (such as metal damage for example), the only data routinely available that is suitable for addressing the question of processing problems and for determining the real yield-limiting factors is the data existing from the electrical testing of the actual circuits of interest.

There are several reasons why the use of production wafers as the source of data is good (and is much better than attempting to use artifacts that are sensitive to only a small part of the total process). The main reason for using the circuit itself for the diagnostic tool is that it represents the actual item of primary interest. It contains the primary data.[6] All the idiosyncracies that might cause a problem are present in the exact form of interest, namely in the circuit itself. A second, very important, reason for using the circuit, if possible, is that a semi-continuous sampling of the actual behavior of the production line is available. This is virtually impossible to obtain from any non-salable artifact, primarily because of the high cost of processing and testing and general unwillingness to commit production resources for this kind of work.

The obvious problem with using the circuit is that it is difficult to perform accurate diagnosis from the test results alone. The hope of finding easier roads to diagnosis of the causes of nonfunctioning or malperforming circuits is the primary motivation for considering using artifacts. Test results from artifacts are ostensibly easier to interpret. In the authors experience, the use of artifacts (or "test" wafers) presents a great many opportunities for obtaining misleading information concerning the applicability of the results to production circuits. It is only with the greatest care and attention to detail in both the processing and testing of artifacts that the results can be effectively used to influence (favorably) the performance of the production circuits. In some cases it may be *necessary* to use structures that are less complicated than the circuit in order to relate to the effects causing the unsatisfactory performance. The use of noncircuit structures should be carefully and deliberately guided by the performance of the production circuits. An example of an effective use of a test wafer is given in the next section.

Since most circuits of interest today are relatively complex, it is reasonable to assume that circuit parametric values, such as stand-by current, access time, etc., are not due to a single unique cause. Within any nominally identical batch of circuits, the most that can be reliably stated is that the circuits do or do not satisfy the criteria for being "good". This statement is made with the assumption that the accept–reject choice is made purely on the basis of electrical tests, and the test criteria remain constant over the period of data sampling.

We recognize that specialized test programs used on a limited sampling of wafers or circuits that attempt to localize the failure site or mode can be very beneficial for diagnosing problems. These are limited, however, in that they are not economical to use in a continuous-volume application and require additional efforts to obtain the data. If visual tests are also used as part of the testing criteria, additional direct diagnostic information is available for the sample of parts examined. Since it is a necessary condition that all the electrical features of the circuits be satisfactory and these electrical features can be relatively easily tested, it is not logical (from a production viewpoint) to perform extensive visual examination before electrical testing. One can therefore assume that detailed visual inspection results will only be available from circuits that work electrically. By the time the visual inspections are performed, all relationship of the circuits to the wafer position is usually lost, so that it is difficult at best to relate the results back to the process.

We are lead to the same conclusion from several routes, that the most effective place in the production process, as it is usually performed, for obtaining the maximum information about the yield-limiting mechanisms is after the wafers have been electrically tested but before the individual circuits are separated. As indicated in a previous section, it is in the spatial relationship of the working circuits to each other that the best clues to the process are found.

This relationship, spatial distribution, or wafer map, can be obtained by a number of methods. The most direct is the recording of the positions during electrical testing. This allows automatic printout of the maps in a easily handled form and allows further manipulation of the data, if desired, using well known computer methods. When automatic position recording is not available, an inexpensive alternate method is to mark the position of the good circuits as they are separated after scribing and breaking. This is done by the operator on the backing paper after all good circuits have been removed. This requires perhaps one minute per wafer for a typical size chip. The backing paper, which is normally discarded, then contains an array of marks that indicate the position of the good circuits. This method was used for all data presented in this paper. Since the wafers are inked during testing the distribution can be recorded by

making "blotter" impressions while the ink is still wet. This method can ruin an entire wafer if the ink is smeared and requires some skill. The inked wafer may be photographed at considerable additional expense and effort. In some cases simple visual observation is sufficient.

3.2 Results

Fig. 1 shows data from three different lots of SOS/CMOS 1K RAM's processed on 3-inch sapphire wafers. The X's indicate the position of the good circuits and the O's indicate the "knock out" positions. Cursory inspection reveals that self-consistency within the same wafer exists within the nonzero yield areas. For these wafers, particularly F, G, H and K, one can identify multiple regions of nonzero yield. The local yield varies strongly from wafer to wafer, thus indicating the unpredictability of effective defect densities even under nominally identical conditions. The most important observation to make, however, is the vast areas of zero yield that exist on most of the wafers. Good circuits are obtained within two or three pellet spacings of the wafer edge on many of the wafers. This clearly defines the handling distance to be on the extreme outer portions of the wafer. For this set of data, parametric incompatibility probably accounts for the major part of the bad circuits. If this process and circuit design were compatible in a parametric sense, the overall yield would be at least tripled.

This data clearly establishes: (a) "real life" production wafers potentially can contain a great deal of information concerning the factors influencing the overall yield; (b) sharp dividing lines between areas of the same wafers where yield is obtained and where zero yield is obtained exist on many wafers; (c) within the nonzero-yield areas of the same wafer, the density of good circuits is reasonably constant; and (d) the zero-yield portion of many wafers can be comparable to or larger than the area of the wafer yielding good circuits.

It is worth noting at this point that plots of radial distribution of yield (such as shown in Fig. 2 for data similar to that of Fig. 1) will nearly always show a dramatic falloff in yield as one proceeds away from the center of the wafer. These plots primarily reflect the nonuniform use of area rather than radial distributions of local yield. Inspection of the actual distributions shows that local yields are quite high near the edges for many of the wafers. Therefore, while radial distribution plots are useful for demonstrating the the poor usage of the outer areas of the wafer, they are not suitable for diagnostic work. Furthermore, the slope of the radial distribution curve indicates that gradual transitions exist between high- and low-yield areas of the wafer. Inspection of the actual maps show this to be completely untrue.

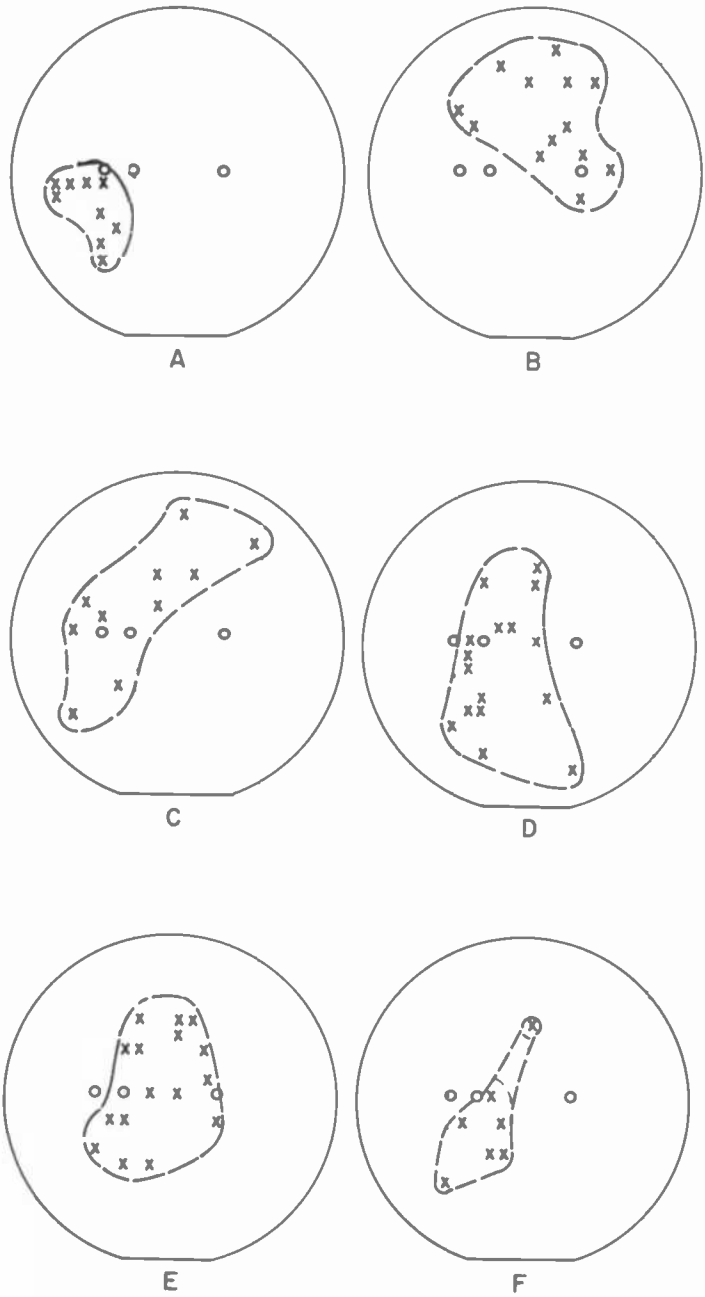

Fig. 1—Yield distribution from three typical lots of SOS 1K RAM parts. The lightly dashed lines in F, G, H, I, and K indicate alternate choices for the loci.

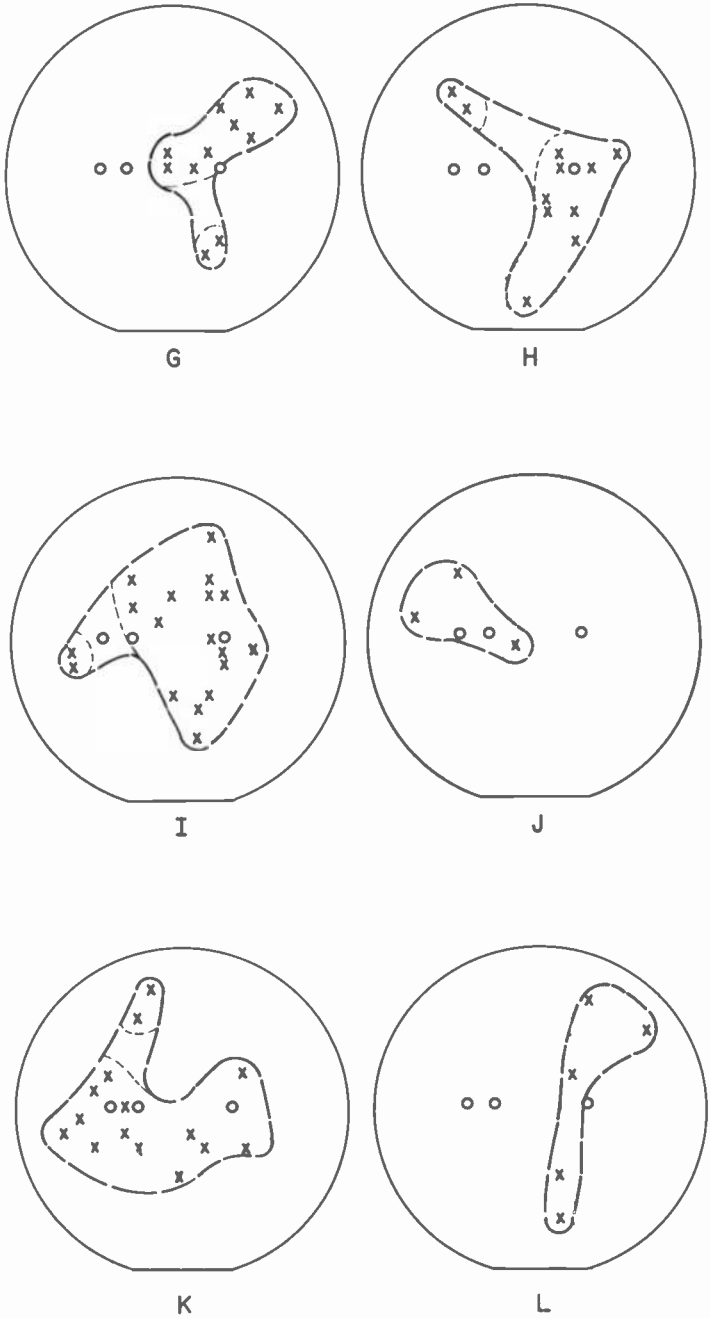

Fig. 1—(*Continued from previous page*)

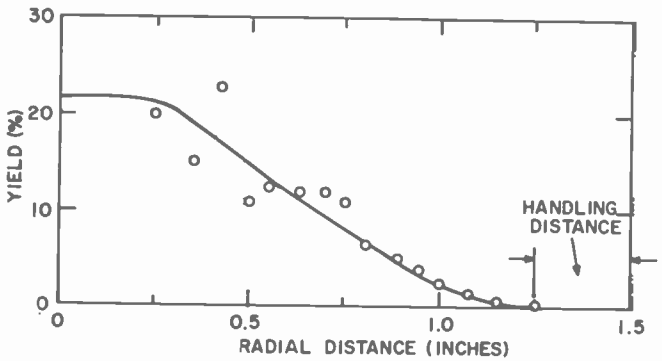

Fig. 2—A typical plot of the radial distribution of yield from data similar to that of Fig. 1.

Radial plots must be made using concentric rings of constant area to determine the yield at a position r from the center if they are to have any value at all. This means that the rings will have smaller width near the wafer edge than near the wafer center. If r_1 is the inner radius and r_2 is the outer radius of any given ring, the total area is kept constant by choosing r_2 such that

$$A = \pi(r_2^2 - r_1^2) \text{ or } r_2 = \sqrt{\frac{A}{\pi} + r_1^2}\,. \qquad [3]$$

If one mistakenly maintains a $r_2 - r_1$ constant, the incremental area sampled near the edge of the wafer is much larger than that sampled near the center.

For the particular case of the SOS RAM data of Fig. 1, reasonable test devices existed on the circuit periphery for acquisition of electrical data. Examination of these devices across entire wafers showed that none of the measured electrical parameters were incompatible with the circuit design. Visual inspection, however, revealed that a good chance existed that the physical tolerances required for the circuit were too demanding for the available technology. An effort to make the circuit more compatible with the process by making only the physical tolerances of the circuits less demanding (not the design or testing procedure) produced dramatic increases in the total fraction of the wafer that produced working circuits. This data is shown in Fig. 3. The major difference between the data of Fig. 3 and that of Fig. 1 is that much more of the non-edge region is used for yield. The local yield on some of the wafers of Fig. 1 is as high as that for most of the wafers of Fig. 3. We can therefore conclude that the increase of wafer area used for yield (AUF) was a direct result of less critical physical tolerances being required.

As additional support for the existence of substantial regions of zero yield on many production wafers and to show that these effects are not

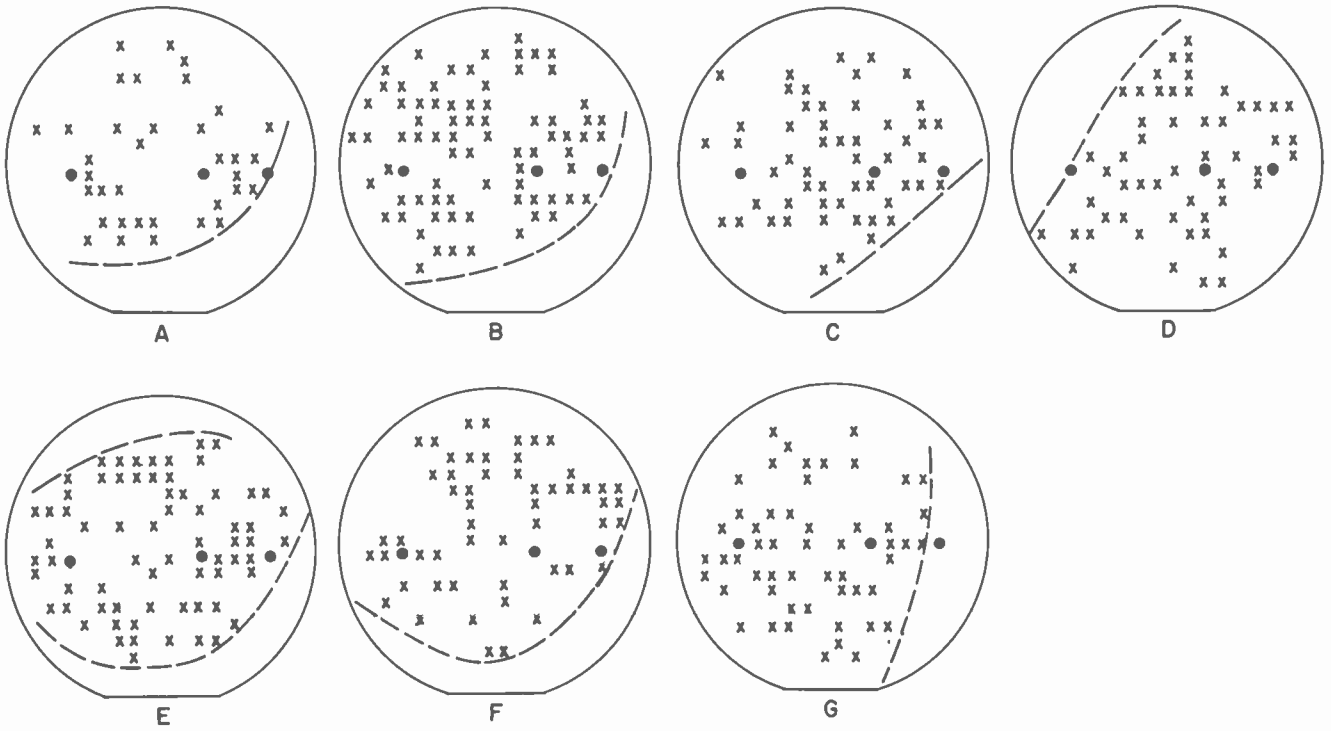

Fig. 3—Yield distribution from a lot of SOS 1K RAM parts. This circuit was of identical design to that of Fig. 1, except that certain physical tolerances were made less demanding.

peculiar to SOS technology, data from a lot of bulk silicon NMOS 1K RAM circuits on three-inch bulk silicon wafers is shown in Fig. 4. Here again the handling distance is confined to the extreme outer edge of the wafers. Within the non-edge portions of the wafers, the local yield of good circuits varies widely from a low on Wafer E to nearly 100% on parts of wafers B, C, G, H, I, J, and K.

For the specific example of the lot shown in Fig. 4, only Wafer F is using almost all of the non-edge wafer to produce circuits. On some wafers, less than half of the available non-edge area is being used, and on the overall lot perhaps 20–25% of the available area is not being used. Thus for this process the AUF is $\simeq .75$. Multiple isolated regions of high yield do not exist on the same wafer, for any of the wafers of Fig. 4.

4. Discussion

The data in the preceding section amply demonstrates that large yield loss can and does occur for reasons other than isolated point defects. Clustering of good circuits can be caused entirely by parametric incompatibility. Multiple, well developed clusters of good circuits were never found, although some wafers showed some clustering of low yield and a very few had some multiple nonzero yield regions. These low-yield clusters could be caused either by defects or by parametric problems. If lack of defects caused the high-yield clusters, why were only single, well developed clusters (which usually occur near the alignment keys) found on the same wafer?

It is, of course, essentially impossible to deconvolve any set of data such as a yield map from a single wafer into a definite cause for the zero-yield portions of the wafers. The zero-yield loci may be caused by different parameters on each side. However, examination of several wafers can produce good reason to suspect specific causes. For example, the data of Figs. 3 and 4 strongly suggest a possible photomask alignment tolerance problem, since most of the zero yield loci are convex (this is discussed in Part II of this paper*) and are randomly oriented. Nonrandom yield-figure edge positions or shapes from wafer to wafer strongly suggests a problem related to part of the process where the wafers are fixed in orientation. Further detective work can probably reveal the exact place where the yield loss is occurring. The data of Fig. 1 suggests severe parametric incompatibility that is only partially caused by simple mask-alignment problems. An etching nonuniformity, for example, coupled with alignment tolerance problems would be a good candidate to investigate for these wafers. (Such a situation was, in fact, found.) In

* Scheduled for a future issue of *RCA Review*

A
B
C
D
E
F
G
H

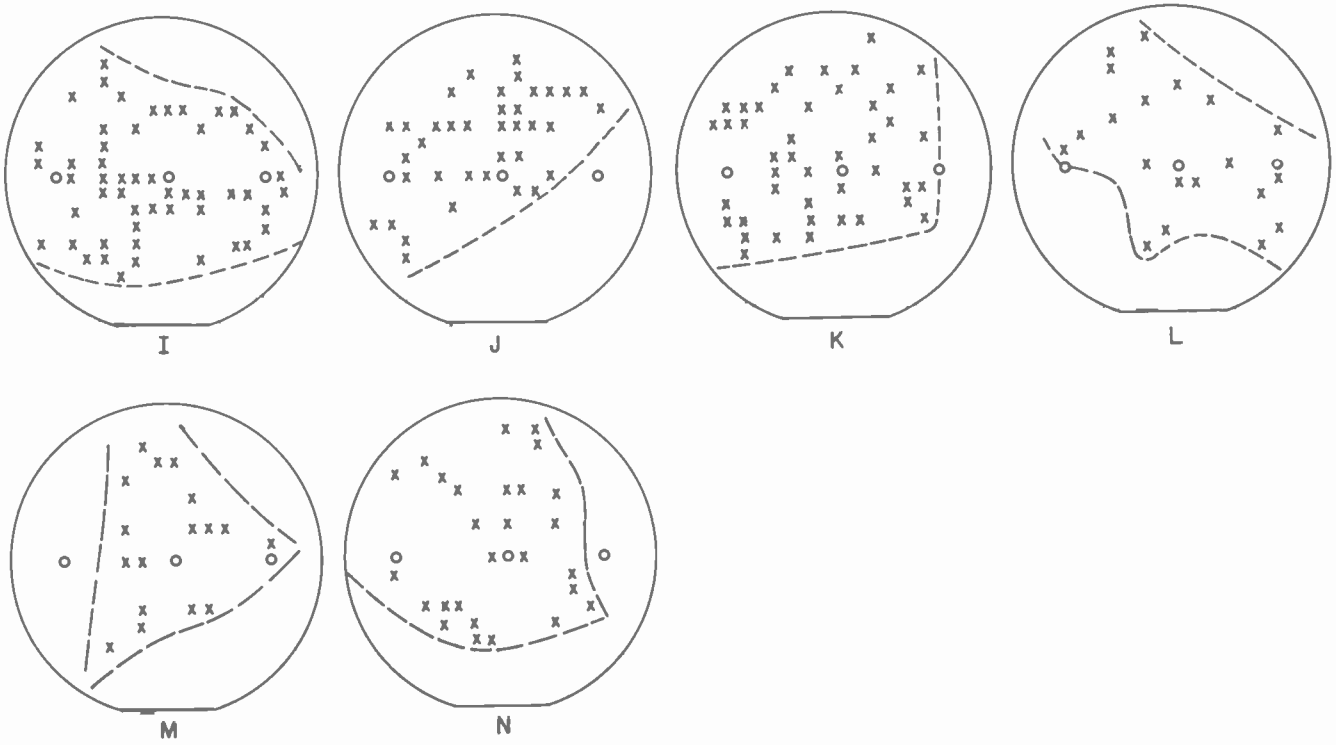

Fig. 4—Yield distribution on a typical lot of bulk silicon 1K RAM parts. The X's indicate the position of the good circuits and the O's indicate the "knock out" positions. The lightly dashed lines in E, F, G, and H indicate alternate choices for the loci.

many cases, particularly where ion implantations using photoresist masks are involved, there is no simple physical way to test the alignment obtained after the wafer is finished since there is no visible evidence of the alignment remaining. (Sophisticated dedicated electrically sensitive test structures can be used for this purpose if available.) For these situations, the zero-yield loci shapes can be the only clue to a possible problem at that processing step.

Some parameters commonly found to cause zero-yield areas of wafers are:

(1) Alignment of photomasks to images already on the wafer. This is perhaps the most common source of zero-yield regions and Part II of this paper will discuss these effects in some detail.
(2) Nonuniform film thicknesses across wafers. This frequently causes problems when batch etching processes are employed.
(3) Nonuniform line widths and/or spaces across wafers. This is usually associated with overetching a nonuniform film or with nonuniform photoresist exposure or development.
(4) Nonuniform distributions of fixed charges. This usually affects the silicon surface inversion potentials (whether the charge is in the dielectric or in the silicon) or the electrical resistance of interconnects.

It is also possible that a slowly varying (across a wafer) physical dimension can allow a smaller size of point defect to become effective. For example, point defects of a certain size might not cause a failure in a narrow poly line on part of the wafer, but could be very effective in producing a failure in another part of the wafer where the poly line was narrower (due to the distribution of the etch rates across the wafer). This type of failure must also be termed a parametric failure.

Detection of the parameters that are the likely cause of the problems is best done using test devices located on the circuits themselves, coupled with careful visual examination of the wafers. In most practical cases in the authors' experience the *on-circuit* test devices are poorly designed for parametric data acquisition, are nearly impossible to test automatically, or do not reflect the features of interest. It is of primary importance that a *sampling of the entire wafer* be available, since we need information on the *continuity* of the data and information concerning the spatial *distribution* of the parameter values. Test devices in centrally located knock-out positions are *manifestly unsuitable* for this job. If the production wafers themselves are not testable because of lack of suitable test devices, then an approach similar to that described in Refs. [1] and [2] can be used. Under this method, entire wafers of test devices that reflect virtually everything of importance, in a parametric sense, to the process are processed as an entire lot through the production fa-

cility. These wafers are then carefully tested to determine which parameters are not compatible with the circuit design being used. This approach requires only a few test wafers if implemented properly. It is difficult, however, in this case (and in every case where non-circuit artifacts are being used) to ensure that the test wafers are processed with the same care and exact methodology that a typical circuit wafer will receive. It is also difficult to test an artifact with the same stimuli that internal circuit devices experience during the circuit operation. Therefore the use of test wafers should be a last resort and extreme care must be exercised in order to simulate the environment experienced by the production circuits. Unfortunately this last resort is frequently required because it is usually impossible to deconvolve circuit electrical data to a point where a specific parametric value that can be related to the process can be unambiguously obtained. It is worthwhile to note that the effort to relate to the defect-limited yield in the nonzero-yield portion with artifacts will require many wafers and a much larger overall effort. The circuits themselves are probably the best vehicle to address improvement of the defect-limited regions.

The efforts to find the offending parameters can be costly and, when they are found, improving the situation may be another matter altogether. However, where to put the effort will become clear, and a very substantial saving in the total energy required to effect a positive change on the product wafers will be realized. If an effective diagnostic and corrective effort is used, it will not take long to recover the costs incurred through improved area yields.

Of course the first step in effecting this improvement is to determine whether areas of zero yield do in fact exist. The creation of the zero-yield loci with any reasonable accuracy requires that a substantial number of good circuits exist on the same wafer. The smaller the die size the better the edge resolution. For some very large area circuits with low yield, it will be difficult to detect zero-yield areas. For these circuits it is absolutely critical that parametric compatibility be designed into the circuit using known expected parametric distributions. As the size of individual wafers continues to grow (currently 5 inches for some applications) factors such as thermal stability and theta accuracy during the photomask alignment process as well as actual dimensional stability of the wafers themselves becomes even more important. The analysis of the wafers of Figs. 3 and 4 on 2-inch diameter wafers would probably show no parametric problems.

It may happen that, as in the case of Fig. 3, expanding the size of a circuit to make it parametrically compatible with the process may actually produce more good circuits per wafer even though fewer possible circuits exist on a given size wafer.

Arbitrarily increasing the size or changing the shape of the wafers to provide more potentially good area does not guarantee more good circuits per wafer. Area-yield analysis can greatly help to determine if parametric incompatibility exists and to point corrective action in the right direction.

5. Conclusions

The data and discussions in this paper have led to the following conclusions:

(1) The primary conclusion is that significant information concerning the processing of wafers is contained in the distribution of good and bad circuits across production wafers and that this data is available with minimal acquisition effort.

(2) Application of yield models which assume that defects are the dominating influence must be restricted to the nonzero-yield portions of the wafer. The result of these models must be multiplied by the appropriate area usage factor (AUF) to arrive at the proper result.

(3) Great value in terms of better circuit design and yield can be expected by understanding the parametric properties of *whole wafers* and using these results to direct process improvement efforts precisely where they belong.

(4) Examination of several lots of actual production wafers has clearly shown the existance of large areas of zero yield with most wafers exhibiting sharp dividing lines between zero- and nonzero-yield areas.

(5) Within the nonzero-yield areas of the same wafer, reasonable uniformity of local yield exists.

(6) The local yield varies widely from wafer to wafer and therefore is essentially *a priori* unpredictable for individual wafers.

(7) In many cases the areas of zero yield are caused by parametric incompatibility—not by defects.

(8) Actual production circuits are one of the best tools for defect related yield studies. Artifacts cannot be processed with the volume required for determining the *true* effects of the *total* process without gross expenditure of possibly uncompensated money and effort.

(9) Plots of radial yield distribution are very likely to be misleading in terms of the sharpness of the transistion between the zero- and nonzero-yield portions of the wafer. Only the original map is truly suitable for detailed examination.

(10) For parametric diagnostic purposes it is important to consider *individual wafers* as opposed to statistical collections. Statistical collections can produce some information concerning constantly failing sites (mask defects) or gross population shifts, but they mask the true shape of the yield loci in much the same way as radial distribution plots.

Acknowledgments

The author thanks N. Turner for making some of the yield data available and N. Goldsmith and C. S. Kim for valuable discussions. Special thanks are due to R. Sunshine for a critical review of this paper.

References

[1] W. E. Ham, "Test Pattern Design and Analysis for SOS/LSI," Interim REPORT No. 1, National Bureau of Standards Contract No. 5-35916 (June 1977).

[2] W. E. Ham, "Intrachip and Spatial Parametric Integrity—An Important Part of IC Process Characterization," *Extended Abstracts IEDM*, Abstract No. 19.6, p. 406, 1977; IEEE Cat. No. 77CH1275-7ED.

[3] R. M. Warner, "Applying a Composite Model to the IC Yield Problem," *IEEE J. Sol. St. Cir.*, **SC-9**, p. 86 (1974) and references therein.

[4] A. Gupta, W. A. Porter, and J. W. Lathrop, "Defect Analysis and Yield Degradation of Integrated Circuits," *IEEE J. Sol. St. Cir.*, **SC-9**, p. 96 (1974) and references therein.

[5] G. H. Schwuttke, "Yield Problems in LSI Technology," *Proc. Reliability Physics Conf.*, p. 274, 1970; IEEE Cat. No. 70C59-PHY.

[6] W. E. Ham, "Test Data Reduction," *IEEE Trans. Manufacturing Technology*, **MFT-5**, p. 24 (1976).

Advances in BiMOS Integrated Circuits

Otto H. Schade, Jr.

Abstract—The transfer characteristics of BiMOS MOSFET's throughout the nanoampere to milliampere range are presented, and their use in current mirrors, amplifiers, oscillators, and timing circuits is discussed. Linear operation in the subthreshold region is shown to provide extended performance in micropower integrated circuits, with transconductance levels similar to those of bipolar devices. Advances in MOSFET pair-matching are analyzed; commercial capabilities are described; and the combination of subpicoampere input-bias levels with protected-gate devices is shown to be practical at elevated temperatures.

Introduction

The introduction of the first commercial monolithic MOS/Bipolar (BiMOS) integrated circuit in 1973 was followed by a series of operational amplifiers that exploited the unique input characteristics of field-effect transistors.[1,2] Subsequent developmental efforts have led to improved techniques for matching the transfer characteristics of MOSFET pairs, and also for reducing the leakage current of the gate-protective network. The availability of both bipolar and complementary MOS devices on a common substrate[3] immediately suggests applications combining both analog and digital functions. Custom circuits for a broad range of applications are under development. Such an evolution of BiMOS circuits could be foreseen some years ago.

Investigations of micropower circuit design have shown, however, that the MOSFET has more to offer than low-power logic and near-ideal buffering. In the subthreshold region of their transfer characteristics, both NMOS and PMOS devices show levels of transconductance nearly equal to those of bipolar transistors, and one can extend the performance capabilities of current mirrors, reference circuits, amplifiers, oscillators, and timers with economical BiMOS designs.

This paper provides a general description of these developments from the viewpoint of an IC designer, so that the reader can more readily appraise the usefulness of this new "mixed technology". Circuit design with enhancement MOSFET's, although not complicated, does require a more

specific knowledge of device parameters than does bipolar design, and it is therefore useful to characterize some BiMOS devices and IC configurations as well.

MOSFET Operation and Characterization

Operation of BiMOS devices is practical over six orders of current magnitude or more, but the square-law voltage/current relationship usually attributed to IGFET's applies to only one-half this range. As linear circuits become more complex and efforts are made to reduce stage currents and power levels, the circuit designer may find it necessary to plan operation at or below the transition region between square-law and exponential transfer characteristics. Furthermore, micropower IC's can use the exponential region to advantage because MOS transconductance approaches that of bipolar devices, and gate-source voltage offsets and instabilities are substantially reduced.

Hayashi and Tarui[4] described MOS tail currents and exponential leakages in 1967; Richman[5] proposed a weak inversion criterion for channel conduction below threshold; and Gosney[6] experimentally investigated the subthreshold region of PMOS transistors. Much of the subsequent literature describes the physical mechanisms[7] and refines analytical models for LSI digital design, but until recently, emphasis has remained on the deleterious effects produced: for example, parasitic leakage in dynamic memories and power consumption in MOS logic. In 1977, Vittoz and Fellrath[8] built silicon-gate CMOS current references, and described an oscillator and amplifier stage designed for operation in the weak-inversion region. Tsividis and Ulmer[9] show an improved CMOS IC voltage reference operated at a total current of 16 μA. A BiMOS monolithic smoke-detector chip incorporating operational amplifier and timer circuit-stages operating at 10 to 100 nA was recently described by the author.[10]

Figs. 1a and 1b show the measured transfer characteristics of typical BiMOS n and p devices. While the high-current range is square-law, a transition is seen to occur at a threshold of approximately 10 μA, and subthreshold currents are exponential. The two regions are better defined in the transconductance curves of Figs. 2 and 3, which have been graphically constructed from the transfer characteristics. Subthreshold operation is essentially independent of device topology, and the transconductances were found to be:

$$g_{mp} \simeq I/40 \text{ mV} \quad [1]$$

$$g_{mn} \simeq I/60 \text{ mV} \quad [2]$$

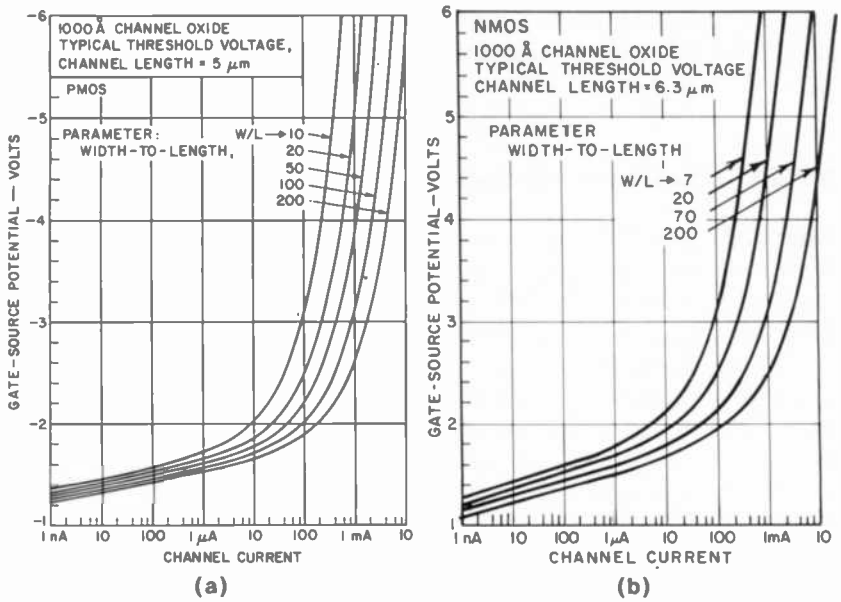

Fig. 1—Typical BiMOS PMOS and NMOS transfer characteristics.

The $w/l = 100$ PMOS and complementary $w/l = 70$ NMOS devices are developmental types having channel portions whose gate metallization can be isolated and biased off, so that channel width is variable. An additional PMOS design having $w/l = 160$ provides further data from which a family of w/l curves may be constructed. Values below $w/l = 25$ have

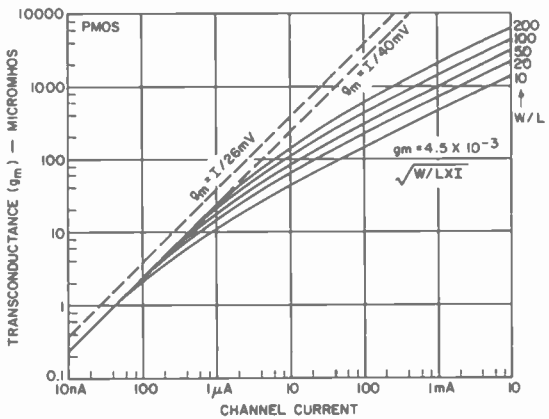

Fig. 2—Typical PMOS transconductance curves.

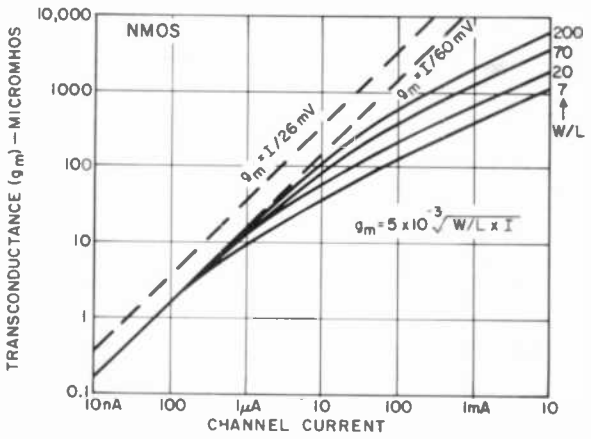

Fig. 3—Typical NMOS transconductance curves.

been extrapolated. Unfortunately, a sufficient quantity of samples having a nominal threshold was not available, so that entire data sets were shifted by fixed amounts up to 0.3 volt in order to present a coherent family. The curves of Figs. 1a and 1b are intended to illustrate the characteristic behavior of typical BiMOS types, but should be further corroborated before being considered as the basis for an analytical model.

BIMOS Current Mirrors, Sources, and References

The current-mirror amplifier is a basic element of integrated-circuit design, providing sources and sinks of both bias and signal currents, impedance-transforming functions, and active loads.[11] Although n-p-n mirrors are dependable building blocks, operating successfully over large current and temperature ranges, their lateral p-n-p counterparts are more limited, and suffer low beta levels and occasional processing difficulties. Note that the base currents of the p-n-p mirror in Fig. 4 detract from the input current drawn from the master diode, particularly with geometrically-larger slave transistors, $n > 1$. In addition, should any member of a multiple-output bipolar mirror become saturated, all of the partners would become affected by loading of the common-base bus. In contrast, the enhancement MOS mirror has essentially infinite beta, and the electrostatic coupling between devices precludes a dc loading of the gates.

Operation of the PMOS mirror of Fig. 5 is readily explained with a voltage-feedback model. Input current tends to pull down the gate-bus

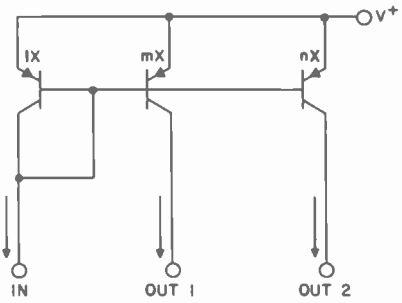

Fig. 4—The basic p-n-p current mirror.

potential until a master-diode channel current is generated sufficient to meet the input demand. The resultant gate–source potential biases all other slaved devices. Current ratios are established by channel geometry, w/l, although (as with bipolar mirrors) dynamic output resistance of the slaved transistors produces a change in ratio with voltage excursions.

A comparison of several 1:1 p-n-p mirrors and cross-coupled PMOS pairs ($w/l = 160$) is shown in Fig. 6. The p-n-p mirror ratio is strongly beta dependent, and peaks in the several-hundred microampere range. Removal of the base currents by a PMOS device, Fig. 7, produces the dashed curve for a cross-coupled p-n-p pair, greatly improving ratio accuracy over an extended range. The PMOS mirror ratio is essentially constant until affected by leakages at approximately 3 nA. In fact, above a 10-μA threshold, the square-law characteristics are an advantage, producing excellent ratio control. The lateral construction of n and p MOSFET's permits consolidation of multiple-output devices in a common isolation boat, thus providing improved packing density.

The MOS mirror amplifier may be cascoded in a manner similar to bipolar mirrors to achieve high output resistance and capacitive decoupling. A Wilson equivalent is shown in Fig. 8, although the gate-source

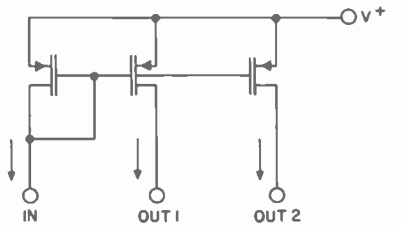

Fig. 5—The basic PMOS current mirror.

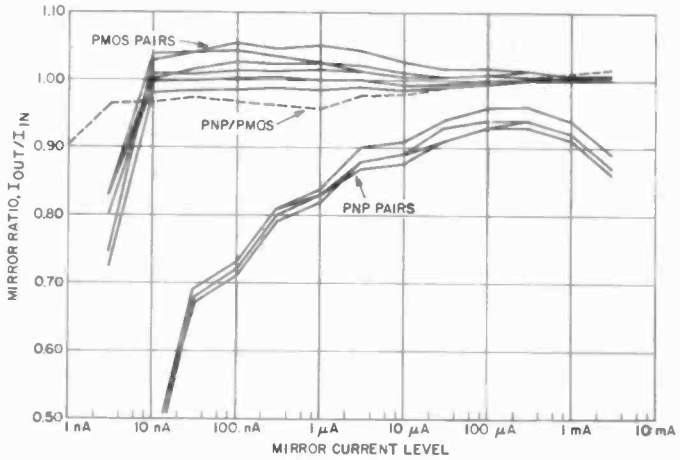

Fig. 6—P-type mirror performance.

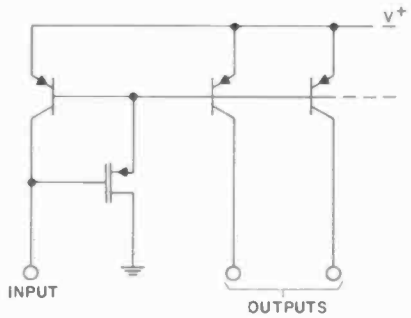

Fig. 7—Removal of p-n-p base currents with a PMOS device.

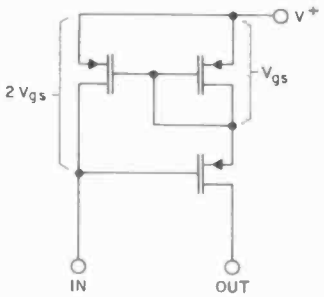

Fig. 8—An MOS Wilson equivalent.

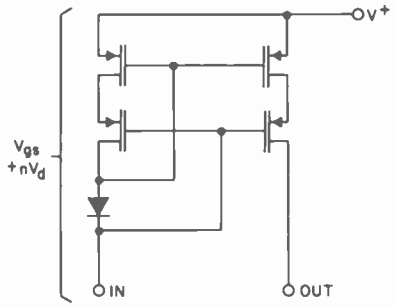

Fig. 9—The nested MOS cascode.

offsets of BiMOS devices may make this configuration insufficient of operating potentials at higher currents. An alternative, based on the considerable difference between gate-source voltage and drain-source saturation potential, is shown in the "nested cascode" of Fig. 9. Here, input current biases the second-gate pair some multiple of bipolar diodes below the first, and reduces the potentials needed to operate both input and output circuits.

Mirror ratios considerably greater (or smaller) than unity can be produced by cascading amplifiers, as in Fig. 10, to minimize required chip area. The n-p-n mrror shown produces a current step-up of $(n + 1)$ beyond that of the PMOS pair, provided that $\beta_n > n$. Removal of the bipolar diode results in a beta-multiplied composite of the PMOS transistor. The MOS and n-p-n devices can even share common diffusions as a true composite structure. The cascaded-amplifier concept can be applied to the cascode potential problem previously discussed. The circuit of Fig. 11 uses an n-p-n mirror to reduce the current density in

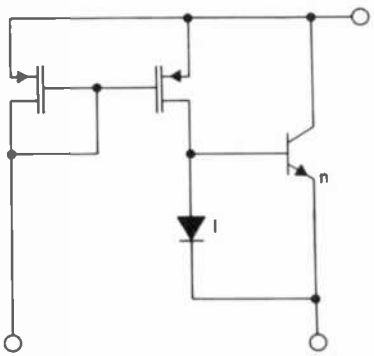

Fig. 10—A cascaded BiMOS mirror.

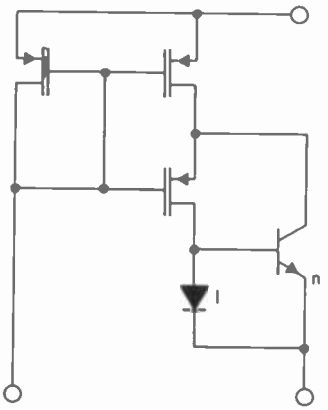

Fig. 11—Channel-current density reduction.

the PMOS cascode transistor, which in turn decreases its gate–source bias and makes available sufficient operating potential for the PMOS-mirror slave transistor.

Mirror-circuit designers have made use of PMOS devices, although their NMOS complements are also suitable. The NMOS mirror offers no better ratio control than can be attained with n-p-n's and, therefore, has not been extensively used in linear applications. In the past, the manufacture of NMOS devices having the necessary stability of transfer characteristics throughout their life has generally been considered more difficult to achieve than PMOS devices with the same qualities, because the gate-oxide field in the NMOS device is more liable to produce impurity migration. Although in an absolute sense the stability of the NMOS device may be inferior to the PMOS, excluding it from more critical applications such as a potential-stressed differential pair, mirror operation puts the same fields on both master and slave gate oxides. Recent BiMOS custom circuits, in which linear signals and saturated logic excursions are intimately associated, have brought about renewed interest in NMOS mirrors.

Fig. 12 shows the 85°C life stability at 100 nA and 100 μA for BiMOS mirrors in plastic packages. The figure shows that operation at the 100 μA levels is very stable for several thousand hours. At the 100 nA level, variations are about $\pm 2\%$, in the order of measurement error and test-fixture stability. Performance is considered to be proven adequate for many applications.

MOS devices also find application in reference and bias circuits, both as high-quality temperature-stable mirrors and as diode-connected

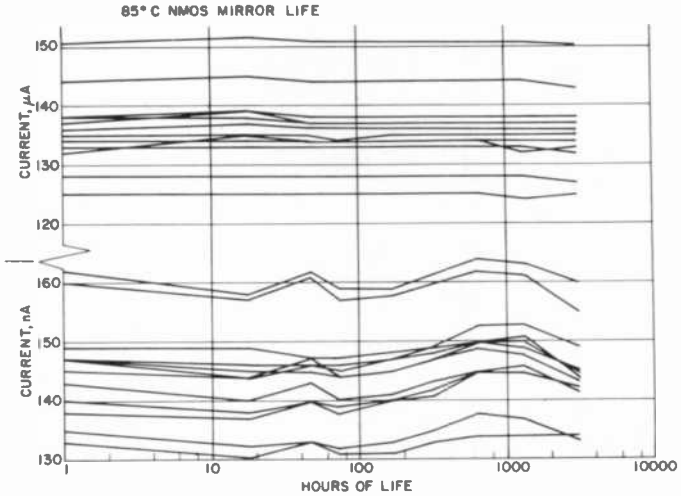

Fig. 12—NMOS mirror life at 85°C.

reference elements, although dynamic resistance may be relatively high. Fig. 13 shows a bias circuit such as is employed in the CA3140 op amp. The resistor/transistor at the far left provides a starting current, and the loop sets up at the current

$$I = \frac{V_{gs} - V_{be}}{R} \qquad [3]$$

Note from Figs. 14 and 15 that the temperature coefficient of V_{gs} can

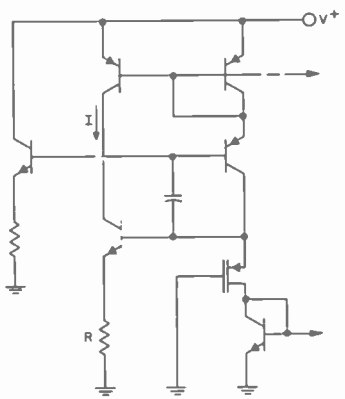

Fig. 13—A PMOS voltage-reference circuit.

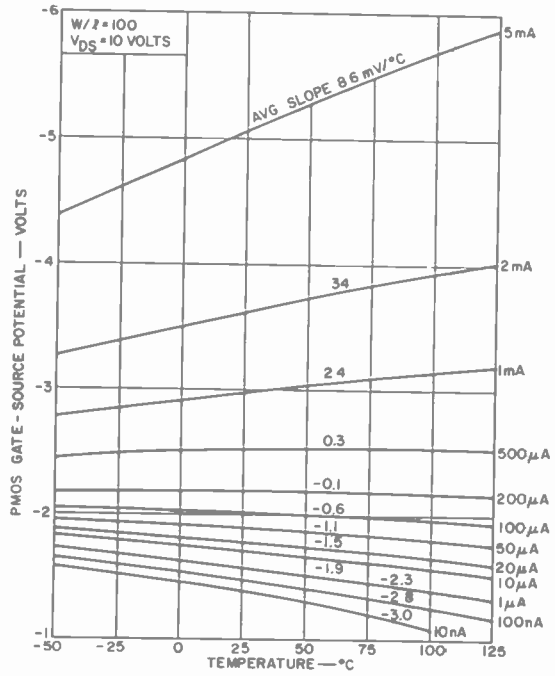

Fig. 14—Temperature dependence of PMOS transfer characteristics.

be chosen to provide negative, near-zero, or positive bias-current coefficients.

The circuit of Fig. 16(a), similarly to a configuration designed by Steckler,[12] derives the major current component from the n-p-n base-emitter potential differences, as in the Brokaw bandgap reference. The 10:1 device geometry shown causes a 60 mV potential, having a positive temperature coefficient, to be impressed across resistor R+. The major influence on the current generated in R− is provided by the negative coefficient of diode D. When the currents in R+ and R− are added in proper ratio, loop current I_o is relatively independent of temperature. Fig. 16(b) shows this current (measured in the 100-μA range) as a function of temperature for various resistor ratios. Over a 0 to 70°C range, it is possible to hold variations within 1%, i.e., in the order of 100 ppm/°C. A larger range of −50 to +125°C, however, is more greatly affected by nonlinearity of the 200 ohm-per-square resistor diffusions, and excursions of 4 to 5% are apparent. Discrete film resistors (which require considerably different R ratios) have produced a coefficient of 10 ppm/°C.

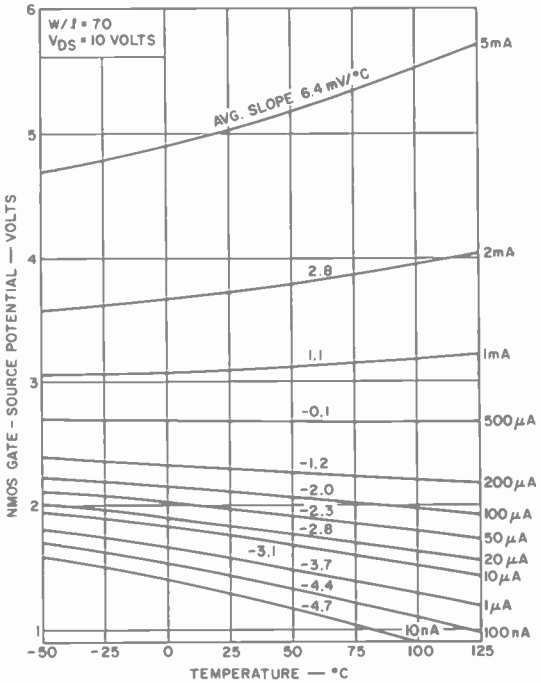

Fig. 15—Temperature dependence of NMOS transfer characteristics.

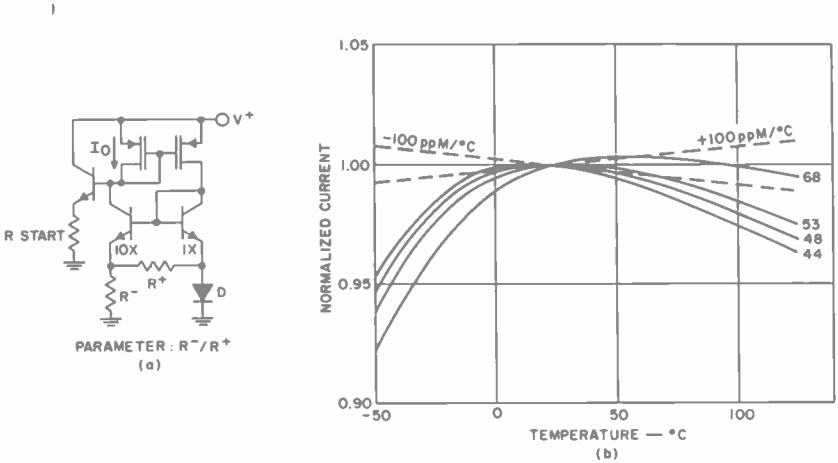

Fig. 16—(a) Current reference circuit using PMOS mirror and (b) loop current as a function of temperature for various resistor ratios.

Advances in Pair Matching

Integrated-circuit technology offers designers the opportunity to match the transfer characteristics of devices to a high degree. Classic applications are found in current mirrors, resistance (ladder) networks, binary switches, and the differential input stages of operational amplifiers. Continued need to improve such circuits has led to various "trimming" techniques when it appeared that the normal limits of layout and processing techniques had been reached. The efficient use of chip area requires that one determine the cause for and nature of matching errors, that is, whether error is the result of variation in device geometry and/or a nonhomogeneity of the diffused regions. In addition, an error histograph should be examined to determine if the distribution is random or if an "ordered" deviation (built-in offset) exists.

For example, the input stage of an op amp may consist of a long-tailed PMOS differential pair terminated in an n-p-n mirror pair. It has been observed, when a linear bipolar process is employed, that, so long as emitters 1.2-mil square (30.5 μm square) or larger are used, ordered errors are insignificant compared to the random offset spread. An electrically cross-coupled n-p-n quad is necessary to reduce the random spread because of diffused variations (to perhaps ± 300-μV offset). In contrast, ordered PMOS pair offset may constitute an unacceptable portion of the total spread (or it may be unpredictable from layout to layout), and efforts to improve matching by a brute-force increase of channel area or further interdigitation may prove inadequate. The remainder of this section will treat the reduction of ordered error in MOSFET pairs.

BiMOS MOSFET's operated above threshold have transfer characteristics:

$$I_d \propto V_{gs}{}^2 \cdot w/l \qquad [4]$$

Channel length l is by far the most critical photomask-defined parameter to reproduce in a match-device pair, since $w/l > 10$ normally applies. It has been observed that an imperfectly matched artwork master-pair A/B, from which a photomask is generated, produces silicon device pairs A_1/B_1, A_2/B_2, A_3/B_3 . . . . . A_n/B_n, which inherit the imperfections.[13] Drain currents for adjacent individual devices show the relationship:

$$i_{A2} \simeq Mi_{A1}, i_{B2} \simeq Mi_{B1}, \qquad [5]$$

from which it may be inferred that:

$$l_{A2} \simeq Kl_{A1}, l_{B2} \simeq Kl_{B1} \qquad [6]$$

Also, it is generally known that ratios of adjacent device parameters

(such as resistance ratios) hold more constant than absolute parameter values, which for the MOSFET can be translated to

$$l_{A1}/l_{B1} \simeq l_{A2}/l_{B2} \simeq N. \qquad [7]$$

More explicitly, silicon-pair mismatch contains both random and ordered components, and the latter may dominate to the extent that an imperfect master must be improved to achieve better-matched pairs.

In theory, better matching may be achieved by a process of electrically cross-coupling adjacent chip pairs that have been generated from the same imperfect master, Fig. 17. Random-pair errors are statistically reduced by $\sqrt{2}$, but ordered errors will be much more dramatically re-

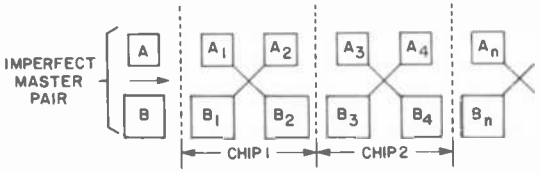

Fig. 17—Schematic diagram of electrically cross-coupled adjacent chip pairs generated from the same imperfect master.

duced. The process of cross coupling also eliminates the first-order terms of variations in sheet resistivity, oxide thickness, and temperature gradients.

In practice, a "composite master" photomask is produced by multiple exposures of the imperfect pair. The resulting composite device pair is fed balanced currents:

$$i_{A1} + i_{B2} = i_{A2} + i_{B1} \qquad [8]$$

to permit observation of the input-voltage offset. From Eq. [4], holding channel-width variations constant,

$$\frac{V_1^2}{l_{A1}} + \frac{V_2^2}{l_{B2}} = \frac{V_2^2}{l_{A2}} + \frac{V_1^2}{l_{B1}}. \qquad [9]$$

Rearranging terms and substituting from Eqs. [6] and [7], the input offset is readily expressed in ratio form:

$$\frac{V_1}{V_2} \simeq \sqrt{\frac{1 + KN}{K + N}}, \qquad [10]$$

where K reflects the process ability to absolutely reproduce adjacent devices from a common master, and N is the master-pair ratio. For an ideal reproduction, $K = 1$, an imperfect master pair $N \neq 1$ is ideally

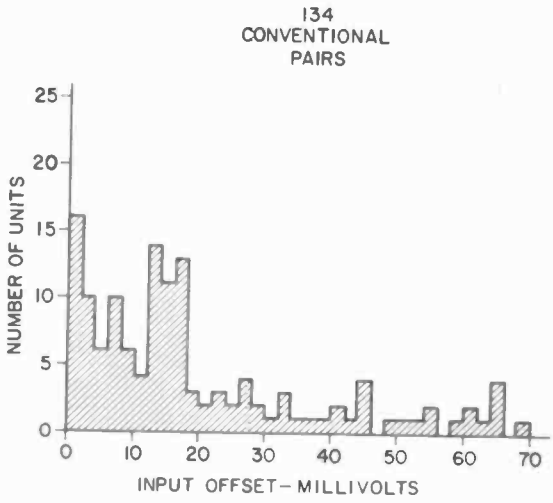

Fig. 18—Offset of conventional side-by-side PMOS pair.

complemented by cross coupling, that is, $V_1/V_2 = 1$, and ordered errors are cancelled. Even for $K = 1.1$ and $N = 1.1$, Eq. [10] shows the ratio to be $V_1/V_2 = 1.0023$ compared to $\sqrt{1.1} = 1.049$ for a single pair.

The channel dimensional tolerances required to achieve a given degree of pair matching depend upon current level above the threshold. For broadband operation, drain currents of at least 100 μA are typical, and a 400-μmho g_m requires a 1-mV signal to respond to a 0.4% current change. A nominal channel length of 5 μm must therefore be reproduced to an accuracy of $0.004 \times 5 = 20$ nm (about $\frac{1}{20}$th the wavelength of light) to maintain a 1-mV offset. Such tolerances are not readily measured on either a photomask or device channel; hence, process performance is not defined until the electrical test of device pairs.

In manufacturing practice, it would be academic to characterize a process by K or N values. Instead, performance is judged by means of offset-distribution histographs. Fig. 18 shows the input-offset voltages of two "conventionally-produced" adjacent units. Fig. 19(a) shows the offset of a pair made by cross-metallizing four separate devices generated by "best design" practices. Fig. 19(b) demonstrates the tighter distributions obtained with the composite-master process. Furthermore, whereas the first two distributions were each taken from one silicon wafer, the last is a random production sampling from more than 50 wafers.

Previous curves of MOSFET transfer characteristics suggest that pair offset is a function of drain current. If a device sample from the histo-

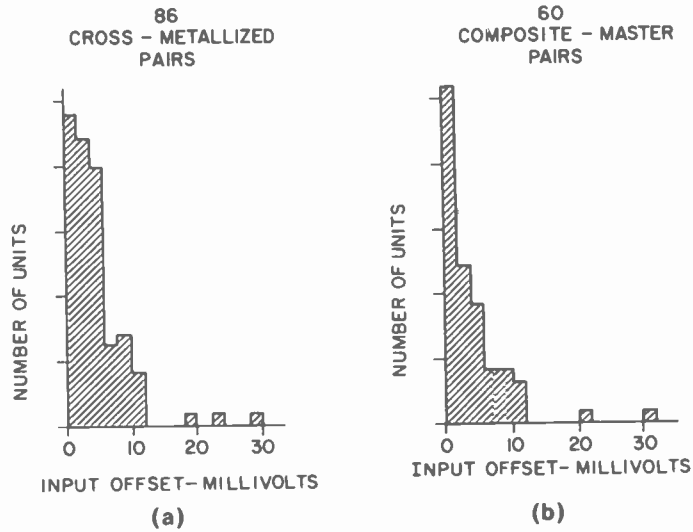

Fig. 19—(a) Cross-metallized "best design" PMOS quad and (b) offset of "composite master" PMOS pair.

graph of Fig. 19(b) is taken, the input offsets can be plotted as a function of drain current. Fig. 20 shows pairs having offsets near 0, 2, 5, and 10 mV at 100 μA operated from the nanoampere to the milliampere region. Discounting the three units that exhibit leakage in the 1 to 20 nA range (no criterion was used to select "good" or "bad" wafers before testing), the subthreshold range does show essentially constant offset. The drain/body junction size is sufficient to the extent that 1-nA leakages are expected to influence offset. The measurements show that a high percentage of devices should be operable at 100 nA up to 80 or 90°C. At current levels above 1 mA, offsets are relatively high, and current-mirror amplifiers may be used to advantage to reduce channel currents for improved performance.

Finally, better matching of pair transfer characteristics also improves common-mode and power-supply rejections. Commercial op amps changed over to a composite-master process have shown 20-dB improvements. The technique is generally applicable to other types of active and passive IC elements as well.

Protective Network Bootstrapping

The very low leakage levels promised by a metal-oxide gate construction have often been difficult to realize in commercial practice. Experience

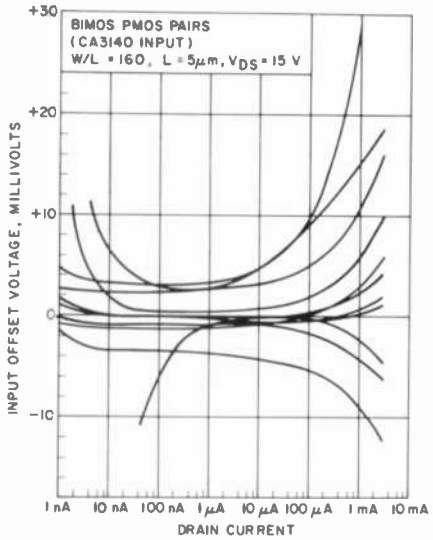

Fig. 20—Input offset as a function of drain current.

with BiMOS op amps indicates that stable input-offset levels are not attained by unprotected pairs. Fortunately, effective networks such as that shown in Fig. 21 can be built with bipolar device diffusions, and their leakages are generally unrelated to the devices they protect. Fig. 22 shows a typical differential input stage utilizing the network of Fig. 21. In a linear operating mode, the common p-diffusion is free to assume a potential dictated by the n-boat/substrate leakage. Because the PMOS source potential is higher than that of its gate, this large-junction substrate leakage is supplied by the common source node, and gate leakages, typically 5 to 10 pA at room temperature, are a function of the reverse bias on the small n+/p diodes. Wittlinger[14] has found that the resultant

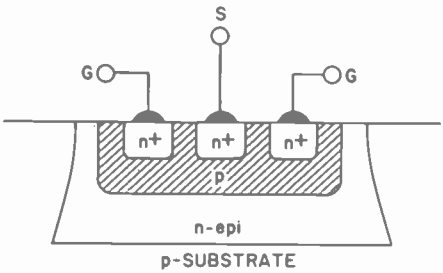

Fig. 21—Diffusion section of protective network.

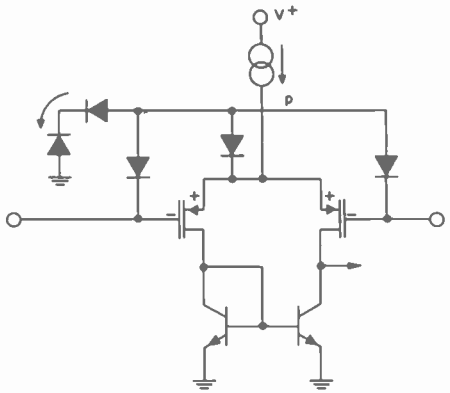

Fig. 22—Protective network applied to differential pair.

input current varies linearly with common-mode voltage, as shown in Fig. 23 for the CA3160 op amp. Note that the network leakages null out at approximately 0.6 volt, a common-mode level at which the amplifier functions well. Such negative-buss-potential biasing permits operation at low picoampere levels[15] when chip heating is minimized (leakage doubles every 10°C).

When the input-gate diode (p-diffusion) is reverse-biased near zero potential, femtoampere leakage levels are practically attainable. The differential PMOS circuit of Fig. 24 features unity-gain buffers that bootstrap the input diodes to their respective gates, assuring a low bias potential in the polarity shown. The diffused protective network is shown

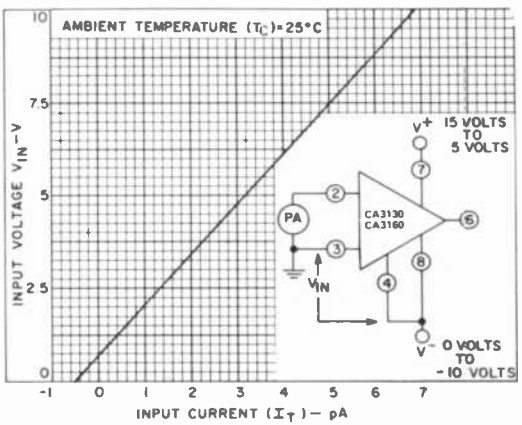

Fig. 23—Network leakage with common-mode bias.

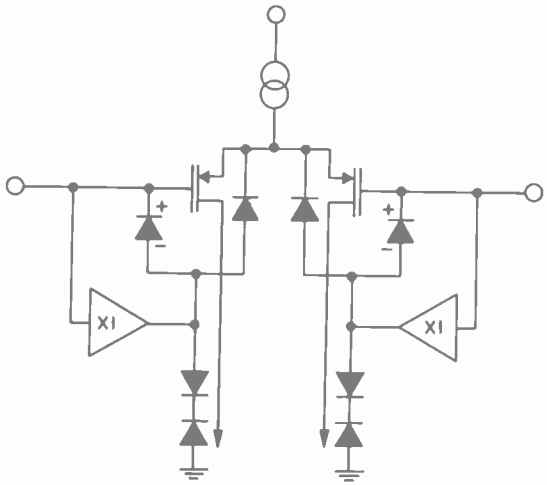

Fig. 24—Unity-gain buffer for bootstrapping protective network.

in Fig. 25 where, for this mode of operation, the isolation n-boat is biased at the positive buss bar. The resulting input bias for a typical network is plotted as a function of temperature in Fig. 26.

Each diode-bias level was established at several common-mode potentials ranging between 0 and 10 volts; the result is the current bands shown. Reverse-bias levels between 0 and 100 mV produce room-temperature currents in the order of 10 to 100 femtoamperes, and maintain a subpicoampere leakage over a 0 to 70°C range. Subpicoampere operation to 100°C requires that diode bias-levels be controlled to better than 10 mV. Conventional types of buffer amplifiers are suitable.

The bootstrapped protective network radically reduces input bias

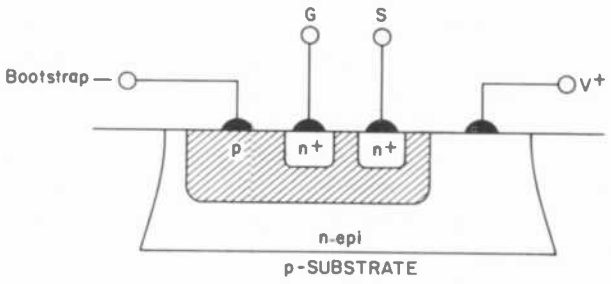

Fig. 25—Assignment of protective network potentials.

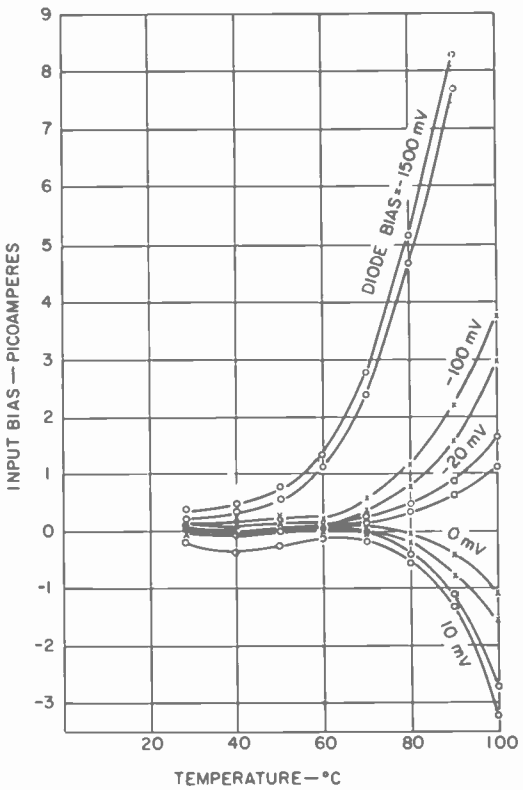

Fig. 26—Typical bootstrapped input bias currents.

levels, and maintains them over large common-mode ranges; the technique permits greater flexibility in protective-network design. Cascaded diodes may be used to provide greater gate-to-source and gate-to-gate voltage ratings. The reduced temperature dependence of leakage provides a greater tolerance to chip power levels and operating ambients and reduces thermal feedback and crosstalk. It should be stressed that the use of integrated circuits, their packages, and the system in which they operate at femtoampere current levels requires special attention to cleanliness, isolation from noise, and a deliberate approach to leakage measurement, particularly at elevated temperatures. Note that a 10-fA current slews a 1-pF parasitic capacitance at 10 mV/s. Test circuitry must be constantly recalibrated, perhaps temperature cycled, to assure measurement confidence.

Micropower Oscillators, Timers, and Amplifiers

A combination of MOS and bipolar devices is highly desirable in micropower linear circuits,[16] and may be used to advantage over broad current ranges. MOS operation at subthreshold currents with moderate gate-source offset potentials and near-ideal buffering qualities simplifies designs and saves IC die area. Several types of circuits that have been used repeatedly in applications in both proprietary and custom circuits are described below. Along with CMOS logic, or speciallized variants thereof, mixed-technology devices provide extended design capability and performance.

An astable multivibrator is shown in Fig. 27; current sources I_0 through I_4 are PMOS devices that may be operated down to nanoampere levels.

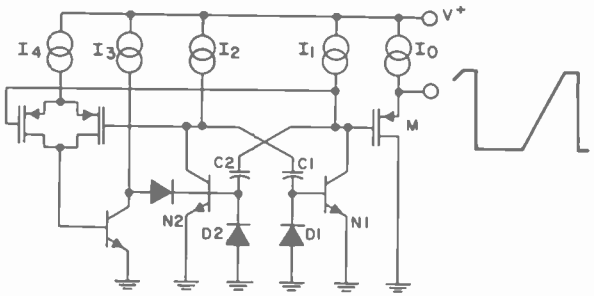

Fig. 27—Astable BiMOS multivibrator and starting circuit.

I_1 and I_2 feed n-p-n transistors and timing capacitors. Operation differs from the more common multivibrator in that capacitor displacement currents directly drive the bases. A cycle may be visualized as beginning with one capacitor, for example C2, essentially discharged. I_1 then provides a charging current that drives N_2 into saturation, so that I_2 is absorbed and N_1 is kept cut off. Potential at the N_1 collector ramps linearly positive until C_2 is fully charged and source I_1 becomes saturated. N_2 conduction then ceases, and N_1 is turned on by C_1 displacement current from I_2. N_1 absorbs not only I_1, but also initially discharges C_2 via D_2. N_1 conducts until the I_2 source saturates, and the cycle repeats.

Timing has been made independent of on-chip resistors, and may be current programmed. Using integrated capacitors, a 10-nA current slews 10 pF at 1 V/ms. A 10-volt supply, therefore, permits 50-Hz operation; the skipped sawtooth drives output buffer M. The simple circuit suffers a limitation in that a mode can exist in which both sources I_1 and I_2 are saturated. No timing function can occur in such a state; hence, a starting

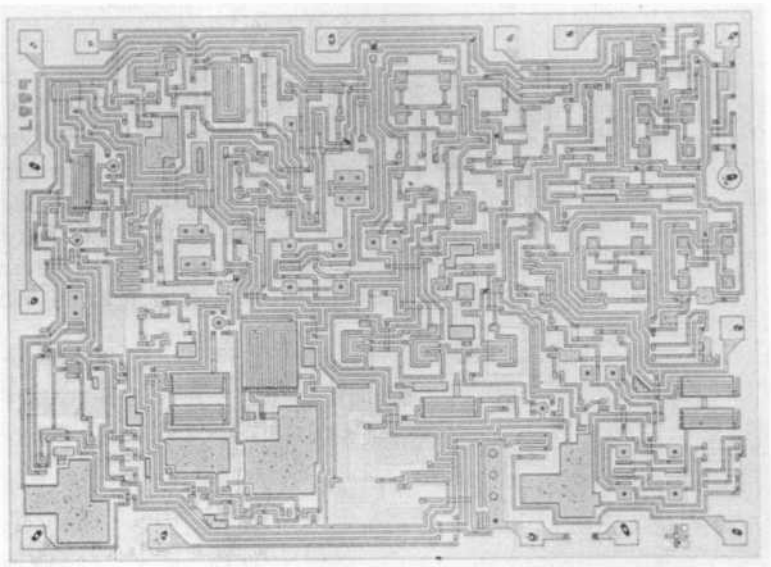

Fig. 28—Custom IC employing oscillator and synchronous detection.

circuit is needed. I_4 feeds two PMOS switches whose function is to open the base circuit of a transistor clamp should their gate potentials be simultaneously high. I_3 can then provide a pulse of current to initiate oscillation. A 93 × 128 mil custom BiMOS IC employing such an 8-kHz oscillator and MOS logic for synchronous detection is shown in Fig. 28.

A simpler BiMOS multivibrator is the one-shot of Fig. 29. A definite monostable condition is produced with source I_1 and the NMOS saturated and with C essentially discharged. Upon application of a positive

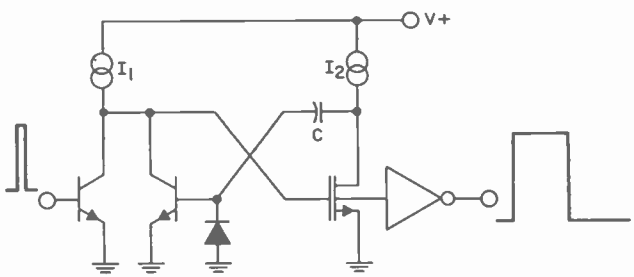

Fig. 29—One-shot BiMOS multivibrator.

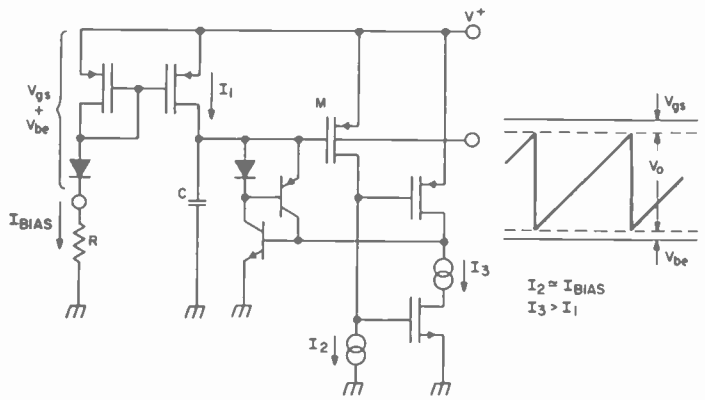

Fig. 30—BiMOS oscillator.

trigger pulse, a timing cycle in which I_2 charges C is initiated. A 10-nA source, integrated 40-pF capacitor, and 10-volt supply provide a 40-ms output pulse, which is readily buffered by a CMOS inverter.

Another approach, well-suited to timing circuits where but a single IC pin is available for an external capacitor, is shown in Fig. 30. Bias current is reflected by a PMOS mirror, and capacitor C is charged. PMOS M is turned off as the voltage ramp approaches the positive power supply buss potential, and I_2 is permitted to reduce the potential of the gates of a modified CMOS inverter. This action turns on a regenerative (SCR-like) bipolar switch, causing rapid discharge of C and reestablishing an M conduction state. If holding current I_3 exceeds charging current I_1, the switch will open upon capacitor discharge, and the timing cycle repeats. Control of the operating mode is readily provided by changing I_3, shorting C, and/or gating the n-p-n base. A p-n-p mirror switch may be used rather than a single p-n-p transistor to increase operating frequency. Parasitic capacitance at the p-n-p base tends to hold the base down during the charging period, producing a current proportional to ramp slope. The diode path prevents a resultant beta-multiplied current at the p-n-p collector, and reduces the required value of I_3. For a variable timing cycle, all current generators can be slaved to the bias current, which may be programmed externally to the IC.

The potential ramp on C approaches the positive power supply buss potential to within the gate-source voltage of M needed to produce level I_2. The discharge switch clamps C to roughly a base-emitter offset above the negative buss potential. If a supply potential is set up:

$$V^+ = V_{gs} + V_{be} + V_o \qquad [11]$$

(where V_o has a well-controlled value, such as that provided by a bandgap reference), both the bias-resistor voltage and ramp excursion will equal V_o. As a result, a resistor with a low temperature-coefficient can provide excellent timing stability. For example, a 3-kHz oscillator employing a 10-pF capacitor and 5 megohm (1% precision) film resistor at a nominal 3-volt supply has shown a period stability of 0.005% per °C, or 50 ppm/°C. This figure is in the order of the coefficient claimed for such resistors.

For greatest reliability, a micropower timer intended for relatively long periods should incorporate a well-defined delay function during cycle reset. Fig. 31 shows a timer employing the capacitor-ramp detection

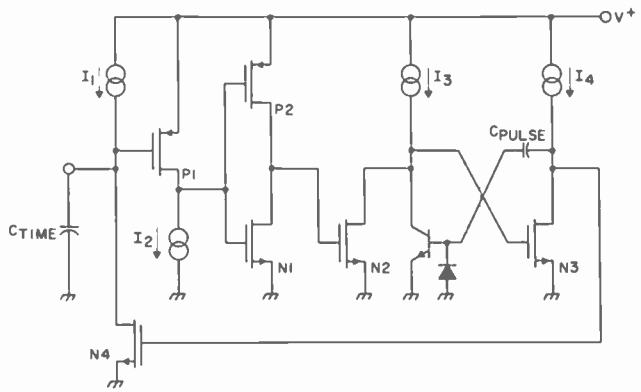

Fig. 31—BiMOS timer.

scheme of Fig. 30 in combination with a one-shot like that shown in Fig. 29. PMOS P_1 drives a CMOS inverter, which triggers the normally-high n-p-n collector of the one-shot. Only after the transition to reduced potential occurs can current source I_4 begin to increase its drain potential. When the drain potential reaches the threshold of NMOS N_4, discharge of the timing capacitor is initiated, reestablishing P_1 conduction and removing the trigger potential. The one-shot continues to supply increasing drive to N_4 until I_4 saturates, after which it resets itself, and a new timing cycle begins. The circuit is more complex than the version in Fig. 30, but does not depend upon an SCR regeneration rate at low current levels. In a smoke-detector application, a 0.1-μF timing capacitor and 10-nA current sources establish a 50-second period after which an integrated delay capacitor produces a 40-ms discharge. The triggering point provides a pulse needed for other system functions. The period

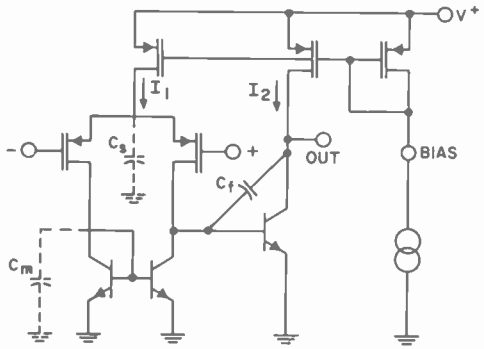

Fig. 32—Two-stage BiMOS amplifier.

is stable to approximately 0.04%/°C and the pulse-width to approximately 0.08%/°C over a 100°C range.

The simple two-stage amplifier of Fig. 32 has a configuration similar to that introduced in the bipolar LM101 by Widlar[17] and the uA741 by Fullagar.[18] Analysis of similar circuits has been treated by Leidich[11] and Solomon[19]. Differential input, level shifting, single-ended conversion of signal, and Miller feedback around a high-gain stage are efficiently provided with relatively few components. As shown, output can be taken directly from the second-stage n-p-n collector, as in Wheatley's operational transconductance amplifier (OTA).[20] Many signal-processing functions do not require that the amplifier have low output impedance, particularly in custom micropower applications, and the (sometimes complex) driver stage commonly provided in commercial products may be omitted.

The open-loop voltage gain of the amplifier is

$$A_{vol} = \beta_n g_m R_l, \qquad [12]$$

where the PMOS transconductance is that value taken at current $I_1/2$ and R_l is the resistance at the output node. Typical values approach 90 dB at subthreshold levels when working into high-impedance loads. The mirror and source frequency poles occur at

$$f_m = \frac{I_1}{0.33\, C_m} \qquad [13]$$

$$f_s = \frac{g_m}{2\pi C_s}. \qquad [14]$$

The unity-gain crossing may be calculated from

Table 1—Typical Measured FET-Amplifier Performance

Total Supply Current	10 nA	100 nA	1 μA	10 μA	100 μA	1 mA
Input Offset (mV)	2	2	2	3	4	8
Open-Loop Gain (dB)	85	88	90	90	83	74
Unity-Gain Slew (V/μs)	0.00004	0.0004	0.004	0.04	0.4	4
Unity-Gain Bandwidth, (kHz)	0.7	7	65	450	2000	7000

$$f_1 = \frac{g_m}{2\pi C_f}. \qquad [15]$$

Table 1 shows the measured performance over six levels of current magnitude with $C_f = 20$ pF.

It is possible to achieve higher voltage gains at the expense of input common-mode range and circuit complexity where the second-stage n-p-n device has been replaced with a Darlington/mirror configuration and the input-stage mirror has been revised to reflect a balanced base-current loading to the differential pair, Fig. 33. The open-loop gain then becomes

$$A_{vol} = m\beta_n g_m R_l \qquad [16]$$

(until the first-stage output resistance limits at large m-values) and, unlike circuits employing resistors in the Darlington, amplifier currents "track" over large bias-current ranges.

For applications in which higher currents must be delivered, or a substantial capacitance driven, the amplifier requires an output stage.

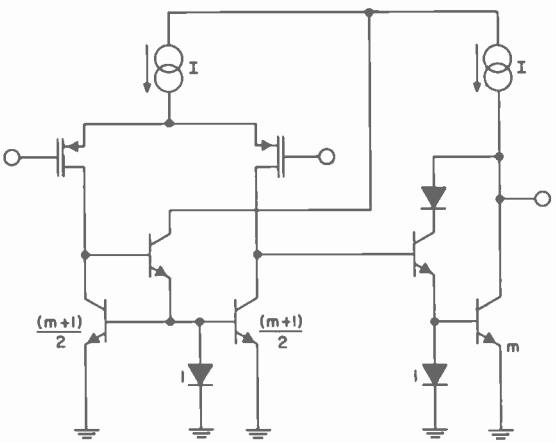

Fig. 33—Amplifier revisions for increased voltage gain.

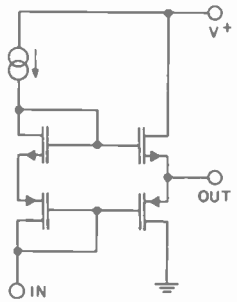

Fig. 34—BiMOS complementary source-follower.

Single-ended transistors or a CMOS inverter provide a simple compact approach, but micropower standby power levels are maintained only in a comparator or similar function where high or low (buss-level) output states occur. A complementary source-follower stage, Fig. 34, is suitable for linear submilliampere outputs, and provides excellent buffering for nanoampere drive levels with moderate gate-source offsets.[10] However, for operation at relatively low supply voltages and milliampere-level outputs, a bipolar stage is generally most practical, and the complementary emitter follower is employed.

An improved bipolar stage developed by Wheatley for the CA3078 micropower op amp is shown in Fig. 35. The n/p and p/n composites have current gains $\beta_n \times \beta_p$, which reduce the drive requirements for a given output and help equalize buffering for widely-varying beta values. Idling

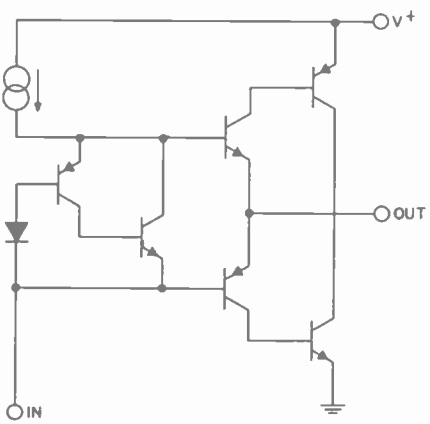

Fig. 35—Improved composite bipolar output (after Wheatley).

currents are related to those of the previous stage by choice of transistor geometries, and they maintain proportionality over the bias current range in a programmable amplifier. Outputs of several milliamperes are produced with a 1-μA drive, and the voltage swing is equivalent to that of a standard complementary emitter follower—within ($V_{be} + V_{sat}$) of either buss potential.

Acknowledgments

Test circuitry for femtoampere-level measurements was designed by H. Wittlinger, and appears in the technical data sheets for the CA3160. Photomask techniques were guided by S. Ahrens and J. P. Keller. Some of the oscillator, timer, and amplifier investigations were carried out by T. DeShazo and the author in support of a developmental contract with Fenwal, Incorporated. MOS test structures were built under the direction of W. Scheraga.

References:

[1] M. A. Polinsky and S. Graf, "MOS Bipolar Monolithic Integrated Circuit Technology," *IEEE Trans. on Elec. Dev.*, p. 239, March 1973. Material originally presented at IEDM70.

[2] O. H. Schade, Jr., "A New Generation of MOS/Bipolar Operational Amplifiers," *RCA Review*, **37**, No. 3, p. 404, Sept. 1976.

[3] M. A. Polinsky, O. H. Schade, Jr., and J. P. Keller, "CMOS-Bipolar Monolithic Integrated Circuit Technology," *Proc. 1973 Inter. Elec. Dev. Mtg.*, p. 229.

[4] Y. Hayashi and Y. Tarui, Tech. Group on Semiconductors and Semiconductor Devices, IECE Japan, SSD-67-6, 1967.

[5] P. Richman, "Theoretical Threshold Voltages for MOS Field Effect Transistors," *Solid State Electron.*, **11**, p. 869, 1968.

[6] W. M. Gosney, "Subthreshold Drain Leakage Current of MOS Devices Under Surface Depletion Conditions," *IEEE Trans. Elec. Dev.*, **ED-19**, p. 213, Feb. 1972.

[7] R. R. Troutman and S. N. Chakravarti, "Subthreshold Characteristics of Insulated-Gate Field-Effect Transistors," *IEEE Trans. on Ckt. Theory*, **CT-20**, No. 6, p. 659, Nov. 1973.

[8] E. Vittoz and J. Fellrath, "CMOS Analog Integrated Circuits Based on Weak Inversion Operation," *IEEE J. Solid State Ckts.*, **SC-12**, No. 3, p. 224, June 1977.

[9] Y. P. Tsividis and R. S. Ulmer, "A CMOS Reference Voltage Source," *ISSCC Digest of Tech. Papers*, Vol. XXI, p. 48, Feb. 1978.

[10] O. H. Schade, Jr., "BiMOS Micropower IC's," *ISSCC Digest of Tech.* Papers, Vol. XXI, p. 230, Feb. 1978.

[11] A. J. Leidich, "Linear Monolithic Circuit Techniques," paper 740012, p. 67–81 of *Automotive Electronics*, New York SAE, Feb. 1974.

[12] S. A. Steckler, "V_{be} Voltage Source Temperature Compensation Network," United States Patent No. 3,886,435, issued May 27, 1975.

[13] O. H. Schade, Jr., "Matching of Semiconductor Device Characteristics," United States Patent No. 3,863,331, issued Feb. 4, 1975.

[14] Technical Data Sheet No. 976 for types CA3160/60A/60B, p. 7, RCA Solid State Division, Nov. 1976.

[15] G. Checkowski, G. Granieri, and H. Wittlinger, "BiMOS and Bipolar IC's Simplify Smoke Detector Designs," ICAN-6629, RCA Solid State Division, May 1977.

[18] M. V. Hoover, "An Introduction to the Characteristics and Applications of COS/MOS Transistors in Linear Service," Eighth Asilomar Conf., Pacific Grove, Calif., Dec. 4, 1974

[17] R. J. Widlar, "Monolithic Op Amp with Simplified Frequency Compensation," *EEE*, **15**, p. 58, July 1967.

[18] D. Fullagar, "A New High Performance Monolithic Operational Amplifier," Fairchild Semiconductor Tech. Paper, 1968.

[19] J. E. Solomon, "The Monolithic Op Amp; A Tutorial Study," *IEEE J. Solid State Ckts.*, **SC-9**, No. 6, p. 314, Dec. 1974.

[20] C. F. Wheatley and H. A. Wittlinger, "OTA Obsoletes Op Amp," *1969 NEC Proceedings*, Dec. 1969.

Chemical Etching of Silicon, Germanium, Gallium Arsenide, and Gallium Phosphide

Werner Kern

RCA Laboratories, Princeton, N.J. 08540

Abstract—The chemical etching processes used in the semiconductor microelectronics technology for silicon, germanium, gallium arsenide, and gallium phosphide are reviewed according to the following classification: Isotropic liquid etching, anisotropic liquid etching, electrochemical and selective etching, gas and vapor phase etching, and chemical-mechanical polishing. The discussion is intended to complement a companion paper (see Ref. [1]) in which detailed etchant compositions and specific etching conditions are presented in tabular form.

1. Introduction

Chemical etching of semiconductor crystalline and amorphous thin films and of single-crystal slices plays on essential role in microelectronic device technology. Etching is used for shaping and polishing, as well as for characterizing structural and compositional features. The purpose of this paper is to survey and discuss the chemical etching processes used for shaping and polishing four important semiconductor materials: silicon, germanium, gallium arsenide, and gallium phosphide.

Silicon is by far the most important semiconductor because of its widespread use in modern microelectronic discrete and photovoltaic devices; it will therefore be treated in more detail. Germanium has become of lesser technical importance. Since its chemical properties are similar to those of silicon, many etching processes can be applied to both materials. It is therefore convenient to discuss both of these group IV elemental semiconductors jointly.

Of the compound semiconductors, gallium arsenide is most widely used in solid-state devices for specialized applications, followed by gallium phosphide, a second group III-IV compound semiconductor material.

The exact composition of etchants, substrate parameters, and detailed etching conditions will be presented in a tabular format for practical applications in a complementary article.[1] Fundamental principles and techniques of etching are also discussed, reviewed, and referenced extensively in that publication, which includes etching processes for additional semiconductors and for many other materials as well.

2. General Etching Mechanisms of Semiconductors

The fundamental reactions underlying semiconductor etching processes are electrochemical in nature. They involve, typically, oxidation-reduction reactions, followed by dissolution of the oxidation products, frequently by complexing. Both anodic and cathodic microscopic sites exist at the semiconductor surface. Oxidation of semiconductor atoms takes place at the anodic sites, while the oxidant is reduced at the cathodic sites. In the case of silicon and germanium the oxidizing agent is frequently HNO_3, and the complexant is HF.

The dissolution process is either reaction-rate limited or diffusion-limited. If the dissolution is a function of the chemical reaction rate, the process is called reaction-rate limited; if it depends on the transport of etchant by diffusion to or from the surface through the liquid or gas, it is called diffusion-limited. Diffusion controlled processes have lower activation energies (of the order of a few Kcal/mol) than reaction-rate limited processes, and are therefore relatively insensitive to temperature variations. They are, however, affected by agitation to a much greater extent, leading to an increase in etch rate because of the increased supply of reactant material to the semiconductor surface. Changes in etching conditions, such as temperature, as well as slight variations in relative proportions of etchant components can change the rate-limiting process. The supply of minority carriers to the semiconductor surface can also limit the dissolution rate in etching reactions that result in a depletion of electrons or holes. Creation of electron–hole pairs on the surface by illumination or by application of electric currents, etc., or providing generation sites, can then increase the rate of etching. Additional factors that determine the rate of etching of crystalline semiconductors include crystal orientation, type and concentration of doping atoms, lattice defects, and surface structure.

Etching reactions of single-crystal gallium arsenide (GaAs), as well as of other compound semiconductors, are complicated because of

crystallographic surface orientation effects. For example, one side of a {111}-oriented GaAs crystal slice is composed of Ga Atoms,* whereas the opposite side consists of As atoms** causing chemical reactions to differ widely for each surface. Most liquid etchants used are therefore orientation-dependent, unless they are combined with mechanical polishing techniques. Chemical etching of GaAs (and of other group III-V and group II-VI compound semiconductors) proceeds by oxidation-reduction-complexing reactions analogous, in principle, to the general mechanism for Si and Ge etching. Other useful group III-V compounds are GaP, GaSb, InAs, InP, and InSb.

Chemical etching processes can be categorized according to the technique by which the reaction is being conducted. This type of classification is particularly practical and useful in surveying semiconductor etching and will be applied to this review. We can distinguish (a) isotropic liquid etching, (b) anisotropic liquid etching, (c) electrochemical and selective chemical etching, (d) gas and vapor phase etching, and (e) chemical–mechanical polishing. Physical–chemical etching processes, such as sputter etching and plasma etching, are beyond the scope of this review (see, for example, Ref. [2]).

3. Etching of Silicon and Germanium

3.1 Isotropic Liquid Etching

Isotropic chemical etching of semiconductors in liquid reagents is the most widely used etching process for removal of work-damaged surfaces, for creating structures or planar surfaces in single-crystal slices, and for patterning single-crystal or polycrystalline semiconductor films. For Si, etchants containing HF, HNO_3 and H_2O are most frequently used; Ge etchants based on HF, H_2O_2 and H_2O are typical.

Turner[3] has reported on electrode potential measurements of Si and Ge in HF-HNO_3 as a function of substrate properties and etching conditions which suggest that excess holes and electrons are produced at the semiconductor surface. At a ratio of 1 HNO_3/4.5 HF† the etch rate of Si reaches a maximum of 28 μm/s, corresponding to the large value of about 190 A/cm^2 average corrosion current. Assuming H_2SiF_6 to be the anode product, the overall etching reaction that fits the data best was formulated as follows:

---

* Denoted as {111} Ga, or {111}A plane, or {111} Ga plane, surface, or face. The crystallographic notations used in this paper are in each case those used by the author(s) of the references cited.

** Denoted as {111} As, or {111}B plane, or {111} As plane, surface or face.

† Refers to volumes of reagents in the normal concentrated form, unless otherwise stated throughout this paper.

$$3Si + 4HNO_3 + 18HF \rightarrow 3H_2SiF_6 + 4NO + 8H_2O + 3(4 - n)e^+ + 3(4 - n)e^-.$$**

Etching studies[4] over the entire range of $HF\text{-}HNO_3$ compositions at various temperatures and stirring rates indicated no differences in the etch rate of n, p, and n-p junction Si ranging in resistivity from 0.05 to 78 ohm-cm. Between 65 and 70% HF, the etch rate reached a maximum, but it was most controllable in etches containing less than 65% HF. Between 10 and 50% HF, the etch rate was directly proportional to the HF concentration. For compositions containing 2 to 50% HF, the etch rate increased linearly with temperature in the range of 0° to 60°C. The activation energy of 3.9 Kcal/mol indicated that the reaction was diffusion-rate limited. Use of large volumes of slow-etching low-HF compositions with controlled agitation made processing at room temperature possible with good dimensional control. The etch rate was directly related to the speed of sample rotation.

Extensive studies on the mechanism of silicon etching in ternary mixtures of $HF\text{-}HNO_3\text{-}H_2O$ and in aqueous $HF\text{-}HNO_3\text{-}CH_3COOH$ compositions have been reported by Schwartz and Robbins in a series of four papers from 1959 to 1977.[5–8] In their first paper[5] they reported that the reaction involves oxidation as the first step, followed by dissolution. In high-HF etchants the HNO_3 concentration determines the etch rate because oxidation is the rate-limiting step. In high-HNO_3 compositions, the etch rates are a function of only the HF because in this case dissolution is the rate-limiting process. The reaction kinetics for the $HF\text{-}HNO_3$ etchants containing either H_2O or CH_3COOH as the diluent were compared next.[6] The two systems are similar qualitatively, but the CH_3COOH-containing etchants have a higher tolerance for the diluent than the H_2O-containing compositions due to decreased ionization of HNO_3 in CH_3COOH. The etch rates of Si in the two etchant types over the temperature range of 0° to 50°C showed that the activation energy of 4 Kcal/mol for high-HNO_3 etchants was indicative of a diffusion-governed reaction.[7] Addition of H_2O or CH_3COOH diluents increased the activation energy. Two values were observed in high-HF compositions, one of 10–14 and one of 20 Kcal/mol, due to several complications in the reaction. Finally, composition-mechanism interactions as manifested in geometrical effects on etching of Si dies in the two types of etchants were discussed.[8] Plots of iso-etch rate contours and plots of etched die geometry as a function of etchant composition were presented. A few of these plots are reproduced here for convenience because of their general usefulness in selecting suitable etch compositions for specific applications. In Figs. 1 and 2 iso-etch rate contours are shown for HF-

** In this equation e^+ represents a hole and n is the average number of holes required to dissolve 1 Si atom (n = 2 to 4 depending on the current multiplication factor).

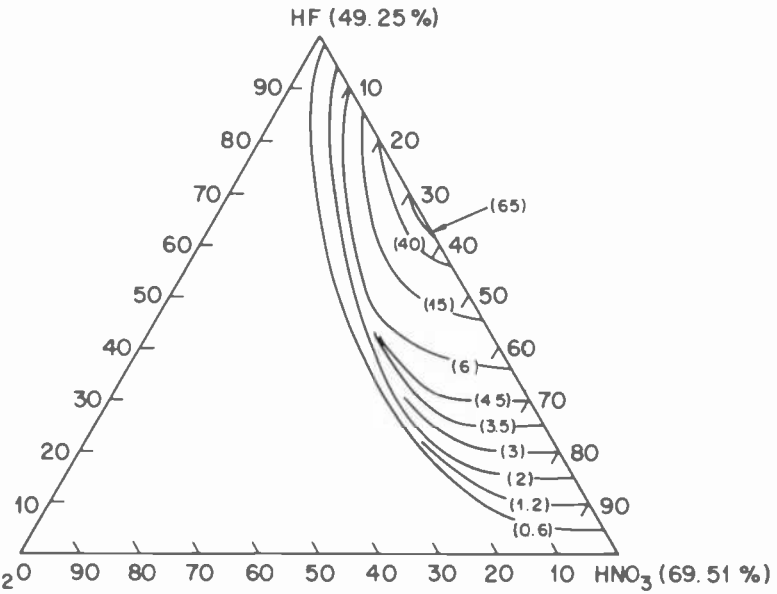

Fig. 1—Curves of constant rate of change of die thickness (mils per minute combined for two Si wafer surfaces) as a function of etchant composition, in the 49% HF-70% HNO_3 system (from Ref. [8], reprinted by permission of the publisher, *The Electrochemical Society*, Inc.).

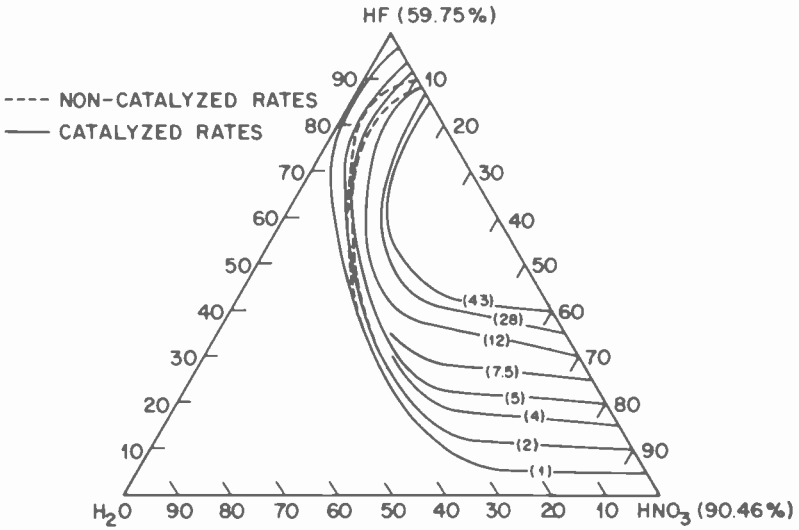

Fig. 2—Curves of constant rate of change of die thickness (mils per minute combined for two Si wafer surfaces) as a function of etchant composition in the 60% HF-90% HNO_3 system; the effect of added catalyst ($NaNO_2$) is shown as the dashed lines (from Ref. [8], reprinted by permission of the publisher, *The Electrochemical Society*, Inc.).

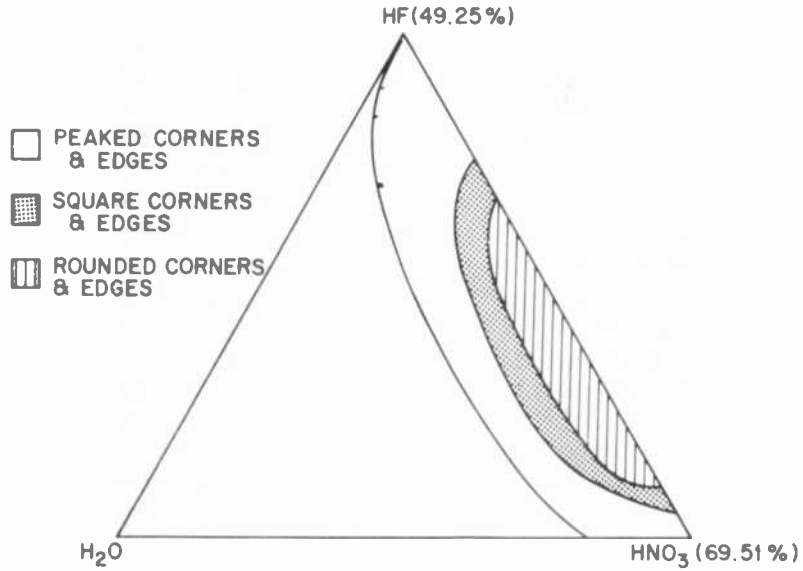

Fig. 3—Resultant geometry of the etched Si die as a function of the etchant composition in the 48% HF-70% HNO_3 system (from Ref. [8], reprinted by permission of the publisher, *The Electrochemical Society*, Inc.).

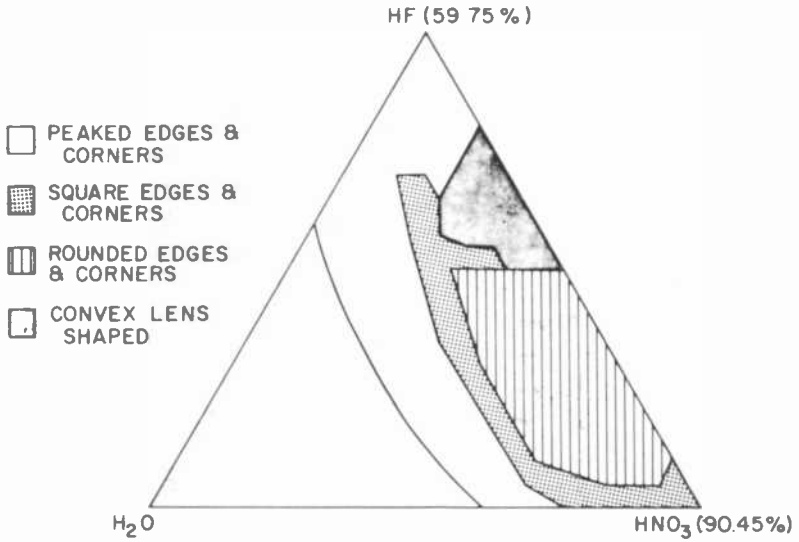

Fig. 4—Resultant geometry of the etched Si die as a function of the etchant composition in the 60% HF-90% HNO_3 system (from Ref. [8], reprinted by permission of the publisher, *The Electrochemical Society*, Inc.).

HNO_3-H_2O in normal and high concentration acids. The resulting geometry effects on initially rectangular (111)-plane parallelepipeds (n-type, 2 ohm-cm dies) are indicated in Figs. 3 and 4 for the same etchants. Very similar effects result if CH_3COOH is used as diluent instead of H_2O. Since oxidation is the rate-limiting step in composition regions very low in HNO_3 and rich in HF, the etching is affected by electron concentration, surface orientation, crystal defects, and catalysis by lower oxides of nitrogen. In compositions where oxide dissolution and associated diffusion are the rate-limiting steps due to low HF, etching is not influenced by crystallographic orientation and conductivity type, but factors such as agitation have pronounced effects. For further details the original paper[8] should be consulted.

Independent measurements of the activation energies for both Si and Ge in typical HF-HNO_3-CH_3COOH etchant confirmed that the processes are indeed diffusion controlled reactions.[9]

Silicon is also soluble to a small extent in HF solutions;* for 48% HF, a rate of 25°C of 0.3 Å/min was observed for n-type 2-ohm-cm (111)-Si. Diluted HF etches at a higher rate because the reaction in aqueous solutions proceeds by oxidation of Si by OH^- ions.[10] A typical buffered HF solution (BHF) has been reported to etch at radiochemically measured rates of 0.23 to 0.45 Å/min, depending on doping type and dopant concentration.[11] However, other observations have shown that typical BHF etchant etches p-type Si at a rate of 17 Å/min.[12] This effect may be due to certain trace impurities, particularly nitrates, which are present in commercial electronic grade NH_4F. For example, 40% NH_4F solutions from three suppliers contained NO_3^- in the range of 3 to 5 ppm, SO_4^{--} at 2 ppm, and Cl^- at 1 ppm. Photomicrographs of patterns that resulted by immersing p-type Si wafers masked with photoresist patterns in BHF ($7NH_4F$ 40%–1HF 48%) prepared with electronic grade NH_4F solutions from three suppliers are shown in Fig. 5 for (111) surface Si and in Fig. 6 for (100) surface Si. The amount of Si dissolved isotropically in 15 minutes at room temperature was 250 Å, corresponding to about 17 Å/min.[12] This effect can lead to serious consequences in the processing of certain types of devices if it is not carefully controlled.

Isotropic liquid etching has been used for thinning of semiconductor slices. Spot-thinning jet techniques, using HNO_3-HF for Si and HNO_3-HF or NaOCl for Ge, have been described for preparing small-area films of less than 1 μm thickness for electron microscopy.[13,14] Similar etchants can be applied for uniform thinning of mounted larger-area Si slices by simple immersion[15] or by floating in the etchant above a stream of CO_2 gas bubbles in an illuminated Teflon container that allows monitoring by observing the color of light transmitted through the

* This was demonstrated not to be due to dissolved O_2.

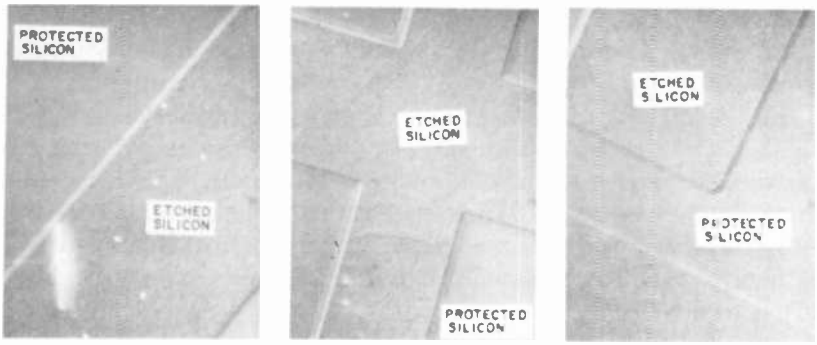

Fig. 5—Isotropic etching of (111) surface p-type silicon with ammonium fluoride buffered HF prepared with electronic grade 40% NH_4F solution from three suppliers A, B, and C (interference contrast photomicrographs, original magnification 350X; from Ref. [12]).

thinned area.[16] Improved uniformity of films as thin as 3 μm can be prepared by this technique if the Si slice is rotated part of the time.[17] An apparatus with planetary wafer movement has been described in which uniform polishing and thinning is attained with a reciprocating jet nozzle dispensing HNO_3-HF-CH_3COOH etchant for Si, or similar etchants for Ge.[18] Thinning has also been accomplished by rotating a wafer attached to a Teflon holder in 95 HNO_3-5HF etchant.[19,20]

An effective prepassivation surface cleanup etch for high-breakdown voltage pnp structures prior to thermal oxidation has been described which consists of $HClO_4$, HNO_3, HF, and CH_3COOH.[21] Stains of

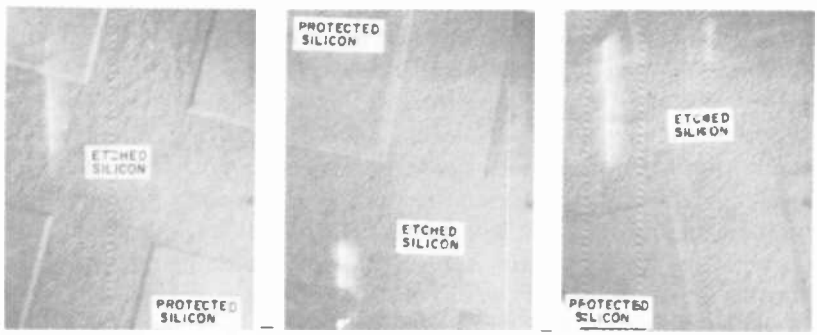

Fig. 6—Isotropic etching of (100) surface silicon with ammonium fluoride buffered HF prepared with electronic grade 40% NH_4F solution from three suppliers A, B, and C (interference contrast photomicrographs, original magnification 350X; from Ref. [12]).

suboxides of Si that may form during certain etching operations can be removed by a 10–15s dip in 2HF-1$KMnO_4$ 6% which at 20–25°C etches n-type and p-type Si at a rate of 0.3 and 0.4 μm/min, respectively.[22]

Pattern etching with a low degree of undercutting of the photoresist mask has been achieved with an etchant of nearly neutral pH, consisting of a diluted NH_4F solution containing a small amount of H_2O_2.[23] Pattern etching of n- and p-type Si (except with high concentrations of boron) using SiO_2 films as the etching mask is possible with etchants consisting primarily of HNO_3 with a small quantity of HF, with or without addition of some $NaNO_2$.[24]

Epitaxial films of Si have been etched with freshly prepared HF-CH_3COOH-$KMnO_4$ etchant.[25]

Several etchant compositions are available for polishing of Si surfaces. Compositions are based on various proportions of HNO_3-HF, with and without CH_3COOH and Br_2.[26–29] A solution of NH_4F in H_2O_2 has also been used.[30]

Widely used etchants for general Si etching are "Planar Etch" (15-HNO_3-5CH_3COOH-2HF)[31] and "Iodine Etch" (110mL CH_3COOH-100mL HNO_3-50mL HF-3g I_2).[32]

Isotropic etching of Ge is in many respects similar to that of Si. Several instances of Ge etching have already been noted in the preceding paragraphs, notably measurements of potentials[3] and activation energies for HF-HNO_3-CH_3COOH,[9] and polish-thinning of single-crystal slices.[13,14,18,33]

Two principal types of Ge etchants are used. In one, the oxidant is HNO_3 and the complexing agents are aqueous HF or HF plus CH_3COOH; in the other, H_2O_2 serves as oxidant and aqueous HF is usually the complexant. Etchants based on HNO_3-HF-H_2O or HNO_3-HF-CH_3COOH are notoriously difficult to control, mainly because of variable induction periods.[34] Studies on mode of action of the components in the widely-used chemical polish C.P.-4A, (comp.: 25HNO_3-15HF-15CH_3COOH) were reported in 1962.[35,36] These studies included effects of compositional variations and of the addition of Br_2 (C.P.-4), and other parameters such as solution aging. Examination of the etch rates of HNO_3-based compositions with various complexing agents (H_2O, sucrose, mannitol, oxalic acid, citric acid, HF, tartaric acid, and HCl) showed that the complexing agents control the etching rate, increasing it in the sequence noted from H_2O to HCl.[37] Replacing HNO_3 with H_2O_2 as the oxidizing agent showed that the oxidant in these etch mixtures controls the type of attack at the surface (shape and orientation of etch pits).[37,38]

The HF-H_2O_2-H_2O etch system affords much better control than the HNO_3-based etchants.[34] Detailed studies[39] of the kinetics and reaction

mechanism led to the conclusion that this sytem is etch-rate limited by a surface-reaction or product desorption step. Divalent Ge appears the prime oxidation state leaving the (100) and (111) surfaces, $GeOF_2$ the prime oxidation state of the (110) surface. The $HF\text{-}H_2O_2\text{-}H_2O$ etch rate is strongly influenced by crystal orientation. Very useful triaxial plots of etch rate as functions of the etchant composition and data on temperature effects and activation energies are available.[39]

A reaction mechanism was proposed more recently[38] for the $HF\text{-}H_2O_2\text{-}H_2O$ system where Ge is oxidized forming the hydroxyl complex $Ge(OH)_2^{++}$ which is then desorbed and reacts with H_2O to form H_2GeO_3, metagermanic acid. The metagermanic acid finally reacts with the complexant forming the end product.

Controlled removal of thin layers of Ge at low etch rate (0.02 μm/min) can be accomplished with 3% H_2O_2 at pH 3.8. Etch rates increase at much lower or much higher pH and are orientation-dependent.[40]

Several additional useful isotropic etchants for Ge have been devised for general applications. Compositions are based on $5H_2O\text{-}1H_2O_2$,[41] $4H_2O\text{-}1HF\text{-}1H_2O_2$,[41–43] $H_2O\text{-}H_2O_2\text{-}NaOH$ (used at 70°C),[44] and $H_2O\text{-}NaOCl$.[33,45]

Other types of etchants frequently used for general Ge etching and chemical polishing are based on $5HNO_3\text{-}3HF\text{-}3CH_3COOH$ with or without 0.06 Br_2,[42,46,47,48] and similar compositions using I_2 instead of Br_2.[49]

3.2 Anisotropic Liquid Etching

In anisotropic or orientation-dependent etching, the etch rate varies within the principal crystallographic directions of the semiconductor single-crystal. Orientation effects during this type of preferential etching have been attributed to crystallographic properties, particularly the density of surface free bonds, the relative etch rate increasing with the number of available free bonds.[50–53]

Anisotropic liquid etchants for Si are usually alkaline solutions used at elevated temperature. As in the case of isotropic acid etchants, the two principal reactions are oxidation of the Si, followed by dissolution of the hydrated silica. The oxidant can be H_2O in aqueous alkaline systems such as NaOH[54,55] or, preferably, KOH,[53,56–64] hydrazine,[64,66] ethylenediamine,[64,67] quaternary ammonium hydroxides,[68] or sodium silicates.[55] The complexing or chelating agent to effect dissolution consists usually of an alcohol, such as isopropanol,[60,66] n-propanol,[59,63] sec-butanol,[59] or pyrocatechol.[67] Water is also of importance in these dissolution reactions which form hydrogen as the gaseous byproduct.[53] Alkalis of high concentration, hydrazine, and quaternary ammonium

hydroxide can act in water solutions as both oxidant and complexant so that no additional dissolving agents may be needed. A unique anisotropic electrodeless pattern etching technique has been described that is based on electrochemical displacement of Si with metal ions, Cu^{++} yielding the most precise patterns.[53]

The essential feature for silicon technology of all these solutions is that their etch rate is up to 100 times higher in the ⟨100⟩ direction than in the ⟨111⟩ direction.[59,60] For example, in the case of the water-ethylene diamine-pyrocatechol etchant the etch rates of (100), (110), and (111) oriented Si are approximately 50:30:3 μm/h, respectively.[67] Anisotropic etching of (100) Si through a patterned SiO_2 mask creates precise V-grooves, with the edges being (111) planes at an angle of 54.7° from the (100) surface. The depth of the grooves is determined initially by the pattern width, since etching will stop automatically at the point where the (111) planes intersect and the (100) bottom surface no longer exists. Application of this preferential etching principle allows fabrication of high-density monolithic integrated circuits[66,69,70] and Si-on-sapphire integrated circuits[24,61] and has found other important uses in silicon device structuring and patterning,[67,71,72] including the preparation of diffraction gratings.[73]

Anisotropic etching of (100) plane Si without a mask leads to a textured surface consisting of microscopic pyramids that are useful in solar cells.[74,75] Deep moats, bars, or parallelograms (having walls almost exactly perpendicular to the surface) can be precision etched with KOH-H_2O into a SiO_2-masked Si wafer having a (110) surface orientation.[57] The preparation of 5-μm-wide and 80-μm-deep moats with 5-μm-wide ridges between was demonstrated to be feasible by this technique, offering interesting possibilities in dielectric isolation[70] and epitaxy refill.[76]

Anisotropic etchants are generally insensitive to resistivity and type of dopant element except in the case of boron at high concentrations.

A few interesting examples of anisotropic etching selected from work done in the author's laboratory are shown in Fig. 7.[77] All samples were (100)-surface polished Si wafers on which arrays of circular SiO_2-masked areas were photolithographically delineated with BHF to serve as etch masks. Exposure to various etchants at elevated temperatures in a reflux apparatus led to etched surfaces featuring regularly shaped pyramids for 64% N_2H_4 (Fig. 7A), pyramids of various size for 19% KOH (Fig. 7B), smooth surfaces with occasional pyramids for aqueous ethylenediamine-pyrocatechol (Fig. 7C), and smooth, structureless surfaces for 9.7% NH_4OH (Fig. 7D). The features resulting from various etchants are strongly affected by the concentration of the etchants and the temperature and duration of etching. These examples demonstrate the variety

of surface structure that can be obtained from (100) surface Si wafers by suitable choice of the etching process.

No comparable papers appear to exist on anisotropic etching of Ge, other than for analytical applications and for early crystallographic studies with single-crystal Ge spheres.[78]

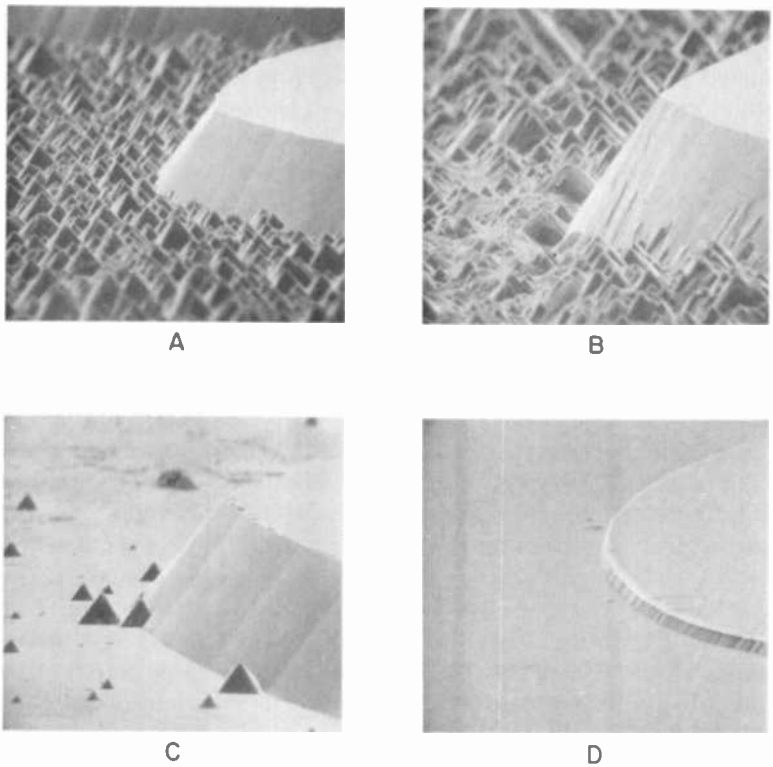

Fig. 7—Anisotropic etching of (100) surface silicon with various etchants. Smooth elevated areas are originally circular patterns of SiO_2-masked Si (SEM, original magnification 500× at 80° specimen tilt; Ref. [77]).

A. 64% hydrazine solution, 30 min at 90°C. Etch rate 1.6 μm/min. Etch depth 48 μm.

B. 19% potassium hydroxide aqueous solution, 120 min at 80°C. Etch rate 0.59 μm/min. Etch depth 71 μm.

C. 30mL ethylenediamine–6g pyrocatechol–20mL H_2O, 67 min at 85°C. Etch rate 0.87 μm/min. Etch depth 58 μm.

D. 1 vol ammonium hydroxide 58%–5 vol H_2O, 60 min at 85–92°C. Etch rate 0.11 μm/min. Etch depth 6.6 μm.

3.3 Electrochemical and Selective Chemical Etching

Etching of semiconductors in liquid reagents by application of an external EMF is used for preparing mirror-like surfaces of semiconductor slices, and for creating very thin films of single-crystal semiconductors. Thin films or selectively etched semiconductor regions can also be created by use of specific etchant compositions whose etching rate is a function of dopant type or concentration. In this section we will briefly discuss specific electrochemical systems and techniques that are useful for etching of Si and Ge films and substrates, with emphasis on the more recent work.

Specific shapes can be imparted to Ge and Si crystals by controlled localized electrochemical etching techniques (such as jet spraying, conductivity enhancement using localized carrier injection, etc.).[79–81]

Electropolishing of Ge and Si has been achieved by scanning jet techniques using diluted acid electrolyte solutions containing HF.[82] Electrochemical polishing machines featuring rotating, narrow-spaced cathode–anode electrodes in conjunction with glycerol-containing electrolytes have been described for Ge[83] and for Si.[83,84] Simple immersion electro-etching can also produce highly polished surfaces. For example, electropolishing to a bright finish is readily achieved in 5% HF with n^+ Si having a resistivity less than 0.05 ohm-cm, if a Si electrode potential of at least 3V and a current density of typically greater than 120 mA/cm^2 is used. Below this current density and potential the surface becomes rough and pitted, or thick brown films may form. Similar results are obtained with p-type Si if the resistivity is not above 1 ohm-cm.[85]

Selective electrochemical etching of single-crystal Si substrates having suitable epitaxial structures has been employed for preparing very thin (0.5–20 μm) Si crystals, since the current density is related to the Si dissolution rate.[86] P-type and heavily doped n-type (n^+) Si can be dissolved anodically in dilute (1–5N) HF at sufficiently low voltages, whereas n-type Si does not dissolve, causing the etching process to stop automatically at the interface. Thin p-type crystals can be prepared from n^+np structures by mounting the wafer p-side down on a carrier, dissolving the n^+ substrate anodically, then stripping the thin, intermediate n-type stopping layer by chemical etching.[86] Substrates and conditions are required that prevent localized avalanche breakdown in the n-type Si.[87–89] A detailed investigation[25] of this preferential electrochemical thinning process showed that, for successful application, the substrate donor concentration in n^+n structures should be higher than 3×10^{18} cm^{-3}, and that etching stops at a donor concentration of about 2×10^{16} cm^{-3}. However, there are porous channels in the high-resistivity n-type layer requiring additional chemical etching to obtain a thin, homogeneously doped epitaxial layer of high perfection.[25] In p-type Si, dissolution occurs at acceptor concentrations about 5×10^{15} cm^{-3}.

The optimal operating point for smooth etch-thinning of Si with a given crystal orientation and resistivity is a potential just beyond the maximum (J_{max}) in a current density/electrode potential curve where electropolishing takes place.[85,87] Numerous applications of various epitaxial layer combinations on selectively etchable substrates have been described for electro-thinning in HF to yield very thin single-crystal films of Si for a variety of solid state devices.[25,86–88,90–92]

Selective electrochemical thinning of n^+-type Si substrates is also possible with alkali solutions (typically 5.5N KOH at 85°C) in which the etch rates are strongly dependent on the electrode potentials.[94,95] Precise control of the etch rates for separation and isolation of patterns in (100) and (110)-oriented Si during anisotropic etching has been attained by applying this principle of Si/solution potential control.[59]

Anodic oxidation of Si in electrolyte solutions based on organic media, particularly ethylene glycol,[96] is also possible. For example, anodization in ethylene glycol with 0.04N KNO_3 plus 2.5% H_2O with $Al(NO_3)_3$, followed by oxide stripping with 8% HF, has been described for analytical sectioning in the precise determination of Si diffusion profiles.[97] This two-step analytical technique is outside the scope of the present paper and is mentioned for completeness only.

Certain anomalies can occur during etching. Objects, both conductive or insulating, in contact with the Si surface during etching can either slow down or enhance the local etch rate very considerably.[98] Furthermore, certain substrate and etching conditions (especially below a critical current density) in the anodic dissolution of Si in aqueous HF may not yield the smooth surface finish desired in electrochemical thinning and polishing but can lead to brown layers, etch pits, and porous channels as noted in preceding paragraphs.[25,79,85,86,88–90,94,99–101] These phenomena appear to be caused by preferential etching and partial dissolution at localized sites.[3,88] For example, anodic dissolution of n^+-type Si in aqueous HF at moderate current densities leads to crystallographically oriented porous channels whose density and depth depend on applied voltage, donor concentration, and exposure time.[88] The direction of channels or tunnels on (111) substrates was determined by Laue x-ray diffraction photographs to be ⟨100⟩.[89] Preferential attack along the ⟨100⟩ direction was also reported for anodically etched Ge.[79] Addition of H_2SO_4 to the HF-H_2O electrolyte increases the etch rate and reduces the bias requirement to less than 0.5V while maintaining high current density. The formation of brown films is thus minimized, and a smooth surface is left on the unetched high-resistivity n-type Si substrate.[80]

Single-crystal films of porous Si, formed purposely from n- and p-type Si by anodic reaction in concentrated HF,[86,101] are very similar or

identical to the channeled brown layers noted above. They are readily soluble in cold alkali solutions, and are quite reactive chemically.[102] These films are of considerable interest in semiconductor isolation technology because p-type regions can be converted selectively without mask at a low anodizing voltage without affecting the n-type regions, and can be oxidized subsequently with ease to form thick insulating layers.[101] The electrochemical reaction can be further enhanced by proton irradiation prior to anodization.[103] No further references will be made to this interesting process, since it is not a homogeneous etching reaction of the type discussed here.

Selective etching to dissolve Si of different dopant types and resistivities can also be achieved by chemical technique without use of external electrodes. The etch rate of (100) Si in hot KOH-H_2O-n- or isopropanol anisotropic etchant is not affected by the resistivity or by the dopant element except in the case of boron. In this case the etch rate decreases strongly with increasing B-concentration (Fig. 8) yielding etch rate differentials as high as 50:1.[24,60,104–106] Etching of n^+,n^-,p^-, and p^+ Si of (100) surface orientation in KOH-isopropanol-CH_3OH at 80°C is represented graphically as a function of time in Fig. 9.[106] The etch rate of B-doped Si in ethylenediamine-pyrocatechol-water etchant is affected similarly.[107] At a critical concentration of $\sim 7 \times 10^{19}/cm^3$ the etching is completely inhibited.[15] In acidic etchant compositions, such as $1HF\text{-}3HNO_3\text{-}8CH_3COOH$, the etch rates are a function of the dopant concentration; resistivities greater than 0.068 ohm-cm are no longer etchable.[24,108–110]

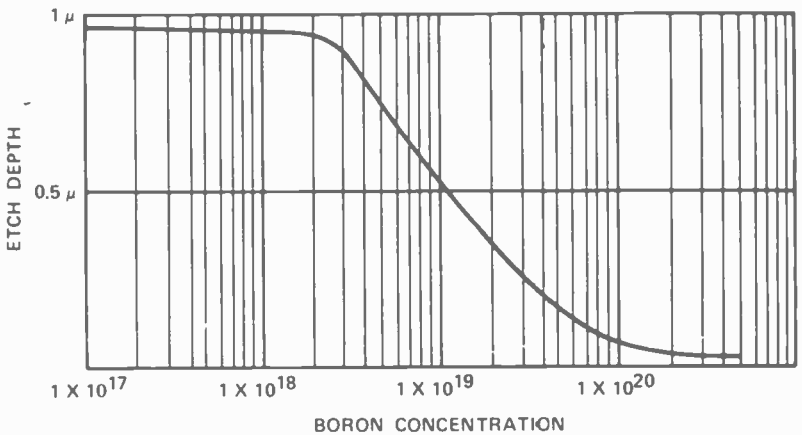

Fig. 8—Selective etching of silicon: Si ⟨100⟩ etch rate per minute versus boron concentration. Etchant system is KOH-H_2O-isopropylalcohol at 80°C. (From Ref. [105], reprinted by permission of the publisher, *The Electrochemical Society*, Inc.).

Preferential etching and shaping of n-type Si through SiO_2 windows can be achieved in refluxing H_3PO_4 containing As_2O_3.[111]

Use of electrodeless selective chemical etching has been made in preparing thin epitaxial Si films on insulating substrates[105] and in dielectric isolation processing of integrated circuits.[70,112,113] The technique has been utilized recently for fabricating micromechanical light modulator arrays,[114] and for electromechanical devices utilizing thin Si diaphragms.[115]

3.4 Gas and Vapor Phase Etching

Gas and vapor phase etching in semiconductor technology are widely used for etching polishing of Si substrate wafers *in situ* prior to epitaxial crystal growth. Substrate surfaces free of mechanical damage and impurities are essential to achieve the high crystallographic perfection required of the epitaxial layer. The most successful reagent to date for vapor polishing silicon is sulfur hexafluoride, SF_6. It produces a smooth, mirror-like surface when reacted in a dilution with H_2 at temperatures as low as 950°C,[116] but more usually above 1050°C, according to the overall reaction:[117]

$$4\,Si(s) + 2\,SF_6(g) \rightarrow SiS_2\,(s \text{ or } l) + 3\,SiF_4(g).$$

Since the free energy of the reaction is −706.81 Kcal/mol at 1400°C[118] the etching proceeds spontaneously and irreversibly, producing volatile sulfides and fluorides of Si as the reaction products.[117] Temperatures of 1050°C and below, over a concentration range of 0.01–0.1 vol%,[119] cause cloudy surfaces due to preferential etching. The temperature is much more critical than the SF_6 partial pressure in determining the quality of the resulting surface. The etch rate as a function of temperature for 0.1% SF_6 is shown in Fig. 9, and as a function of SF_6 partial pressure in Fig. 10.[119] The advantages of SF_6 over other reagents lie in the noncorrosiveness, nontoxicity, relatively low temperature needed (minimizing the generation of dislocations), in addition to the excellent planarity attainable which makes the process applicable to wafer thinning. No detectable sulfur contamination results; however, SiO_2 is attacked and cannot be used as a masking layer for pattern etching.[119] For best results, the wafers should be mechanically and chemically polished prior to vapor etching at 1050°–1100°C.[117]

Other vapor etchants used for polishing silicon are HCl,[120–127] HBr,[122] HI,[128,129] Cl_2,[130] H_2S,[131] HI-HF,[132,133] and H_2O.[134,135] Of these, anhydrous HCl in H_2 at 1100°–1200°C has been most commonly used; it is capable of producing surface finishes superior to those obtainable by liquid chemical polishing, and the rate of etching is not strongly de-

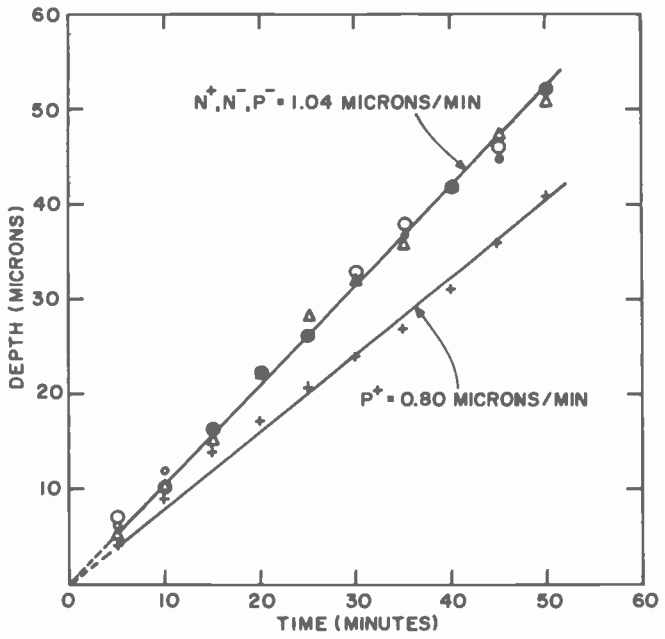

Fig. 9—Etch depth of variously doped (100) surface Si as a function of time at 80 ± 2°C in 50g KOH–100ml isopropyl alcohol–50mL methyl alcohol. Dopant concentrations are as follows: $n^+ = 4 \times 10^{18}$ Sb cm^{-3}, $n^- = 5 \times 10^{14}$ P cm^{-3}, $p^- = \sim 3 \times 10^{14}$ B cm^{-3}, and $p^+ = 5 \times 10^{18}$ B cm^{-3} (from Ref. [106]).

pendent on temperature or Si resistivity.[120] $HBr-H_2$ is easier to control because of its moderate temperature dependence, and leads to surface polishes at least as good as $HCl-H_2$.[122] Etching with H_2S-H_2 is two orders of magnitude faster than with $HCl-H_2$ at comparable temperature.[131] However, all of these etchants have certain disadvantages that make them less desirable than SF_6.

In general, gas and vapor phase etching is not dependent on the resistivity level or type of the Si. Isotropic etching is essential in creating smooth, reflective surfaces of high quality, and can usually be attained under optimal conditions of substrate temperature, reactant concentration, and type of diluent gas. Less than optimal conditions lead to rough or cloudy surfaces due to preferential etching. Conditions can be selected in the $HCl-H_2$ etching system that result in anisotropic etching of various crystal planes at different rates; this has been exploited technologically in selective epitaxy[124,126,127] involving (100) plane Si.

A few papers have been published on the gas phase etching of Ge. Similar to Si, (111)-Ge can be polish-etched with anhydrous HCl at 820°–830°C.[136] $HI-H_2$ mixtures at 911°C have been used for vapor polishing Ge wafers that were prepolished with NaOCl solution.[128]

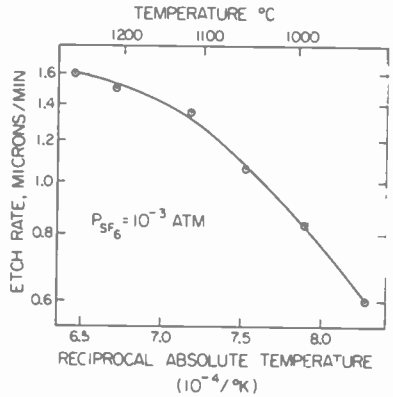

Fig. 10—Vapor phase etching of silicon with sulfur hexafluoride: Si etch rate versus substrate temperature for a SF_6 pressure of 10^{-3} atm (from Ref. [119], reprinted by permission of the publisher, *The Electrochemical Society*, Inc.).

Polish-etching of Ge is also possible with H_2S-H_2 at 650°–900°C.[137] However, superior results are obtainable with H_2-H_2O vapor as the etchant at 900°C. It produces clean, structureless, and polished surfaces with n- and p-type Ge of (111), (110), and (100) orientations.[137]

3.5 Chemical-Mechanical Polishing

Polishing by combined chemical and mechanical processes is now usually the last step in preparing flat and specular semiconductor wafers that are sufficiently free of mechanical work damage to be suitable as starting substrates in semiconductor device manufacture. The generally preferred

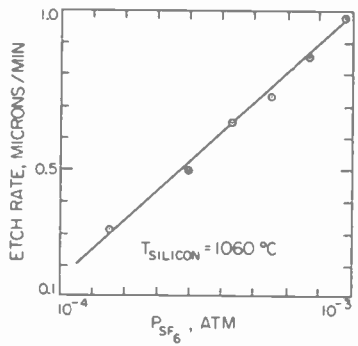

Fig. 11—Vapor phase etching of silicon with sulfur hexafluoride: Si etch rate versus SF_6 partial pressure at 1060°C. (From Ref. [119], reprinted by permission of the publisher, *The Electrochemical Society*, Inc.).

technique for silicon processing that yields nearly perfect surfaces is the silica-sol (Syton[138]) polishing method.[139,140] The medium consists of a colloidal suspension of silica gel in aqueous NaOH solution of controlled pH and is dispensed on the polishing pad of a rotating polishing machine. Silicon removal proceeds by oxidation of the surface by water in the presence of the alkali ions and continuous dissolution of the surface oxide, aided by the silica gel which serves as a mild abrasive. Factors affecting the final surface quality are composition and pH of the slurry, pad material, temperature, and pressure.[141] The polishing rate depends on both polishing parameters and dopant type and concentration of the Si, and may be as high as 30 μm/h.[140] The total Si thickness removed by slurry polishing is 38 to 50 μm.[142] The process can also be used for polishing Ge wafers, but H_2O_2 must be added to the dispersion to achieve a smooth surface finish.[140] A double-sided prepolishing method has been described recently which simultaneously slurry-polishes both sides of a Si slice and reduces the slice taper.[143]

Another type of chemical-mechanical polishing process for Si employs an aqueous solution containing copper and fluoride ions.[144,145] In this process Cu^{+2} ions are reduced by Si to metallic Cu,and Si is oxidized to Si^{+4} according to the overall reaction:

$$Si^{\circ} + 2Cu(NO_3)_2 + 6NH_4F \rightarrow (NH_4)_2SiF_6 + 4NH_3 + 2Cu^{\circ} + 4HNO_3.$$

The Cu deposit on the wafer surface is removed by a polishing cloth, while the oxidized Si dissolves as fluosilicate. The wafers are finally rinsed in diluted HNO_3 to extract any residual Cu. Stock removal rates at 375 to 500 are much faster than those of slurry-polishing, and clean damage-free surfaces result.

Mechanical polishing without use of abrasive particles can be combined with liquid chemical etching by moving the semiconductor wafer with uniform pressure on a polishing cloth soaked with the etchant liquid. This technique has been used for Ge[146] and GaAs.[147,148]

After chemical-mechanical polishing, Si wafers are frequently thermally oxidized prior to device fabrication. The oxidation process results in conversion of approximately 0.5 μm of the Si surface to SiO_2 per μm of SiO_2 and, thus, serves as a final etching step.[149]

Several reviews are available on semiconductor slicing, lapping, polishing, and the damage introduced by these operations which may be consulted for additional details.[139–142,144,150–153]

4. Etching of Gallium Arsenide

Reviews by Tuck[154] in 1970 and by Stirland and Straughan[155] in 1976 thoroughly cover important aspects of compound semiconductor etching in detail and discuss the relevent literature. Rather than attempting a complete survey of early work, we shall emphasize papers published after 1974.

4.1 Isotropic Liquid Etching

The most commonly employed etchants for GaAs are various compositions of Br_2-CH_3OH,[156–159] NaOH-H_2O_2,[160,161] H_2SO_4-H_2O_2-H_2O,[161–164] and NH_4OH-H_2O_2-H_2O.[159,165] High-viscosity etchants such as 8glycerol-1HCl-1HNO_3 are preferred for chemical polishing because the attack tends to be preferential at the high points resulting in a leveling action.[166]

Jet etching combined with slice rotation using 2Br_2-98CH_3OH is nonpreferential and can produce polished $(\overline{111})$ and (100) surfaces.[18] Rotating jet etching has also been used to thin and polish GaAs specimens for electron microscopy.[167] Stationary jet etching at 20°C with 40 HCl-4H_2O_2-1H_2O has been described for thinning GaAs samples in preparation for electron microscopic analysis.[168]

Nonpreferential polishing to obtain high-quality GaAs surfaces has also been achieved by rotating the slices on a polishing pad impregnated with 700 H_2O_2-1NH_4OH,[159] with H_2O-NaOCl solutions,[147,148,155,169] or with 3H_2SO_4-1H_2O_2-1H_2O.[169]

Investigations of variations in the commonly used and generally selective H_2SO_4-H_2O_2-H_2O system have shown that good polishing is possible for all low-index orientations except for the {111} Ga surface.[163] Rapid etching of (100) GaAs in 4H_2SO_4-1H_2O_2-1H_2O at 50°C for 3 minutes leads to a residual oxide film thickness of 50 Å. Successive rapid etching followed by HF dips and slow etching in diluted NaOH-H_2O_2-H_2O solution at 30°C can reduce the oxide residue to less than 10 Å.[161]

Rapid polish-etching of GaAs can be obtained with 3HNO_3-2H_2O-1HF mixture.[170] Very fast polishing etching of (100) plane and all other low-index planes of GaAs results with 4H_2O_2-1H_2SO_4-1HF etchant at 25 or 40°C; thinning of wafers and removal of disturbed surface layers from GaAs crystals have also been achieved successfully with this new etchant.[171]

It should also be noted that very thin (≤100 Å) carbon films are impervious to common etchants for III-V compounds and can therefore be used as excellent pattern-definition masks, even with highly corrosive etchants such as cold H_2SO_4-H_2O_2-H_2O mixtures. The carbon films can be readily removed by plasma etching or oxidation.[172]

4.2 Anisotropic Liquid Etching

Crystallographically preferential etching of GaAs can be attained with Br_2 at low concentration (<3 wt %) in CH_3OH. The relative etch rates in 1 wt % Br_2 in CH_3OH are {110} ≳ {111} B ≳ {100} ≫ {111} A.[158] Concentrations greater than 1–2 vol % Br_2 preferentially etch the {332} Ga plane.[173] Gannon and Nuese[165] have described selective preferential etching through SiO_2 masks with 0.3N NH_4OH–0.1N H_2O_2 aqueous solution. This etchant greatly reduces the attack of the SiO_2, and hence minimizes undercutting leading to well-defined, flat profiles. Enhanced etch rate resulted for the {111} B surface, and retarded etch rates for the {111} A surface, similar to the effect observed for the H_2SO_4-H_2O_2-H_2O[163] and the Br_2-CH_3OH system[158] noted above. The {111}Ga (or {111} A) plane of GaAs etches unsatisfactorily in most etchants but it can be polish-etched like the {111} As surface with $49H_3PO_4$-$11HNO_3$ at 60°C.[174]

The orientation dependent etching characteristics of $8H_2O_2$-$1H_2SO_4$-$1H_2O$ have been utilized for etching channels of various geometries for Gunn-effect logic circuits[175] and for superlattice structures deposited by molecular beam epitaxy.[176] GaAs double heterostructure lasers have been fabricated using a smoothly acting preferential structural etchant consisting of $3CH_3OH$-$1H_3PO_4$-$1H_2O_2$.[177]

Otsubo *et al.*[178] have recently introduced a preferential etchant which does not erode photoresist masking layers and provides improved pattern resolution. It consists of citric acid-H_2O_2-H_2O, typically 10 vol citric acid (50 wt %)-1 vol H_2O_2 (30 wt %). The rate of etching ranges from 1 to 100Å/s, depending on crystal plane, solution composition, temperature, and agitation. The etching time is a factor for low concentrations of citric acid which are limited by both the chemical reaction rate and the diffusion rate. The etching process for the higher and more typical acid compositions is chemical reaction rate limited.

4.3 Electrochemical and Selective Chemical Etching

Etching of GaAs can also be accomplished by electrolytic techniques. Anodic etching of (100), (110), and (111) faces of n-type GaAs in flowing 10% KOH solution has been described.[179,180] Electropolishing of p-type and heavily doped n-type GaAs in 10–40% KOH or NaOH solutions at a current density of 1–5 A/cm^2 has been reported.[181] Selective removal of p-type GaAs substrates from 2 to 10-μm thick n-type GaAs or $GaAs_{1-x}P_x$ epitaxial layers can be accomplished by electrolytic spray etching with 3N NaOH at a current density of 100 mA/cm^2.[182] Anodic electropolishing of n-type GaAs under illumination with NaOH-EDTA-H_2O solution serving as the electrolyte has been reported in detail by Yamamoto and Yano.[183] Uniformly thin layers of n-GaAs from excessively thick and

nonuniform material can be obtained by electrolytic etching in HCl-H_2O in which the anodic oxide dissolves as rapidly as it is being formed.[184] Alternatively, 0.01–0.1N HNO_3 can be used at current densities of 10–20 mA/cm^2 at 2–3 V for controlled electrolytic etching of n-GaAs or n-GaP.[185]

Thinning of GaAs specimens for electron microscopy has been accomplished electrolytically in flowing $25HClO_4$-$75CH_3COOH$ etchant.[186]

Greene[187] has utilized the photovoltaic properties of GaAs in contact with solutions containing Fe(III) compounds for preferential etching according to conductivity type, with rates of $n^- > n^+ > p^+ \sim p^-$. This photochemical selective etching technique was carried out with an aqueous solution of $FeSO_4$ and EDTA under illumination similar to direct sunlight and resulted in etched surfaces with a matt finish.

Tijburg and van Dongen have shown[188,189] that redox solutions of appropriate composition and pH can etch GaAs selectively with respect to $Ga_{1-x}Al_xAs$, or *vice versa*. The same behavior is displayed by GaP-InGaP and GaP-GaAlAs heterostructures. This method can be very useful for structuring or characterizing complex semiconductor layer combinations in solid state device technology Other redox systems etch selectively with respect to the dopant type in these materials. Suitable aqueous redox solutions that are stable in both acid and alkaline solutions include I_2-KI, $K_3Fe(CN)_6$-$K_4Fe(CN)_6$, and $C_6H_4O_2$-$C_4H_6O_2$ (quinone-hydroquinone). Redox etchants that are stable only in the acid pH range are $FeCl_3$-$FeCl_2$ and $Ce(SO_4)_2$-$Ce(NO_3)_3$.

4.4 Gas and Vapor Phase Etching

A few papers have been published on vapor phase etching of GaAs substrates in preparation for epitaxy. H_2-H_2O-N_2 vapors at a substrate temperature of 1000°C etches $(\overline{1}11)$ and (100) GaAs but does not produce highly polished surfaces.[190] High-quality specular surfaces for $(\overline{1}\overline{1}\overline{1})$ and (100) GaAs can be obtained with anhydrous HCl in a H_2-As ambient at temperatures above 870°C, as reported recently by Bhat, Baliga, and Ghandhi, where the etch rate is mass-transport limited for all orientations.[191] Etch rates under optimal conditions at 900°C ranged from 7–11 μm/min. Since this latter process is critically dependent upon the purity of the vapor etchant, $AsCl_3$, which is available at very high purity, is an attractive alternate etchant.[192]

5. Etching of Gallium Phosphide

The same or similar etchant as described for GaAs are used for solution-etching GaP. No papers on gas phase etching could be found.

5.1 Isotropic and Anisotropic Liquid Etching

Chemical polishing can be accomplished with 1%Br_2-99%CH_3OH at an etch rate of about 0.25 μm/min,[154] in 2HCl-2H_2O-1HNO_3 at 60°C,[193] in CH_3OH saturated with Cl_2,[157] or in a saturated Cl_2-water solution which etches at a rate of about 0.5 μm/min and produces excellent results.[194] Polishing by Pellon cloth techniques is done with 1–20% Br_2 in CH_3OH.[157]

Polishing and mesa etching of n- and p-type (100) and (111) GaP has been achieved with an aqueous mixture of 1.0M $K_3Fe(CN)_6$ and 0.5M KOH. An etch mask of SiO_2 or Ti was used for patterning at temperatures ranging from 60–95°C.[195]

Preparation of ⟨111⟩ GaP for epitaxial growth involves work damage removal by chemical-mechanical techniques using mixtures of Br_2-H_3PO_4-CH_3OH.[196,197] The surfaces can then be etched in 5H_2SO_4-1H_2O_2 at 80°C for 5 min, immediately followed by a final etch with Br_2-CH_3OH-H_3PO_4 at 50°C at an optimal etch rate of 1.5 μm/min.[197]

The P {111} surface of GaP can be polish etched with HCl-HNO_3-H_2O at normal[198] or elevated temperature,[193,199] or with 2HCl-2H_2SO_4-2H_2O-1HNO_3.[200] Polishing of some orientations has been reported for 2HNO_3-1HCl.[201]

Groove and pattern etching of $(\bar{1}\bar{1}\bar{1})$ GaP is possible with aqua regia (3HCl-1HNO_3) at room temperature using an SiO_2 film as the etch mask.[202] Concentrated H_3PO_4 at 150–200°C is an excellent etchant for producing V-shaped grooves in $(\bar{1}\bar{1}\bar{1})$, (111), and (100) GaP using a gold film as the etch mask.[193]

5.2 Selective Chemical and Electrochemical Etching

Surface etching for saw damage removal can be accomplished with 3H_2SO_4-1H_2O_2 (33%)-1H_2O at 60°C for 5 min; p-type GaP etches preferentially.[203,204]

Flat, smooth surfaces for {111} GaP have been obtained by chemical jet etching in the case of the P $(\bar{1}\bar{1}\bar{1})$ plane, and by jet electrochemical etching in the case of the Ga (111) plane.[18]

Localized thinning of specimens for transmission electron microscopy has been done for Ga (111) and P (111) planes of GaP by electrolytic jet techniques using solutions of Cl_2 in CH_3OH.[205]

Selective etching of p-type GaP on n-type GaP substrates can be achieved by chemical etching at room temperature in 1HF-1H_2O_2.[203]

The opposite, selective dissolution of n-type GaP on p-type GaP substrates, can be done electrochemically in 7NaOCl (5.25 wt%)—1HCl, making the n-type the anode and using Pt as the cathode.[203] Anodic dissolution of n-type ($\overline{111}$) GaP by electrolytic etching with 0.1N HNO_3 at a current density of 20 mA/cm^2 has also been described.[185]

Anodic dissolution and selective etching of Ga (111) and P ($\overline{111}$) GaP results by electrolytic treatment with 3N NaOH at 20°C using a Pt cathode.[206]

6. Conclusions

Chemical etching reactions have been reviewed for the two most important pairs of elemental and compound semiconductors: (1) silicon and germanium and (2) gallium arsenide and gallium phosphide. Emphasis has been placed on a descriptive discussion of specific etching systems that is intended to complement detailed data in a tabular format published in a separate paper.[1]

Modern semiconductor microelectronic technology has numerous requirements for polishing, accurate patterning, and selective removal of specific types of semiconductor materials under highly controlled conditions. A considerable variety of etching systems has been developed over the years, as evident from this review, that can meet most of the demands of these sophisticated requirements.

Acknowledgment

I wish to thank E. Cave and J. P. White for allowing me to use Figs. 5, 6, and 9 from their work, C. E. Tracy for assisting me in compiling the literature references, and G. L. Schnable and C. W. Horsting for reviewing the manuscript.

References:

[1] W. Kern and C. A. Deckert "Chemical Etching," in: *Thin Film Processes*, J. L. Vossen and W. Kern, Eds., Academic Press, New York (Late Fall 1978).

[2] C. M. Melliar-Smith and C. J. Mogab, "Plasma Assisted Etching Technique for Pattern Delineation," in: *Thin Film Processes*, J. L. Vossen and W. Kern, Eds., Academic Press, New York (Late Fall 1978).

[3] D. R. Turner, "On the Mechanism of Chemically Etching Germanium and Silicon," *J. Electrochem. Soc.*, **107**, p. 810 (1960).

[4] D. L. Klein and D. J. D'Stefan, "Controlled Etching of Silicon in the HF-HNO_3 System," *J. Electrochem. Soc.*, **109**, p. 37 (1962).

[5] H. Robbins and B. Schwartz, "Chemical Etching of Silicon—I. The System HF, HNO_3, and H_2O," *J. Electrochem. Soc.*, **106**, p. 505 and 1020 (1959).

[6] H. Robbins and B. Schwartz, "Chemical Etching of Silicon—II. The Systems HF, HNO_3, H_2O and $HC_2H_3O_2$," *J. Electrochem. Soc.*, **107**, p. 108 (1960).

[7] B. Schwartz and H. Robbins, "Chemical Etching of Silicon—III. A Temperature Study in the Acid System," *J. Electrochem. Soc.*, **108**, p. 365 (1961).

[8] B. Schwartz and H. Robbins, "Chemical Etching of Silicon—IV. Etching Technology," *J. Electrochem. Soc.*, **123**, p. 1903 (1976).

[9] A. F. Bogenschütz, W. Krusemark, K-H. Locherer, and W. Mussinger, "Activation Energies in the Chemical Etching of Semiconductors in HNO_3-HF-CH_3COOH," *J. Electrochem. Soc.*, **114**, p. 970 (1967).

[10] S. M. Hu and D. R. Kerr, "Observation of Etching of n-Type Silicon in Aqueous HF Solutions," *J. Electrochem. Soc.*, **114**, p. 414 (1967).

[11] W. Hoffmeister, "Determination of the Etch Rate of Silicon in Buffered HF Using a ^{31}Si Tracer Method," *Internatl. J. Appl. Radiation and Isotopes*, **2**, p. 139 (1969).

[12] E. Cave et al., RCA Solid State Division, private communication.

[13] G. R. Booker and R. Stickler, "Method of Preparing Si and Ge Specimens for Examination by Transmission Electron Microscopy," *Brit. J. Appl. Phys.*, **13**, p. 446 (1962).

[14] C. J. Schmidt, P. V. Lenzo, and E. G. Spencer, "Preparation of Thin Windows in Silicon Masks for X-ray Lithography," *J. Appl. Phys.*, **48**, p. 4080 (1975).

[15] G. Das and N. A. O'Neil, "Preparation of Large-Area Electron-Transparent Samples from Silicon Devices," *IBM J. Res. Dev.*, **18**, p. 76 (1974).

[16] A. I. Stoller, R. F. Speers and S. Opresko, "A New Technique for Etch Thinning Silicon Wafers," *RCA Review*, **31**, p. 265 (1970).

[17] J. Freyer, "Investigations of the Fabrication of Thin Silicon Films for Microwave Semiconductor Transit Time Devices," *J. Electrochem. Soc.*, **122**, p. 1238 (1975).

[18] B. A. Unvala, D. B. Holt, and A. San, "Jet Polishing of Semiconductors-III. Polishing and Shaping of Si, Ge, GaAs, and GaP Slices," *J. Electrochem. Soc.*, **119**, p. 318 (1972).

[19] R. J. Jacodine, "Interaction of Diffusion and Stacking Faults in Si Epitaxial Material," *J. Appl. Phys.*, **36**, p. 2811 (1965).

[20] M. V. Sullivan and R. M. Finne, "A New Technique for the Preparation of Flat Germanium Surfaces," Abstract 156, ECS Meeting, Houston, Texas (Oct. 1960).

[21] R. E. Blaha and W. R. Fahrner, "Passivation of High Breakdown Voltage p-n-p Structures by Thermal Oxidation," *J. Electrochem. Soc.*, **123**, p. 515 (1976).

[22] D. G. Schimmel and N. J. Elkind, "An Examination of the Chemical Staining of Silicon," *J. Electrochem. Soc.*, **125**, p. 152 (1978).

[23] M. Chappey and P. Meritet, "Etching Solutions for the Surface Treatment of Semiconductor Materials," French Pat. 1, 266,612 (Dec. 26, 1961).

[24] Ch. Raetzel, S. Schild and H. Schlötterer, "Etching Properties of Different Silicon Films," *Electrochem. Soc. Ext. Abstr.*, **74-2**, p. 336 (1974).

[25] M. J. J. Theunissen, J. A. Apples, and W. H. C. G. Verkuylen, "Application of Preferential Electrochemical Etching of Silicon to Semiconductor Device Technology," *J. Electrochem. Soc.*, **117**, p. 959 (1970).

[26] R. R. Stead, "Etching of Semiconductor Material," U.S. Pat. 2,973,253 (Feb. 28, 1961).

[27] P. J. Holmes, p. 370 in *The Electrochemistry of Semiconductors*, P. J. Holmes, Ed., Academic Press, London (1962).

[28] *1967 Book of ASTM Standards*, ASTM, Philadelphia, PA (1967).

[29] P. J. Holmes, Proc. Inst. Elect. Engrs., **B106**, (Suppl. 17), p. 861 (1959).

[30] Telefunken AG, "An Improved Etching Fluid and Method of Etching," Brit. Pat. 962,335 (July 1, 1964).

[31] W. R. Runyan, p. 199 (Table 7.3) in *Semiconductor Measurements and Instrumentation*, McGraw-Hill Book Co., New York (1975).

[32] *Integrated Circuit Silicon Device Technology; X-Chemical Metallurgical Properties of Silicon*, ASD-TDR-63-316, Vol. X; AD 625,985; Research Triangle Institute, North Carolina (Nov. 1965).

[33] B. A. Irving, "The Preparation of Thin Films of Germanium and Silicon," *Brit. J. Appl. Phys.*, **12**, p. 92 (1961).

[34] B. Schwartz and H. Robbins, "Chemical Etching of Germanium in Solutions of HF, H_2O, and HC_2-H_3O_2," *J. Electrochem. Soc.*, **111**, p. 196 (1964).

[35] B. A. Irving, "The Chemical Polishing of Germanium," *J. Electrochem. Soc.*, **109**, p. 120 (1962).

[36] D. R. Turner, "Experimental Information on Electrochemical Reactions at Germanium and Silicon Surface," Chapter 4 in *The Electrochemistry of Semiconductors*, P. J. Holmes, Ed., Academic Press, London (1962).

[37] J. W. Faust, Jr. "A Study of the Reaction of Germanium with Various Oxidizing and Complexing Agents," pp. 337–343 in *Reactivity of Solids*, J. W. Mitchell, R. C. De Vries, R. W. Roberts, and P. Cannon, Eds., John Wiley and Sons, New York 1969.

[38] M. F. Ehman, J. W. Faust, Jr., and W. B. White, "The Influence of the Complexing Agent Concentration on the Etch Rate of Germanium," *J. Electrochem. Soc.*, **118**, p. 1443 (1971).

[39] B. Schwartz, "Chemical Etching of Germanium in the System HF-H_2O_2-H_2O," *J. Electrochem. Soc.*, **114**, p. 285 (1967).

[40] W. Primak, R. Kampwirth, and Y. Dayal, "Peroxide Etching of Germanium," *J. Electrochem. Soc.*, **114**, p. 88 (1967).

[41] P. R. Camp, "A Study of the Etching Rate of Single Crystal Germanium," *J. Electrochem. Soc.*, **102**, p. 586 (1955).

[42] J. P. McKelvey and R. L. Longini, *J. Appl. Phys.*, **25**, p. 634 (1954).

[43] B. W. Batterman, "Hillocks, Pits and Etch Rate in Germanium Crystals," *J. Appl. Phys.*, **28**, p. 1236 (1957).

[44] P. J. Holmes, p. 369 in *The Electrochemistry of Semiconductors*, P. J. Holmes, Ed., Academic Press, London (1962).

[45] G. A. Geach, B. A. Irving, and R. Phillips, *Research, Lond.*, **10**, p. 411 (1957).

[46] F. L. Vogel, W. G. Pfann, H. E. Corey, and E. E. Thomas, "Observations of Dislocations in Lineage Boundaries in Germanium," *Phys. Rev.*, **90**, p. 489 (1953).

[47] R. D. Heidenreich, U.S. Pat. No. 2,619,414 (1952).

[48] P. J. Holmes, p. 368 in *The Electrochemistry of Semiconductors*, P. J. Holmes, Ed., Academic Press, London (1962).

[49] P. Wang, *Sylv. Tech.*, **11**, p. 50 (1958).

[50] H. C. Gatos and M. C. Lavine, "Chemical Behaviour of Semiconductors: Etching Characteristics" pp. 1–46 in *Progress in Semiconductors*, Vol. 9, A. G. Gibson and R. E. Burgess, Eds., Temple Press Books Ltd., LONDON (1965).

[51] W. W. Harvey and H. C. Gatos, "The Reaction of Germanium with Aqueous Solution-I. Dissolution Kinetics in Water Containing Dissolved Oxygen," *J. Electrochem. Soc.*, **105**, p. 654 (1958).

[52] H. C. Gatos and M. C. Lavine, "Etching Behavior of the [110] and [100] Surfaces of InSb," *J. Electrochem. Soc.*, **107**, p. 433 (1960).

[53] W. K. Zwicker and S. K. Kurtz, "Anisotropic Etching of Silicon Using Electrochemical Displacement Reactions," pp. 315–326 in *Semiconductor Silicon 1973*, H. R. Huff and R. R. Burgess, Eds., The Electrochemical Soc., Princeton, N.J. (1973).

[54] W. R. Runyan, Chapters 1, 2, 7, and 9 in *Semiconductor Measurements and Instrumentation*, McGraw-Hill Book Co., New York (1975).

[55] I. J. Pugacz-Muraszkiewicz and B. R. Hammond, "Application of Silicates to the Detection of Flaws in Glassy Passivation Films Deposited on Silicon Substrates," *J. Vac. Sci. Technol.*, **14**, p. 49 (1977).

[56] H. A. Waggener, R. C. Kragness and A. L. Tyler, "Anisotropic Etching for Forming Isolation Slots in Silicon Beam Leaded Integrated Circuits," *Internatl. Electron Devices Mtg.* Abstr. No. 11.1, p. 68 (Oct. 1967).

[57] A. I. Stoller, "The Etching of Deep Vertical-Walled Patterns in Silicon," *RCA Review*, **31**, p. 271 (1970).

[58] H. A. Waggener, R. C. Kragness, and A. L. Tyler, "Two-way Etch," *Electronics*, **40**, No. 23, p. 274 (Nov. 13, 1967).

[59] H. A. Waggener and J. V. Dalton, "Control of Silicon Etch Rates in Hot Alkaline Solutions by Externally Applied Potentials," *Electrochem. Soc. Ext. Abstr.*, **72-2**, p. 587 (1972).

[60] J. B. Price, "Anisotropic Etching of Silicon with KOH-H_2O-Isopropyl Alcohol," pp. 339–353 in *Semiconductor Silicon 1973*, H. R. Huff and R. R. Burgess, Eds., The Electrochemical Society, Princeton, N. J. (1973).

[61] D. P. Clemens, "Anisotropic Etching of Silicon on Sapphire," *Electrochem. Soc. Ext. Abstr.*, **73-2**, p. 407 (1973).

[62] D. F. Weirrauch, "Correlation of the Anisotropic Etching of Single-Crystal Silicon Spheres and Wafers," *J. Appl. Phys.*, **46**, p. 1478 (1975).

[63] K. E. Bean, R. L. Yeakley and T. K. Powell, "Orientation Dependent Etching and Deposition of Silicon," *Electrochem. Soc. Ext. Abstr.*, **74-1**, p. 68 (1974).

[64] M. J. Declercq, J. P. DeMoor, and J. P. Lambert, "A Comparative Study of Three Anisotropic Etchants for Silicon," *Electrochem. Soc. Ext. Abstr.*, **75-2**, p. 446 (1975).

[65] D. B. Lee, "Anisotropic Etching of Silicon," *J. Appl. Phys.*, **40**, p. 4569 (1969).

[66] M. J. Declercq, L. Gerzberg, and J. D. Meindl, "Optimization of the Hydrazine-Water Solution for Anisotropic Etching of Silicon in Integrated Circuit Technology," *J. Electrochem. Soc.*, **122**, p. 545 (1975).

[67] R. M. Finne and D. L. Klein, "A Water-Amine-Complexing Agent System for Etching Silicon," *J. Electrochem. Soc.*, **114**, p. 965 (1967).

[68] M. Asano, T. Cho and H. Muraoka, "Application of Choline in Semiconductor Technology," *Electrochem. Soc. Ext. Abstr.*, **76-2**, p. 911 (1976).

[69] I. Bassou, H. N. Yu, and V. Maniscalco, "Topology of Silicon Structures with Recessed SiO_2," *J. Electrochem. Soc.*, **123**, p. 1729 (1976).

[70] K. E. Bean and W. R. Runyan, "Dielectric Isolation: Comprehensive, Current and Future," *J. Electrochem. Soc.*, **124**, p. 56 (1977).

[71] T. R. Payne and H. R. Plumlee, "Thermal Printer," *IEEE J. Solid-State Circ.*, **SC-8**, p. 71 (1973).

[72] K. E. Bean and J. R. Lawson, "Application of Silicon Crystal Orientation and Anisotropic Effects to the Control of Charge Spreading in Devices," *IEEE J. Solid State Circ.*, **SC-9**, p. 111 (1974).

[73] W. Tsang and S. Wang, "Preferentially Etched Diffraction Gratings in Silicon," *J. Appl. Phys.*, **46**, p. 2164 (1975).

[74] C. R. Baraona and H. W. Brandhorst, "V-Grooved Silicon Solar Cells," *11th IEEE Photovoltaic Spec. Conf.*, Scottsdale, Ariz., p. 44 (1975).

[75] F. Restrepo and C. E. Backer, "On Black Solar Cells or the Tetrahedral Texturing of a Silicon Surface," *IEEE Trans. Electron Devices*, **ED 23**, No. 10, (Oct. 1976).

[76] R. K. Smeltzer, "Epitaxial Deposition of Silicon in Deep Grooves," *J. Electrochem. Soc.*, **122**, 1666 (1975).

[77] W. Kern, unpublished work.

[78] R. C. Ellis, Jr., *J. Appl. Phys.*, **25**, p. 1497 (1954); and **28**, p. 1068 (1957).

[79] A. Uhlir, "Electrolytic Shaping of Germanium and Silicon," *Bell System Tech. J.*, **35**, p. 333 (1956).

[80] G. L. Schnable and W. M. Lilker, "Jet-Electrolytic Etching and Plating," *Electrochem. Technol.*, **1**, p. 203 (1963).

[81] J. I. Pankove, pp. 297–308 in *The Electrochemistry of Semiconductors*, P. J. Holmes, Ed., Academic Press, London (1962).

[82] G. R. Booker and R. Stickler, "Large-Area 'Jet' Electrolytic Polishing of Ge and Si," *J. Electrochem. Soc.*, **109**, p. 1167 (1962).

[83] M. V. Sullivan, D. L. Klein, R. M. Finne, L. A. Pompliano, and G. A. Kolb, "An Electropolishing Technique for Germanium and Silicon," *J. Electrochem. Soc.*, **110**, p. 412 (1963).

[84] C. E. Hallas, "Electropolishing Silicon," *Solid State Technol.*, **14**, No. 1, p. 30 (Jan. 1971).

[85] R. L. Meek, "Anodic Dissolution of N^+ Silicon," *J. Electrochem. Soc.*, **118**, p. 437 (1971).

[86] H. J. A. van Dijk and J. de Jonge, "Preparation of Thin Silicon Crystals by Electrochemical Thinning of Epitaxially Grown Structures," *J. Electrochem. Soc.*, **117**, p. 553 (1970).

[87] R. L. Meek, "Electrochemically Thinned N/N^+ Epitaxial Silicon—Method and Applications," *J. Electrochem. Soc.*, **118**, p. 1240 (1971).

[88] M. J. J. Theunissen, "Etch Channel Formation during Anodic Dissolution of N-Type Silicon in Aqueous Hydrofluoric Acid," *J. Electrochem. Soc.*, **119**, p. 351 (1972).

[89] M. J. Hill, "Electron Microscopy of Electrolytically Etched Silicon Surfaces," *J. Electrochem. Soc.*, **120**, p. 142 (1973).

[90] C. P. Wen and K. P. Weller, "Preferential Electrochemical Etching of P^+ Silicon in an Aqueous HF-H_2SO_4 Electrolyte," *J. Electrochem. Soc.*, **119**, p. 547 (1972).

[91] T. I. Kamins, "A New Dielectric Isolation Technique for Biopolar Integrated Circuits Using Thin Single-Crystal Silicon Films," *Proc. IEEE*, **60**, p. 915 (1972).

[92] T. I. Kamins, "Diffusion in Thin Silicon Films Formed by Electrochemical Etching," *J. Electrochem. Soc.*, **121**, p. 286 (1974).

[93] T. I. Kamins, "Hall Mobility in Dielectrically Isolated Single-Crystal Silicon Films Defined by Electrochemical Etching," *Solid-State-Electronics*, **17**, p. 667 (1974).

[94] H. A. Waggener, "Electrochemically Controlled Thinning of Silicon," *Bell System Tech. J.*, **49**, p. 473 (1970).

[95] H. A. Waggener and J. V. Dalton, "Electrochemically Controlled Thinning," *Electrochem. Soc. Ext. Abstr.*, **70-2**, p. 450 (1970).

[96] A. Manara, A. Ostidich, G. Pedroli, and G. Restelli, "Anodic Oxidation as Sectioning Technique for the Analysis of Impurity Concentration Profiles in Silicon," *Thin Solid Films*, **8**, p. 359 (1971).

[97] H. D. Barber, H. B. Lo, and J. E. Jones, "Repeated Removal of Thin Layers of Silicon by Anodic Oxidation," *J. Electrochem. Soc.*, **123**, p. 1404 (1976).

[98] J. Sanada, K. Furuno, K. Shima, and T. Momoi, "A Modified Method of Etching Semiconductors and Manufacturing Thin Silicon Solid State Detectors," *Japan. J. Appl. Phys.*, **16**, p. 299 (1977).

[99] R. Memming and G. Schwandt, "Anodic Dissolution of Silicon in Hydrofluoric Acid Solutions," *Surface Sci.*, **4**, p. 109 (1966).

[100] P. H. Bellin and W. K. Zwicker, "Observation of Surface Defects in Electrolytically Etched Silicon by Infrared Microscopy," *J. Appl. Phys.*, **42**, p. 1216 (1971).

[101] Y. Watanabe, Y. Arita, T. Yokoyama, and Y. Igaraschi, "Formation and Properties of Porous Silicon and its Application," *J. Electrochem. Soc.*, **122**, p. 1351 (1975).

[102] Y. Arita and Y. Sunohara, "Formation and Properties of Porous Silicon Films," *J. Electrochem. Soc.*, **124**, p. 285 (1977).

[103] Y. Yashiro, K. Saito, and T. Suzuki, "Isolation Layer Formation by Enhanced Anodization of N-Type Silicon," *Electrochem. Soc. Ext. Abstr.*, **74-2**, p. 351 (1974).

[104] J. B. Price and W. C. Roman, "The Resistivity Dependence of the Silicon Etch Rate and Anisotropy in KOH-IPA-H_2O," *Electrochem. Soc. Ext. Abstr.*, **72-2**, p. 584 (1972).

[105] G. L. Kuhn and C. J. Rhee, "Thin Silicon Film on Insulating Substrate," *J. Electrochem. Soc.*, **120**, p. 1563 (1973).

[106] J. P. White, RCA Solid State Division, Private Communication.

[107] A. Bohg, "Ethylene Diamine-Pyrocatechol-Water Mixture Shows Etching Anomaly in Boron-Doped Silicon," *J. Electrochem. Soc.*, **118**, p. 401 (1971).

[108] H. Muraoka, T. Ohhashi, and Y. Sumitomo, "Controlled Preferential Etching Technology," pp. 327–338 in *Semiconductor Silicon 1973*, by H. R. Huff and R. R. Burgess, Eds., The Electrochemical Soc., Princeton, NJ (1973).

[109] Y. Yasuda and T. Moriya, "Marked Effects of Boron-Doping on the Growth and Properties of Polycrystalline Silicon Films," pp. 271–284 in *Semiconductor Silicon 1973*, H. R. Huff and R. R. Burgess, Eds., The Electrochem. Soc., Princeton, NJ (1973).

[110] R. E. Chappelow and P. T. Lin, "Characteristics of P-Doped Polysilicon with B_2H_6 used as In-Situ Dopant Source," *J. Electrochem. Soc.*, **123**, p. 913 (1976).

[111] J. Lawrence, "Preferential Etching of Silicon," Canad. Pat. No. 903,650 (June 27, 1972).

[112] Y. Sumitomo, K. Niwa, H. Sawazaki and K. Sakai, "A New Technology for Dielectric Isolation," pp. 893–904 in *Semiconductor Silicon 1973*, H. R. Huff and R. R. Burgess, Eds., The Electrochemical Soc., Princeton, NJ (1973).

[113] Y. Sumitomo, T. Yasui, H. Nakatsuka, T. Oohashi, H. Tsutsumi and H. Muraoka, "Applications of Selective Epitaxy and Preferential Etching of Silicon to Semiconductor Devices," *Electrochem. Soc. Ext. Abst.*, **72-1**, p. 74 (1972).

[114] K. E. Peterson, "Micromechanical Light Modulator Array Fabricated on Silicon," *Appl. Phys. Lett.*, **31**, p. 521 (1977).

[115] H. Guckel, S. Larsen, M. G. Lagally, G. Moore, J. B. Miller, and J. D. Wiley, "Electrochemical Devices Utilizing Thin Si Diaphragms," *Appl. Phys. Lett.*, **31**, p. 618 (1977).

[116] Y. S. Chiang and G. Looney, RCA Laboratories, private communication.

[117] P. Rai-Choudhury, "Sulfur Hexafluoride as an Etchant for Silicon," *J. Electrochem. Soc.*, **118**, p. 266 (1971).

[118] S. E. Mayer and D. E. Shea, "Epitaxial Deposition of Silicon Layers by Pyrolysis of Silane," *J. Electrochem. Soc.*, **111**, p. 550 (1964).

[119] L. J. Stinson, J. A. Howard, and R. C. Neville, "Sulfur Hexafluoride Etching Effects in Silicon," *J. Electrochem. Soc.*, **123**, p. 551 (1976).

[120] G. A. Lang and T. Stavish, "Chemical Polishing of Silicon with Anhydrous Hydrogen Chloride," *RCA Review*, **24**, p. 488 (1963).

[121] W. H. Shepherd, "Vapor Phase Deposition and Etching of Silicon," *J. Electrochem. Soc.*, **112**, p. 988 (1965).

[122] L. V. Gregor, P. Balk and F. J. Campagna, "Vapor-Phase Polishing of Silicon with H_2–H Br Gas Mixture," *IBM J. Res. Dev.*, **9**, p. 327 (1965).

[123] W. Runyan, pp. 72–73 in *Silicon Semiconductor Technology*, McGraw-Hill Co., New York (1965).

[124] K. E. Bean and P. S. Gleim, "The Influence of Crystal Oriertation on Silicon Semiconductor Processing," *Proc. IEEE*, **57**, p. 1469 (1969).

[125] K. Sugawara, Y. Nakazawa, and Y. Sugita, "Observation of Facets on Silicon Surface Formed by Hydrochloric Acid Selective Vapor Etching," *Electrochem. Technol.*, **6**, p. 295 (1968).

[126] K. Sugawara, "Facets Formed by Hydrogen Chloride Vapor Etching on Silicon Surfaces through Windows in SiO_2 and Si_3N_4 Masks," *J. Electrochem. Soc.*, **118**, p. 110 (1971).

[127] M. Druminski and R. Gessner, "Selective Etching and Epitaxial Refilling of Silicon Wells in the System SiH_4/HCl/H_2," *J. Crystal Growth*, **31**, p. 312 (1975).

[128] A. Reisman and M. Berkenblit, "The Etching and Polishing Behavior of Ge and Si with HI," *J. Electrochem. Soc.*, **112**, p. 812 (1965).

[129] L. D. Dyer, "HI-H_2 Vapor Etch for Low Temperature Silicon Epitaxial Manufacturing," *Amer. Inst. Chem. Engrs. J.*, **18**, p. 728 (1972).

[130] J. P. Dismukes and R. Ulmer, "Gas-Phase Etching of Silicon with Chlorine," *J. Electrochem. Soc.*, **118**, p. 634 (1971).

[131] P. Rai-Choudhury and A. J. Noreika, "Hydrogen Sulfide as an Etchant for Silicon," *J. Electrochem. Soc.*, **116**, p. 539 (1969).

[132] J. P. Dismukes and E. R. Levin, "Low-Temperature Gaseous Etching of Silicon with HI-HF Mixtures," p. 135 in *Chem. Processing of Microelectronic Materials*, L. D. Locker and J. Childress, Eds., Am. Inst. Chem. Engrs., New York (1970).

[133] E. R. Levin, J. P. Dismukes and M. D. Coutts, "Electron Microscope Characterization of Defects on Gaseous-Etched Silicon Surfaces," *J. Electrochem. Soc.*, **118**, p. 1171 (1971).

[134] T. L. Chu, G. A. Gruber and R. Stickler, "In Situ Etching of Silicon Substrates Prior to Epitaxial Growth," *J. Electrochem. Soc.*, **113**, p. 156 (1966).

[135] W. G. Oldham and R. Holmstrom, "The Growth and Etching of Si through Windows in SiO_2," *J. Electrochem. Soc.*, **114**, p. 381 (1967).

[136] J. A. Amick, E. A. Roth, and H. Gossenberger, "The Etching of Germanium Substrates in Gaseous Hydrogen Chloride," *RCA Review*, **24**, p. 473 (1963).

[137] T. L. Chu and R. W. Kelm, "The Etching of Germanium with Water Vapor and Hydrogen Sulfide," *J. Electrochem. Soc.*, **116**, p. 1261 (1969).

[138] R. J. Walsh and A. H. Herzog, U.S. Patent No. 3,170,273 (Jan. 10, 1963).

[139] E. Mendel, "Polishing of Silicon," *SCP and Solid State Technol.*, **10**, p. 27 (1967).

[140] T. M. Buck and R. L. Meek, "Crystallographic Damage to Silicon by Typical Slicing, Lapping, and Polishing Operations," pp. 419–430 in *Silicon Device Processing*, C. P. Marsden, Ed., National Bureau of Stds. Special Publ. 337 (Nov. 1970).

[141] J. T. Law, "Silicon Wafer Technology—State of the Art," *Solid-State Tech.*, **14**, No. 1, p. 25 (1971).

[142] R. B. Herring, "Silicon Wafer Technology—State of the Art 1976," Solid-State Tech., **19**, No. 5, p. 37 (1976).

[143] E. L. Kern, G. L. Gill, and P. Rioux, "Double Sided Free Polishing of Silicon," pp. 182–186 in *Semiconductor Silicon 1977*, H. R. Huff and E. Sirtl, Eds., The Electrochem. Soc., Inc., Princeton, N.J. (1977).

[144] E. Mendel and K. Yang, "Polishing of Silicon by the Cupric Ion Process," *Proc. IEEE*, **57**, p. 1476 (1969).

[145] L. H. Blake and E. Mendel, "Chemical-Mechanical Polishing of Silicon," *Solid-State Tech.*, **13**, No. 1, p 42 (1970).

[146] A. Reisman and R. Rohr, "Room Temperature Chemical Polishing of Ge and GaAs," *J. Electrochem. Soc.*, **111**, p. 1425 (1964).

[147] V. L. Rideout, "An Improved Polishing Technique for GaAs," *J. Electrochem. Soc.*, **119**, p. 1778 (1972).

[148] H. Hartnagel and B. L. Weiss, "A Contribution to Etch Polishing of GaAs," *J. Mater. Sci.*, **8**, 1061 (1973).

[149] G. L. Schnable and P. F. Schmidt, "Applications of Electrochemistry to Fabrication of Semiconductor Devices," *J. Electrochem. Soc.*, **123**, p. 310C (1976).

[150] P. J. Holmes, "Practical Applications of Chemical Etching", Chapter 8 *in The Electrochemistry of Semiconductors*, P. J. Holmes, Ed., Academic Press, New York (1962).

[151] T. M. Buck, "Damaged Surface Layers; Semiconductors," pp. 107–135 in *The Surface Chemistry of Metals and Semiconductors*, H. C. Gatos, Ed., John Wiley & Sons, Inc., New York (1960).

[152] R. B. Soper, "Mechanical Damage—Its Role in Silicon Surface Preparation," pp. 412–418 in *Silicon Device Processing*, C. P. Marsden, Ed., National Bureau of Stds. Special Publ., 337 (Nov. 1970).

[153] A. C. Bonora, "Silicon Wafer Process Technology: Slicing, Etching, Polishing," pp. 154–169 in *Semiconductor Silicon 1977*, H. R. Huff and E. Sirtl, Eds., The Electrochem. Soc., Inc., Princeton, N. J. (1977).

[154] B. Tuck, "The Chemical Polishing of Semiconductors," *J. Materials Science*, **10**, p. 321 (1975).

[155] D. J. Stirland and B. W. Straughan, "A Review of Etching and Defect Characterization of Gallium Arsenide Substrate Material," *Thin Solid Films*, **31**, p. 139 (1976).

[156] M. V. Sullivan and G. A. Kolb, "The Chemical Polishing of Gallium Arsenide in Bromine-Methanol," *J. Electrochem. Soc.*, **110**, p. 585 (1963).

[157] C. S. Fuller and H. W. Allison, "A Polishing Etchant for III-V Semiconductors," *J. Electrochem. Soc.*, **109**, p. 880 (1962).

[158] Y. Tarui, Y. Komiya, and Y. Harada, "Preferential Etching and Etched Profile of GaAs," *J. Electrochem. Soc.*, **118**, p. 118 (1971).

[159] J. C. Dyment and G. A. Rozgonyi, "Evaluation of a New Polish for Gallium Arsenide Using a Peroxide-Alkaline Solution," *J. Electrochem. Soc.*, **118**, p. 1346 (1971).

[160] D. W. Shaw, "Enhanced GaAs Etch Rates Near the Edges of a Protective Mask," *J. Electrochem. Soc.*, **113**, p. 958 (1966).

[161] I. Shiota, K. Motoya, T/ Ohmi, N. Miyamoto, and J. Nishizawa, "Auger Characterization of Chemically Etched GaAs Surfaces," *J. Electrochem. Soc.*, **124**, p. 155 (1977).

[162] M. V. Sullivan and L. A. Pompliano, "A Chemical Polishing Technique for Gallium Arsenide," *J. Electrochem. Soc.*, **108**, p. 60C (1961).

[183] S. Iida and K. Ito, "Selective Etching of Gallium Arsenide Crystals in H_2SO_4-H_2O_2-H_2O System," *J. Electrochem. Soc.*, **118**, p. 768 (1971).

[164] See Ref. [54], p. 199, Table 7.4.

[165] J. J. Gannon and C. J. Nuese, "A Chemical Etchant for the Selective Removal of GaAs Through SiO_2 Masks," *J. Electrochem. Soc.*, **121**, p. 1215 (1974).

[166] R. D. Packard, "Notes on the Chemical Polishing of Gallium Arsenide Surfaces," *J. Electrochem. Soc.*, **112**, p. 871 (1965).

[167] R. W. Bicknell "A Simple Rotating Jet-Thinning Apparatus for Producing Taper Sections and Electron Microscope Specimens from Silicon and Compound Semiconductors," *J. Phys. D*, **6**, p. 1991 (1973).

[168] E. Biedermann and K. Brack, "Preparation of GaAs Specimens for Transmission Electron Microscopy," *J. Electrochem. Soc.*, **113**, p. 1088 (1966).

[169] E. W. Jensen, "Polishing Compound Semiconductors," *Solid State Tech.*, **16**, (8) p 49 (1973).

[170] J. L. Richards and A. J. Crocker, "Etch Pits in Gallium Arsenide," *J. Appl. Phys.*, **31**, p. 611 (1960).

[171] F. Kuhn-Kuhnenfeld, "A Polishing Dislocation Etch for GaP and GaAs," *Inst. Phys. Conf. Ser. No. 33a*, Chapter 3, p. 158–160 (1977).

[172] G. H. Olson and V. S. Ban, "Use of Thin Carbon Films for Selective Chemical Etching and Epitaxial Deposition of III-V Semiconductors," *Appl. Phys. Lett.*, **28**, p. 734 (1976).

[173] L. A. Koszi and D. L. Rode, "{332} Ga Habit Planes Formed on GaAs During Br_2:CH_3OH Etching," *J. Electrochem. Soc.*, **122**, p. 1676 (1975).

[174] J. G. Grabmaier and C. B. Watson, "Dislocation Etch Pits in Single Crystal GaAs," *Phys. Status Solidi*, **32**, p. K13 (1969).

[175] O. Wada, S. Yanagisawa, and H. Takanashi, "Process for GaAs Monolithic Integration Applied to Gunn-Effect Logic Circuits," *J. Electrochem. Soc.*, **123**, p. 1546 (1976).

[176] W. T. Tsang and A. Y. Cho, "Growth of GaAs-$Ga_{1-x}Al_xAs$ over Preferentially Etched Channels by Molecular Beam Epitaxy: A Technique for Two-Dimensional Thin-Film Definition," *Appl. Phys. Lett.*, **30**, p. 293 (1977).

[177] J. L. Merz and R. A. Logan, "GaAs Double Heterostructure Lasers Fabricated by Wet Chemical Etching," *J. Appl. Phys.*, **47**, p. 3503 (1976).

[178] M. Otsubo, T. Oda, H. Kumabe, and H. Miki, "Preferential Etching of GaAs Through Photoresist Masks," *J. Electrochem. Soc.*, **123**, p. 676 (1976).

[179] T. Ambridge, C. R. Elliot, and M. M. Faktor, *J. Appl. Electrochem.*, **3**, p. 1 (1973); ibid, **4**, p. 135 (1974).

[180] M. M. Fakter, D. G. Fiddyment, and M. R. Taylor, "Electrochemically Etched Tunnels in Gallium Arsenide," *J. Electrochem. Soc.*, 122, p. 1566 (1975).

[181] Y. V. Pleskov, *Dokl. Akad. Nauk SSSR*, **143**, p. 1399 (1962).

[182] C. J. Nuese and J. J. Gannon, "Electrolytic Removal of P-Type GaAs Substrates from Thin N-Type Semiconductor Layers," *J. Electrochem. Soc.*, **117**, p. 1094 (1970).

[183] A. Yamamoto and S. Yano, "Anodic Dissolution of N-Type Gallium Arsenide Under Illumination," *J. Electrochem. Soc.*, **122**, p. 260 (1975).

[184] D. L. Rode, B. Schwartz, and J. V. DiLorenzo, "Electrolytic Etching and Electron Mobility of GaAs for FETs," *Solid State Electron.*, **17**, p. 1119 (1974).

[185] B. Schwartz, F. Ermanis, and B. H. Brastad, "The Anodization of GaAs and GaP in Aqueous Solutions," *J. Electrochem. Soc.*, **123**, p. 1089 (1976).

[186] E. S. Meieran, "Transmission Electron Microscope Study of Gallium Arsenide," *J. Appl. Phys.*, **36**, p. 2544 (1965).

[187] P. D. Greene, "Preferential Photochemical Dissolution of n-GaAs in Fe(III)-Based Etches," *Inst. Phys. Conf. Ser. No. 33a*, Chapter 3, pp. 141–149 (1977).

[188] R. Tijburg, "Advances in Etching of Semiconductor Devices," *Physics in Technology.*, pp. 202–207 (Sept. 1976).

[189] R. P. Tijburg and T. van Dongen, "Selective Etching of III-V Compounds with Redox Systems," *J. Electrochem. Soc.*, **123**, p. 687 (1976).

[190] C. Lin, L. Chow, and K. Miller, "Vapor Phase Etching of GaAs in the H_2-H_2O Flow System," *J. Electrochem. Soc.*, **117**, p. 407 (1970).

[191] R. Bhat, B. J. Baliga, and S. K. Ghandhi, "Vapor-Phase Etching and Polishing of Gallium Arsenide Using Hydrogen Chloride Gas," *J. Electrochem. Soc.*, **122**, p. 1378 (1975).

[192] R. Bhat and S. K. Ghandhi, "Vapor-Phase Etching and Polishing of GaAs Using Arsenic Trichloride," *J. Electrochem. Soc.*, **124**, p. 1447 (1977).

[193] T. Uragaki, H. Yamanaka, and M. Inoue, "Selective Etching of GaP Crystals with Hot Phosphoric Acid," *J. Electrochem. Soc.*, **123**, p. 580 (1976).

[194] A. Milch, "Etch Polishing of GaP Single Crystals by Aqueous Solutions of Chlorine and Iodine," *J. Electrochem. Soc.*, **123**, p. 1256 (1976).

[195] L. Rankel Plauger, "Controlled Chemical Etching of GaP," *J. Electrochem. Soc.*, **121**, p. 455 (1974).

[196] W. G. Oldham, "Chemical Polishing of GaP," *Electrochem. Technol.*, **3**, p. 57 (1965).

[197] A. G. Sigai, C. J. Nuese, R. E. Enstrom, and T. Zameroski, "Vapor Growth of $In_{1-x}Ga_xP$ for P-N Junction Electroluminescence," *J. Electrochem. Soc.*, **120**, p. 947 (1973).

[198] R. H. Saul, "The Defect Structure of GaP Crystals Grown from Gallium Solutions, Vapor Phase and Liquid Phase Epitaxial Deposition," *J. Electrochem. Soc.*, **115**, p. 1185 (1968).

[199] M. I. Val'kevskaya and Y. S. Boyarskaya, "Revelation of Dislocations and Dislocation Structure, Emerging During the Deformation of Gallium Phosphide Single Crystals," *Sov. Phys.-Solid State*, **8**, p. 1976 (1967).

[200] E. Hájková and R. Fremunt, "Effective Chemical Polishing of n-Type GaP Surfaces," *Phys. Status Solidi*, **A10**, p. K35 (1972).

[201] See Ref. [50], p. 38.

[202] N. E. Schumaker, M. Kuhn, and R. A. Furnage, "Gallium Phosphide Beam Lead Electroluminescent Devices," *IEEE Trans. Electron Devices*, **ED-18**, p. 627 (1971).

[203] W. H. Hackett, Jr., T. E. McGahan, R. W. Dixon, and G. W. Kammlott, "A Scanning Electron Microscope Investigation of Etching Phenomena in GaP Electroluminescent Diodes," *J. Electrochem. Soc.*, **119**, p. 973 (1972).

[204] N. E. Schumaker and G. S. Rozgonyi, "Dicing Damage in GaP Electroluminescent Diodes," *J. Electrochem. Soc.*, **119**, p. 1233 (1972).

[205] B. D. Chase, D. B. Holt, and B. A. Unvala, "Jet Polishing of Semiconductors—I. Automatic Jet Thinning of GaP for Transmission Electron Microscopy," *J. Electrochem. Soc.*, **119**, p. 310 (1972).

[206] R. L. Meek and N. E. Schumaker, "Anodic Dissolution and Selective Etching of Gallium Phosphide," *J. Electrochem. Soc.*, **119**, p. 1148 (1972).

A Versatile High-Voltage Bias Supply for Extended Range MIS C(V) and G(V) Measurements*

Paul Kuczer, Harvey O. Hook, and Alvin M. Goodman

RCA Laboratories, Princeton, N.J. 08540

Abstract—Recently developed technology has enabled the measurement of MIS C(V) and G(V) at bias-voltage magnitudes as large as 25 kV. This report describes the design of a versatile high-voltage power supply intended for use as a bias source in carrying out such measurements.

The design allows the user a wide variety of options in the selection of the sweep function (waveform), sweep time, initial bias voltage, and the amplitude of the bias sweep. There are six possible sweep functions: (1) increasing ramp, (2) decreasing ramp, (3) positive polarity half-wave sawtooth (increasing ramp followed by decreasing ramp), (4) negative polarity half-wave sawtooth (decreasing ramp followed by increasing ramp), (5) full-wave sawtooth starting with increasing ramp, and (6) full-wave sawtooth starting with decreasing ramp. Either single or repetitive sweeps may be selected. The sweep time from the initial value to the end of the first ramp segment may be varied from 1 to 2000 s. Operator convenience is enhanced by certain features of the design; among these are light-emitting diodes which display the state of the sweep and automatic pen control for use of the sweep with an x-y recorder.

I. Introduction

A recently developed modification of the MIS C(V) measurement technique[1] has extended its useful range, allowing it to be used for the measurement of samples with insulating layers more than two orders

* This research was partially funded by the Advanced Research Projects Agency Order 2397 through the National Bureau of Standards' Semiconductor Technology Program Contract 5-35912 and is not subject to copyright. A version of this paper, containing mechanical details not included here, is available as NBS Special Publication 400-41.

of magnitude greater in thickness than was previously possible. This use, however, often requires the application of a very large bias voltage to the capacitor sample being measured. In the first high-voltage MIS C(V) measurements of Al-sapphire-Si capacitors,[2] the bias voltage was controlled manually and an attempt was made to keep the rate of change of bias as uniform as possible. It was clear, however, that the measured capacitance, C, depended to some extent upon the rate of change of the bias voltage, V, and that a constant (operator selected) sweep rate of the bias voltage would be required. It was also required that the bias supply be suitable for use with other equipment developed concurrently and described in References [3]–[5].

2. General Description of Equipment, Operation, and Performance

The high-voltage bias supply is constructed in the form of two modules: (1) a control module with digital logic,* a D/A (digital to analog) converter, and amplifiers to provide the required waveforms at low- and medium-voltage levels, and (2) a high-voltage amplifier module (HVAM).

2.1 Control Module

The control module allows the user a wide variety of options in the selection of the (1) sweep function (waveform), (2) sweep time, (3) initial bias voltage, and (4) amplitude of the bias sweep. The front panel of the control module (illustrated in Fig. 1) shows the location of all of the user-operated controls.

Operating power for the digital logic and low-voltage analog circuitry is controlled by the switch marked LOGIC† located above the POWER label. The function of the two adjacent switches will be discussed later.

Any one of six basic sweep functions may be selected by depressing (to its latched position) the appropriate pushbutton under the heading SWEEP FUNCTION. Each sweep function is composed of one or more "ramp" segments. The term ramp is understood here to mean a voltage which varies linearly with time.

The time, T, required for the sweep is controlled by selection of INITIAL TO PEAK SWEEP TIME; i.e., the time required for the sweep voltage to change from its initial value to the first maximum excursion from that initial value.

* Digital sweep generation was chosen to provide the very slow sweeps and the "hold" function described later. In addition, digital sweep generation provides excellent linearity and is easier to program than most analog sweep generators.

† Words corresponding to labels, callouts, or titles in the figures are capitalized in the text.

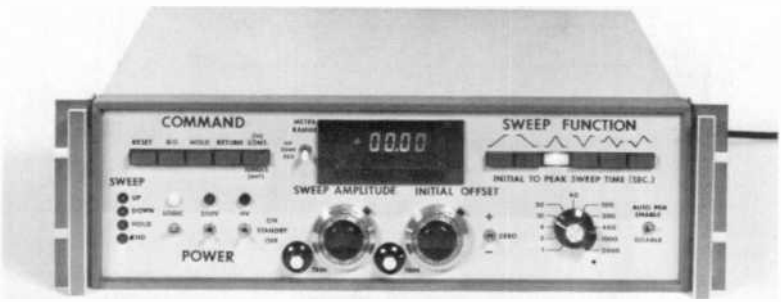

Fig. 1—Control module front panel.

If a single sweep is desired the CONT.-SINGLE pushbutton under the COMMAND heading should be in the released (out) position. If continuous operation is desired, this pushbutton should be in the depressed (in) position.

The magnitude of the starting value of the sweep is set by adjusting the INITIAL OFFSET control and the adjacent TRIM control. The polarity of the initial offset voltage is selected by setting the adjacent three-position toggle switch to the appropriate + or − position or to ZERO if no initial offset is desired.

The maximum one-way excursion of the sweep voltage from its initial value is set by adjusting the SWEEP AMPLITUDE control and the adjacent TRIM control.

Before initiating a sweep, the digital control circuits must be reset by momentarily depressing the RESET pushbutton. The sweep may then be initiated by depressing the GO pushbutton. The resulting sweep voltage as a function of time will be one of the waveforms shown in Fig. 2. If the sweep has been started in the continuous operation mode, the sweep will continue indefinitely until it is ended in one of two ways: (1) the continuous sweep may be ended by depressing the RESET pushbutton; this will bring the sweep voltage back to its initial value almost instantaneously, i.e., at slewing speed. (2) Alternatively, the continuous sweep may be ended by releasing the CONT.-SINGLE pushbutton to its out position. This causes a reversion to single sweep operation, which ends with the completion of the sweep function then in progress.

The sweep may be stopped at any time by pressing the pushbutton marked HOLD. The output voltage will then remain constant until the

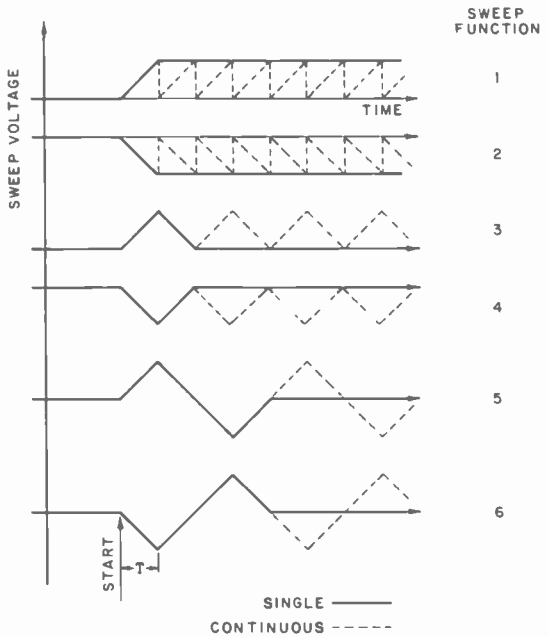

Fig. 2—Sweep function waveforms.

GO, RETURN or RESET pushbutton is pressed. If the GO pushbutton is pressed, the sweep will continue from the point at which it had been stopped. If the RETURN pushbutton is pressed the sweep voltage will return to the initial value by the shortest possible path (single ramp) at a sweep rate corresponding to the selected SWEEP AMPLITUDE and INITIAL TO PEAK SWEEP TIME. If the RESET pushbutton is pressed, the sweep voltage will return to its initial value almost instantaneously, i.e., at slewing speed.

The operating state of the sweep is indicated by four monitor lights (light-emitting diodes) located under the label SWEEP. The UP or DOWN light will be on if the sweep is on an up or down ramp, respectively. The HOLD light indicates that the sweep has been stopped by pressing the HOLD pushbutton. The END light indicates completion of a sweep. The absence of any light indicates that the RESET pushbutton has been pressed.

The sweep waveforms generated in the control module are available from connectors on the rear panel at three different voltage levels. The nominal low-voltage output of the control module has a maximum magnitude of about 10 V with respect to ground. The output is "antisymmetric" with respect to ground. That is, the voltages which appear at the two output terminals (BNC connectors labeled +10 V and −10

V) have equal magnitudes but are of opposite polarity with respect to ground. Thus, when one terminal is at +10 V, the other is at −10 V and the total output magnitude is 20 V. The voltage at each output terminal is also bipolar; i.e., it can have either polarity with respect to ground. When the +10 V terminal is positive with respect to ground, the output polarity is said to be positive or normal; if the +10 V terminal is negative with respect to ground, the output polarity is said to be negative or reversed. This total voltage is indicated on the digital voltmeter when the METER RANGE switch is in the 20-V position.

There are two additional outputs available from the control module. (1) A single-ended output proportional to the previously described bipolar output, but equal to one-half of its magnitude. This voltage appears at a single BNC connector labeled ANALOG OUTPUT and is intended for use with grounded instruments to monitor the bipolar output. (2) A medium level bipolar output equal to ten times the low-voltage bipolar output. This voltage appears at two terminals (UHF connectors) labeled +100 V and −100 V. The total voltage between these terminals is indicated on the digital voltmeter when the METER RANGE switch is in the 200-V position. Operating power for this medium level analog circuitry is controlled by the switch marked 200 V located above the POWER label.

A BNC connector (labeled PEN) located on the rear panel provides a connection to the contacts of a pen-control relay. When the AUTO PEN switch on the front panel is in the ENABLE position, the pen relay is energized during sweep operation; this provides a closed circuit at the PEN connector. If the AUTO PEN switch is in the DISABLE position the relay cannot be energized and the PEN control circuit remains open.

The operating power for the digital logic and all levels of analog circuitry is provided through the connector labeled AC POWER INPUT.

The operating power for the high-voltage amplifier module is provided through a connector labeled TO HV POWER INPUT and is controlled by the switch marked HV located above the POWER label. This connector also provides a circuit connection between the control and high-voltage modules that may be used for monitoring the voltage level (attenuated 1000:1) at the output of the HV amplifier module. This voltage may be read on the digital voltmeter of the control module when the METER RANGE switch is in the HV position.

2.2 High-Voltage Amplifier Module

The high-voltage amplifier module (HVAM) is designed to accept the nominal low-voltage bipolar output of the control module at its input, and to produce a high-voltage bipolar output at levels up to ~25 kV (~12.5 kV between each output terminal and ground). The input terminals are located on the rear panel and labeled +10 V DRIVE and −10 V DRIVE. The output terminals are located on the front panel and are labeled + and −.

The gain of the HVAM is controlled by a switch located on the front panel and labeled HIGH VOLTAGE RANGE (kV). There are five positions of this switch, ±1, ±2.5, ±5, ±10, ±25 each indicating the nominal maximum output voltage magnitude in kilovolts corresponding to that switch position with the maximum nominal input voltage. The corresponding amplifier gain figures are 50, 125, 250, 500, and 1250.

Operating power is supplied through a connector labeled POWER INPUT and located on the rear panel. The operating power is controlled by the switch on the control module labeled HV. This switch has three positions. The OFF position indicates that all power to the HVAM is shut off. The STANDBY position indicates that power is connected to all of the low-voltage control circuitry, the regulator tube heaters, and a time-delay relay, but not to the plate supplies for the regulator tubes. The regulator tube plate supplies are powered only when the time-delay (~8 s) relay has closed. High voltage can reach the output terminals only if the time-delay relay has closed, the HV switch is in the ON position, and the safety interlock circuit[4] is complete. The latter occurs when a low resistance is connected across the insulated BNC connector marked SAFETY INTERLOCK.

Three pilot lights on the front panel of the HVAM and one on the front panel of the control module monitor the status of operating power for the HVAM. When the HV switch is in the OFF position, all pilot lights should be off. When the HV switch is in the STANDBY position, the pilot light above the HV switch and the one marked CONTROL POWER should be on. The pilot light marked STANDBY will be on only after the time-delay relay has closed. When the HV switch is in the ON position, the OUTPUT pilot light on the HVAM will be on if the safety interlock circuit is complete. Note that it is possible for the OUTPUT light to be on when the STANDBY light is off during the ~8 s waiting period of the time-delay relay; there will be no voltage at the output terminals during this time.

3. Circuit Details of Control Module

A functional block diagram of both modules of the high-voltage bias supply is shown in Fig. 3. This section deals with the circuits comprising the control module portion of that figure. These circuits are described in five schematic diagrams, one for each of the blocks in the digital portion (Figs. 4 to 7) and a fifth (Fig. 8) for the three blocks representing the analog portion. The circuit elements shown in these figures are described in Table 1.

In brief, a crystal-oscillator-derived square wave (clock) is fed to an up/down counter producing a binary number that varies linearly with time. A digital-to-analog (D/A) converter uses this number to produce a ramp voltage that varies linearly with time. The details of these circuits, the digital control circuits and the analog circuits necessary to produce the desired output levels are described in the remainder of this section.

3.1 Sweep Time Selection

The initial-to-peak sweep time, T, is the time necessary for a CLOCK output of the SWEEP TIME SELECTION circuit (see Fig. 4) to accumulate 2048^{+0}_{-1} counts on the UP/DOWN COUNTER to be described in Sec. 3.2. The functions of the SWEEP TIME SELECTION circuit are to produce a stable high-frequency clock signal and to modify it using "divide by n" circuits to produce the appropriate CLOCK output.

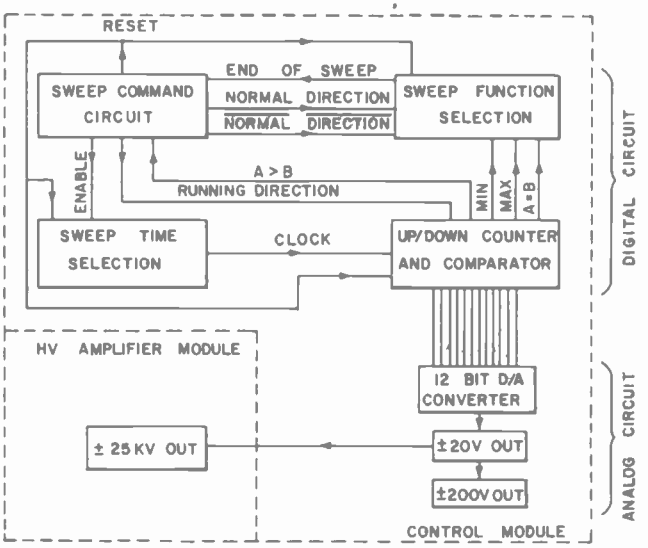

Fig. 3—Functional block diagram of high-voltage bias supply.

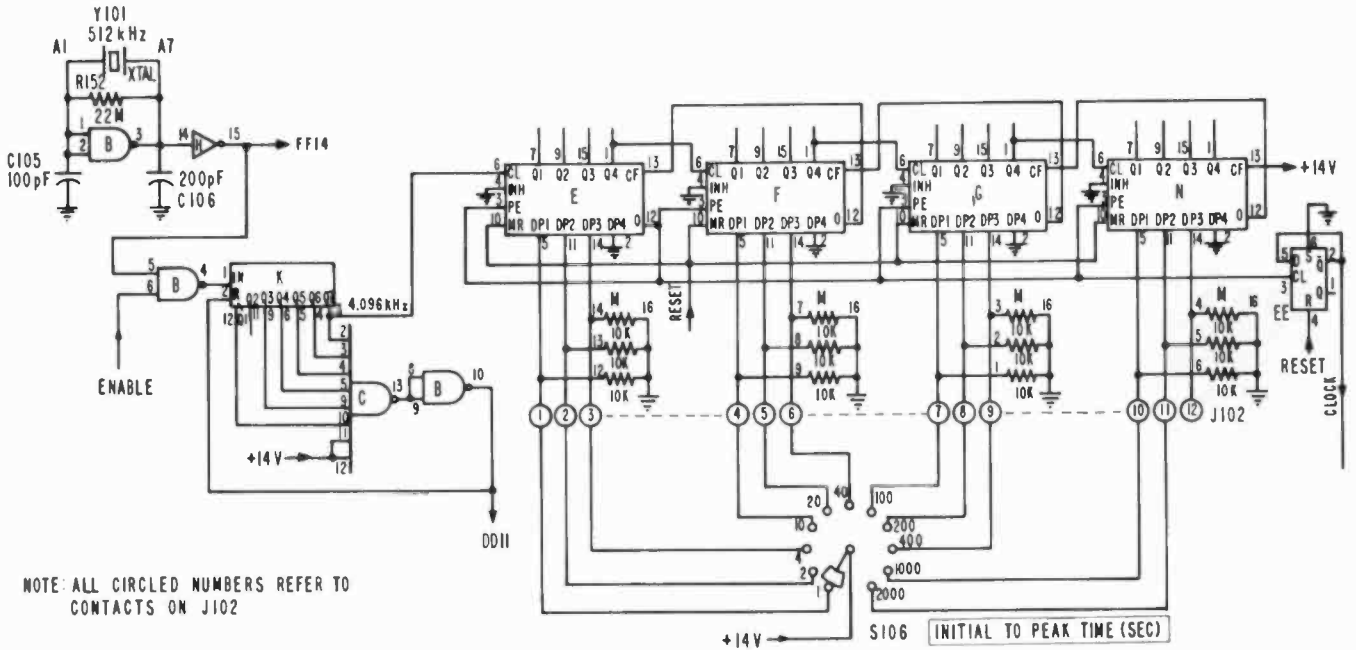

Fig. 4—Sweep time selection circuit.

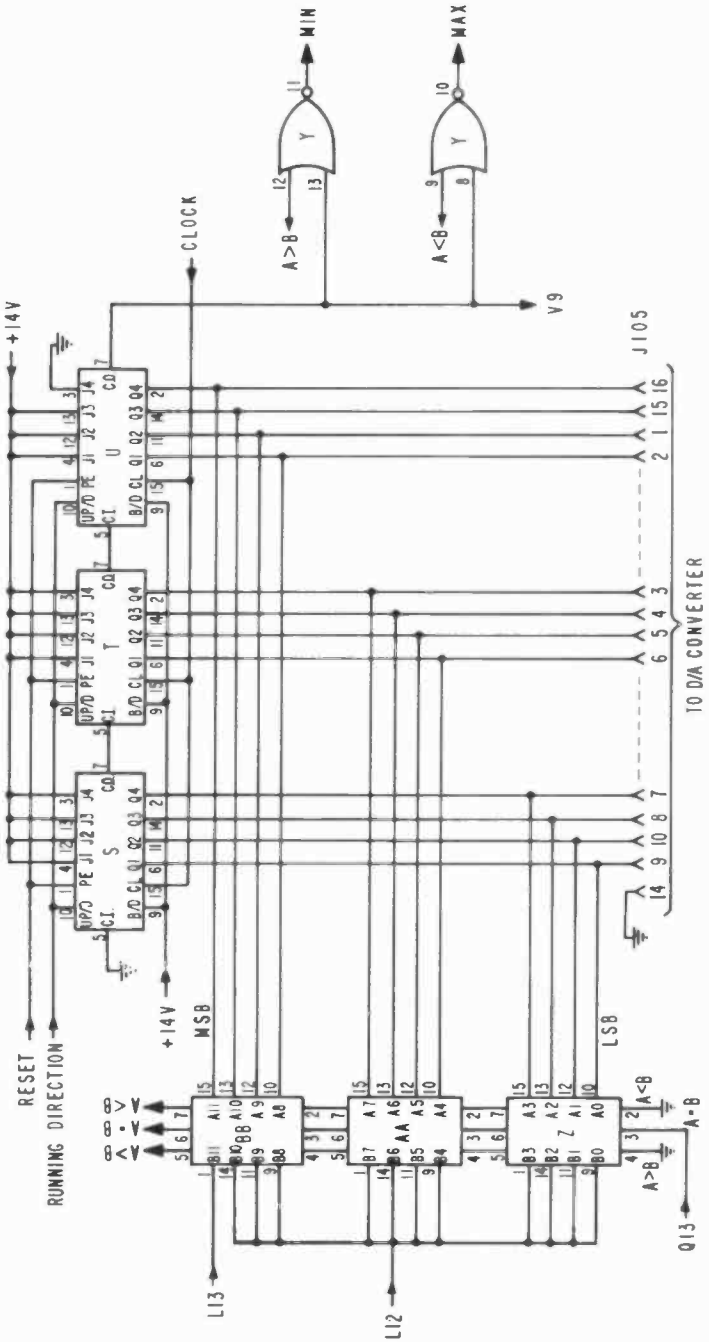

Fig. 5—Up/down counter and comparator circuit.

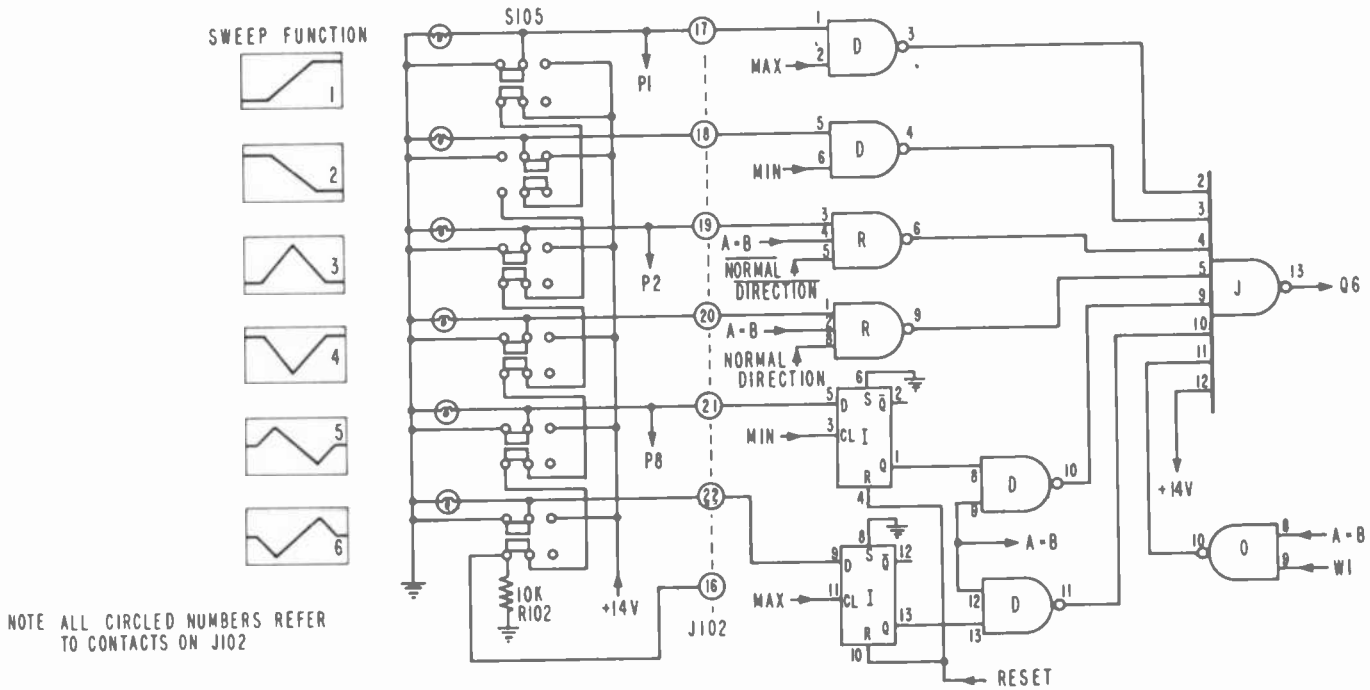

Fig. 6—Sweep function selection circuit.

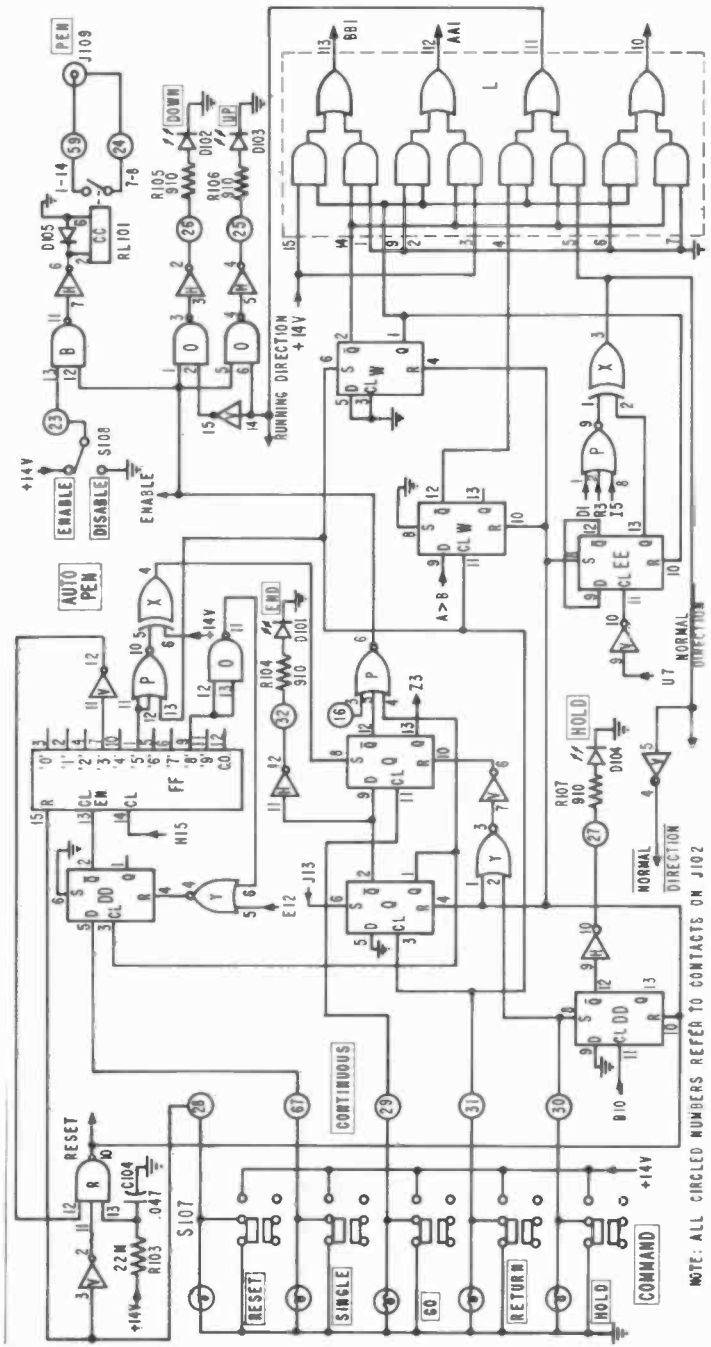

Fig. 7—Sweep command circuit.

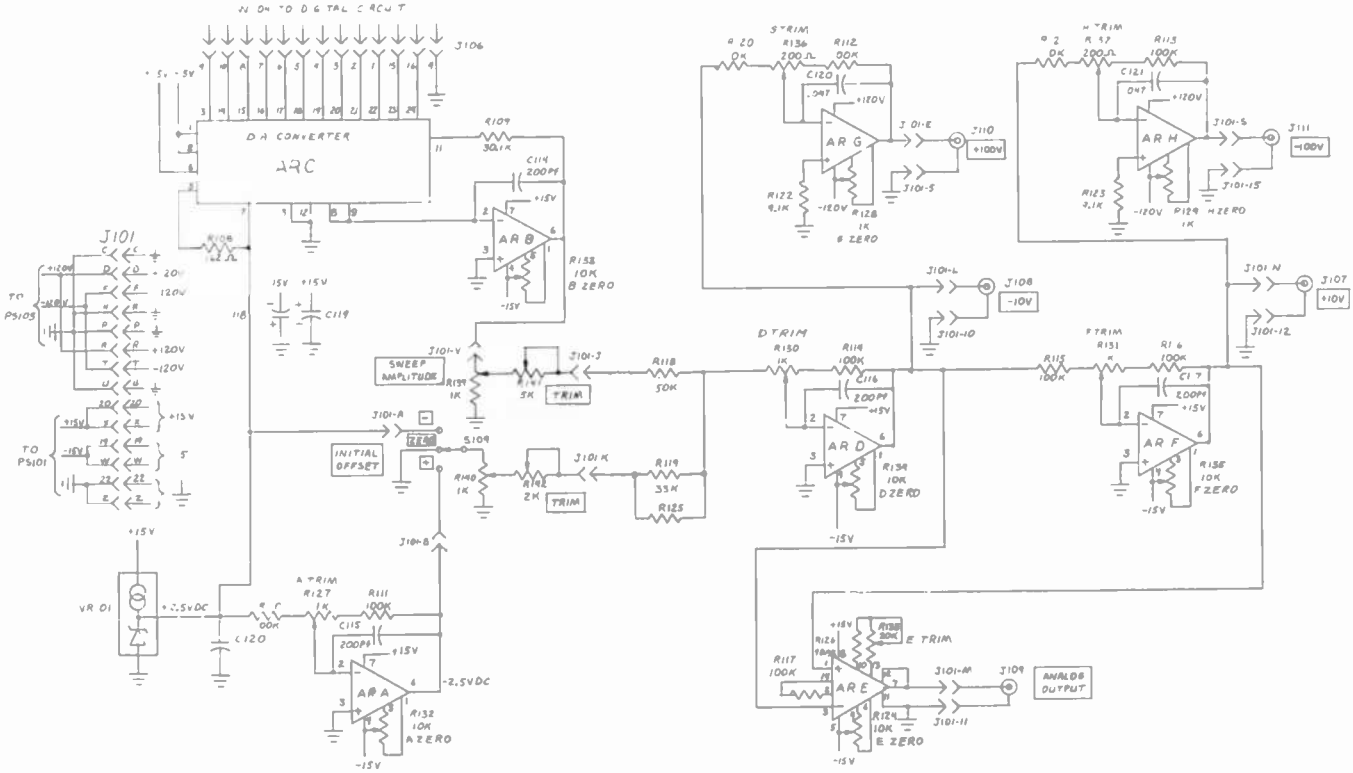

Fig. 8—Analog circuit bias supply control module.

Table 1—Electrical Parts List for Bias Supply Control Module

Schematic Reference	Description (Commercial description,* if appropriate, shown in parentheses)
AR A, AR B, AR D, AR F	General purpose FET operational amplifier. (Teledyne Philbrick, Type 1421).
AR C	Digital to analog converter (Analog Devices, Type AD562 KD/BIN)
AR E	Precision instrumentation amplifier (Analog Devices Type AD 521)
AR G, AR H	High-voltage FET operational amplifier (Teledyne Philbrick, Type 1022).
B	CD4011AE
C	CD4068AE
D	CD4011AE
E	MC14522
F	MC14522
G	MC14522
H	CD4049AE
I	CD4013AE
J	CD4068AE
K	CD4024AE
L	CD4019AE
M	Dual-In-line, 10 kΩ resistor network [contains 15 resistors; 12 used in circuit, figure 7].
N	MC14522
O	CD4011AE
P	CD4025AE
Q	CD4013AE
R	CD4023AE
S	CD4029AE
T	CD4029AE
U	CD4029AE
V	CD4049AE
W	CD4013AE
X	CD4030AE
Y	CD4001AE
Z	CD4063AE
AA	CD4063AE
BB	CD4063AE
CC	12 V Relay SPST [RL101]
DD	CD4013AE
EE	CD4013AE
FF	CD4017AE
C101	0.22 μF, 50 V, mylar capacitor
C102, C103	0.1 μF, 50 V, mylar capacitor
C104	0.047 μF, 25 V, ceramic capacitor
C105	100 pF, silvered mica capacitor
C106	200 pF, silvered mica capacitor
R120, R121	10 kΩ, ½ W
R122, R123	9.1 kΩ, ½ W
R125	Optional. Used to calibrate initial offset dial.
R126	90.9 kΩ
R127, R128, R129, R130, R131	1 kΩ, ¼ W, trim-pot.
R124, R132, R134, R135	10 kΩ, ¼ W, trim-pot.
R136, R137	200 Ω, ¼ W, trim-pot.
R138	20 kΩ, ¼ W, trim-pot.

* Certain commercial equipment, instruments, or materials are identified in this report in order to adequately specify the experimental procedure. In no case does such identification imply recommendation or endorsement by the National Bureau of Standards, nor does it imply that the material or equipment identified is necessarily the best available for the purpose.

R133	10 kΩ, 2 W, 15-turn, potentiometer
R139, R140	1 kΩ, 2 W, 10-turn, potentiometer
R141	5 kΩ, 2 W, 10-turn, potentiometer
R142	2 kΩ, 2 W, 10-turn, potentiometer
R152	22 MΩ, ¼ W, 5%
RL101	Dual-in-line reed relay, contact rating 10 VA (MRR1AD 12 VDC Struthers-Dunn, Inc.) Inc.)
S101, S102	Double-pole, double-throw, 3-A @ 125 VAC toggle switch
S103, S104, S109	4 P 3-pos. toggle switch (Alco MTA406A) [S104 wired to form 2-pole, 3-pos. switch].
S105	6-section, reciprocal release, pushbutton switch. (LEDEX Series 1400 pushbutton switch) [The contacts operate when the pushbutton is depressed, contacts remaining operated and the button depressed until any other button is pressed, which causes release. The first button resets before the second button has been pressed far enough to operate its own contacts.]
S106	1-pole, 12-pos. shorting type rotary switch
S107	5-section pushbutton switch. (LEDEX Series 1400 pushbutton switch.) [4-sections: momentary interlock, normal pushbutton action. Contacts operate only for the time the pushbutton is depressed. Interlock is incorporated so that only one pushbutton can be depressed at a time. 1-section: push-push. Two-step action. When this button is pressed once, the contacts operate and the button remains in. When the button is pressed a second time, the contacts and the button reset. No interlock.]
S108	SPDT toggle switch
VR 101	Constant voltage source (Analog Devices, Type AD580).
C107, C108, C109, C110, C111, C112, C113	1.5 μF, 50 V, ceramic capacitor
C114, C115, C116, C117	200 pF, silvered mica capacitor
C118, C119	22 μF, 15 V, tantalum capacitor
C120, C121	.047 μF, 200 V mylar capacitor
D101, D102, D103, D104	Light-Emitting Diode
D105	IN914
DS101	Panel lamp, neon, clear (Leecraft type 44N-0117)
DS102	Panel lamp, neon, amber (Leecraft type 44N-0113)
DS103	Panel lamp, neon, red (Leecraft type 44N-0111)
F101, F102	¼ A. SB, 3 AG Type
F103	2A, SB, 3 AG Type
J101	22/44 PCB connector 0.156 contact centers (Amphenol 225-22221-101)
J102	35/70 PCB connector 0.100 contact centers (Stanford Applied Engineering CPH 7000-70)
J103	"Jones" plug (Cinch P-308-AB)
J104	15/30 PCB connector 0.156 contact centers (Amphenol 225-21521-101)
J105, J106	16-pin dual-in-line socket
J107, J108, J109	Insulated BNC chassis connector

J110, J111	UHF chassis connector
J112, J113	"Grabber" clip (Pomona Electronics Type 3925)
M101	4½ digit panel meter (Ballantine 8310-05)
P101	16-pin dual-in-line plug
P102, P103	8-pin cable male connector, "Jones" type (Cinch Type P308-CCT)
PS101	Power supply, regulated with 14 V @ 0.5A (Power One Type HA15-5)
PS102	Power supply, regulated with +15 V and −15 V outputs @ 10 mA (Acopian type D15-10A)
PS103	Power supply, regulated with +120 V and −120 V outputs @ 40 mA (Teledyne-Philbrick type 2217)
R101, R102	10 kΩ, 5%, ½ W
R103	22 MΩ, 5%, ¼ W
R104, R105, R106, R107	910 Ω, 5%, ½ W
R108	162 Ω, 1%, ½ W
R109	30.1 kΩ, 1%, ½ W
R110, R111, R112, R113, R114, R115, R116, R117	100 kΩ, 1% ½ W
R118	50 kΩ, ½ W
R119	33 kΩ, ½ W
W101	2 ft. long, 8-conductor, shielded cable. (Shield is used as one of the conductors.)
W102, W103	RG58 coax cable, 2 ft. long, male BNC connector on both ends.
W104	16-pin, ribbon cable, male plugs on both ends.
W105	Adapter cable for use with DVM (M101) as a calibration instrument. This cable can also be used to check the DVM for accuracy. Cable is RG174 coax, 3 ft. long.
Y101	Crystal, specifications: AT cut, 32 pf load capacity, parallel resonant, 512.000 kHz.
Enclosure	(Optima Type K051716R, beige w/bronze trim, Scientific Atlanta, Inc.)
Digital circuit board	(Cambion Type 715-1115-01, Cambridge Thermionic Corporation)

The heart of the circuit is a crystal-controlled oscillator which can be seen as gate† B3 in Fig. 4. The output is buffered by an inverter (H15) which, in turn, feeds a "divide-by-125" circuit through a NAND gate (B4). The "divide by 125" circuit employs a 7-stage binary counter (K) and two NAND gates (C13 and B10). Gate C13 detects the count of 125 (all inputs "high" ††) and resets the counter through inverter B10. The 4.096-kHz clock signal is obtained from pin K3 because the signal there (MSB output) is most nearly symmetrical. Note that the gate B4 will pass the oscillator output signal only when the level at pin B6 (marked ENABLE) is "high"; a "low" ENABLE level will terminate the sweep.

The 4.096-kHz clock signal is further modified by four programmable stages of frequency division (E, F, G, and N) and a "divide by 2" flip-flop

† Gates and inverters will be designated by their output terminals.
†† Positive logic is assumed.

(EE) to produce the desired CLOCK output at pin EE2. Switch S106 is used to program, E, F, G, and N, i.e., to select the frequency division factor necessary to produce the desired value of T.

3.2 Up/Down Counter and Comparator

In this circuit an up/down counter is used to convert the CLOCK output to a binary number which will be the input for the D/A converter to be described in Sec. 3.5. A binary number comparator is used to determine the points at which the counter reaches "critical numbers," i.e., numbers corresponding to sweep maxima, sweep minima, or end of sweep.

The UP/DOWN COUNTER AND COMPARATOR circuit is shown in Fig. 5. Three stages of presettable 4-bit up/down counters (S, T, and U) are used to generate a 12-bit binary number. The starting point for all sweeps is the number $2047_{(10)}$. This is illustrated in Fig. 9 which shows the counter number sequences for the six sweep functions in the single sweep mode. In the continuous mode of operation, the selected sweep function is automatically and continuously repeated.

Three 4-bit binary comparators (Z, AA, and BB) are cascaded to form a 12-bit comparator which monitors the magnitude of the counter output (A inputs) relative to a reference number (B inputs). The truth table for the comparator is shown in Table 2.

Under certain conditions, operation of the sweep function switches can produce a false $A = B$ signal. To prevent this from causing a false

SWEEP FUNCTION	SEQUENCE
1	A TO B, STOP
2	C TO D, STOP
3	A TO B TO C, STOP
4	C TO D TO E, STOP
5	A TO B TO C TO D TO E, STOP
6	C TO D TO E TO F TO G, STOP

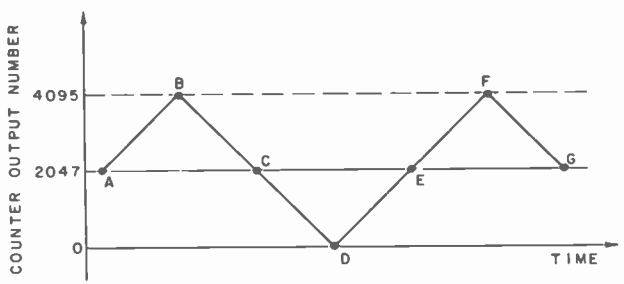

Fig. 9—Counter number sequences for the six sweep functions (single sweep).

Table 2—Truth Table for UP/DOWN Comparator.

CONDITION	Output Line and Location		
	"A > B" pin 5 of BB	"A = B" pin 6 of BB	"A < B" pin 7 of BB
A > B	1	0	0
A = B	0	1	0
A < B	0	0	1

end of sweep condition, the $A = B$ signal input to comparator Z is inhibited except during sweep by keeping Q13 "high" during sweep and "low" at all other times.

Two additional gates are used to monitor the state of the 12-bit output. NOR gate Y11 detects a minimum count of zero (all 12 outputs "low") in the following manner. The carry-out signal at pin 7 of counter U is normally "high" and goes "low" only when the counter reaches its maximum count in the up mode, or its minimum count in the down mode. Since a low at the carry-out terminal indicates *either* a minimum or a maximum count, additional information is required to resolve the ambiguity. A "low" at the $A > B$ output indicates that the counter is below the starting count of $2047_{(10)}$. A "high" output at gate Y11 (labeled MIN) indicates that both of its inputs are "low" and that the counter output is at a minimum. A maximum count of $4095_{(10)}$ (all twelve outputs "high") is detected in a similar manner by gate Y10. A "low" at the $A < B$ output indicates that the counter is above the starting count of $2047_{(10)}$. A "high" output at gate Y10 (labeled MAX) indicates that both of its inputs are low and that the counter is at a maximum.

3.3 Sweep Function Selection

The selection of the SWEEP FUNCTION is made by depressing one of the pushbuttons of S105 (see Fig. 6). A mechanical interlock prevents the depression of more than one pushbutton at a time. Fig 6 also shows the logic circuitry that determines when a single sweep is terminated (assuming that it is *not* terminated by pressing the RESET button). The termination of a single sweep will occur when any of the 8 inputs to NAND gate J13 becomes "low." The required conditions for termination of each of the sweep functions is shown in Table 3. They may be checked by referring to Fig. 6. The first six entries in the table are termination conditions for single sweeps during which neither the RETURN nor RESET buttons have been depressed. The 7th entry describes the termination conditions after the RETURN button has been pushed.

Table 3—Conditions Required for the Termination of a Single Sweep

Sweep Function	Required Conditions
1	(a) S105-1 depressed (b) MAX level "high"
2	(a) S105-2 depressed (b) MIN level "high"
3	(a) S105-3 depressed (b) $\overline{\text{NORMAL DIRECTION}}$ level "high" (c) A = B level "high"
4	(a) S105-4 depressed (b) NORMAL DIRECTION level "high" (c) A = B level "high"
5	(a) S105-5 depressed (b) MIN level "high" (c) A = B level "high"
6	(a) S105-6 depressed (b) MAX level "high" (c) A = B level "high"
Any	(a) RETURN button previously depressed (b) A = B level "high"

3.4 Sweep Command Circuit

The SWEEP COMMAND CIRCUIT in Figure 7 comprises five subscripts which can be considered according to the purpose served by each: (a) COMMAND pushbutton switches and associated circuitry, (b) DIRECTION circuit (c) SWEEP STATUS indicators, (d) AUTO PEN control circuit, and (e) CONTINUOUS restart circuit.

(a) Command Pushbutton Switches and Associated Circuitry

To initiate a sweep, all inputs to gate P6 must be "low"; this will produce a "high" on the ENABLE line. A "low" at input pin 3 of gate P6 indicates that one of the SWEEP FUNCTION pushbuttons is depressed. When the RESET button is depressed, the flip-flop Q1 is reset producing a "low" at input pin 4 and a "high" at input pin 5 of gate P6. This "high" at input pin 5 is changed to a "low" by depressing the GO pushbutton.

The HOLD pushbutton stops the sweep by resetting Q13, thereby changing the previous "low" at input pin 5 of P6 to a "high."

The RETURN pushbutton unconditionally returns the sweep to its starting point at its previous rate. Here, flip-flop Q1 is "clocked" to its reset state and flip-flop Q13 is set through gates P10 and X4.

The RESET command may be generated in three ways:

(1) When the LOGIC POWER switch is initially turned on, the charging time constant of (R103) × (C104) produces a low at input pin 3 of gate R10 for a time sufficient to insure reset.
(2) From the continuous restart circuit gate V12.
(3) From the RESET pushbutton switch.

(*b*) *Direction Circuit*

The direction of the UP/DOWN COUNTER is determined by the level of the RUNNING DIRECTION line; a "high" corresponds to UP, a "low" corresponds to DOWN. When the RESET pushbutton is depressed, RUNNING DIRECTION control is transferred to gate X3 by flip-flop W1. At this time the output W1 is "high" and the output X3 is gated to the RUNNING DIRECTION line through AND/OR-select gate L11. The RUNNING DIRECTION at the start of a sweep is controlled by the gate P9; for SWEEP FUNCTIONS 1, 3, and 5 it is UP; for 2, 4, and 6 it is DOWN. When the maximum or minimum count is reached, the RUNNING DIRECTION is reversed by flip-flop EE13 and exclusive-OR gate X3. Pin 2 of gate X3 determines whether the level at pin 1 becomes inverted at the output or not; i.e., a "high" level at pin 2 causes inversion, a "low" level does not. During a sweep in which the RETURN pushbutton is *not* depressed, the NORMAL DIRECTION and RUNNING DIRECTION are the same.

If the RETURN pushbutton is depressed, the NORMAL DIRECTION level is not altered, but the RUNNING DIRECTION level *may* be altered through flip-flop W13, depending upon the levels of the $A > B$ and NORMAL DIRECTION lines. If the levels of $A > B$ and NORMAL DIRECTION are *both* "high" or *both* "low," the RUNNING DIRECTION level will be changed; if the $A > B$ level is "high" and the NORMAL DIRECTION level is "low" (or vice versa), the RUNNING DIRECTION will not be changed.

It should be noted that although pressing the RESET pushbutton sets the COMPARATOR reference number to $2047_{(10)}$, pressing the RETURN pushbutton changes the COMPARATOR reference number to 2048 by setting flip-flop W1, which, in turn, changes the output levels of the AND-OR gates L12 and L13. This is done for the following reason. In the RESET condition, $A = B$ and the $A > B$ level is "low." This would cause the COUNTER to immediately start to count up if the RETURN pushbutton were pressed after RESET. In this case, the SWEEP is forced to end after one count by the change of reference number just described. This results in an offset of the output voltage at any END of SWEEP caused by pressing the RETURN pushbutton. The amount of this offset (1 part in 2047) is undetectable in normal use.

(*c*) *Sweep Status Indicators*

Four LEDs (light-emitting diodes) are used to indicate the operating state of the SWEEP. Only one LED can be on at a given time during a normal sequence of COMMANDs. When the ENABLE level from gate P6 is "high," either the UP or DOWN LED will be on, depending upon whether the RUNNING DIRECTION level is "high" or "low." De-

pressing the HOLD pushbutton sets flip-flip DD13, which turns on the HOLD LED through inverter H10. If the sweep is resumed by pressing the GO pushbutton, flip-flop DD13 is clocked to its "reset" state. The END LED turns on at termination of the SWEEP when flip-flop Q1 is set by a "high" level at J13. When the RESET pushbutton is depressed, any previously turned-on LED is turned off. Thus, if all LEDs are off, it signifies that RESET has taken place.

(*d*) *Pen Control Circuit*

Both inputs to NAND gate B11 must be "high" to energize the PEN relay RL101. Thus, the PEN control circuit will be closed only when the AUTO PEN switch is in the ENABLE position *and* the ENABLE level is "high."

(*e*) *Continuous Restart Circuit*

A decade counter, FF, is used to generate (1) a RESET pulse via inverter V12 and NAND gate R10; (2) a "start" pulse via NOR gate P10 and exclusive-OR gate X4; and (3) a "self-turnoff pulse: via NAND gate O11 and NOR gate Y4 in the following manner. A "high" from NAND gate J13 signifies the end of a sweep and sets flip-flop Q1. A "high" from pin Q1 clocks flip-flop DD1 to its "set" state when the SINGLE-CONTINUOUS switch is in the CONTINUOUS position. This causes a "low" level at the clock enable line FF13 of the counter* FF, allowing it to start counting 512 kHz pulses from the oscillator buffer-inverter H15. First, the RESET pulse is generated from gate V12; next, the "start" pulse from X4 sets Q13 (this is equivalent to a GO command), and finally, the counter is "disabled" by the "self-turnoff pulse" from Y4, which resets flip-flop DD1, bringing the clock enable line of the counter "high" preventing further counts. This sequence is repeated continuously if the SINGLE-CONTINUOUS switch remains in the CONTINUOUS position.

If, at the end of the sweep, this switch is in the SINGLE position, flip-flop DD1 is clocked to its "reset" state, in which case, the counter-clock enable level is "high," and counting is inhibited.

* For this counter, the "clock enable level" must be "low" to enable counting; a "high" level disables or inhibits counting.

3.5 Analog Circuit

The digital circuit produces a 12-bit binary number which varies linearly with time. The heart of the analog circuit is a digital to analog (D/A) converter which uses this time-varying binary number to produce a voltage which varies linearly with time. A series of dc amplifiers is used to make this voltage bipolar and to obtain the desired amplitude and offset levels.

Current from the D/A converter AR C (Fig. 8) develops a voltage across an internal load resistor. This voltage is passed through amplifier AR B to provide a basic single-ended sweep signal. A portion of this is selected by adjusting the SWEEP AMPLITUDE and TRIM potentiometers and then summed with an INITIAL OFFSET voltage through amplifier AR D to produce the −10 V output. This output is inverted by amplifier AR F to produce the +10 V output. Amplifiers AR G and AR H are used to raise these levels to provide the +100 V and −100 V outputs, respectively. The voltage regulator VR101 provides a stable 2.5-V (nominal) level which is used as a reference for the D/A converter and a source of INITIAL OFFSET voltage. An INITIAL OFFSET voltage of opposite polarity is obtained by using an inverter, AR A. The instrumentation amplifier AR E is used to provide a single-ended output labeled ANALOG OUTPUT which is equal to one-half the difference between that at the +10 V and −10 V output terminals. It is intended for monitoring the sweep generator output with an instrument having one input terminal grounded.

3.6 Power Supplies and Metering

The schematic circuit diagram of the POWER SUPPLIES AND METERING for the control module is shown in Fig. 10.

4. Circuit Details of High-Voltage Amplifier Module

A schematic diagram of the power supplies and switching for the HVAM is shown in Fig. 11. The high-voltage amplifier consists of two substantially identical amplifiers, one of which is driven in the positive direction while the other is driven in the negative direction, and vice versa. The two amplifiers are represented schematically in the right and left halves of Fig. 12. The electrical components shown in Figs. 11 and 12 are described in Table 4. The circuit description which follows refers only to the right side of Fig. 12, the amplifier which produces the "+" output. Overall feedback from the output to the input of the driver section linearizes the transfer function and "regulates out" power supply variations.

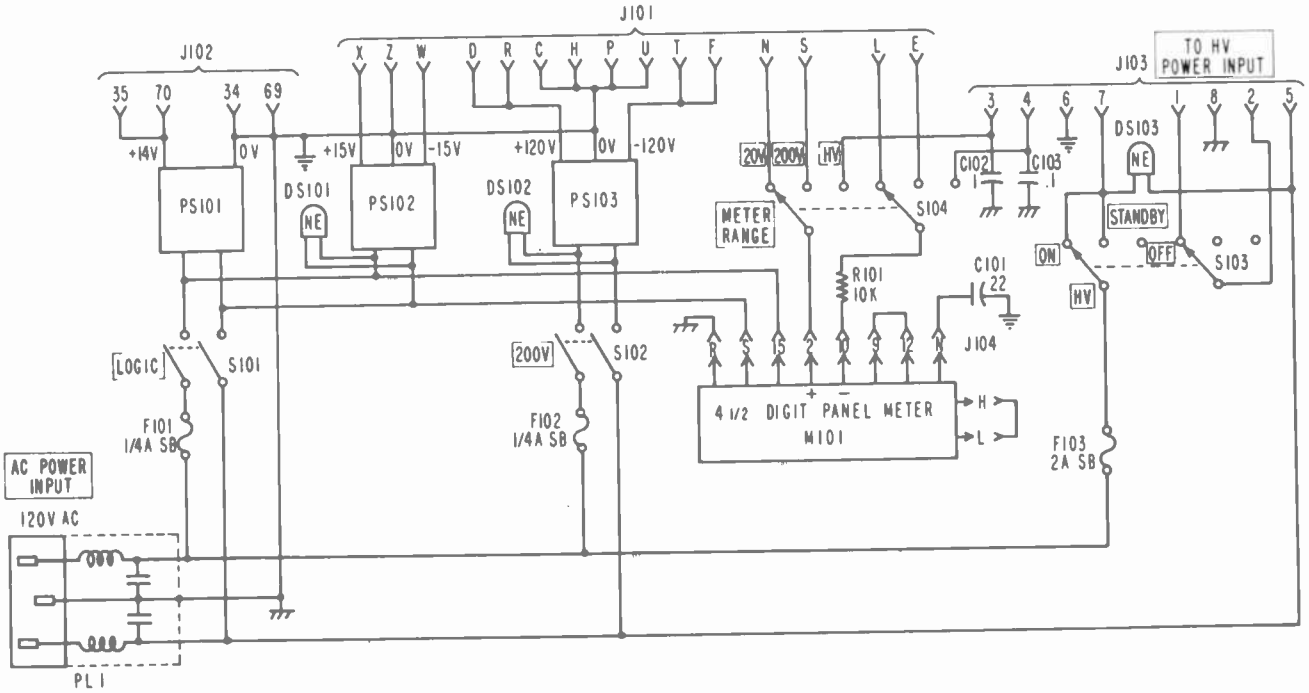

Fig. 10—Power supplies and metering for bias supply control module.

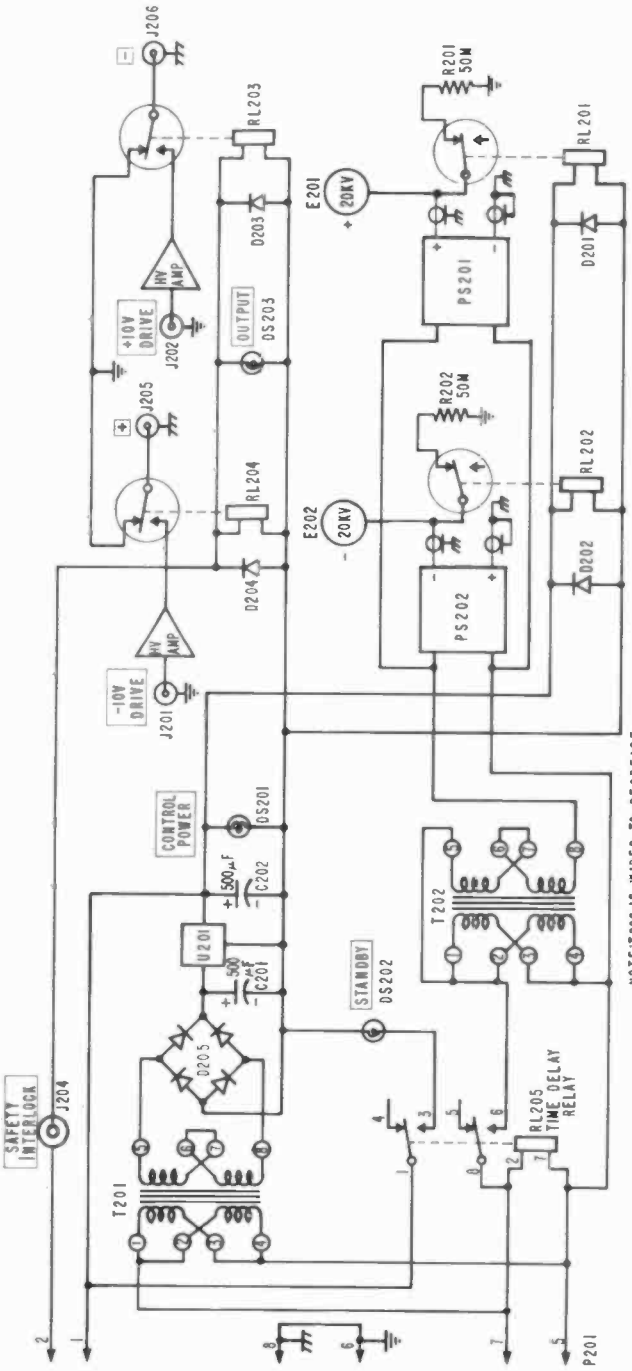

Fig. 11—Power supplies and switching for high-voltage amplifier module.

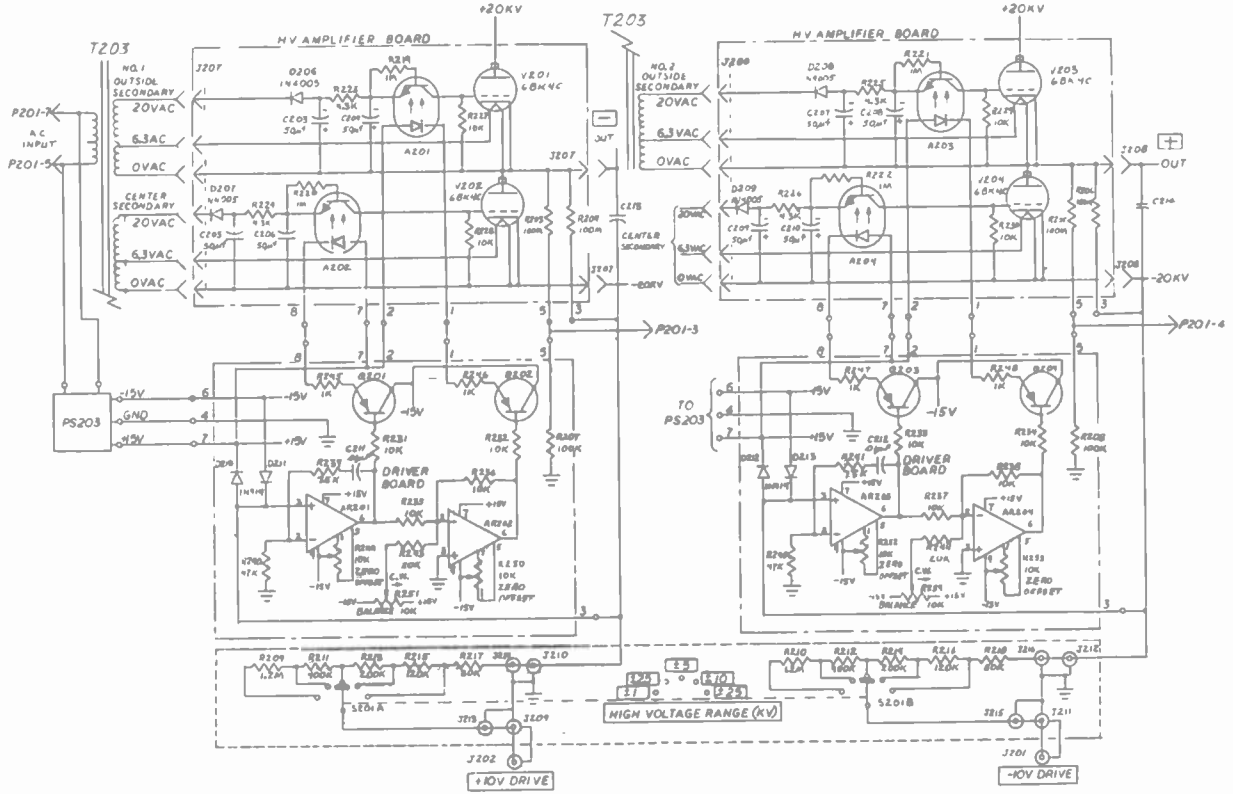

Fig. 12—High-voltage amplifier schematic.

Table 4—Electrical Parts Lists for HV Amplifier Module

Schematic Reference	Description (Commercial description,* if appropriate, shown in parentheses)
A201, A202, A203, A204	50 kV optical coupler (Optron, Type OPI-150)
AR201, AR202, AR203, AR204	General purpose FET operational amplifier (Teledyne Philbrick, type 1421).
C201, C202	500 μF, 50 VDC, electrolytic capacitor
C203, C204, C205, C206 C207, C208, C209, C210	50 μF, 50 VDC electrolytic capacitor
C211, C212	0.01 μF, 50 V, ceramic capacitor
C213, C214 (Gimmick capacitor)	¾" length of copper braid around output cable covered with shrinkable tubing.
D201, D202, D203, D204	IN4005 diode
D206, D207, D208, D209	IN4005 diode
D205	18DB6A bridge rectifer
DS201, DS202, DS203	24 V indicator lights
E201, E202	Corona ring, high-voltage terminal
J201, J202, J204	Insulated BNC coaxial chassis connector (Amphenol type 31-010)
J203	8-pin cable female connector, "Jones" type (Cinch type S-308-CCT)
J205, J206	High-voltage coaxial chassis connector (Amphenol type 97-3102A-18-420S)
J207, J208	HV connector assembly
J209, J210, J211, J212	Sub-minax coax cable connector (Amphenol type 27-7)
J213, J214, J215, J216	Sub-minax receptacle (Amphenol type 27-9).
P201	8-pin chassis male connector, "Jones" type (Cinch type P-308-AB)
PS201, PS202	Power supply, 20 kV, dc, @ 1 mA (Del, type 20-1-4)
PS203	Power supply, regulated with +15 V and −15 V outputs @ 100 mA (Acopian type D15-10A).
Q201, Q202, Q203, Q204	2N4036 transistor.
R201, R202	50 MΩ, 5 W, 5%.
R203, R204, R205, R206	100 MΩ, 15 W, 1%.
R209, R210	1.2 MΩ, ¼ W, 1%.
R211, R212	400 kΩ, ¼ W, 1%.
R213, R214	200 kΩ, ¼ W, 1%.
R215, R216	120 kΩ, ¼ W, 1%.
R217, R218	80 kΩ, ¼ W, 1%.
R219, R220, R221, R222	1 MΩ, ½ W, 5%.
R223, R224, R225, R226	4.3 kΩ, ½ W, 5%.
R227, R228, R229, R230	10 kΩ, ½ W, 5%.
R231, R232, R233, R234	10 kΩ, ½ W, 5%.
R235, R236, R237, R238	10 kΩ, ½ W, 5%.
R239, R241	7.5 kΩ, ½ W, 5%.
R240, R242	47 kΩ, ½ W, 5%.
R243, R244	20 kΩ, ½ W, 5%.
R245, R246, R247, R248	1 kΩ, ½ W, 5%.
R249, R250, R251, R252, R253, R254	10 kΩ, ¼ W, 15-turn trim-pot
RL201, RL202, RL203, RL204	High-voltage relay (ITT Jennings, Type RE6B-26-N300).
RL205	Time-delay relay (Guardian TD062C30-115A).
S201	2-pole, 5-position, shorting type rotary switch
T201, T202	Transformer (Stancor, Type P6377).
T203	High-voltage isolation transformer. (NWL Model 24543).
U201	24 V dc regulator.
V201, V202, V203, V204	Vacuum tube 6BK4C/6EL4A.

* Certain commercial equipment, instruments, or materials are identified in this report in order to adequately specify the experimental procedure. In no case does such identification imply recommendation or endorsement by the National Bureau of Standards, nor does it imply that the material or equipment identified is necessarily the best available for this purpose.

4.1 General Specifications

Each amplifier, when driven by a signal in the range −10 to +10 V, produces a signal proportional to the negative of the input signal. The proportionality constant (gain) is selectable such that −10 V input will produce outputs of 0.5, 1.25, 2.5, 5.0, and 12.5 kV. The circuit is such that +10 V input produces the negatives of the voltages indicated above. The amplifiers can drive a 50-MΩ external load. The anticipated voltage sweep rates are low so that capacitive loading of the output is not a problem.

4.2 Output Stage

The output stage is the most challenging design because it must be constructed within the capabilities of available tubes and, in addition, controls what the drivers must do. Therefore, the output stage is the starting point of the description. The 6BK4C tube is rated to withstand 27 kV plate to cathode and has maximum power and current ratings of 40 W and 1.6 mA, respectively. With the effective load resistance of 25 MΩ (50-MΩ external load in parallel with 50-MΩ internal load) the current limitation is more restrictive than the power limit. The 6BK4C is intended as a shunt regulator tube so it is not designed to have a low voltage drop at zero grid bias. With a 20-kV power supply and 25-MΩ load the zero-bias plate voltage is about 5 kV. With a ±20 kV power supply, the swing is thus limited to ±15 kV. With the usual 20% tolerance on tube specifications the actual limits may be as low as ±13 kV. When one tube is zero-biased, the other one is cut off and has approximately 35 kV across it. The rated voltage is then exceeded substantially at the limits of the swing. This undesirable condition is accepted because no higher-rated tube is available; there is no other choice. However, the rating is based on leakage developing due to glass electrolysis when the voltage is maintained constant as it is in shunt regulator service. In the present service, the time the tube spends under maximum voltage stress is much less than 100% of the operating time, so electrolysis will proceed more slowly. Also, the temperature of the bulb is low since the dissipation is much less than the allowed 40 W. Electrolysis rate increases as an exponential of the absolute temperature. Life of the 6BK4C's is expected to be reasonable, and the chance of having to replace tubes was taken as a calculated risk. Although our use experience is limited, no failures have occurred thus far.

Each tube has a floating 28-V negative-bias supply that is derived from a 20-V winding of T203. The 6.3-V tap on the same winding provides heater current. V204 and V202 (of the other amplifier) share the winding which is connected to the −20 kV power supply. The 10-kΩ grid resistors (R229 for V203, and R230 for V204) bias the grids to 0 V when the

transistors of the optical couplers, A203 and A204, are cut off. The high-voltage power supplies should not be turned on with the driver circuits disconnected since approximately 3 mA per tube would flow, exceeding both the current and power ratings of the tubes. With properly adjusted drivers connected, tubes are not both biased on simultaneously. When the light-emitting diode of A203 is turned on, the transistor of A203 conducts, biasing V203 toward cutoff. When A203 is driven on, A204 is driven off so that V203 is turned off as V204 is turned on, and the output voltage swings in a negative direction. The maximum negative bias for the tubes is about 20 V when the optical coupler transistors are full on. R225 and R229 form a voltage divider with a ratio of approximately 0.7. Thus, 20 V should be enough to cut the tubes off at 35-kV plate voltage.

The optical couplers, A203 and A204, have a transfer ratio of about 0.7. About 2 mA must flow to bias the tube off, which means about 3 mA in the LED.

The bias supplies are typical half-wave RC-filtered power supplies and are unregulated. Bias variations are taken out by the overall feedback.

4.3 Driver Circuit

The dividing line between the output circuit and the driver circuit bisects the optical couplers and the feedback resistors. All parts of the driver circuits are within ±15 V of ground.

Transistors Q203 and Q204 drive the LED's of the optical couplers. The LED currents are limited by series resistors R247 and R248 to less than 30 mA. R233 and R234 limit the base currents of Q203 and Q204 to conservative values; these allow adequate drive and also avoid loading the outputs of AR203 and AR204.

AR204 is operated as a unity-gain inverter to provide the inverted drive for V203. Additional current is injected into the node at pin 2 to provide a dc offset in the output. This offset is used to establish the quiescent current in the output tubes at 0 V output. See Section 5.2 for a discussion of how the quiescent point is established.

AR203 is operated locally open-loop at dc to provide a high gain near the overall amplifier summing point. The local feedback provided by R240, R241, and C212 produces a high-frequency roll-off to stabilize the overall feedback amplifier. The overall feedback summing point is at the + input of AR203, which is the − input if the whole amplifier is considered as a unit. Diodes D212 and D213 protect this input from the inevitable surges that occur in high-voltage equipment. R206 and C214 in parallel provide the overall feedback path and establish main fre-

quency roll-off of the whole amplifier. C214 is about 2.5 cm of braid around the output cable which produces a capacitance of about 2 pF.

The overall gain is established by choosing the input resistor using switch S201B (a make-before-break selector switch). The input resistance is varied from 80 kΩ to 2 MΩ.

5. Calibration and Adjustments

5.1 Bias Supply Control Module

Since no adjustments in the digital circuit are necessary or possible, we need concern ourselves only with the analog circuit.

Either the digital voltmeter (DVM) in the control module or a separate DVM may be used for calibration and adjustment of the ANALOG CIRCUIT. If the internal DVM is to be used, it will be necessary to install an adaptor cable to receptacle J103 at the rear of the control module; the leads will be connected to the DVM when the METER RANGE switch is in the HV position. The internal DVM should be checked for accuracy before proceeding with the ANALOG CIRCUIT calibration.

(a) Initial Offset Calibration

(1) Remove VR101 from its socket and connect a short jumper wire between the output and ground terminals on the socket. This will provide 0.00 V input to AR A.
(2) Measure the output of AR A with a DVM and adjust trim-pot A-ZERO for 0.00 V output.
(3) Remove the jumper wire and reinstall VR101.
(4) Measure the output from VR101 and record the value.
(5) Reverse the meter leads*, measure the output of AR A, and adjust trim-pot A-TRIM to obtain the same reading as that previously obtained from the output of VR101.

(b) Zero adjustment of AR B

(1) Connect the DVM to the output of AR B (pin 6).
(2) Momentarily depress the RESET button; then adjust B-ZERO for 0.00 V output.

* For best results (greatest accuracy) the polarity of the DVM should be the same for all measurements.

(c) *Zero adjustments of AR D, AR E, AR F, AR G, and AR H*

(1) Set SWEEP AMPLITUDE and associated TRIM completely counterclockwise (CCW). Set INITIAL OFFSET and associated TRIM completely CCW; set INITIAL OFFSET switch to ZERO. Carry out following steps *in sequence.*
(2) Connect DVM to AR D output at −10 V BNC connector (J108) or pin 6 of AR D and adjust D-ZERO for 0.00 V meter reading.
(3) Connect DVM to AR F output at +10 V BNC connector (J107) or pin 6 of AR F and adjust F-ZERO for 0.00 V meter reading.
(4) Connect DVM to AR E output at ANALOG OUTPUT connector (J109) or pin 7 of AR E and adjust E-ZERO for 0.00 V meter reading.
(5) Connect DVM to AR G output at +100 V UHF connector (J110) and adjust G-ZERO for 0.00 V meter reading.
(6) Connect DVM to AR H output at −100 V UHF connector (J111) and adjust H-ZERO for 0.00 V meter reading.

(d) *Amplifier tracking (gain) adjustments*

(1) Select SWEEP FUNCTION 1 (Ascending single ramp) and a low INITIAL TO PEAK SWEEP TIME. Momentarily depress the RESET pushbutton and then the GO pushbutton. After the END of SWEEP light comes on, adjust the SWEEP AMPLITUDE CONTROL for exactly −5.00 V at the −10 V output connector. Carry out the following steps *in sequence.*
(2) Connect DVM to AR F output and adjust F-TRIM for a +5.00 V meter reading.
(3) Connect DVM to AR G output and adjust G-TRIM for a +50.00 V meter reading.
(4) Connect DVM to AR H output and adjust H-TRIM for a −50.00 V meter reading.
(5) Connect DVM to ANALOG OUTPUT and adjust E-TRIM for a +5.00 V meter reading.

This completes the ANALOG CIRCUIT calibration.

5.2 High-Voltage Amplifier Module

Only 3 adjustments are provided. The zero-offset adjustments provided on AR203 and AR204 by R252 and R253, respectively, are noncritical and need adjustment only if the associated amplifier is replaced. The quiescent point adjustment may be needed if the replacement of one or more tubes is required. The overall feedback maintains the amplifier performance quite well even if these adjustments are not optimized.

Remove the time-delay relay from its socket. This will prevent accidental high-voltage turn-on.

(*a*) *Zero-Offset Adjustment*

(1) Remove both operational amplifiers from high-voltage driver board.
(2) Measure the slider voltage of R254 (it can be accessed at the end of R244 that is nearer the center of R254) with an external DVM and adjust the center trim-pot, R254, for 0.00 V.
(3) Temporarily connect a 10-MΩ resistor between pin 2 and 6 of operational amplifier AR203. Switch high-voltage range switch to 25 kV and install both operational amplifiers.
(4) Short the −10 V driver connector, J201. Turn power switch to standby and adjust zero-offset trim-pot, R252, near operational amplifier AR203 for 0.00 V reading on operational amplifier output pin 6.
(5) Measure the output of operational amplifier AR204 pin 6 and adjust the trim-pot nearest AR204 for 0.00 V. Remove short across −10 V drive and 10-MΩ resistor. Reinsert the time delay relay.

As mentioned in Sec. 4.3, the balance potentiometer R254 is used to establish the quiescent operating current in V203 and V204. With zero volts input the overall feedback will cause the intermediate signal voltages in the amplifier to assume values which will bring the output voltage to zero. When the output voltage is zero equal currents flow in V203 and V204. Producing an offset voltage in *that portion of the circuit which drives V203 but does not affect the drive to V204* will change the quiescent current. The adjustments described above reduced this offset to zero.

(*b*) *Quiescent Point Adjustment*

While the voltages and currents in the high-voltage amplifier may not be lethal, contact with them would certainly be unpleasant. Before starting the following procedure, power must be off and grounds connected to the output and both power supply corona caps. Remove the connector from the plate V204 and connect a 200-μA meter from the connector to the cap. The + of the meter goes to the connector. The meter must be insulated so there is at least 2 or 3 cm between it and the nearest conductors. The measurement will be made with the meter near ground but transients may occur during turn-on. With the meter securely in place where it can be read easily, remove the grounds and go through normal turn-on procedure. With the input at 0.00 V, check that the output is also 0.00 V. Then, adjust R254 for a quiescent current between 5 and 50 μA. The output should still be 0.00 V. Shut off the power, ground the output and power supplies, remove the meter, and replace the plate cap.

The procedure for adjusting the amplifier on the left side of Fig. 12

is exactly the same except, of course, that the component designation numbers are different.

Acknowledgment

The authors acknowledge with gratitude many helpful suggestions contributed by Jeremiah Avins during the development of the equipment described in this paper.

References:

[1] A. M. Goodman, "A Useful Modification of the Technique for Measuring Capacitance as a Function of Voltage," *IEEE Trans. Electron Devices*, **ED-21**, p. 753 (1974).

[2] A. M. Goodman, "An Investigation of the Silicon-Sapphire Interface Using the MIS Capacitance Method," *IEEE Trans. Electron Devices*, **ED-22**, p. 63 (1975).

[3] A. M. Goodman, *Semiconductor Measurement Technology:* Safe Operation of Capacitance Meters Using High Applied-Bias Voltage, NBS Special Publication 400-34 (December 1976); also, *RCA Review*, **37**, p. 491 (1976).

[4] A. M. Goodman, *Semiconductor Measurement Technology:* Suppression of Premature Dielectric Breakdown for High-Voltage Capacitance Measurements, NBS Special Publication 400-37 (July 1977); also, *RCA Review*, **38**, p. 436 (1977).

[5] A. M. Goodman, *Semiconductor Measurement Technology:* A 25-kV Bias-Isolation Unit for 1-MHz Capacitance and Conductance Measurements, NBS Special Publication 400-40 (September 1977); also, *RCA Review*, **38**, p. 512 (1977).

Extending the Content and Expanding the Usefulness of the Simple Gaussian Lens Equations

Part 2: Derivation and Application of Gaussian Design Relationships That Are Inherent in the Primitive Relay Optical System

L. T. Sachtleben

RCA Advanced Technology Laboratories, Camden, N.J.

Abstract—A primitive relay optical system comprises a fixed, coaxially mounted pair of separated lenses. It functions to produce finite images of two fixed objects that are spaced apart at finite distances along the system axis, and to do so in a manner that conserves the light. In its most general form, the primitive system is called the "basic relay" system. The Gaussian relationships are developed for this form. Other relationships are easily obtained for the more specialized forms of the primitive system. One such form that has a special interest is infinite in focal length, and is called the "afocal relay" system. In order to illustrate application of the relay theory, it is used to design an increase in the length of a short, hand-held type telescope to more than 30 feet by the efficient use of relaying stages. Lens locations, lens focal lengths, and diameters of all objects and images are determined on the basis of the simple or first-order lens equations developed in the earlier Part 1 (see Ref. [11]), and in this paper.

13. Introduction*

The primitive relay optical system is a very simple structure that com-

* Editors Note: This paper is a continuation of the author's previous paper in the December 1976 issue of *RCA Review*. The numbering system employed here is, therefore, a continuation from that paper.

prises a fixed, coaxially mounted pair of separated thick lenses. In the theory of its general or basic form it is assumed to produce images of two fixed object points that are spaced apart at perfectly general object locations along the system axis. The assumed locations and focal lengths of the lenses are suited to project the two images at general image locations and magnifications.

Two or more relay systems may be placed in serial or sequential order along a common optical axis, in order to form an extended relay system. The design procedure for a single relay system may either be applied repeatedly to produce the extended system, or the design relationships may be solved for any of their variables and the solutions applied as required. Suitable substitutions can always be made to change any of the variables. No attempt is made in this paper to discuss all of the possibilities.

In Part I of this paper it was stated that the (conjugate displacement) theorem is an indispensable adjunct to the equations for two separated lenses or systems, when developing the general Gaussian theory of relay optical systems.† It is the purpose of Part 2 to develop such theory, and to apply it to a representative example. General theories can be applied in endless ways, and study of the example should challenge and stimulate readers' normal ingenuity sufficiently to enable them to apply the theory to a wide variety of problems.

14. Note On the Organization of This Paper

The equations of Part 1 were numbered through Eq. [104]. As they are needed, these equations will be called up by their original numbers, notationally transposed as described in Sec. 18, and then assigned new numbers in a serial order that is continuous with, and an extension of the equation numbering of Part 1. Thus, an equation with a number that precedes [105] can be found, together with its actual or indicated derivation, in Part 1. Extension of the numbering in Part 2 will begin with Eq. [105].

15. The Relay Concept in Optical Literature

The literature of optics appears to give very little explicit attention to relay systems and the relaying of images. This is true in spite of the fact that the relay principle is an important feature of many optical systems. The brief bibliography on the subject that is available to the writer is cited in References [12]–[15].

Webster's New International Dictionary (1961), and the Random

† See Ref. [11], p. 451.

House Dictionary of the English Language (1966), do not mention the optical use of the word "relay." A paraphrase of one of their definitions of the word as a noun, might give it an optical meaning as follows: an arrangement by which images may be formed or repeated, by one or a succession of optical systems, in order to make the images appear at required locations in a manner that conserves the light.

One of the dictionary definitions of the noun is, "The act of passing along (a message, a signal, a ball) by stages; also: one of such stages." In the case of a relay optical system, images are the "messages" or "signals" that are passed along by one or more optical "stages."

A dictionary definition of the word as a transitive verb is, "To pass along by relays (. . . to distant points . . .)." Application of this meaning of the word to the optical case will become clear from the reading of this paper.

Recent optical glossaries, and handbooks, dictionaries, and encyclopedias of physics that are available to the writer, do not mention an optical use of the word "relay." [16–20]

16. The Primitive Relay Optical System in Its General or Basic Form

The arrangement and manner of functioning of a basic relay stage is illustrated by Fig. 9. The light direction is indicated by the shaded arrow. Under the Gardner sign convention, the light direction in the optical system is always the optically positive direction. By using this device the formulas of Gaussian optics need not be tied to or dependent upon the positive direction in any particular or fixed Cartesian reference frame. L is the anterior lens, and L′ is the posterior lens of the system.* The common axis is *c-c*. O is the anterior object, and is shown real.** O′, the posterior object, is shown virtual (a virtual object *must always* be formed by projection). If O′ were anterior to L it would be real. I is the anterior image, and is shown virtual. If I were posterior to L′ it would be real. I′, the posterior image, is shown real. Image I, whether real or virtual, *must*

* ANTERIOR denotes spacing or displacement away from a counterpart, or from some other reference, in a sense opposed to the direction of light travel. POSTERIOR denotes removal in the direction of light travel. The anterior lens or surface of an optical system is the lens or surface that is anterior to all of the other lenses or surfaces of the system. Similarly, the posterior lens or surface of a system is posterior to all the other lenses or surfaces of the system. An image or object at infinity can be considered either anterior or posterior to a local reference; however, the choice may involve the sign of the magnification.

** The following are practical definitions of REAL and VIRTUAL objects and images. When referred to a specific lens, optical system, or portion thereof, an object is real in the practical sense providing it is either anterior to or coincident with the anterior refracting (or reflecting) surface; otherwise the object is virtual. Similarly, an image formed by a specific lens, optical system, or portion thereof is real in the practical sense providing it is either posterior to or coincident with the posterior refracting (or reflecting) surface; otherwise the image is virtual. An image or object at infinity can be considered either real or virtual; however, the choice may involve the sign of the magnification. See the preceding definitions of the terms ANTERIOR and POSTERIOR.

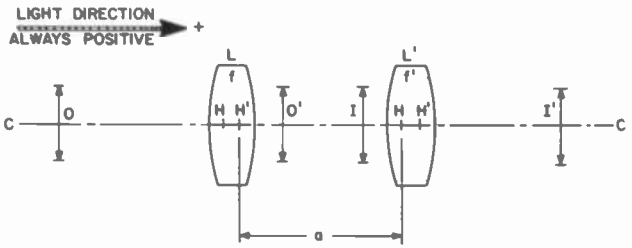

Fig. 9—The generalized or basic primitive relay optical system.

be regarded as the image of O that is produced by system L-L′, with *both* L and L′ taking part in its formation. I′ *must similarly* be the image of O′, whether O′ be real or virtual. The distance a from H′ of L to H of L′ is the optical separation of lens L′ from lens L.

It is important to further emphasize the last portion of the preceding paragraph. In the equations that will be derived in the appendix for a general, or basic primitive relay optical system, the object-distances that will make an explicit appearance are *only those* for lens L. Similarly, the image-distances that will make an explicit appearance are *only those* for lens L′. No image-distances for L, and no object-distances for L′, will make any explicit appearance whatsoever in the equations. Thus, in considering the primitive relay of Fig. 9, the relationships of object-points O and O′ to the system are wholly restricted to their respective signed distances from the first principal point H of anterior lens L. Similarly, the relationships of image-points I and I′ to the system are wholly restricted to their respective signed distances from the second principal point H′ of posterior lens L′. It is important to understand the matters emphasized in this paragraph very clearly, in order not to become confused over the relations of the points O, O′, I, and I′ to the primitive relay system.

If the two images that have been produced by a relay stage, such as shown in Fig. 9, become in turn the objects for a second or succeeding relay stage, two further relayed images of the original objects will result. The second stage may then be followed by a third relaying stage, and so on. Repeated extension of a corrected relaying system by adding stages cannot be continued indefinitely, for it is limited by the progressive buildup of residual aberrations in the images; notably field curvature and secondary color.[13] However, discussion of the aberration problem is not a purpose of this paper.

In the case of any relay stage, it is a common practice for the posterior object O′ to be located near to or possibly within the anterior lens L. The image I of the anterior object O, is likewise located near to or possibly within the posterior lens L′. This kind of relationship is maintained

approximately in every stage of an extended relay system, but it is usual to have the object or image a little outside the associated lens so that dirt or defects in the lenses will not appear sharply focused in one or both of the final two images. The anterior object that precedes the anterior stage of an extended relay system and the posterior image that follows the posterior stage of the system are not always required to be in close proximity to any lens of the optical system.

With the last-mentioned two exceptions, the purpose of proximity of the objects and the images to the lenses is to assure geometric conservation of the light. It does this by keeping light rays from being directed to points that lie outside the entrance pupils of the lenses. Proximity of the lenses to the objects and the images within the system becomes increasingly important as both the field angle and the convergence angle at any object or image become greater. When the field angle is very small, or the convergence angle is very small, proximity may be of less importance. This may be the case, for example, in systems that receive their light directly from a laser source.

A special case arises when the axially centered diameter of the object O, Fig. 9, substantially equals the axially centered diameter of object O′. In this case all of the light rays are contained in a cylindrical space between O and O′. The location of lens L is then photometrically immaterial providing it lies between a real O and a virtual O′, and also providing it is not smaller in diameter than O or O′. Other considerations that are not gone into here may then govern its adopted location.

Proximity to objects and images, however, usually helps to minimize the diameters of the lenses and their entrance pupils. Failure of geometric conservation of the light usually results in reduction of illumination in the image of outlying parts of the field; a condition that is known as vignetting, or as shading in television parlance, and is often moderately acceptable.

The relationships that are developed here, however, are purely Gaussian or paraxial in character, and they are not directly concerned with photometric matters. Their initial development, for the basic relay system, assumes perfectly general degrees of proximity of O′ to L, and of I to L′.

17. Design Considerations in the General or Basic Primitive Relay Optical System

The entitites and quantities that are involved in developing the design relationships that are inherent in the basic relay stage, are illustrated by Fig. 10. The light direction is indicated by the shaded arrow. By the Gardner sign convention, it is the positive direction in the system. All algebraic relationships in this paper are based on the Gardner sign

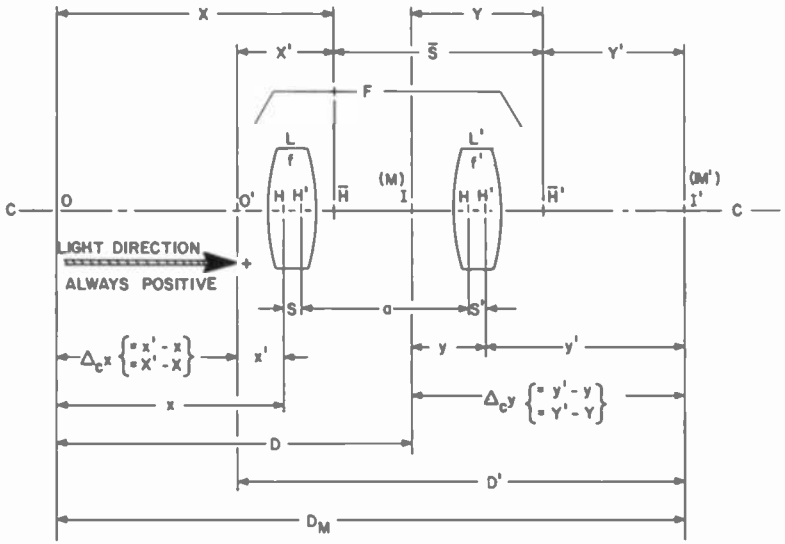

Fig. 10—Quantities that are involved in the Gaussian design relationships for the basic primitive relay optical system.

convention.[11] L and L′ are two thick lenses or subsystems, respectively anterior and posterior, on the common axis c-c. The principal points of L are H and H′, with separation S; those of L′ are H and H′ with separation S'. The focal length of L is f; that of L′ is f'. The focal length of system L-L′ is F. The principal points of system L-L′ are $\overline{H}$ and $\overline{H}'$, with separation $\overline{S}$.

O and O′ are the two axial object-points, respectively anterior and posterior. I is the image of O, the magnification being M. I′ is the image of O′, with magnification M'. The object-points and image-points have perfectly general (but not infinitely distant) locations along the axis with respect to the lenses.

The object-points O and O′ are respectively at object-distances x and x' from H of lens L. Their separation or displacement in the object-space is $\Delta_c x = x' - x$. The image-points I and I′ are respectively at image-distances y and y' from H′ of lens L′. Their separation or displacement in the image-space is $\Delta_c y = y' - y$. The optical separation of L′ from L is a, and is measured from H′ of L to H of L′. The distance of I from O is D; of I′ from O′ is D'. The distance of I′ from O is D_M, and is called the major object-to-image distance. This distance ordinarily sets the general scale of the basic relay stage.

The object-points O and O′ are respectively at object-distances X and X' from $\overline{H}$. Image-points I and I′ are respectively at image-distances Y and Y' from $\overline{H}'$. It is also true that $\Delta_c x = X' - X$, and that $\Delta_c y = Y' - Y$ (see the preceding paragraph).

In the particular statements of the inherent Gaussian relationships that are derived for the basic relay system in Appendix 3, the primary parameters to which arbitrary or required numerical values might be assigned are:

(1) D_M, to establish the overall scale of the system.

(2) Either of $\Delta_c x$ and $\Delta_c y$; either of x and x'; either of y and y'; to partially establish the internal scale of the system. (It is usual, in the interest of controlling proximity, to assign values to x' and y, rather than to x and y'.)

(3) Either of M and M'.

(4) Principal point separations S and S', of lenses L and L′.

The unknown primary parameter to be determined might be:

(5) The nonarbitrary one of M and M'. (In general, the choice of unknown parameters depends on the problem or on the judgment of the designer. For example, both M and M' may be given arbitrary values, and D_M may be an unknown parameter; in which case the overall scale of the system will not be controlled.)

Secondary, unknown parameters, whose determination usually constitutes the main objective of the design work, are:

(6) The optical separation a, of L′ from L, to complete the internal scale of the system.

(7) The respective focal lengths f and f', of L and L′.

Auxiliary parameters whose values may be determined, either as required in calculating the foregoing unknown primary or secondary parameters, or as a matter of convenience in making check calculations, etc., are:

(8) The non-arbitrary ones of $\Delta_c x$ and $\Delta_c y$; of x and x'; of y and y'; (see item 2, above).

Ancillary parameters that it may be necessary or useful to calculate, are:

(9) D, D', F, $\overline{S}$, X, X', Y, and Y'.

18. Using the Equations From Part 1

In the discussion of the system of two separated coaxial lenses or subsystems contained in Part 1,[11] some of the literal quantities employed numerical subscripts. The subscripts were used to designate the order in which the light reached and passed through the two lenses or systems, and to designate the distances, principal points, focal lengths, and other quantities immediately associated with each of them. For example, f_2 was the focal length of the second (posterior) lens or system L_2 through which the light passed, and its principal points were H_2 and H_2'.

Such use of subscripts is convenient where a single object and its image are being considered. Where two separate objects and their images are

being considered in relation to a system of two separated lenses, numerical subscripts tend to become awkward and restrictive. They are, therefore, dropped in this part of the paper. In the examples and appendices that follow, all numerically subscripted quantities called up from Part 1 are conveniently transposed by dropping the numerical subscripts and leaving the letters plain where the subscript had been 1, but priming the letters where the subscript had been 2. The one exception to this is the designation of the principal points. Principal points H_1 and H_2 both become H; principal points H_1' and H_2' both become H′. To distinguish between the two H′s, for example, we speak of either H′ of L, or of H′ of L′.

Where primitive relay systems follow one another to form an extended relay system on a common axis, the literal quantities that pertain to a single one of such systems may all be identified by using the same *literal* subscript for each of them. In a P-, U-, or Z-relay stage, for example, every literal quantity may be subscripted with a P, U, or Z. In what follows, except in the perfectly general equations of the appendices, these literal subscripts are used.

19. A Representative Application

The design equations for the general or basic primitive relay optical system are developed in Appendix 3. In the example to follow, these equations will be called up only by equation number and used in a variety of ways according to need. Auxiliary equations, whose derivation might unduly complicate the discussion and solution of the example, will be developed in Appendix 4, and called up from there only by equation number.

Example 9: Increasing the Length of a Hand-held Type Telescope by the Use of Relays, While Also Changing the True Field Angle

An arbitrary Keplerian or astronomic telescope of hand-held proportions and 10× power is assumed to be constructed as follows* (this telescope corresponds to lens L_Z' and the associated Erfle eyepiece, bottom right of Fig. 11):

(1) Nominal magnifying power (hand-held telescope)	10×
(2) Apparent angular semi-field of view	$\epsilon = 30°$
(3) Exit-pupil diameter (inches)	0.200
(4) Minimum eye relief (inches)	0.787

* See Ref. [13], p. 14-14 and pp. 15-1 through 15-3.

(5) Objective lens focal length (y'_Z, Fig. 11) (inches) 10.000
(6) Objective lens diameter (Lens L'_Z, Fig. 11) (inches) 2.000
(7) Eyepiece type Erfle
(8) Eyepiece focal length (inches) 1.000
(9) Eyepiece field-stop diameter (d'_{iZ}, Fig. 11) (inches) 1.155
(10) A Porro erecting prism may or may not be used, according to circumstance.
(11) True semi-field angle (hand-held telescope) $\phi = \tan^{-1} \frac{1.155}{20}$

It is desired to modify the focal length of the objective lens L'_Z of this telescope, at the same time keeping its diameter and location relative to the eyepiece fixed, by making it a part of an extended relay optical system. The purpose of the relay system is to permit location of a new objective lens at a distance of 360 inches from the field-stop of the Erfle eyepiece (assuming no Porro prism to be present). The greatly lengthened system, represented by Fig. 11, is to also have a new true semi-field angle $\theta = \tan^{-1} 0.07$, that is about 21% greater than ϕ. The apparent angular semi-field of view is to remain $\epsilon = 30°$, as provided by the original eyepiece.

The general plan of approach to the problem will begin by following the new objective lens L_0 (top left of Fig. 11) of focal length f_o, with an expansion relay P. The anterior object-point O_P for the P-relay will be the second principal point H'_0 of the new objective L_0. The posterior object-point O'_P will be the posterior principal focal point of the new objective, the object that is being viewed by L_0 being assumed at infinity.

The aperture-stop diameter (assumed intrinsically positive) associated with O_P will be d_{oP}, and for practical purposes will be the free diameter of lens L_0. The field-stop diameter (also assumed intrinsically positive) associated with O'_P will be d'_{oP}. The image of O_P at I_P will be formed at magnification $M_P < 0$, and the image diameter (intrinsically positive) associated with I_P will be d_{iP}, where

$$-d_{iP} = M_P d_{0P} \qquad [105]$$

The image of O'_P at I'_P will be formed at magnification $M'_P < 0$, the image diameter (intrinsically positive) associated with it being d'_{iP}, where

$$-d'_{iP} = M'_P d'_{oP} \qquad [106]$$

In the planned approach, the diameters (intrinsically positive) of the two images projected by the P-relay will be arbitrarily equal, or

$$-d'_{iP} = -d_{iP} = M'_P d'_{oP} = M_P d_{oP} \qquad [107]$$

The images formed by the P-relay will both be inverted, because the magnifications are both negative. Positive magnifications are excluded because they result in intermediate inverted images between the two

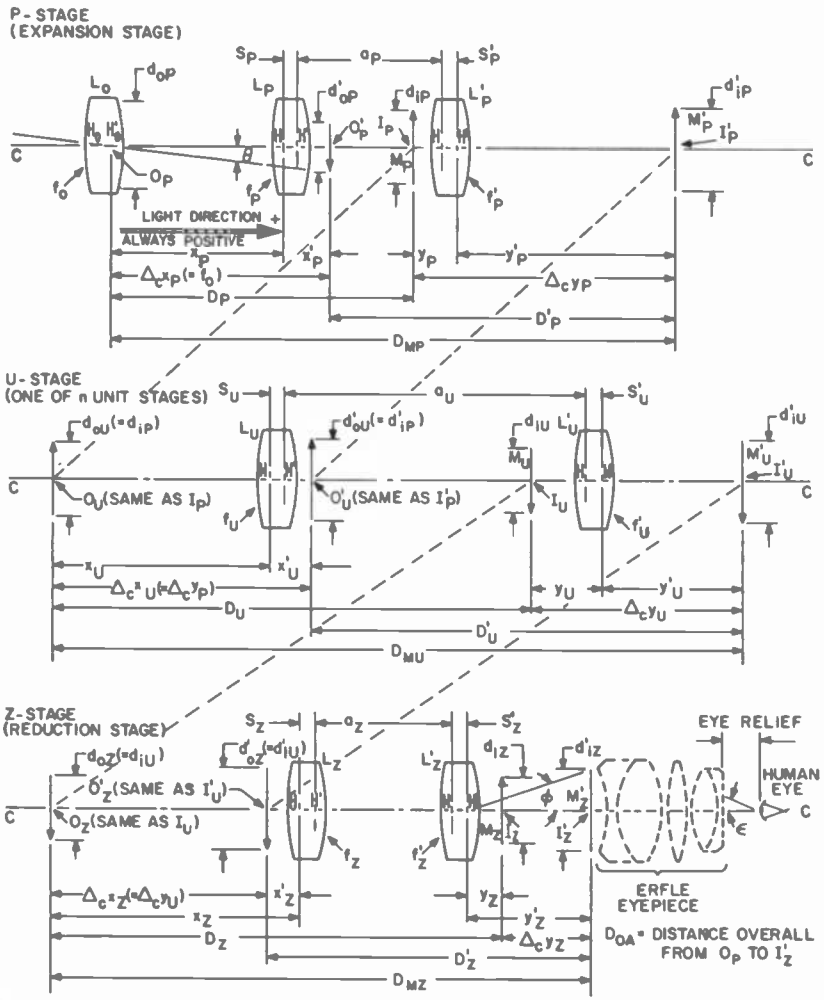

Fig. 11—Schematic diagram for an extended relay optical system (Example 9).

lenses L_P and L'_P, and in a much shorter focal length for one or both of those lenses.

The images at I_P and I'_P become, respectively, the anterior and posterior objects O_U and O'_U for the next stage of the system.

The general plan continues by following the P-relay or expansion stage, with one or a succession of U-relays or unit stages (middle of Fig. 11). A unit stage will relay both images that are presented to it by the P-relay to points further along the axis in the direction of light travel, without changing the sizes of the images. Thus O_U, which is coincident with I_P, will be imaged at I_U, the magnification being

$$M_U = -1. \qquad [108]$$

Similarly O'_U, which is coincident with I'_P, will be imaged at I'_U, the magnification also being

$$M'_U = -1. \qquad [109]$$

From Eqs. [47], [108], and [109] we may write

$$M'_U = M_U = M_A = -\frac{f'_U}{f_U} = -1, \qquad [110]$$

or

$$f'_U = f_U. \qquad [111]$$

Thus, a unit relay ($M_A = -1$) comprises two lenses of the same focal length. That the focal length of the unit relay must be infinite is shown by writing Eqs. [13] and [45] respectively as

$$M = \frac{ff'}{(f + f' - a)x + f(f' - a)} \qquad [112]$$

and

$$F = \frac{ff'}{f + f' - a} \qquad [113]$$

When $M = -1$ and $f = f'$, in Eq. [112], it follows that $a = f + f' = 2f = 2f'$.* Then, by Eq. [113], it follows that $F = \infty$, or the system is afocal. (Note that a unit relay with positive magnification $M_A = M'_A = +1$, is an entirely different kind of system that the reader can easily investigate to advantage in the same manner, using Eqs. [110] through [113]. In this case, contrary to Eq. [111], $-f'_U = f_U$).

It is obvious that unit relays can be arranged in serial or sequential order to relay the images along the axis in a sequence of steps, without changing their sizes. Where unit relays are thus used in a series of n steps or stages, the overall magnification of the U-relay sequence is

$$M_{Un} = M'_{Un} = (-1)^n. \qquad [114]$$

Following the unit stage (or stages) the general plan terminates the system with a final reduction relay Z (bottom of Fig. 11). The diameters (all intrinsically positive) of the images that have been formed by the U-relay (or U-relays), and which are the objects for the Z-relay, are

$$d'_{oZ} = d_{oZ} = d'_{iU} = d_{iU} = d'_{oU} = d_{oU} = d'_{iP} = d_{iP} \qquad [115]$$

The anterior image of the two images that are formed by the U-relay (or

* The solution $x = -f = -f'$ is irrelevant because x does not appear in Eq. [113].

U-relays) is at O_Z, and the posterior one is at O'_Z. These are the two objects for the Z-relay.

The image-points I_Z and I'_Z for the Z-relay are located respectively at the second principal point of posterior lens L'_Z, and at the center point of the field-stop of the Erfle eyepiece. The aperture-stop diameter (intrinsically positive) associated with I_Z is d_{iZ}, which is substantially equal to the free diameter of L'_Z. The field-stop diameter (intrinsically positive) associated with I'_Z is d'_{iZ}, equal to the diameter of the field-stop of the eyepiece.

The given, or assumed, and the unknown quantities that are involved in the design of the proposed relay system are symbolized, and their values and status given in Table 9. The order in which the symbols appear in the table corresponds roughly to the order in which their referents are encountered by the light as it passes through the system along axis *c-c* (see Fig. 11).

Of the 59 quantities of Table 9 that are pertinent to the solution of the proposed problem, over half (51%) of them are unknown. Not all of these unknowns are distinct, in the sense of having no simple relation, or equivalence, to each other; but final numerical values are not easily computed for (or assignable to) any of them at the onset, as they are for items 8, 17, 49, and 53.

The solution to this example is carried through on a step-by-step basis, which has been found preferable to setting up and solving elaborate equations. There is no routine theoretical basis for deciding the order in which the unknowns of Example 9, or the unknowns of any other relay system problem, must be evaluated. That the order of procedure is important is evident from the fact that evaluation of certain of the unknowns depends on the prior evaluation of certain of the others. The order that is followed in any particular case will depend on the character of the problem, and on the experience that the designer acquires in handling such problems.

It happens that in the problem of Example 9 the pattern of the given quantities is such that preliminary designs of the P-relay and the Z-relay can be completed independently of the design of the unit or U-relay (except for knowing its magnification to be $M_{Un} = (-1)^n$ for its n stages where $n > 0$ is an integer; and also knowing that $d_{oU} = d_{iP}$, $d'_{oU} = d'_{iP}$, and $d_{iP} = d'_{iP}$).

If the relay sequence between the P-stage and the Z-stage had been made up of afocal relays that were not unit relays, but had each some negative magnification $M_A = M'_A \neq -1$; and if the diameters d_{iP} and d_{oZ} as well as the diameters d'_{iP} and d'_{oZ} had been given (none of the four given diameters having necessarily been equal to each other or to their counterparts in Table 9), then it would also have been possible to design

Table 9—Serially numbered symbols of quantities that are pertinent to the statement and solution of Example 9. Symbols are given numerical values, where known; assigned a status (known or unknown, etc.); and are given identifying names.

Item	Symbol & Value	Status	Name
The following items 1 through 20 pertain to the objective lens L_0 and the expansion relay P of Fig. 11. The numerical values of eleven of the twenty items are initially unknown.			
1.	$\theta = \tan^{-1} 0.07$	Given	True semi-field angle
2.	$f_o = 8.000''$	Given	Focal length of lens L_0
3.	d_{oP}	Unknown	Diameter of lens L_0 in planes of H_0 & H'_0
4.	$\Delta_c x_P (= f_o = 8'')$	Given	Object-space displacement for P-relay
5.	$x'_P = -0.500''$	Given	Object-distance for object O'_P
6.	f_P	Unknown	Focal length of lens L_P
7.	$S_P = 0.200''$	Given	H-H′ spacing for lens L_P
8.	$d'_{oP} = 2f'_o \tan\theta = 1.120''$	(computed)	Diameter of object at O_P
9.	a_P	Unknown	Separation between lenses L_P & L'_P
10.	$S'_P = 0.333''$	Given	H-H′ spacing for lens L'_P
11.	$y_P = 0.500''$	Given	Image distance for image I_P
12.	$d_{iP} = 4.500''$ Trial diameter	(assumed)	Diameter of image at I_P
13.	M_P (<0)	Unknown	Magnification from object O_P to image I_P
14.	f'_P	Unknown	Focal length of lens L'_P
15.	$\Delta_c y_P$	Unknown	Image-space displacement for P-relay
16.	$d'_{iP} = 4.500''$ Trial diameter	(assumed)	Diameter of image at I'_P
17.	$M'_P = -d'_{iP}/d'_{oP} = -4.017857$	(computed)	Mag'n. from object O'_P to image I'_P
18.	D_P	Unknown	Distance from object O_P to image I_P
19.	D'_P	Unknown	Distance from object O'_P to image I'_P
20.	D_{MP}	Unknown	Distance from object O'_P to image I'_P
The following items 21 through 38 pertain to the first (or any one) of an arbitrary number n of identical unit relay stages. Ten of these items are initially unknown, including the number n which is determined by trial. The n stages are assumed to succeed each other in serial order, with the anterior or object-space displacement interval of each succeeding stage coinciding with the posterior or image-space displacement interval of the preceding stage.			
21.	$d_{oU} (= d_{iP}) = 4.500''$ Trial diam.	(assumed)	Diam. at object O_U and image I_P
22.	$\Delta_c x_U (= \Delta_c y_P)$	Unknown	Object-space displacement for U-relay, that coincides with image-space displacement for P-relay
23.	x'_U(assumed = $-y_U$)	Unknown	Object-distance for object O'_U
24.	$f_U (= f'_U)$	Unknown	Focal lengths of lenses L_U & L'_U
25.	$S_U = 0.333''$	Given	H-H′ spacing for lens L_U
26.	$d'_{oU} (= d'_{iP}) = 4.500''$ Trial diam.	(assumed)	Diam. at object O'_U and image I'_P
27.	a_U	Unknown	Separation between lenses L_U and L'_U
28.	$S'_U = 0.333''$	Given	H-H′ spacing for lens L'_U
29.	y_U (assumed = $-x'_U$)	Unknown	Image-distance for image I_U
30.	$d_{iU} (= d_{oU}) = 4.500''$ Trial diam.	(assumed)	Diam. at image I_U and object O_U

31.	$M_U = -1$	Given	Magnification from object O_U to image I_U
32.	$\Delta_c y_U$ (= $\Delta_c x_U$)	Unknown	Image-space displacement for U-relay
33.	d'_{iU} (= d'_{oU}) = 4.500″ Trial diam.	(assumed)	Diam. at image I'_U and object O'_U
34.	$M'_U = -1$	Given	Magnification from object O'_U to image I'_U
35.	n (Probably, by trial, $n = 1$ or $n = 2$)		Number of unit stages
36.	D_U	Unknown	Distance from object O_U to image I_U
37.	D'_U	Unknown	Distance from object O'_U to image I'_U
38.	D_{MU}	Unknown	Distance from object O_U to image I'_U

The following items 39 through 58 pertain to the final or reduction relay Z. Nine of them are initially unknown. Item 59 pertains to the overall system.

39.	d_{oZ} (= d_{iU}) = 4.500″ Trial diam.	(assumed)	Diam. at object O_Z and image I_U
40.	$\Delta_c x_Z$ (= $\Delta_c y_U$)	Unknown	Object-space displacement for Z-relay, that coincides with image-space displacement for U-relay
41.	$x'_Z = -0.500''$	Given	Object-distance for object O'_Z
42.	f_Z	Unknown	Focal length of lens L_Z
43.	S_Z 0.333″	Given	H-H′ spacing for lens L_Z
44.	d'_{oZ} (= d'_{iZ}) = 4.500″ Trial dia.	(assumed)	Diam. at object O'_Z and image I'_U
45.	a_Z	Unknown	Separation between lenses L_Z and L'_Z
46.	$S'_Z = 0.333''$	Given	H-H′ spacing for lens L'_Z
47.	$y_Z = 0.000''$ (Zero)	Given	Image distance for image I_Z
48.	$d_{iZ} = 2.000''$	Given	Diameter of lens L'_Z in its planes H and H′
49.	$M_Z = -d_{iZ}/d_{iU}$ = −0.444444	(computed)	Mag'n. from object O_Z to image I_Z
50.	f'_Z	Unknown	Focal length of lens L'_Z
51.	$\Delta_c y_Z = 10.000''$	Given	Image-space displacement for Z-relay
52.	$d'_{iZ} = 1.155''$	Given	Diameter at image I'_Z (Erfle field-stop dia.)
53.	$M'_Z = -d'_{iZ}/d'_{IU}$ = −0.256667	(computed)	Mag'n. from object O'_Z to image I'_Z
54.	$\phi = \tan^{-1}$ (1.155/20)	Given	True semi-field angle of Z-relay, and of original hand-held telescope
55.	$\epsilon = 30°$	Given	Apparent semi-field angle of overall system, and of original hand-held telescope
56.	D_Z	Unknown	Distance from object O_Z to image I_Z
57.	D'_Z	Unknown	Distance from object O'_Z to image I'_Z
58.	D_{MZ}	Unknown	Distance from object O_Z to image I'_Z
59.	$D_{OA} = 360.000''$	Given	Distance over all from object O_P to image I'_Z

the P-stage and the Z-stage independently of the design of the intervening afocal relay stages. This is because the overall magnifications of the afocal relay sequence are calculable as $M_{An} = M'_{An}$ where

$$M_{An} = \frac{\pm d_{oZ}}{d_{iP}} \qquad [116]$$

and

$$M'_{An} = \frac{\pm d'_{oZ}}{d'_{iP}} \qquad [117]$$

The signs of M_{An} and M'_{An} are unimportant because they are the same, for the two magnifications appear as the product $M_{An}M'_{An}$ in the derivation of equations that are used in designing the P- and Z-stages (see derivations of Eqs. [192] and [197], Appendix 4).

Other aspects of solving for the unknowns in a relay system design problem will be discussed in Secs. 20, 21, and 22.

Each of the 18 principal steps in solving Example 9, will be identified by a Roman numeral to make later reference easy. Referenced equations, from Eq. [1] through Eq. [104] will be found in Part 1.[11] Eqs. [105] through [158] appear in the present paper (Part 2). Eqs. [159] through [189] are developed in Appendix 3. Eqs. [190] through [204] are derived in Appendix 4.

In the following steps I through XVIII, only numerical solutions are given where the relevant expressions are taken directly from Appendices 3 or 4. In the remaining cases both literal and numerical solutions are developed for each step. The order of the steps is suited to Example 9 and the item numbers refer to Table 9.

I. Item 3; (d_{oP})—Unknown: Numerical solution only.
By Eq. [194], and Table 9

$$d_{oP} = \frac{d_{iZ}d'_{iZ}}{2\Delta_c y_Z \tan\theta} \qquad [118]$$

$$d_{oP} = \frac{2(1.155)}{2(10)(.07)} = 1.650 \text{ inches.}$$

II. Item 13; (M_P)—Unknown: Literal and numerical solutions.
By Eq. [107], and the fact that $d_{oU} = d_{iP}$ in Eq. [115], we have

$$M_P \triangleq \frac{-d_{oU}}{d_{oP}} = \frac{-d_{iP}}{d_{oP}} \qquad [119]$$

whence, by Table 9 and the just computed value of d_{oP}

$$M_P = \frac{-4.500}{1.650} = -2.727273.$$

III. Item 49; (M_Z)—Unknown: Literal and numerical solutions.
In the reduction, or Z-stage, remembering that the image-diameters are intrinsically positive while the magnifications must be negative, and noting that $d_{oZ} = d_{iU} = d_{oU} = d_{iP}$, as in Eq. [115],

$$M_Z = \frac{-d_{iZ}}{d_{oZ}} = \frac{-d_{iZ}}{d_{iU}} = \frac{-d_{iZ}}{d_{iP}}. \qquad [120]$$

Therefore, by Table 9 and Eq. [120] it is computed that

$$M_Z = \frac{-2.000}{4.500} = -0.444444.$$

IV. Item 53; (M'_Z)—Unknown: Literal and numerical solutions. Depending on Eq. [115], in a manner similar to the derivation of Eq. [120]

$$M'_Z = \frac{-d'_{iZ}}{d'_{oZ}} = \frac{-d'_{iZ}}{d'_{iU}} = \frac{-d'_{iZ}}{d'_{iP}}. \qquad [121]$$

By Table 9 and Eq. [121] it is computed that

$$M'_Z = \frac{-1.155}{4.500} = -0.256667.$$

V. Item 15; $(\Delta_c y_P)$—Unknown: Numerical solution only. By Eq. [199], Table 9, and the computed values of M_Z and M'_Z under steps III and IV

$$\Delta_c y_P = \frac{\Delta_c y_Z}{M_Z M'_Z} \qquad [122]$$

or

$$\Delta_c y_P = \frac{10}{-.444444(-.256667)} = 87.662312 \text{ inches.}$$

VI. Item 17; (M'_P)—Unknown: Literal and numerical solutions. Since O'_P is imaged at I'_P, which in turn coincides with O'_U, the magnification from O'_P to I'_P is (by noting Eq. [115])

$$M'_P = \frac{-d'_{iP}}{d'_{oP}} = \frac{-d_{oU}}{d'_{oP}} \qquad [123]$$

and by substitution from Eq. [193]

$$M'_P = \frac{-d_{oU}}{2\Delta_c x_P \tan\theta} \qquad [124]$$

or, by Table 9

$$M'_P = \frac{-4.500}{2(8)(.07)} = -4.017857.$$

VII. Item 40; $(\Delta_c x_Z)$—Unknown: Numerical solution only. By Eq. [198], and noting Eq. [122] and the numerical solution that follows it

$$\Delta_c x_Z = \Delta_c y_P \qquad [125]$$

or

$$\Delta_c x_Z = 87.662312 \text{ inches.}$$

VIII. Item 45; (a_Z)—Unknown: Numerical solution only.

By Eq. [180], after applying Z-subscripts; and noting Table 9, and Eqs. [120], [121], and [125] and the numerical solutions that follow them

$$a_Z = \frac{M_Z(x'_Z - \Delta_c x_Z)(y_Z + M_Z M'_Z \Delta_c x_Z) - M'_Z x'_Z y_Z}{M_Z M'_Z \Delta_c x_Z}. \qquad [126]$$

In making numerical substitutions, the conjugate displacement theorem offers a simplification by allowing the known value of $\Delta_c y_Z$ to be substituted for $M_Z M'_Z \Delta_c x_Z$ (see Eq. [166]). Thus,

$$a_Z = \frac{-.444444(-.50 - 87.662312)(0 + 10) - (-.256667)(-.50)(0)}{10}$$

$$a_Z = 39.183211 \text{ inches.}$$

IX. Item 58; (D_{MZ})—Unknown: Numerical solution only.
By Eq. [182], after applying Z-subscripts; and noting Table 9, and Eqs. [120], [121], [125], and [126] and the numerical solutions that follow them

$$D_{MZ} = (1 + M_Z M'_Z)\Delta_c x_Z + a_Z - x'_Z + y_Z + S_Z + S'_Z. \qquad [127]$$

Thus

$$D_{MZ} = [1 + (-.444444)(-.256667)](87.662312) + 39.183211 - (-.50) + 0 + .333 + .333$$

$$D_{MZ} = 138.011523 \text{ inches.}$$

X. Item 9; (a_P)—Unknown: Numerical solution only.
By Eq. [180], after applying P-subscripts; and noting Table 9, and Eqs. [119], and [124] and the numerical solutions that follow them

$$a_P = \frac{M_P(x'_P - \Delta_c x_P)(y_P + M_P M'_P \Delta_c x_P) - M'_P x'_P y_P}{M_P M'_P \Delta_c x_P}. \qquad [128]$$

Once again, in making numerical substitutions the conjugate displacement theorem offers a simplification (see Eq. [166]) by allowing the known value of $\Delta_c y_P$, from Eq. [122] and the numerical solution that follows it, to be substituted for $M_P M'_P \Delta_c x_P$. Thus

$$a_P = \frac{-2.727273(-.50 - 8)(.50 + 87.662312) - (-4.017857)(-.5)(.5)}{87.662312}$$

$$a_P = 23.302584 \text{ inches.}$$

XI. Item 20; (D_{MP})—Unknown: Numerical solution only.
By Eq. [182], after applying P-subscripts; and noting Table 9, and Eqs. [119], [124], and [128] and the numerical solutions that follow them

$$D_{MP} = [1 + M_P M'_P)\Delta_c x_P + a_P - x'_P + y_P + S_P + S'_P. \qquad [129]$$

Thus

$$D_{MP} = [1 + (-2.727273)(-4.017857)](8) + 23.302584 - (-.5) + .5 + .2 + .333$$

$$D_{MP} = 120.497927 \text{ inches.}$$

XII. Items 18, 6, and 14; (D_P, f_P, and f'_P)—Unknowns: Literal and numerical solutions.

By Eqs. [172], [181], [7], [8], [159], and [160], with P-subscripts added; and noting Table 9 and the numerical solutions that follow Eqs. [119], [122], [128], [129], [130], and [131], we have by Eq. [172]

$$x_P = x'_P - \Delta_c x_P \qquad [130]$$

or

$$x_P = -.500 - 8.000 = -8.500 \text{ inches.}$$

By Eq. [181]

$$D_P = D_{MP} - \Delta_c y_P \qquad [131]$$

or*

$$D_P = 120.497927 - 87.662312 = 32.835615 \text{ inches. (Item 18).}$$

By Eq. [159]

$$f_P = \frac{-M_P a_P x_P}{(1 - M_P)(x_P - a_P) + D_P - S_P - S'_P} \qquad [132]$$

or

$$f_P = \frac{-(-2.727273)(23.302584)(-8.5)}{[1 - (-2.727273)(-8.5 - 23.302584)] + 32.835615 - .2 - .333}$$

$$f_P = 6.264286 \text{ inches (Item 6).}$$

By Eq. [160]

$$f'_P = \frac{a_P(x_P + D_P - S_P - S'_P - a_P)}{(1 - M_P)x_P + D_P - S_P - S'_P} \qquad [133]$$

or

$$f'_P = \frac{23.302584(-8.5 + 32.835615 - .2 - .333 - 23.302584)}{[1 - (-2.727273)](-8.5) + 32.835615 - .2 - .333}$$

$$f'_P = 18.769519 \text{ inches (Item 14).}$$

* If needed, D'_P and D_Z may be computed in an exactly similar manner, using the relationships of continued Eq. [181].

XIII. Items 57, 42, and 50; (D'_Z, f_Z, and f'_Z)—Unknowns: Literal and numerical solutions.

By Eqs. [181], [7], [8], [159], and [160], with Z-subscripts added; and noting Table 9 and the numerical solutions that follow Eqs. [121], [125], [126], [127], and [134], we have by Eq. [181]

$$D'_Z = D_{MZ} - \Delta_c x_Z \qquad [134]$$

or

$$D'_Z = 138.011523 - 87.662312 = 50.349211 \text{ inches (Item 57).}$$

By Eq. [159], in accordance with the note that follows Eq. [162],

$$f_Z = \frac{-M'_Z x'_Z a_Z}{(1 - M'_Z)(x'_Z - a_Z) + D'_Z - S_Z - S'_Z} \qquad [135]$$

or

$$f_Z = \frac{-(-.256667)(-.50)(39.183211)}{[1 - (-.256667)](-.5 - 39.183211) + 50.349211 - .333 - .333}$$

$$f_Z = 27.126822 \text{ inches (Item 42).}$$

By Eq. [160], in accordance with the note that follows Eq. [162],

$$f'_Z = \frac{a_Z(x'_Z + D'_Z - S_Z - S'_Z - a_Z)}{(1 - M'_Z)x'_Z + D'_Z - S_Z - S'_Z} \qquad [136]$$

or

$$f'_Z = \frac{39.183211(-.5 + 50.349211 - .333 - .333 - 39.183211)}{[1 - (-.256667)](-.5) + 50.349211 - .333 - .333}$$

$$f'_Z = 7.987628 \text{ inches (Item 50).}$$

XIV. Item 38; (D_{MU})—Unknown: Literal and numerical solutions.

In the particular problem that we are dealing with, the posterior or image-space displacement $\Delta_c y$, of each stage, coincides with the anterior or object-space displacement $\Delta_c x$ of the succeeding stage. It follows from this, and from the nature of each of the n identical unit relay stages, that the overall length of the system from O_P to I'_Z is

$$D_{OA} = D_{MP} + D_{MZ} + nD_{MU} - (n + 1)\Delta_c y_P, \qquad [137]$$

or, by solving Eq. [137] for D_{MU},

$$D_{MU} = \frac{D_{OA} - D_{MP} - D_{MZ} + (n + 1)\Delta_c y_P}{n}, \qquad [138]$$

By assigning $n = 1$; and noting the numerical solutions that follow Eqs. [129], [127], and [122], and the value of D_{OA} in Table 9 (Item 59)

$$D_{MUa} = 360 - 120.497927 - 138.011523 + 2(87.662312)$$
$$D_{MUa} = 276.815174 \text{ inches } (n = 1).$$

By similarly assigning $n = 2$, for example

$$D_{MUb} = \frac{360 - 120.497927 - 138.011523 + 3(87.662312)}{2}$$

$$D_{MUb} = 182.238743 \text{ inches } (n = 2),$$

XV. Item 22; $(\Delta_c x_U)$—Unknown: Literal and numerical solutions. The necessary coincidence of $\Delta_c y_P$ and $\Delta_c x_U$ dictates directly that

$$\Delta_c x_U = \Delta_c y_P. \qquad [139]$$

By noting the numerical solution that follows Eq. [122],

$$\Delta_c x_U = 87.662312 \text{ inches.}$$

XVI. Item 23; (x'_U)—Unknown: Numerical solutions only.
By Eq. [204], Table 9, and the numerical solutions that follow Eqs. [138] and [139],

$$x'_U = \frac{3\Delta_c x_U - D_{MU} + S_U + S'_U}{4} \qquad [140]$$

or, for $n = 1$ (see Item 38, under step XIV for $n = 1$)

$$x'_{Ua} = \frac{3(87.662312) - 276.815174 + .333 + .333}{4}$$

$$x'_{Ua} = -3.290560 \text{ inches } (n = 1).$$

Similarly, by assigning $n = 2$ (see Item 38, under step XIV for $n = 2$)

$$x'_{Ub} = \frac{3(87.662312) - 182.238743 + .333 + .333}{4}$$

$$x'_{Ub} = 20.353548'' \ (n = 2).$$

(Note: It is desired that x'_U shall be positive, in the interest of conserving the light. It is also desired that n, the number of unit stages, be as small as possible in the interest of keeping the speeds and the accumulated net power of the lenses in the unit relay sequence, as small as possible.[13] The present sequence of solutions for the unknowns of Table 9 is continued beyond this point, both on the basis of $n = 1$, and $n = 2$, to show the effects of a change in the number of unit stages when the overall length D_{OA} of the system remains unchanged. The question of how to introduce a change that will make x'_U positive, when $n = 1$, will be discussed briefly after the present sequence of solutions is completed. The numerical results of such a change is given in Table 10 of Sec. 20.)

XVII. Item 27; (a_U)—Unknown: Numerical solutions only.

By Eq. [201], and the numerical solutions that follow Eqs. [139] and [140]

$$a_U = \Delta_c x_U - 2x'_U \qquad [141]$$

For $n = 1$ (see Item 23, under step XVI for $n = 1$)

$$a_{Ua} = 87.662312 - 2(-3.290560) = 94.243432 \text{ inches } (n = 1)$$

For $n = 2$ (see It. 23, under step XVI for $n = 2$)

$$a_{Ub} = 87.662312 - 2(20.353548) = 46.955216 \text{ inches } (n = 2)$$

XVIII. Items 24; ($f_U = f'_U$)—Unknown: Numerical solutions only. By Eq. [51], after transposition to eliminate numerical subscripts (see Sec. 18), and on adding U-subscripts; and noting Table 9 and the numerical solutions following Eq. [141]

$$f_U = \frac{a_U}{1 - M_U} \qquad [142]$$

or, for $n = 1$ (see Item 27, under step XVII for $n = 1$)

$$f_{Ua} = \frac{94.243432}{1 - (-1)} = 47.121716 \text{ inches } (n = 1).$$

Similarly, for $n = 2$ (see Item 27, under step XVII for $n = 2$)

$$f_{Ub} = \frac{46.955216}{1 - (-1)} = 23.477608 \text{ inches } (n = 2).$$

It is shown by Eq. [111] that for any unit ($M_U = -1$) relay, $f'_U = f_U$. Therefore

$$f'_{Ua} = f_{Ua} = 47.121716 \text{ inches } (n = 1),$$

and

$$f'_{Ub} = f_{Ub} = 23.477608 \text{ inches } (n = 2).$$

20. Criticism and Adjustment of the Solutions of Example 9

The two solutions of Example 9 have been arrived at by the immediately preceding steps. The two solutions are identical except that in one, the unit-relay has only a single stage (two lenses of focal length $f_{Ua} = f'_{Ua} =$ 47.1217 inches), while in the other the unit-relay has two stages (four lenses of focal length $f_{Ub} = f'_{Ub} = 23.4776$ inches). Both solutions have the same expansion or P-relay, and they have the same reduction or Z-relay. A further difference between the two solutions is that in the single-stage unit-relay $x'_{Ua} = -y_{Ua} = -3.2906$ inches, whereas in the two-stage unit-relay $x'_{UB} = -y_{Ua} = +20.3535$ inches.

It is desirable that $x'_U = -y_U$ be a positive quantity, but not at the cost both of doubling the number of lenses in the U-relay system, and at the same time dividing their focal lengths and therefore their focal ratios or *f*/numbers substantially in half. The *f*/numbers vary directly as the focal lengths, because the diameters $d_{iP} = d_{oU} = d'_{iP} = d'_{oU} = 4.500$ inches remain unchanged between the two solutions (see Fig. 11).

The problem that now arises is to find a way to change the solution of the example in such a way that $x'_{Ua} = -y_{Ua}$ shall be a positive quantity of a suitable magnitude, the *Ua*-subscripts indicating, as in steps XIV through XVIII, that the unit relay shall have only a single stage and therefore contain only two lenses.

A number of approaches to the solution of this problem are possible:

(1) Trial calculations show that if the system is modified by reducing distance D_{OA} (overall from O_P to I'_Z) by 60 inches, to $D_{OA} = 300$ inches, while keeping $n = 1$, the value of $x'_{Ua} = -y_{Ua}$ is increased algebraically to $x'_{Ua} = +11.7$ inches approximately. No changes occur in the P- and Z-stages, but focal lengths are reduced by about 32% in the U-stage, to $f_{Ua} = f'_{Ua} = 32.122$ inches approximately. This reduces the *f*/numbers of the two lenses in the U-relay system to $32.122/4.5 = f/7.14$, from their former value of *f*/no. $= 47.122/4.5 = f/10.47$.

(2) Trial calculations that are based on keeping $D_{OA} = 360$ inches but increasing the focal length $f_o (= \Delta_c x_P)$ by one-half inch, to $f_o = 8.500$ inches (while keeping $n = 1$), show that the value of $x'_{Ua} = -y_{Ua}$ is increased algebraically by a little less than 0.50 inches. This does not indicate a practical solution to the problem because the change in f_o produces such a small change in x'_{Ua}. Any change in f_o also results in changes in the designs of both the P-stage and the Z-stage.

(3) Trial calculations that assume $d_{iP} = d_{oU} = d'_{iP} = d'_{oU}$ to be increased by one-quarter of an inch, from $d_{iP} = 4.5$ to $d_{iP} = 4.75$ inches, show that the value of $x'_{Ua} = -y_{Ua}$ immediately increases to $x'_{Ua} = +5.0762$ inches approximately, which may be regarded as of a reasonable magnitude. The focal lengths $f_{Ua} = f'_{Ua}$ decrease about 7%, to $f_{Ua} = 43.760$ inches. The *f*/number decreases to $43.76/4.75 = f/9.21$ from the former value of *f*/10.47, and is not considered an extreme reduction. The increase of d_{iP} to 4.750 inches also requires changes in the designs of the P-stage and the Z-stage. The lenses whose diameters must be increased to 4.750 inches are L'_P, L_U, L'_U, and L_Z (see Eq. [115]). All in all, the increase of these diameters in order to make $x'_{Ua} = -y_{Ua}$ positive, while keeping $n = 1$, appears to be an attractive and practical tradeoff.

Table 10 shows the values of all quantities that are affected by a

Table 10—Sensitivity of image locations in the unit-relay stage to the magnifications M_P and M'_P in the expansion or P-stage of the example of Sec. 19.

Item	Symbol & Value ($d_{oU} = 4.50''$; $n = 1$)	Symbol & Value ($d_{oU} = 4.75''$; $n = 1$)
3.	$d_{oP} = 1.650''$	$d_{oP} = 1.650''$
4.	$\Delta_c x_P\ (= f_0) = 8.000000''$	$\Delta_c x_P\ (= f_0) = 8.000000''$
6.	$f_P = 6.264286''$	$f_P = 6.349514''$
8.	$d'_{oP} = 1.120''$	$d'_{oP} = 1.120''$
9.	$a_P = 23.302584''$	$a_P = 24.584106''$
12.	$d_{iP} = 4.500''$	$d_{iP} = 4.750''$
13.	$M_P = -2.727273$	$M_P - 2.878788$
14.	$f'_P = 18.769519''$	$f'_P = 20.007245''$
15.	$\Delta_c y_P = 87.662312''$	$\Delta_c y_P = 97.673032''$
16.	$d'_{iP} = 4.500''$	$d'_{iP} = 4.750''$
17.	$M'_P = -4.017857$	$M'_P = -4.241071$
18.*	$D_P = 32.835615''$	$D_P = 34.117228''$
19.*	$D'_P = 112.497927''$	$D'_P = 123.790260''$
20.	$D_{MP} = 120.497927''$	$D_{MP} = 131.790260''$
21.	$d_{oU} = 4.500''$	$d_{oU} = 4.750''$
22.	$\Delta_c x_U = 87.662312''$	$\Delta_c x_U = 97.673032''$
23.	$x'_{Ua}\ (= -y_{Ua}) = -3.290560''$	$x'_{Ua}\ (= -y_{Ua}) = +5.076094''$
24.	$f_{Ua}\ (= f'_{Ua}) = 47.121716''$	$f_{Ua}\ (= f'_{Ua}) = 43.760422''$
26.	$d'_{oU} = 4.500''$	$d'_{oU} = 4.750''$
27.	$a_{Ua} = 94.243432''$	$a_{Ua} = 87.520844''$
29.	$y_{Ua}\ (= -x'_{Ua}) = +3.290560''$	$y_{Ua}\ (= -x'_{Ua}) = -5.076094''$
30.	$d_{iU} = 4.500''$	$d_{iU} = 4.750''$
32.	$\Delta_c y_U\ (= \Delta_c x_U) = 87.662312''$	$\Delta_c y_U\ (= \Delta_c x_U) = 97.673032''$
33.	$d'_{iU} = 4.500''$	$d'_{iU} = 4.750''$
35.	$n = 1$	$n = 1$
36.*	$D_{Ua} = 189.152862''$	$D_{Ua} = 175.707690''$
37.*	$D'_{Ua} = 189.152862''$	$D'_{Ua} = 175.707690''$
38.	$D_{MUa} = 276.815174''$	$D_{MUa} = 273.380722''$
39.	$d_{oZ} = 4.500''$	$d_{oZ} = 4.750''$
40.	$\Delta_c x_Z\ (= \Delta_c y_P) = 87.662312''$	$\Delta_c x_Z\ (= \Delta_c y_P) = 97.673032''$
42.	$f_Z = 27.126822''$	$f_Z = 29.088316''$
44.	$d'_{0Z} = 4.500''$	$d'_{0Z} = 4.750''$
45.	$a_Z = 39.183211''$	$a_Z = 41.336050''$
49.	$M_Z = -0.444444$	$M_Z = -0.421053$
50.	$f'_Z = 7.987628''$	$f'_Z = 8.071166''$
53.	$M'_Z = -0.256667$	$M'_Z = -0.243158$
56.*	$D_Z = 128.011523''$	$D_Z = 140.175082''$
57.*	$D'_Z = 50.349211''$	$D'_Z = 52.502050''$
58.	$D_{MZ} = 138.011523''$	$D_{MZ} = 150.175082''$
59.	$D_{OA} = 360.000000''$	$D_{OA} = 360.000000''$

* Note that D_P, D'_P, D_{Ua}, D'_{Ua}, D_Z, and D'_Z are all calculated from the relationships of Eq. [181], Appendix 3.

change in the diameters $d_{iP} = d_{oU} = d'_{iP} = d'_{oU}$, both when these diameters are equal to 4.500 inches and after they are increased to 4.750 inches. The number of unit-stages is maintained at $n = 1$, and diameter d_{oP} is maintained at $d_{oP} = 1.650$ inches. The overall distance D_{OA} is maintained at $D_{OA} = 360$ inches, and there is no change in the value of $f_o\ (= \Delta_c x_P) = 8.000$ inches.

All quantities in Table 10 are assigned the same item numbers that they were given in Table 9. The unknown quantities that were calculated in steps I through XVIII of Sec. 19 are given in the left hand "Symbol & Value" column. The corresponding quantities that have been calculated by the same formulas, after the diameters have been increased to $d_{iP} = d_{oU} = d'_{iP} = d'_{oU} = 4.750$ inches, are given in the right hand column.

21. Checking Correctness of a Solution

Correctness of a solution depends on use of proper equations, and on correct numerical substitutions into those equations. Procedures for checking correctness of a solution usually depend on making further independent calculations that are based on the results of the solution.

The primary purpose, in the problem of Sec. 19, has been to determine the focal lengths and locations of the lenses in the system, such that the performance of the system will meet certain conditions. The performance conditions to be met are stated principally in terms of the locations of images, and the magnifications of those images.

The right hand "Symbol & Value" column of Table 10, presents the solution for the case where $d_{oU} = d_{iP} = 4.750$ inches. The solution includes determination of the focal lengths of all of the six lenses in the three relaying stages, determinations of the separations of the lenses in each stage, and determination of all supplementary linear quantities that, together with quantities given in Table 9, can be used to determine the separation of lens L_U from lens L'_P, and of lens L_Z from lens L'_U.

With all lens focal lengths and all lens separations known, the locations and magnifications of all of the images of O_P and of O'_P, can be calculated and compared with the values that are given or calculated for them in Secs. 19 and 20.

The worker may want to develop equations of his own for doing this, but if not it can be done by applying Eqs. [1] through [3], either directly or by means of their solutions that are given in Table 1, Part 1.[11]

A more highly independent check can be made by means of paraxial ray traces through the system, starting from object-points O_P and O'_P.[21,22] The paraxial ray tracing procedure is much simplified in this case by assuming:

(1) The thickness of each lens is zero, each lens to be represented by a surrogate thin lens whose principal points coincide at its center.
(2) The index of refraction of the material of each thin lens is $n = 1.5$.
(3) The distance from O_P to the first surface of the thin lens that represents lens L_P, is equal to distance x_P from H (of L_P) to O_P (and

similarly for O'_P, and for all other object-points and lenses throughout the system).

(4) The distance from the second surface of the thin lens that represents lens L_P, to the image that the lens forms of O_P, is equal to the distance y_P from H′ (of L'_P) to I_P (and similarly for I'_P, and for all other lenses and image-points throughout the system).*

Assumptions (3) and (4) follow directly from assumption (1).

(5) The radius of curvature of the first (anterior) surface of the thin lens that represents lens L_P, is equal to f_P of lens L_P (and similarly for the anterior surfaces of all other surrogate *thin* lenses in the relay stages of the system).

(6) The radius of curvature of the second (posterior) surface of the thin lens that represents L_P, is equal to $-f_P$ of lens L_P (and similarly for the posterior surfaces of all other surrogate *thin* lenses in the relay stages of the system).

Assumptions (5) and (6) follow directly from assumptions (1) and (2), because the radii of curvature of the surfaces of an equiconvex or equiconcave thin lens of focal length f, are given by the expression $r_1 = -r_2 = 2(n-1)f$,† and $n = 1.5$.

The distance (based on the thick lenses) from any object-point, such as O_P, to any of its subsequent images (in air or vacuum), such as I_U or I_Z, is equal to the corresponding distance (based on the surrogate *thin* lenses) that is calculated from the paraxially computed object-distances and image-distances, plus the sum of the principal point separations S_P, S'_P, S_U, etc., of all the (thick) lenses that are involved in forming the image.

The magnification, from any object-point (in air or vacuum) such as O_P, to any of its subsequent images (also in air or vacuum), such as I_U or I_Z, is equal to the ratio u/u', of the arbitrary paraxial inclination angle u of the initial paraxial ray through the object-point, to the paraxially computed value u' of the inclination angle of the corresponding refracted paraxial ray where it passes through the image-point. The inclination angles are measured between the paraxial rays and the system axis. An inverted image is indicated by a negative magnification.

Checks that were made of the solutions in the right hand column of Table 10, using the paraxial ray tracing methods, showed all images to be located within 0.001 or 0.002 inch of their required positions, and all magnifications to be within 0.01 or 0.02% of their required values. The solution of the example (in the right hand column of Table 10) is

* It is noted here, in reference to the second paragraph of Sec. 16, that either the object-distances or the image-distances for each lens (but not both) were involved in the solution of example 9. In the paraxial ray traces, both are determined for each lens of the system.

† See Ref. [21], p. 65.

therefore considered formally and numerically correct. Details of the check calculations are omitted, as they would introduce an unwarranted congestion into this paper.

22. The Special Cases, and the Excluded Cases

Special Case I: The Afocal Relay

The special case that has already been discussed, and used in the solution of the example, is that of the afocal relay at unit ($M_A = -1$) magnification. It is not mandatory in an afocal relay that $M_A = -1$, for in the general case Eqs. [47] and [110] state simply

$$M_A = -\frac{f'}{f}. \qquad [143]$$

In any two-lens afocal system, the F of Eq. [113] must be $F = \infty$, which is obviously the case if $a = f + f'$ in the denominator, independently of the value of M_A. It is true also if $f = f' = \infty$, but this is ordinarily a trivial case, providing that only purely Gaussian relationships are of interest.

So there are in general two practical cases of the afocal relay; that in which $M_A = -1$, and that in which $0 > M_A \neq -1$. By Eq. [48]

$$D = (M_A^2 - 1)x - (M_A - 1)(f + f') + S + S' \ (F = \infty), \qquad [144]$$

and by differentiation

$$\frac{dD}{dx} = (M_A^2 - 1) = \text{constant} \ (F = \infty). \qquad [145]$$

This derivative is constant and independent of x, for the reason that Eq. [143] shows M_A to be independent of x. The derivative therefore has the same value for every object-point on the axis of a given, fixed afocal system.

If from the continued Eq. [181] we write

$$D' + \Delta_c x = D + \Delta_c y, \qquad [146]$$

then, by differentiation with respect to x,

$$\frac{dD'}{dx} + \frac{d\Delta_c x}{dx} = \frac{dD}{dx} + \frac{d\Delta_c y}{dx}. \qquad [147]$$

Since, by Eq. [144], D' is distinguished from D merely by a change of x to x', it follows from Eq. [145] that $dD'/dx = dD/dx$, and therefore

$$\frac{d\Delta_c x}{dx} = \frac{d\Delta_c y}{dx}. \qquad [148]$$

It follows that if $\Delta_c x$ is held constant and independent of x, $\Delta_c y$ must be constant and independent of x. This conclusion is independent of M_A.

Returning now to Eq. [145], if $M_A = -1$ for an afocal system then dD/dx = constant = 0. This means that the afocal, unit-relay system can be shifted along the optical axis without changing the axial location of the image of either object. Since the magnification remains constant, a shift of the relay along the axis has no effect on the Gaussian imagery whatsoever, and the location of the relay along the axis is not critical in any way. By the same token no optical adjustment of a Gaussian nature can be accomplished by moving the unit-relay along its axis.

Once again returning to Eq. [145], if $0 > M_A \neq -1$ for an afocal system, then dD/dx = constant $\neq 0$, and D changes as the relay moves along the axis. But from the conclusion that followed Eq. [148], the value of $\Delta_c y$ is unaffected by the motion of the relay. In fact, since M_A is constant, the value of $\Delta_c y$ is given by the conjugate displacement theorem of Eq. [166] as

$$\Delta_c y = M_A^2 \Delta_c x. \qquad [149]$$

This must be constant if $\Delta_c x$ is held constant.

The conclusion from the preceding paragraph is that some internal adjustment of focus is possible by moving an afocal relay stage axially within an optical system, providing $0 > M_A \neq -1$. Magnifications, and the image displacement ($\Delta_c y$) are unaffected by the adjustment. Example 3 of Part 1, is a special case of this kind of adjustment, although applied to a system that is not a relay in the sense of the present discussion.

Special Cases II: The Classical Relay, and the Semi-classical Relays

The equations of Appendix 3 are directly applicable to the cases of the classical and the semi-classical relays, by making simple substitutions.

A classical relay stage results when we substitute $x' = y = 0$ into the equations. There will then be an image in the second principal plane of each of the two lenses of the stage. No classical stage appeared in the solution of Example 9.

An x'-semiclassical stage results when x' is given a finite value, but $y = 0$. An image is then located in the second principal plane of the posterior lens, but none in that of the anterior lens. The Z-stage of the system of the example was an x'-semiclassical relay.

A y-semiclassical relay stage calls for y to be given a finite value, but $x' = 0$. An image then appears in the second principal plane of the an-

terior lens, but none in that of the posterior lens. No y-semiclassical stages appeared in the solution.

The term "classical" is used because it applies to the simplest form in which the relay concept is explained in terms of thin lenses. A very thin scattering of such explanation appears here and there in technical literature. Justification for the use of the term "semiclassical" is obvious.

Some obvious simplification of the equations of Appendix 3, that precede Eq. [184], results from their specialization to the classical and semiclassical cases. There is little practical economy in restating the equations to cover such cases, and this will not be done.

Special Case III: The Pseudo-relay

This case bears a resemblance to the basic primitive relay optical system, the important difference being that while one of the two object-points is imaged by both lenses of the relay, the other object-point is imaged by only the posterior lens of the system. The Gaussian theory of the pseudo-relay has not been developed as a part of this paper.

The Excluded Cases: General

In any field of technology it is important to know how to solve its typical problems by application of suitable general methods, as far as that is possible. It is important to also have the kind of understanding of its theory that enables us to easily identify those of its problems that do not have solutions, and that therefore fall in the category of "Excluded Cases." Time and money that are spent on a problem that has no solution are usually wasted. If an "excluded case" can be identified at once, it may be practical to change the parameters of the problem in an acceptable manner, so that a solution will be possible.

In Gaussian optics the excluded cases are typically revealed by theoretical studies that involve the solution of quadratic equations. Such solutions typically involve radical expressions. These identify the excluded cases as being any sets of the problem parameters for which any applicable radicand of an even root is less than zero. In such cases the solutions of the quadratics are complex. In Gaussian optics there can be no physical optical system that corresponds to chosen values of the problem parameters that result in complex solutions.

For example, if we write the radicand of the expression for $M_{1,2}$ in Sec. 4, Part 1, in the form of the inequality

$$\left(1 - \frac{D - S}{2f}\right)^2 - 1 < 0 \qquad [150]$$

we display thereby the characteristic form or the type form of an excluded case. It applies to any image-forming system (in air or vacuum) that conforms to the relationships of Gaussian imagery, when it produces an image of a single object. A little care has to be taken in interpreting the symbols D and S in systems that involve mirrors, whether spherical or plane. Ineq. [150] results from the solution of any of the three quadratic equations in Row 9 of Table 1, Part 1.

In particular, Ineq. [150] applies when we deal with a coaxial system of two separated lenses that produces an image of a single object. Moreover, there are other relationships that define the excluded cases for such a system. The relationship of exclusion that is implicit in Eq. [16] is

$$\frac{(D - S_1 - S_2)^2}{4} - (D - S_1 - S_2)(f_1 + f_2) - \frac{(1 - M)^2}{M} f_1 f_2 < 0 \qquad [151]$$

Eqs. [22] and [16] both define the exclusion of Ineq. [151].

The exclusion defined by Eq. [21] for a system of two separated lenses is obviously

$$\frac{(x_1 + D - S_1 - S_2)^2}{4} - f_2[(1 - M)x_1 + D - S_1 - S_2] < 0 \qquad [152]$$

By the replacement transformation of Part 1, Table 3, the exclusion of Ineq. [152] takes the form

$$\frac{(-y_2 + D - S_1 - S_2)^2}{4} - f_1\left(\frac{1 - M}{M} y_2 + D - S_1 - S_2\right) < 0 \qquad [153]$$

By the replacement transformation, Ineq. [151] transforms into itself. Ineqs. [152] and [153] are complementary, with one being expressed in terms of x_1 and f_2, and the other in terms of y_2 and f_1.

The exclusion relationships for the case of the primitive two-lens relay optical system are complicated by the fact that they must define exclusions for a system that is producing images of two objects that are separated by the displacement $x' - x = \Delta_c x$ along the axis of the relay.

While many of these relationships are defined by the radicand of an even root being less than zero, there is another case that is so obvious that it can be easily overlooked. This is any case, including the case of the afocal relay, that contravenes the conjugate displacement theorem of Eqs. [41] and [166]. The exclusion relationship is expressed as the inequality

$$\frac{y' - y}{x' - x} = \frac{\Delta_c y}{\Delta_c x} \neq MM' \quad \text{(see Fig. 10).} \qquad [154]$$

Any proposal to design an optical system in air or vacuum, that satisfies Ineq. [154], is a proposal to waste time and money. Ineq. [154] assumes that x and x' are measured in the same object-space, and y and y' are measured in the same image-space of the optical system, or of any limited portion of the optical system for which the medium of the object-space and of the image-space is air or vacuum. It also assumes a centered system of spherical surfaces.

Example 7, Part 1, dealt with a system proposal that satisfied Ineq. [154]. It introduced changes that made the proposal satisfy the conjugate displacement theorem, before proceeding with the solution.

Apart from the exclusion defined by Ineq. [154], other conditions that define exclusion in the case of all primitive relays, are based on Eqs. [185], [186], [188], and [189]. Other equations that define exclusions result when eq. [184] is solved for $\Delta_c x$, and when Eq. [187] is solved for $\Delta_c y$. Solutions for $\Delta_c x$ and $\Delta_c y$ are not included in Appendix 3, and are left to the reader. As will be seen, all of these equations are needed in order to define all of the possible exclusions when the relay problem is set up in terms of the favored primary parameters that are set forth in Sec. 17.

In the following excluded cases that apply to the basic primitive relay, it is assumed that the given primary parameters are: either of M or M'; either of $\Delta_c x$ or $\Delta_c y$; x'; y; D_M; S; and S' (see Sec. 17).

Excluded Cases I: M and $\Delta_c x$ Given

From Eq. [185], the excluded cases (that result in complex values of M') are defined by the inequality

$$[M\Delta_c x(D_M - S - S' + x' - y - \Delta_c x) - M^2\Delta_c x(x' - \Delta_c x) + x'y]^2 - 4M^3 y\Delta_c x^2 (x' - \Delta_c x) < 0. \qquad [155]$$

By Ineq. [155], there is *no* exclusion when $y = 0$; this applies to the case of the classical relay (where also $x' = 0$), and to the case of the x'-semiclassical relay. When $x' = 0$ in the case of the y-semiclassical relay, there is exclusion *only if* Ineq. [155] is satisfied. In all the other cases where M and $\Delta_c x$ are given, exclusion occurs *only if* Ineq. [155] is satisfied. Note the complementarity with Excluded Cases IV.

Excluded Cases II: M' and $\Delta_c x$ Given

From Eq. [186], the excluded cases (that result in complex values of M) are defined by the inequality

$$[M'\Delta_c x(D_M - S - S' + x' - y - \Delta_c x) - y(x' - \Delta_c x)]^2 + 4M'^2 x'y[\Delta_c x(x' - \Delta_c x) + M'\Delta_c x^2] < 0. \qquad [156]$$

By Ineq. [156], there is *no* exclusion when $y = 0$; this applies to the case

of the classical relay (where also $x' = 0$), and to the case of the x'-semiclassical relay. There is also *no* exclusion when $x' = 0$, which is the case of the y-semiclassical relay and also that of the classical relay. In all other cases where M' and $\Delta_c x$ are given, exclusion occurs *only if* Ineq. [156] is satisfied. Note the similarity with Excluded Cases III.

Excluded Cases III: M and $\Delta_c y$ Given

From Eq. [188], the excluded cases (that result in complex values of M') are defined by the inequality

$$[M^2x'(y + \Delta_c y) - M\Delta_c y(D_M - S - S' + x' - y - \Delta_c y)]^2 - 4Mx'y\Delta_c y[M(y + \Delta_c y) - \Delta_c y] < 0. \quad [157]$$

By Ineq. [157], there is *no* exclusion when $y = 0$; this applies to the case of the classical relay (where also $x' = 0$), and to the case of the x'-semiclassical relay. There is also *no* exclusion when $x' = 0$, which is the case of the y-semi-classical relay and also that of the classical relay. In all other cases where M and $\Delta_c y$ are given, exclusion occurs *only if* Ineq. [157] is satisfied. Note the similarity with Excluded Cases II.

Excluded Cases IV: M' and $\Delta_c y$ Given

From Eq. [189], the excluded cases (that result in complex values of M) are defined by the inequality

$$[M'^2x'y + M'\Delta_c y(D_M - S - S' + x' - y - \Delta_c y) + \Delta_c y(y + \Delta_c y)]^2 - 4M'x'(y + \Delta_c y)\Delta_c y^2 < 0. \quad [158]$$

By Ineq. [158], there is *no* exclusion when $x' = 0$; this applies to the case of the classical relay (where also $y = 0$), and to the case of the y-semiclassical relay. When $y = 0$ in the case of the x'-semiclassical relay, there is exclusion *only if* Ineq. [158] is satisfied. In all other cases where M' and $\Delta_c y$ are given, exclusion occurs *only if* Ineq. [158] is satisfied. Note the complementarity with Excluded Cases I.

Development of the excluded cases that involve the solution of Eq. [184] for $\Delta_c x$, and the solution of Eq. [187] for $\Delta_c y$, is left to the reader.

Proposed systems that satisfy the foregoing applicable inequalities will be impossible to design unless changes are made. A system that does not satisfy an applicable inequality is a design possibility, but this does not guarantee that a practical design can be arrived at. For example, lens separation a might turn out to be negative which, in most cases will not be practical unless a special design effort (which in itself might not be practical) is applied to the lenses and subsystems in order to make their optical separation have the required negative value.

The theory of the excluded cases has been worked out in perfectly general Gaussian terms, and it does not depend in any way on x' and y having the relatively small values that are usually required by a primitive relay that is designed for optimum geometrical conservation of the light (optimum freedom from vignetting). They will, therefore, apply to any special case of the primitive relay that may not conform to the usual, or ideal scheme of light conservation.

23. Conclusion

This theoretical study is designed to reveal the Gaussian relationships that prevail in the basic two-lens relay optical system. It brings out the importance of the displacement intervals $\Delta_c x$ and $\Delta_c y$, and the importance of the mandatory conjugate displacement theorem that relates them. They are key analytical entities in the theory, and provide the basis for much simplification.

Object-to-image distances, that were formally made a part of simple Gaussian lens theory in Part 1 of this paper, play an important theoretical role when basic relays are combined into extended relay systems. It is through them that the required scale of the system is taken into account.

The basic relay problem involves many more variables than a system that produces an image of only one object. Its real solutions (as distinguished from its complex solutions) are subject to complicated limitations. The theory develops analytical means for identifying the complex or "excluded" cases. These can save much time and effort when considering the proposed solution of a relay problem in numerical terms.

Extended relay systems are conceived of as axial sequences of building-block-type modules, each module being a basic two-lens relay system. A simple, arbitrary, extended relay system problem is studied. It is solved in Gaussian terms for its lens locations and focal lengths. This is done to give the reader an elementary insight into the nature of relationships that are inherent between basic relay modules. The system studied is an extended relay optical system that meets representative ideal requirements, specifically substantial freedom from vignetting.

Relay optical systems of many forms and degrees of complexity have been studied by optical designers for many decades. Their methods are undoubtedly a mixture of theory with practical rules-of-thumb that enable them to get the design work completed in the most expedient manner. The methods and approaches that are used have not been made a systematic part of the generally available literature on the subject of lenses and lens systems.

It is true that designers cannot be limited to Gaussian considerations

in their work, and must be concerned with the correction of aberrations, and with vignetting, and with all of the relationships within the system that are involved in producing an image of acceptable quality that also has an acceptable level and distribution of illumination within its boundary. All of this must be consistent with a variety of nonoptical specification requirements, such as imposed mechanical boundary conditions and demands for economy. Many compromises have to be made in the solution of such a problem, the solution process being, because of its complexity, as much an art as it is a science.

It is difficult to make full acquaintance with the professional approach to optical design, without actually working with professional designers. Fortunately, there are few people who need such full acquaintance. But in a period when technical interest in optics is at an all-time high, there are growing numbers of people engaged in the disciplines of research, engineering, and technology who encounter optical problems.

Where these problems are lens problems the worker often needs to have only a preliminary insight into them on the simplest terms, namely the Gaussian terms. But this insight does need to be as complete as Gaussian terms will permit. The worker who cannot muster a thoroughgoing preliminary insight into his lens system problem, on the most effective Gaussian terms, is handicapped and at a definite disadvantage. In later stages of the solution of a lens system problem, optical design sophistication may be required. The worker may then develop a deeper understanding of lens system design, or rely on the help of competent professionals.

Once a system is initially conceived in Gaussian terms, its later elaboration and development may cause it to differ widely from its preliminary form in order to meet all requirements. It will usually, however, continue to conform closely in paraxial terms to the broadly stated Gaussian relationships of the initial solution.

As far as possible preliminary Gaussian solutions should be carried through with due consideration for the relationships of optical system photometry which, while not discussed here, are well developed in the literature.[23–28]

The writer must emphasize that the studies reported here are open ended. They are subject to whatever further development and extension that readers may see fit to pursue. The studies are suggestive of more effective Gaussian approaches to lens problems than previously developed in optical literature. There is no suggestion here that these studies are final and complete. Even greater usefulness of the Gaussian equations may be developed by pursuing their study further.

Appendix 3—Gaussian Development of Design Equations for the Basic Relay Optical System (See Note on Organization, Sec. 14)

With Secs. 16 and 17 and Figs. 9 and 10 as introduction, the analytical development begins by substituting, from Eqs. [7] and [8] (see transpositions of Sec. 18)

$$f = \frac{-Mxa}{(1-M)(x-a)+D-S-S'} \qquad [159]$$

and

$$f' = \frac{a(x+D-S-S')-a^2}{(1-M)x+D-S-S'} \qquad [160]$$

into

$$F = \frac{ff'}{f+f'-a} \quad \text{(see Eqs. [45] and [113])} \qquad [161]$$

with the result that

$$a = \frac{M[x^2+(D-S-S')x]-F[(1-M^2)x+D-S-S']}{Mx-F(1-M)}. \qquad [162]$$

Note: Given any optical system, of two separated lenses or sub-systems, defined by the set $\{f, f', a\}$ then for any conjugate displacement set $\{\Delta_c x, \Delta_c y\}$ defined for the system, with anterior end-points defined by the set $\{x, M, D\}$, and posterior end-points defined by the set $\{x', M', D'\}$, both of the end-point sets will satisfy Eqs. [159] and [160].

From Eq. [5], by simple transposition

$$a = x - y + D - S - S'; \qquad [163]$$

and by elimination of a between Eqs. [162] and [163]

$$D = S + S' - (1-M)x + \frac{Mx - F(1-M)}{FM}y. \qquad [164]$$

By eliminating $D - S - S'$ between Eqs. [163] and [164]

$$a = x - y - (1-M)x + \frac{Mx - F(1-M)}{FM}y. \qquad [165]$$

It will now be useful to express optical separation a in terms of x, x', y, y', M, and M', without involving F in the expression.

From Eq. [41]

$$\Delta_c y = MM'\Delta_c x \text{ (the conjugate displacement theorem)}, \qquad [166]$$

or

$$M' = \frac{\Delta_c y}{M\Delta_c x}. \qquad [167]$$

From Eq. [43]

$$F = \frac{-\Delta_c y}{M' - M}, \qquad [168]$$

and by substituting from Eq. [167] into Eq. [168], the result is

$$F = \frac{M\Delta_c x \Delta_c y}{M^2 \Delta_c x - \Delta_c y}. \qquad [169]$$

When F is eliminated between Eqs. [165] and [169]

$$a = Mx - y + \frac{x(M^2\Delta_c x - \Delta_c y) - (1 - M)\Delta_c x \Delta_c y}{M\Delta_c x \Delta_c y}. \qquad [170]$$

When Eq. [170] is put into the standard form of a quadratic in M

$$M^2 x \Delta_c x(y + \Delta_c y) - Ma\Delta_c x \Delta_c y - y\Delta_c y(x + \Delta_c x) = 0. \qquad [171]$$

By inspection of Fig. 10

$$\Delta_c x = x' - x \qquad [172]$$

and

$$\Delta_c y = y' - y. \qquad [173]$$

From the relationships of Eqs. [172] and [173], Eq. [171] reduces to

$$M^2 x y' \Delta_c x - Ma\Delta_c x \Delta_c y - x' y \Delta_c y = 0. \qquad [174]$$

By dividing Eq. [174] by the coefficient of M^2, and substituting

$$\Delta_c x = \frac{\Delta_c y}{MM'} \quad \text{(from Eq. [166])} \qquad [175]$$

the result is

$$M = \frac{a\Delta_c y + x' y M'}{xy'}. \qquad [176]$$

By then substituting the expression for $\Delta_c y$ from Eq. [173],

$$(M'x' - a)y = (Mx - a)y', \qquad [177]$$

from which the solution for a is

$$a = \frac{Mxy' - M'x'y}{y' - y}, \qquad [178]$$

and upon making substitutions from Eqs. [172] and [173] in order to reintroduce $\Delta_c x$ and $\Delta_c y$

$$a = \frac{Mx(y + \Delta_c y) - M'y(x + \Delta_c x)}{\Delta_c y}, \qquad [179]$$

In Eq. [179] *it is mandatory that the relation* of the conjugate displacement theorem *of Eq. [166] shall hold.* Otherwise Eq. [179] is without optical meaning. By eliminating $\Delta_c y$ between Eqs. [166] and [179],

$$a = \frac{M(x' - \Delta_c x)(y + MM'\Delta_c x) - M'x'y}{MM'\Delta_c x}. \qquad [180]$$

There is another procedure for deriving Eq. [180] from Eqs. [165], [166], [168], and [172], that involves long algebraic reductions that do not reveal interesting and always potentially useful relationships, such as those of Eqs. [169], [171], [174], and [177]. The relative advisibility of adopting various alternative procedures is not debated here.

By simple inspection of Fig. 10, and noting Eq. [166]

$$D_M = OI' = D' + \Delta_c x = D + \Delta_c y = D + MM'\Delta_c x. \qquad [181]$$

Further inspection of Fig. 10 shows that D_M may be expressed in more detailed terms, such as

$$D_M = \Delta_c x - x' + S + a + S' + y + MM'\Delta_c x, \qquad [182]$$

from which

$$a = D_M - S - S' + x' - y - (MM' + 1)\Delta_c x. \qquad [183]$$

Some Relations and Solutions Among the Primary Parameters of Sec. 17.

By eliminating a between Eqs. [180] and [183] there results

$$MM'\Delta_c x(D_M - S - S' + x' - y) - M^2M'^2\Delta_c x^2 - MM'\Delta_c x^2 = My(x' - \Delta_c x) + M^2M'\Delta_c x(x' - \Delta_c x) - M'x'y. \qquad [184]$$

Eq. [184] relates only primary parameters of Sec. 17. It is quadratic in M,M', and $\Delta_c x$, and linear in D_M, x', y, S, and S'.

The solution of Eq. [184] for M' is

$$M' = \frac{M\Delta_c x(D_M - S - S' + x' - y - \Delta_c x) - M^2\Delta_c x(x' - \Delta_c x) + x'y}{2M^2\Delta_c x^2} \pm \left[\left(\frac{M\Delta_c x(D_M - S - S' + x' - y - \Delta_c x) - M^2\Delta_c x(x' - \Delta_c x) + x'y}{2M^2\Delta_c x^2}\right)^2 - \frac{My(x' - \Delta_c x)}{M^2\Delta_c x^2}\right]^{1/2}. \qquad [185]$$

Similarly, the solution of Eq. [184] for M is

$$M = \frac{M'\Delta_c x(D_M - S - S' + x' - y - \Delta_c x) - y(x' - \Delta_c x)}{2[M'\Delta_c x(x' - \Delta_c x) + M'^2\Delta_c x^2]}$$

$$\pm\left[\left(\frac{M'\Delta_c x(D_M - S - S' + x' - y - \Delta_c x) - y(x' - \Delta_c x)}{2[M'\Delta_c x(x' - \Delta_c x) + M'^2\Delta_c x^2]}\right)^2\right.$$

$$\left. + \frac{M'x'y}{M'\Delta_c x(x' - \Delta_c x) + M'^2\Delta_c x^2}\right]^{1/2}. \qquad [186]$$

By substituting into Eq. [184] the value of $\Delta_c x$ in Eq. [175], there results

$$MM'\Delta_c y(D_M - S - S' + x' - y - \Delta_c y) - \Delta_c y^2$$
$$= M^2M'x'(y + \Delta_c y) - M\Delta_c y(y + \Delta_c y) - MM'^2x'y. \qquad [187]$$

As in the case of Eq. [184], Eq. [187] also relates only primary parameters of Sec. 17, they being in this case the parameters M, M', and $\Delta_c y$ in each of which the equation is quadratic, and D_M, x', y, S, and S' in each of which the equation is linear.

The solution of Eq. [187] for M' is

$$M' = \frac{M^2x'(y + \Delta_c y) - M\Delta_c y(D_M - S - S' + x' - y - \Delta_c y)}{2Mx'y}$$

$$\pm\left[\left(\frac{M^2x'(y + \Delta_c y) - M\Delta_c y(D_M - S - S' + x' - y - \Delta_c y)}{2Mx'y}\right)^2\right.$$

$$\left. - \frac{M\Delta_c y(y + \Delta_c y) - \Delta_c y^2}{Mx'y}\right]^{1/2}. \qquad [188]$$

Similarly, the solution of Eq. [187] for M is

$$M = \frac{M'^2x'y + M'\Delta_c y(D_M - S - S' + x' - y - \Delta_c y) + \Delta_c y(y + \Delta_c y)}{2M'x'(y + \Delta_c y)}$$

$$\pm\left[\left(\frac{M'^2x'y + M'\Delta_c y(D_M - S - S' + x' - y - \Delta_c y) + \Delta_c y(y + \Delta_c y)}{2M'x'(y + \Delta_c y)}\right)^2\right.$$

$$\left. - \frac{\Delta_c y^2}{M'x'(y + \Delta_c y)}\right]^{1/2}. \qquad [189]$$

Solution of Eq. [184] for $\Delta_c x$, and of Eq. [187] for $\Delta_c y$, is left to the reader.

Appendix 4—Development of Auxiliary Equations That Are Employed in the Solution of Example 9 (Sec. 19)

All of the equations of this appendix are based on Fig. 11, and many of them also on the relationships derived in Appendix 3 or in the body of the text. Each of them is keyed to the applicable Roman-numeral step-number of the solution of Example 9, and to the item-number of Table 9 to which it applies.

I. Item 3; (d_{oP})—Unknown: Literal solution only.

The magnification from O_P to I_Z is the overall magnification for the anterior object O_P. This may be expressed as

$$M_{OAP} = M_{Un}\frac{d_{iZ}}{d_{oP}} = (-1)^n\frac{d_{iZ}}{d_{oP}} \quad \text{(See Eq. [114]).} \qquad [190]$$

where n is the number of unit-relay stages. The overall magnification from O'_P to I'_Z, for the posterior object O'_P, is similarly expressed as

$$M'_{OAP} = M'_{Un}\frac{d'_{iZ}}{d'_{oP}} = (-1)^n\frac{d'_{iZ}}{d'_{oP}} \quad \text{(See Eq. [114]).} \qquad [191]$$

By the conjugate displacement theorem of Eqs. [41] and [166], it follows from Eqs. [190] and [191] that

$$\frac{\Delta_c y_z}{\Delta_c x_P} = M_{OAP}M'_{OAP} = \frac{d_{iZ}d'_{iZ}}{d_{oP}d'_{oP}}. \qquad [192]$$

But

$$d'_{oP} = 2\Delta_c x_P \tan\theta. \qquad [193]$$

On substitution into Eq. [192] from Eq. [193], and solving for the value of d_{oP} in the P-stage of Fig. 11,

$$d_{oP} = \frac{d_{iZ}d'_{iZ}}{2\Delta_c y_Z \tan\theta}. \qquad [194]$$

V. Item 15; ($\Delta_c y_P$)—Unknown: Literal solution only

When we consider the n identical unit-stages of the U-relay as a group, the magnifications of the group are, from Eq. [114],

$$M_{Un} = (-1)^n \qquad [195]$$

and

$$M'_{Un} = (-1)^n. \qquad [196]$$

By the conjugate displacement theorem of Eqs. [41] and [166]

$$\Delta_c y_{Un}\ (= \Delta_c y_U) = M_{Un}M'_{Un}\Delta_c x_{Un} = \Delta_c x_{Un} = \Delta_c x_U. \qquad [197]$$

But by Table 9, Item 22, $\Delta_c x_U = \Delta_c y_P$, and by Item 40, $\Delta_c y_U = \Delta_c x_Z$, by which it follows from Eq. [197] that

$$\Delta_c y_P = \Delta_c x_Z. \qquad [198]$$

But since also by the conjugate displacement theorem

$$\Delta_c y_Z = M_Z M'_Z \Delta_c x_Z,$$

it follows from Eq. [198] that

$$\Delta_c y_Z = M_z M'_Z \Delta_c y_P$$

and that consequently

$$\Delta_c y_P = \frac{\Delta_c y_Z}{M_Z M'_Z}. \qquad [199]$$

XVI. Item 23; (x'_U)—Unknown: Literal solution only

By Eq. [180], after applying U-subscripts

$$a_U = \frac{M_U(x'_U - \Delta_c x_U)(y_U + M_U M'_U \Delta_c x_U) - M'_U x'_U y_U}{M_U M'_U \Delta_c x_U}. \qquad [200]$$

Then, upon substituting $M_U = M'_U = -1$, and also substituting from Table 9, Item 23, $y_U = -x'_U$, it follows that

$$a_U = \frac{(x'_U - \Delta_c x_U)^2 - x'^2_U}{\Delta_c x_U},$$

or

$$a_U = \Delta_c x_U - 2x'_U, \qquad [201]$$

By Eq. [182], after applying U-subscripts

$$D_{MU} = (1 + M_U M'_U)\Delta_c x_U - x'_U + y_U + a_U + S_U + S'_U. \qquad [202]$$

Then by substituting the following

$$M_U = M'_U = -1,$$

$$y_U = -x'_U,$$

$$a_U = \Delta_c x_U - 2x'_U \quad \text{(from Eq. [201])},$$

Eq. [202] becomes

$$D_{MU} = 2\,\Delta_c x_U - x'_U - x'_U + \Delta_c x_U - 2x'_U + S_U + S'_U. \qquad [203]$$

When Eq. [203] is solved for x'_U

$$x'_U = \frac{3\Delta_c x_U - D_{MU} + S_U + S'_U}{4}. \qquad [204]$$

References:

[11] L. T. Sachtleben, "Extending the Content and Expanding the Usefulness of the Simple Gaussian Lens Equations, (Part 1)," *RCA Review*, 37, No. 4, p. 437, Dec. 1976.

[12] R. Glazebrook, Ed., *Dictionary of Applied Physics*, Vol. IV, pp. 350–374, Macmillan, New York (1922); reprinted through arrangement with the Macmillan Company by Peter Smith, New York (1950).

[13] MIL-HDBK-141 (5 October 1962) FSC-6650, *Military Standardization Handbook—Optical Design*, pp. 12-1 to 12-4 and 7-8 through 7-11, Defense Supply Agency, Washington, D. C. (1962).

[14] R. Kingslake, Ed., *Applied Optics and Optical Engineering*, Vol. I, pp. 229–231 and pp. 240–242, and Vol. II, pp. 227–228, Academic Press, Inc., New York (1965).

[15] W. J. Smith, *Modern Optical Engineering*, pp. 212–213, McGraw-Hill Book Co., New York (1966).

[16] *Glossary of Terms Frequently Used in Optics and Spectroscopy*, Report of Seminar on Optics and Spectroscopy for Science Writers, Pick-Roosevelt Hotel, Pittsburgh, Pa., March 1, 1961, Dr. Stanley S. Ballard, Chairman. American Institute of Physics, New York (1961).

[17] E. U. Condon and Hugh Odishaw, Eds., *Handbook of Physics*, 2nd Edn., McGraw-Hill Book Co., New York (1967).

[18] *The International Dictionary of Physics and Electricity*, 2nd Edn., D. Van Nostrand Company, Inc., Princeton, N. J. (1961).

[19] *American Institute of Physics Handbook*, 3rd Edn., McGraw-Hill Book Company, New York (1972).

[20] Robert M. Besançon, Ed., *The Encyclopedia of Physics*, 2nd Edn., Van Nostrand Reinhold Company, New York (1974).

[21] A. E. Conrady, *Applied Optics and Optical Design—Part One*, pp. 37–43, Oxford University Press (1929) and Dover Publications, Inc. (1957).

[22] D. H. Jacobs, *Fundamentals of Optical Engineering*, pp. 387–388, McGraw-Hill Book Co., New York (1943).

[23] L. C. Martin, *An Introduction to Applied Optics*, Vol. II, Ch. VI - Photometry of Optical Systems, Isaac Pitman, Ltd., London (1932). (In later editions, titled *Technical Optics*.)

[24] H. A. E. Keitz, *Light Calculations and Measurements*, Chaps. VIII, IX, and XI, Philips' Technical Library, N. V. Philips' Gloeilampenfabrieken, Eindhoven, Holland (1955).

[25] Francis Weston Sears, *Principles of Physics III - Optics*, Ch. 12 - Photometry, Addison-Wesley Press, Inc., Cambridge, Mass. (1945).

[26] A. E. Hardy and F. H. Perrin, *The Principles of Optics*, Ch. XIX - The Design of Optical Instruments, McGraw-Hill Book Co., New York (1932).

[27] R. Kingslake, Ed., *Applied Optics and Optical Engineering*, Vol. II, Ch. 5 - Illumination in Optical Images, Academic Press, New York (1965).

[28] R. S. Longhurst, *Geometrical and Physical Optics*, 2nd Edn., Ch. XVIII - Photometry, John Wiley & Sons, Inc., New York (1967).

Patents Issued to RCA Inventors First Quarter 1978

January

A. A. Ahmed Current Mirror Amplifier (4,068,184)
C. H. Anderson Flat Panel Display with Beam Injection Cleanup (4,069,439)
L. J. Bazin Signal Comparator Circuit (4,069,432)
D. J. Carlson Remote Control Transmitter with an Audible Battery Life Indicator (4,067,000)
A. G. Dingwall Method of Making a Semiconductor Device (4,069,577)
A. G. Dingwall and B. D. Rosenthal Direct-Coupled Cascade Amplifier with Automatically Adjusted Quiescent Output Signal Level (4,068,182)
A. H. Firester and J. L. Walentine Defect Plotting System (4,069,484)
R. A. Geshner and L. J. Sciambi Photomask Dryer (4,068,390)
N. F. Gubitose, R. J. Satriano and L. Gawelko Method for Cracking Brittle Material (4,068,788)
J. R. Hall Circuit for Inhibition of Autogenetic False Alarms in a Collision Avoidance System (4,067,010)
E. P. Herrmann Four-Quadrant Multiplier (4,071,777)
R. C. Heuner, S. J. Niemiec, G. I. Morton, and M. B. Goldman Protection Circuitry for Insulated-Gate Field-Effect Transistor (IGFET) Circuits (4,066,918)
R. N. Hurst Arrangements for Testing Color Television Systems (4,069,500)
R. H. Isham Sawtooth Voltage Generator for Constant Amplitude Sawtooth Waveform from Varying Frequency Control Signal (4,071,776)
J. R. Jasinski, J. B. Pickard, and C. E. Doner Double Tuned Input Circuit for Television Transmitter Amplifier (4,070,627)
P. J. Kannam Transistor Having Improved Junction Breakdown Protection Integrated Therein (4,071,852)
R. M. Kongelka Releasable Mounting System (4,071,217)
H. Kressel, R. V. D'Aiello, and P. H. Robinson Polycrystalline or Amorphous Semiconductor Photovoltaic Device Having Improved Collection Efficiency (4,070,206)
N. Kucharewski Amplifier Circuit (4,069,431)
W. G. McGuffin Adaptive Delta Modulation System (4,071,825)
L. Muhlfelder, J. E. Keigler, and B. Stewart Momentum Biased Active Three-Axis Satellite Attitude Control System (4,071,211)
M. N. Norman Automatic Beam Current Limiter (4,067,048)
L. P. Orchard Food Product Extrusion Apparatus and Method (4,068,008)
J. I. Pankove and D. E. Carlson Electroluminscent Semiconductor Device Having a Body of Amorphous Silicon (4,069,492)
H. L. Pinch, B. Abeles, and J. I. Gittleman Method of Making High Resistance Cermet Film (4,071,426)
W. Rosnowski and R. Denning Method of Fabricating a Semiconductor Device (4,066,485)
J. M. Shaw and K. H. Zaininger Method of Manufacturing Apertured Aluminum Oxide Substrates (4,069,094)
A. Waksberg Automatic Optical Bias Control for Light Modulators (4,071,751)
C. S. Warren Data Packets Distribution Loop (4,071,706)

February

C. H. Anderson Flat Display Device with Beam Guide (4,076,994)
F. Aschwanden Secam Identification Circuit (4,072,983)
D. S. Binge Spacecraft Component Rotation Means (4,076,191)
R. S. Crandall and B. W. Faughnan Method of Storing Optical Information (4,075,610)
B. Crowle Voltage Regulators of a Type Using a Common-Base Transistor Amplifier in the Collector-to-Base Feedback of the Regulator Transistor (4,074,181)
A. R. Dholakia VideoDisc Player Employing a Spring Loaded Stylus Apparatus (4,077,050)
A. G. F. Dingwall and B. D. Rosenthal Voltage Controlled Oscillator Having Equally Controlled Current Source and Current Sink (4,072,910)
R. A. Dischert, A. C. Luther, JR., and R. N. Hurst Signal Processor Using Charge-Coupled Devices (4,074,307)
J. G. Endriz System for Modulating a Flat-Panel Image Display Device (4,077,054)
W. G. Gibson Delay Line Network for Processing a Composite Electrical Signal (4,074,308)

H. A. Gross, R. F. Walters, H. R. Frey, and J. D. Messner Method for Inspecting Cathode-Ray-Tube Window for Objectionable Cord (4,076,426)
A. C. Ipri Silicon Resistive Device for Integrated Circuits (4,072,974)
H. C. Johnson Monotonically Ranging FM-CW Radar Signal Processor (4,072,947)
G. Kaganowicz and J. W. Robinson VideoDisc with a Dielectric Layer Formed from Styrene and Nitrogen (4,072,985)
J. M. Keelty Wide Acquisition Range MSK Demodulator Input Circuit (4,072,905)
M. Lysobey Wide Range Frequency Modulation of Narrow Loop Bandwidth Phase-Locked Oscillators (4,074,209)
J. T. Mark Vacuum Tube Grid Structure of Phosmic Bronze Having Copper and Copper Alloy Conical Supports (4,076,992)
J. E. Miller Multi-Target Tracker (4,072,943)
A. M. Morrell Corrugated Shadow Mask Assembly for a Cathode Ray Tube (4,072,876)
J. R. Oberman and R. G. Stewart Write Enhancement Circuit (4,075,690)
J. I. Pankove Apparatus and Method for Maskless Ion Implantation (4,074,139)
B. M. Pradal Network for Temperature Compensation of an AT Cut Quartz Crystal Oscillator (4,072,912)
R. M. Rast and J. Tults Phase Locked Loop Tuning System with Station Scanning Provisions (4,077,008)
J. H. Reisner, W. H. Morewood, and G. H. N. Riddle Electron Beam Disc Recorder (4,074,313)
T. J. Robe Input Stage for Fast-Slewing Amplifier (4,074,205)
T. J. Robe Input Stage for Fast-Slewing Amplifier (4,075,575)
J. M. Shaw and K. H. Zaininger Method of Making Planar Silicon-on-Sapphire Composite (4,076,573)
R. J. Treadwell Brushless Phase Locked Servo Drive (4,072,884)
J. L. Vossen, Jr. VideoDisc with a Conductive Layer Having an Oxygen Content Gradient (4,077,051)
J. L. Vossen, Jr. VideoDisc Capacitive Recording Means with a Conductive Bilayer (4,077,052)
M. H. Wardell, Jr., and B. G. Marks High Voltage Electron Tube Base with Separate Dielectric Fill-Hole (4,076,366)

March

G. W. Albrecht Inverter Circuit Control Circuit for Precluding Simultaneous Conduction of Thyristors (4,078,247)
B. Crowle Differential Amplifier (4,078,206)
W. Denhollander Power Supply Arrangement with Minimum Interaction Between Plural Loads (4,079,295)
R. S. Engelbrecht and K. Knop Diffractive Phase Filter Having Near Field Wavelength Dependent Focusing Properties (4,079,411)
R. D. Faulkner and R. E. McHose Phototube Having Apertured Electrode Recessed in Cup-Shaped Electrode (4,079,282)
P. E. Haferl Switched Mode Vertical Amplifier with Elimination of Feedback Ringing (4,079,293)
P. E. Haferl Conduction Overlap Control Circuit for Switched Output Stages (4,081,721)
P. E. Haferl Signal Processor for Switched Vertical Deflection System (4,081,722)
W. E. Ham Apparatus and Method for Cleaning and Drying Semiconductors (4,079,522)
R. N. Hurst and F. W. Huffman Discontinuous Motion Special Effects Generator for Television (4,080,626)
H. Kawamoto and E. J. Denlinger Pickup Circuitry for a VideoDisc Player with Printed Circuit Board (4,080,625)
J. B. Klatskin and A. Rosen Method of Manufacturing Semiconductor Devices Having a Copper Heat Capacitor and/or Copper Heat Sink (4,080,722)
A. J. Leidich Push-Pull Transistor Amplifier with Drive Circuitry Providing Over-Current Protection (4,078,207)
J. A. Lunsford and L. W. Martinson Output Buffer Synchronizing Circuit Having Selectively Variable Delay Means (4,079,456)
F. J. Marlowe Line Scan Converter for an Image Display Device (4,080,630)
D. D. Mawhinney and Z. Turski Microwave Frequency Discriminator (4,079,325)
E. S. Naidus and C. W. Clark Skylight Frame Construction (4,080,763)
S. E. Ozga Microprocessor Architecture (4,079,455)
H. Popp Grid Having Reduced Secondary Emission Characteristics and Electron Discharge Device Including Same (4,079,286)

R. N. Rast Dual Mode Frequency Synthesizer for a Television Tuning Apparatus (4,078,212)
J. M. Ratay Slurry Process for Coating Particles Upon the Viewing-Window Surface of a Cathode-Ray Tube (4,078,095)
O. H. Schade, Jr. Ground Fault Detector (4,080,641)
J. J. Serafini Automatic Transient Beam Current Limiter (4,079,424)
S. A. Steckler Low Distortion Signal Amplifier Arrangement (4,081,758)
R. G. Stewart Level Shift Circuit (4,080,539)
S. Weisbrod Apparatus for Measuring the Current-Voltage Characteristics of a Trapatt Diode (4,080,571)

AUTHORS

Alvin Malcolm Goodman received a B.S.E. in 1952 from the Drexel Institute of Technology, Philadelphia, Pa. He did graduate work at Princeton University, receiving an M.A. in 1955 and a Ph.D. in 1958. From June 1956 to January 1957, he served as Research Assistant at Princeton University, and then as Assistant Professor of Electrical Engineering at the Case Institute of Technology up to June 1959. He performed research at RCA Laboratories as a summer employee in 1954, 1955, and 1958. He has been a Member of the Technical Staff since June 1959. During the year 1970–71, Dr. Goodman engaged in postdoctoral studies at the Swiss Federal Institute of Technology. Dr. Goodman has specialized in solid-state physics. His thesis subject was "Dember Effect and Trap Levels in Silver Chloride," and he has worked extensively in the areas of photoconductivity, metal-semiconductor contacts, metal-insulator contacts, insulator properties, and tunnel diodes. He has received two RCA Laboratories Achievement Awards for work on metal-semiconductor contacts (1963) and MNOS (metal-nitride-oxide-silicon) memory devices (1969). Dr. Goodman is a member of the American Physical Society, Sigma Xi and the Institute of Electrical and Electronics Engineers.

William E. Ham received the B.S. and M.E. (electrical engineering) degrees from the University of Oklahama, at Norman, both in 1966, and the Ph.D. degree in electrical engineering from Southern Methodist University, Dallas, Texas, in 1970. Since graduation he has been a Member of the Technical Staff at RCA Laboratories. His primary work has been with electrical characterization and MOS device properties of heteroepitaxial semiconductors for integrated circuit applications.His research has involved the effective use of various types of vehicles, testing, and data presentation for the optimal understanding of integrated circuit processes. Specifically, this involves the study of electrical instabilities and dielectric properties, correlation of physical structure with electrical performance, and spatial distribution of device parameters. Current studies also include dimensional instabilities of wafers and photomasks and analysis of circuit yield distributions.

Dr. Ham is a member of the Electrochemical Society, Sigma Xi, Eta Kappa Nu, Sigma Tau, and Tau Beta Phi.

Harvey O. Hook received the BA with a chemistry major from Elon College in 1947, and the BEE and MSEE degrees from North Carolina State College in 1949 and 1950, respectively. Since 1950, Mr. Hook has been with RCA Laboratories in Princeton, N. J. He has worked on display storage tubes, opto-electronic computer components, vacuum technology, color picture tubes, the optics of displays, high-resolution photolithography, and research instrumentation. Mr. Hook is a Senior Member of IEEE and a member of the Instrument Society of America, the American Vacuum Society, and the Society of Photographic Scientists and Engineers.

Werner Kern received his education in chemistry at schools in Switzerland and the U. S. A., including the University of Basle and Rutgers University. His thesis, published in 1947, was on the chromatographic isolation and characterization of fluorescing polynuclear hydrocarbons which he discovered in soil. He was analytical research chemist with Hoffmann-LaRoche in Switzerland until 1948 when he transferred to their research division in New Jersey to develop radioactive tracer methods for biochemical applications. In 1958 he joined Nuclear Corporation of America where he became chief chemist directing applied research in nuclear radiation chemistry. He joined RCA in 1959, working primarily on the investigation of semiconductor processes by radiochemical methods. He was project scientist and consultant on several research projects, and was in charge of radiological safety. Since 1964 he has been at RCA Laboratories, where he has specialized in semiconductor process research, chemical vapor deposition technology, and the development of new analytical methods for characterizing dielectric films. In 1974–1975 he was project scientist for two government-sponsored research contracts on integrated-circuit glass passivation.

Mr. Kern is a member of the American Chemical Society, the Electrochemical Society, the Research Honorary Society of Sigma Xi, and Geological Society of New Jersey. He received an RCA Achievement Award in 1966 for his work in integrated-circuit process research. Mr. Kern is the recipient of the T. D. Callinen Award for 1971 of the Dielectrics and Insulation Division of the Electrochemical Society, in recognition of his pioneering work in chemical-vapor-deposition research. He received a 1973 RCA Laboratories Outstanding Achievement Award for his team contributions to glass passivation of silicon device structures.

Paul Kuczer graduated from Trenton Technical Institute in 1961. In 1962, he joined RCA Laboratories, Princeton, N. J. His initial responsibilities as a member of the Engineering Service Department were layout, design, and fabrication of electronic instruments. In 1968, he completed the Electronic Computer Systems Course at RCA Technical Institutes and assumed the responsibilities of maintaining the Gerber Automated Coordinatograph, for which he received an award in 1970. His present responsibilities are in the design and fabrication of instruments to aid in the manufacture of television picture tubes.

L. T. Sachtleben received his B.Sc. Degree in Physics and Mathematics from Antioch College. His work in optics began in the field of sound motion pictures, from which point he has continued as a generalist in optics at the RCA plants in Camden and Indianapolis. He has written or co-authored numerous published papers, and 47 patents have been issued to him. He is a member of SMPTE, OSA, AAPT, and NCTM.

Otto H. Schade, Jr., received the BEE degree from Rensselaer Polytechnic Institute in June, 1953. Following graduation, he joined the Electron Tube Division of RCA at Harrison, N. J. where he designed damper and deflection tubes for TV applications, Nuvistors, and mechanical filters; he also conducted studies of heat flow and electron optics in vacuum tubes. During 1961, Mr. Schade was responsible for the electrical performance and testing of silicon-germanium thermoelectric converters for Atomics International's SNAP 10A space program. He then built high-efficiency gaseous and liquid-fuel radiant burners for 1000 to 60,000 BTU/hr applications, and performed the system design of a 100-watt portable thermoelectric generator.

In 1966, Mr. Schade transferred to the Solid State Division, Somerville, N. J., where he developed silane-deposition systems, studied means for automatic photomask alignment and projection printing, and analyzed steady-state and transient thermal behavior in rf power transistors. During 1968, he assisted O. H. Schade Sr. in the development of the high-definition TV art by designing solid-state horizontal and vertical deflection amplifiers for a 100-MHz, 4000-line system. In 1969, Mr. Schade joined the Linear IC group at Somerville, where he has been active in the design, development, and application of industrial and custom circuits for uses such as CMOS D/A conversion, ground-fault interrupt service, and conventional and radiation-hard operational amplifiers. He is the architect of RCA's first MOS/bipolar monolithic amplifiers.

Mr. Schade is a member of Eta Kappa Nu.